EPIDEMIOLOGY, EVOLUTION, AND DIAGNOSTICS OF TUBERCULOSIS IN HUMANS AND ANIMALS

By

Syeda Anum Hadi

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Comparative Medicine and Integrative Biology- Doctor of Philosophy

2022

ABSTRACT

EPIDEMIOLOGY, EVOLUTION, AND DIAGNOSTICS OF TUBERCULOSIS IN HUMANS AND ANIMALS

By

Syeda Anum Hadi

Tuberculosis control in animals and humans alike requires early detection of *Mycobacterium tuberculosis* complex as well as current knowledge about the transmission patterns of the disease in the respective populations. These two building blocks provide the foundation on which the disease control programs can build their policies to expediate control efforts. In this thesis we amalgamate molecular epidemiology, genomics, and proteomics. We studied the transmission pattern of *M. tuberculosis* and its evolution within a marginalized population. The patterns led to the identification of gaps in TB control policies in marginalized populations with little access to healthcare. Similarly, we studied the genomewide polymorphisms in a naturally attenuated strain - *M. bovis* strain Ravenel to elucidate possible mechanisms for its reduced virulence and pathogenicity. Insights gained from genome sequence analysis in conjunction with pathogenesis study for *M. bovis* Ravenel paved the pathway to defining the complex and multi-faceted reasons for attenuation of the oldest bacteria of the world. Next, pathogen-specific biomarkers were evaluated to assist in unambiguous disease detection across multiple host species. Discovery and validation of biomarkers work facilitated the field diagnostic applications for TB in animals and humans. This three-pronged approach developed in this study, understanding the genomic basis of attenuation, and enhanced field TB diagnostics in the animal-human interface.

This work is dedicated to my teachers who taught me how to read, write and think critically, especially my Late Grandmother Razia Begum.

ACKNOWLEDGEMENTS

Dr. Sreevatsan once said to me ‘It takes a village to raise a PhD’.

I will run out of space acknowledging everyone who lifted my spirit at my lowest… celebrated my joys with me…or just stayed silent with me when I wanted stillness...

To my PhD advisor, life coach, mentor, friend and spiritual advisor – Dr. Srinand Sreevatsan…

Funding agencies- Fulbright scholarship, United States Department of Agriculture, Graduate School…

Administration and Program Officers both at MSU and Fulbright...

Core support– Mama, Papa, Aiman, Alizeh, Iraj, Ayzal, Alina, Babbi, Olivia, Evan, Eram, Livia, Jamie, Devika, Sylvia, Nicole, Taveesak, Jannani, Kamonpan, Azrina, Fernanda, Carine, Sarad, Margie, Madiha, Sara, Shumaila, and Vish…

And to all grocery stores, dollar stores, meat shops, bakeries, restaurants, cafes, ice-cream parlors, dining halls as well as the bus drivers who were always on time and saved me from freezing cold…

I thank you, with all my heart!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

TB	Tuberculosis
MTC	*Mycobacterium tuberculosis* Complex
MTB	*Mycobacterium tuberculosis variant tuberculosis*
MBO	*Mycobacterium tuberculosis variant bovis*
BTB	Bovine tuberculosis
PSB	Pathogen-specific biomarkers
WHO	World Health Organization
VNTR	Variable-number tandem repeats
RFLP	Restriction fragment length polymorphism
NZG	National Zoological Gardens of South Africa
NHP	Non-human primates
MIRU-VNTR	Mycobacterial-interspersed-repetitive-units of VNTR
LSP	Large Sequence Polymorphism
WGS	Whole genome sequencing
SNP	Single nucleotide polymorphism
LAM	Latin American and Mediterranean
LTBI	Latent TB infection
WGS	Whole-genome sequencing
MS	*Mato Grosso do Sul*

SESAI	Special Secretariat for Indigenous Health of the Ministry of Health
LACEN-MS	Central Public Health Laboratories of *Mato Grosso do Sul ((Laboratório Central de Saúde Pública de Mato Grosso do Sul)*
CTAB	Cetyltrimethylammonium bromide
ML	Maximum likelihood
RRDR	Rifampicin resistance-determining region
MDR	Multi drug resistant
INH	Isoniazid
QRDR	Quinolone resistance-determining region
RD	Regions of differences
PGAP	Prokaryotic Genome Annotation Pipeline
CDS	Coding sequences
SDG	Sustainable Development Goals
O.D. $_{450nm}$	Optical density at 450nm wavelength
NHP	Non-human primates
DPP	Dual path platform
BSL-3	Biosafety level 3
LC-MS/MS	Liquid chromatography mass-spectrometry mass-spectrometry
NCBI	National Center for Biotechnology Information
BLASTp	Basic local alignment search tool for proteins
AFZ	Accredited free zone

MAZ	Modified accredited zone
PATRIC	Pathosystems Resource Integration Center
IPTG	Isopropylthio-β-galactoside
IMAC	Immobilized metal ion affinity chromatography
BSA	Bovine serum albumin

CHAPTER 1 INTRODUCTION

Tuberculosis (TB) is a disease of global health emergency. TB in several animal species is caused by organisms that belong to the *Mycobacterium tuberculosis* Complex (MTC), a complex that share 99.9% identities across their genomes. *Mycobacterium tuberculosis variant tuberculosis* (MTB) is the primary causative agent of TB in humans whereas *Mycobacterium tuberculosis variant bovis* (MBO) is the primary causative agent of TB in cattle otherwise known as bovine TB (BTB). MBO is a zoonotic agent that can be transmitted from animals to humans through consumption of infected dairy products, inhalation of contaminated air-droplets from animals and to a lesser extent by inhalation of infected fecal dust. Bi-directional transmission of both MTB and MBO has been documented in ecosystems where animal-human contact is intimate. The interconnectedness of domesticated animals (especially cattle, sheep, and goats) and humans makes for facile interspecies transmission and almost dictates the need for a multi-disciplinary expertise including epidemiologists, veterinarians, physicians, sociologists, and environmental biologists to work together to inform policy on TB control and eradication. To date, vaccines, aggressive treatment strategies and culling of infected animals have all been employed to eliminate the disease, with limited success.

This thesis utilizes the commonalities between human and animal TB and is built upon three major facets of TB - epidemiology, evolution, and diagnostics.

Specific Aim 1: Genomic epidemiology of MTB in an isolated human indigenous community reveals a unique evolutionary pattern of MTB.

1.1. First, genomes of locally circulating MTB in the indigenous population of the Guarani-Kaiowá (in Southern Brazil) were analyzed to assess if TB was caused by multiple introductions from outside the community, or a result of unchecked transmission within.

1.2 Genomic determinants of drug resistance among the MTC isolates circulating in the indigenous population of the Guarani-Kaiowá, in Brazil were analyzed.

Specific Aim 2: Genomic analysis to evaluate markers of an in-lab attenuated strain of MBO in comparison to pathogenic strains of MBO.

2.1 Perform genetic analysis of the attenuated MBO Ravenel strain in comparison to pathogenic and non-pathogenic strains of MBO.

Specific Aim 3: Evaluation and employment of pathogen-specific biomarkers (PSB) in the development of a point-of-care screening tool applicable across species.

3.1 Validate previously discovered pathogen-specific biomarkers in non-human-primates.

3.2 Validate pathogen-specific biomarkers in infected animals using immunoenrichment assay.

3.3 Develop DNA aptamers against 3 pathogen specific biomarkers, Pks5, MB2515c & MB1895c, to enable a point-of-care (cow-side) testing for TB.

Each sub-aim forms the foundation of an individual chapter in the present work. In this body of work, unique disease patterns, genomic fingerprints as well as more reliable methods of detecting the elusive *Mycobacterium*, were discovered.

CHAPTER 2 REVIEW OF LITERATURE

Disease

Disease History

On 24th March 1882, Dr. Robert Koch discovered the causative agent of the world's oldest recorded disease TB(1–5). This disease does not differentiate between kings and terrorists or between intellectuals and common man. From Kings like Napoleon II (1811-1832) & Grand Duke George Alexandrovich of Russia (1871-1899); to leaders like Eleanor Roosevelt (1884-1962) & Muhammad Ali Jinnah (1876-1948); to acclaimed writers like Jane Austen (1775- 1817), Emily Brontë, (1818-1848) & George Orwell (1903-1950); to poets like John Keats (1795-1821); to musicians like Frederic Chopin (1810-1849); to inventors like Louis Braille (1809-1852); to intellectuals like Khalil Jibran (1883-1931); to physicians like Dr. Anandi Gopal Joshi (1865-1887); to scientists like Andres Celsius (1701-1744), Friedrich Miescher (1844-1895); to terrorists like Mohammad Omar (1960-2013) no one was spared by TB (6).

TB has been defined as a disease that primarily affects the respiratory system(7, 8). It is caused by a complex of organisms called MTBC. The complex comprises of nine mycobacterial members that share high genomic similarity(9). Two of the most prominent members of this complex will be studied in detail in this thesis- MBO and MTB. MTB is the causative agent of TB in humans, whereas TB in bovines is caused by MBO. Even though MTB primarily infects humans, elephants(10–24) and non-human primates(25, 26) cases of transmission of MTB to animals have been reported around the world (27–31). MBO on the other hand primarily infects bovines and other warm-blooded animals but is also identified as a zoonotic agent (32–44). It is transmitted

to humans via infected dairy products, fomites as well as cohabitation with tuberculosed animals (45–53).

TB in numbers

It is estimated that almost one-third of the world's population (~2.5 billion people) has latent or subclinical TB. Furthermore, in 2018, 10 million people were infected and clinically ill with TB, according to WHO. Of these TB cases an estimated 1.2 million mortalities occurred from TB among HIV-negative people, and an additional 251,000 deaths were among HIV-positive people (33% of the total number of deaths caused by HIV/AIDS). Among those manifesting TB signs, World Health Organization (WHO) estimates, 1.43% people are infected with MBO - amounting to approximately 36 million MBO infections worldwide.

Transmission to humans

MBO can be transmitted among animals and humans via two main routes- direct and indirect. Direct transmission occurs when the bacterium from animals infects humans by direct inhalation (63, 64); whereas handling of contaminated meat or milk products causes cutaneous infections such as 'butcher's wart' (63). Indirect transmission occurs via consumption of infected milk, milk products (cheeses derived from cheeses prepared from unpasteurized milk), or meat products. The latter mode of transmission usually results in extrapulmonary forms (gastrointestinal or bone) of TB (50–53, 63, 65, 66).

Infectious animals shed MBO via milk, saliva, feces, and open granulomas. It is generally accepted that human beings get infected either by inhalation of bacteria-containing dust-particles and aerosols shed by infected animals or by ingestion of contaminated animal products

(e.g., raw milk). The main route of infection in cattle is through aerosol exposure, facilitated by close contact between animals. In cattle, ingestion of contaminated products (e.g., pasture and water) is generally considered to be a secondary, less important route of transmission. BTB is spread from cattle to humans primarily by contaminated milk consumed without pasteurizing it. In United Kingdom it was reported that pasteurization of milk helped control the transmission of MBO to humans. The review paper also stated that people who had professional exposure to infected animals such as those working in the abattoir or farms were more prone to getting infected than others (8, 54). Susceptible population also included old people and the immuno-compromised (53, 55, 56). In 2011 alone, the European Union reported that 132 cases of MBO were identified (57). Similarly, in developing countries major routes of transmission of MBO from animals to humans include consumption of infected meat and milk products, co-sharing of space with animal herds as well as inhalation of infected fecal dust that is released during preparation and burning of dung cakes (58–62).

Direct Transmission-Studies have shown that MBO can be transmitted directly from infected animals to humans. Wilkins et al., 2008 reported two cases of human pulmonary TB in Michigan, USA caused by deer or cattle MBO strains circulating in the area (63). Of the two cases, one was reported in 2002 in a 74-year-old US man, who likely contracted it from hunted deer carcass. The second case occurred in 2004 in a 29-year-old hunter who infected his finger while field dressing a white-tailed-deer. The deer had tan nodules in its chest cavity that were identified by the hunter. Sunstrum et al., (2019) reported the case of a 77-year-old man with pulmonary TB diagnosed in May 2017 (64). This patient had been a deer-hunter for over 20 years of his life and

resided in the Lower Peninsula of Michigan that is a known enzootic focus of MBO in free-ranging deer. In 2017 a patient in Spain was reported to have pulmonary TB, caused by multi-drug resistant MBO (67). Genomic analysis showed that the person had been infected from his TB infected sheep herd. A study by Otchere et al., (2019) showed that only 15 of the 1,755 human TB samples were infected with MBO, all showing pulmonary signs of infection(56). Direct contact with livestock or their unpasteurized products was significantly associated with BTB in infected humans ($p<0.0001$, OR = 124.4,95% CI = 30.1–508.3). Another report from UK by O'Connor et al., (2019) showed direct human transmission of MBO from household cats (68). In this study a cluster of 7 cats confirmed to be infected with MBO come in contact with their respective owners, amongst which two became infected with TB (latent). In a retrospective genomic analysis study, Wanzala et al., (2019) reported a case of MBO infection in a patient with pulmonary TB in Uganda (69).

Though this is not a comprehensive list of all human cases caused by direct transmission of MBO from animals, it does capture the idea that direct transmission is neither a one-time phenomenon nor is it specific to a geographical location.

Indirect transmission- Research has shown that exposure to bovine tuberculosis through consumption of contaminated milk causes extra-pulmonary Tuberculosis in humans (54). Rodwell et al., (2008) examined TB case surveillance data from 1994 till 2005 in San Diego county, USA(48). They calculated prevalence of TB caused by MBO to be 10%. Of those who had MBO TB, 54% were children. More than 96% of the people infected with MBO were of Hispanic ethnicity

and it was suggested that consumption of unpasteurized milk or milk products was the major source of MBO in these individuals.

Using Spacer oligonucleotide typing, or spoligotyping, (a rapid, polymerase chain reaction based method for genotyping strains of the MTBC), human MBO isolates (n = 106) from Southern California (50) revealed that over 91% (97/106) of the human MBO isolates carried genotypes that were identical to those in Mexican cattle. It was suggested that milk and milk products coming from Mexico needed to be quality controlled and illegal distribution of unpasteurized products such as queso fresco must be eliminated. Building upon Rodwell et al., (2008)'s work, Scott et al., (2016) published a descriptive analysis of human TB caused by MBO in USA. They reported similar trends as those published earlier. They concluded that children, foreign-born persons, Hispanics, and females are disproportionately affected by MBO, which was independently associated with extrapulmonary disease. Davidson et al., (2017) reported trends of MBO infection in humans in England, Wales and Northern Island from 2002 till 2014. In their report, consumption of unpasteurized milk was one of the risk factors associated with MBO infection(70).

Grandjean-Lapierre et al., (2018) diagnosed recurrent MBO epididymitis in a 63-year old Caucasian man without any signs of pulmonary or disseminated disease(71). The person had been exposed to camel milk consumption during extended visits in Niger before his clinical symptoms appeared. They concluded that isolated epididymis infection was a new manifestation of MBO disease in humans.

Prevalence of MBO in human and cattle population

According to WHO ~1.4% of human TB patients are infected with MBO (72). This value changes substantially depending on geographical location, prevalence of BTB in cattle population as well as lack of access to pasteurized milk and uninfected meat products. Rodwell et al., (2008) reported a 10% prevalence in California, USA(48). Whereas in industrialized countries like the United-kingdom 0.5% and 1.5% prevalence amongst all culture confirmed cases has been reported from 1990-2003 for infections caused by MBO (73).

In cattle the prevalence of MBO varies around the world and is subject to the control programs in each country. Some countries have eliminated the disease like Cook Islands, New Caledonia, French Polynesia; while some have been successful in drastically limiting the disease like USA, Canada, Australia, while others are struggling with high burden of BTB. In Pakistan the prevalence of the disease in buffaloes based on comparative intradermal tuberculin test is 10.06% (74). In yet another study the prevalence of BTB alone in buffaloes in Punjab (Pakistan), was 11.3 % (75). In Ethiopia, Tschopp et al., (2009) showed that the disease burden comparable to the scenario in Pakistan(76). According to them a 7.9% prevalence was recorded in the zebu (indigenous breed) cattle that are traditionally reared in the area for milk production.

Prevalence of MTB in animal population

Reverse zoonosis is the transmission of MTB from infected humans to animals. It is rare but not un-documented. Pavlik et al., (2003) reported all the cases of animal MTB from 1990 to 1999 in six Central European countries (Croatia, the Czech Republic, Hungary, Poland, Slovakia and Slovenia)(27). An estimated 11.8 million heads of cattle were analyzed. MTB was isolated from

tuberculous lesions of 16 animals in two countries (Poland and Slovak Republic) that included – 9 cattle, 4 swine, 3 African elephants, 1 Agouti, and 1 Tapir. Michel et al., (2013) investigated MTB cases in captive wild animal facilities in South Africa from 2002 - 2011(28). They used variable-number tandem repeats (VNTR) and IS*6110* restriction fragment length polymorphism (RFLP) typing to identify strain of MTB circulating in these wild animals. The study included cases from four hot spots- (i) the National Zoological Gardens of South Africa (NZG) (2002-2011), (ii) Johannesburg Zoo (2007 onwards) (iii) a rehabilitation center for vervet monkeys (*Chlorocebus pygerythrus*) (2008 onwards) and (iv) incidental cases in other facilities. Eleven different animal species were identified to be infected with MTB at NZG with each having a different VNTR pattern. The vervet monkeys were also shown to carry two different MTB strains. Apart from these, a sable antelope (*Hippotragus niger*), two chacma baboons (*Papio ursinus*) and a colony of capuchin monkeys (*Cebus capucinus*) were infected with MTB. The study suggested significant increase in TB cases at the NZG (upto 60%) between 2002 and 2011, as compared to the previous decade, due to a country wide increased spillover from humans to animals. In another study Pereira et al., (2018) reported the first diagnosed case of MTB in a captive Tapir *(Tapirus terrestris)* in the Brazilian Amazon region(30). Miller et al., (2015) reported TB in a 33-yr-old male black rhinoceros (*Diceros bicornis*) that died in 2013(31). They confirmed MTB infection in retrospectively collected and preserved serum samples using Elephant TB STAT-PAK, dual-path platform VetTB, and multi-antigen print immunoassay.

MTB is not a natural pathogen of the above-described animals, but it does cause infection naturally in humans, elephants, non-human primates, guineapigs, swine and dogs. Multiple

reports exist cataloging the infection in these natural hosts. Extensive work has been done on elephants where MTB was isolated from infected individuals(12, 14–16, 22–24, 77, 78). This poses threat to conservation of elephants as well as the health of the caretakers at conservation facilities. Such facilities not only have to be alert about MTB infecting the precious elephants but also ensure that the workers neither get infected nor transmit the infection back to elephants. This is because regular testing and administration of treatment is not only extremely expensive but also very difficult to sustain for 9 long months in these intelligent creatures.

Seo et al., (2017) reported a case of MTB infection in a domesticated Korean wild boar (*sus scrofa coreanus*)(79). Dogs are also prone to getting infected from their owners(80–84). Non-human primates (NHP) are no less susceptible than humans to MTB infection. There are reports of MTB infection in a female gorilla housed in a zoo in Nigeria(25) and in cynomolgus monkeys(26) brought in from Cambodia to be used in a transplantation research in a Korean facility. The NHPs also make good animal models for experimental pathogenesis and vaccination efficacy studies of MTB infection.

The section would not be complete without referring to the accidental cases of MTB in animals getting transmitted back to humans, and those of MBO in cattle to humans and back to cattle. Fritsche et al. (2004) reported the first case of TB in cattle, exposed to a 72-year old person infected with MBO as a child in Switzerland(85). On genotyping both isolates, the strain isolated in the cattle and the patient were identical. The history of the patient revealed that he had been exposed to MBO as a child when he worked at the farm. When in Dec 1999 he died and his animals were tested, 19 were positive for TB. In 2011, Posthaus et al., reported an accidental case

in Switzerland of three veterinary pathologists who got infected while performing necropsy on a dog infected with MTB(86). They were likely exposed to aerosols of MTB while using an electric saw to open the brain cavity of the dog at necropsy. Investigation revealed that the dog had not infected its owners during its lifetime, yet one of the two cats that he had lived with might have exposed him to MTB. Croatia reported its first case of MTB transmission from man to cattle in 2012(87). On genotyping the cattle and human isolates revealed identical Mycobacterial-interspersed-repetitive-units of VNTR (MIRU-VNTR) profiles.

MTBC has a complex epidemiology due to its wide range of hosts, virulence, and interspecies transmission cycle. The above compiled literature gives us a peak into the amazing adaptability of this insidious bacterium.

TB Epidemiology: Tools for tracking

Understanding the evolutionary patterns and virulence of a disease is interlinked with the impact it has on the public health policy. The insight helps improve and refine tools for disease diagnosis as well as drug therapy. MTBC is a slow growing, monoclonal complex of bacteria that does not undergo horizontal gene transfer. These traits prevent it from evolving as fast as other disease-causing bacteria. Studies have estimated the rate of mutation of MTB to be as slow as 0.3-0.5 mutations/genome/year (88–90). This allows us to use the mutations in the genome to determine phylogeny, as the mutations get passed down to next generation cells (generally) unaltered. To better understand the phylogenesis of MTBC, molecular techniques such as IS*6110* based IS*6110* RFLP, MIRU-VNTR, Spoligotyping, Large Sequence Polymorphism (LSP), whole genome sequencing (WGS) based single nucleotide polymorphisms (SNPs) are used. Strain typing

and differentiation studies have enabled definition of population genetic structures and transmission patterns of the illusive mycobacterium in both- its human and animal host.

In 2011, Comas and Gagneux expanded and explained the importance of systems epidemiology for effective control of TB (91, 92). The link between increasing human population (host density), changes in demographics of the world, as well as overcrowding (access to host) was reviewed. Comas & Gagneux hypothesized that these factors played a role in the evolution as well as increased virulence of mycobacterium that has led to a higher number of active human TB cases being caused by the more modern lineages of TB (lineage 2- East Asia, lineage 3- East African Indian, lineage 4-Euro-American). This is in comparison to the ancient lineages of TB (Lineage 1- Indo-oceanic, Lineage 5- *M. africanum* -West Africa 1, Lineage 6- *M. africanum*-West Africa 2 and Lineage 7-Ethiopia) that are less virulent leading to more latent cases of TB. Building upon this and through use of high-resolution genomic tools as well as using marginalized or confined communities as study subjects, we are in unique position to significantly modify our approach towards TB control.

Studies have been done to identify evidence of recent and ongoing TB transmission in marginalized human communities in Brazil. Cunha et al., (2014) assessed TB transmission pattern among indigenous people by MTB genotyping with IS*6110* RFLP(93). Of the 3,093 cases identified in Brazil from 1999-2001, 610 (~20%) were indigenous patients. Of the genotyped isolates from 52 indigenous patients, 33 (63.5%) belonged to cluster RFLP patterns, indicating recently transmitted TB. With the help of the molecular tool RFLP, Cunha et al., demonstrated high, on-going TB transmission rates among the indigenous people in Brazil.

Malla et al., (2012) showed that of the 261 human isolates genotyped, 11.5% of the humans infected with TB in Nepal, had the Indo-Oceanic lineage(94). Soon after, genotyping studies in elephant TB isolates were done to determine transmission events between humans and elephants. Interestingly, Paudel et al., (2014) genotyped five MTB isolates from Asian elephants of Nepal using spoligotyping, VNTR, and LSP(95). Their work revealed that all the isolates belonged to Indo-oceanic lineage. Paudel et al., (2019)(14) were also able to report the presence of mixed infection in two of the five elephants that similarly represented the main lineages circulating in adjacent human population as reported earlier by Malla et al., (2012).

In USA, Higgins (2011), reported analyzing 33 elephants infected with MTB between 1997-2010 using 48 of their isolates(96). 14 spoligotypes were identified from the isolates. Comparison with International Type Strains database showed that 11 of these 14 spoligotypes were identical to those circulating in the human population including Beijing, Euro-American, Latino-American and Mediterranean (LAM). The use of genotyping techniques not only deepened our understanding of human TB genomics, but they also provided strong evidence to support circulation of same MTB lineages in two different host population.

WGS has also been used intensively since 2013 in USA for typing, tracing, and evaluation of the bovine TB eradication program. Review by Orloski et al., (2018) based on WGS of cattle isolates infected with MBO from 1989 till 2018 gave insights into the effectiveness of the eradication program(97). Analysis of 1,248 isolates (cattle and farmed cervids herds) showed that nearly 70% of isolates were European clonal complex 1 and 30% were European clonal complex 2. Source of infection was identifiable for 49% of the cases. The study went further and described

SNP divergence as well as multiple possible sources of infection for the remaining 51% of the cases. Similarly, Salvador et al., (2019) explored the role of free roaming elk in transmission of MBO in Michigan(98). Their study was based on WGS of isolates from elk, deer and cattle showed clustering to occur between cattle and elk suggestive of same source of infection. They were able to deduce that elk was not acting as a reservoir for the disease at present, rather deer is the only source of inter-species transmission in Michigan.

The recent boom of molecular and genomic epidemiology has provided us with answers to previously un-answered questions on genetic structure and interspecies transmission events. Techniques such as RFLP, MIRU-VNTR and Spoligotyping are still the main typing method applicable in high burden TB countries because they are mostly under-developed and lack the resources to apply WGS extensively in their disease control programs. As the cost of WGS decreases, we will be able to apply this at levels previously un-approachable and thus improve eradication efforts.

TB Diagnostics: Screening Tests

A screening test or medical surveillance is defined as "a medical test or procedure performed on members (subjects) of a defined https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4389712/ -fn1 asymptomatic population or population subgroup to assess the likelihood of their members having a particular disease"(99). Examples of screening tests include a pap smear for cervical cancer(100, 101), mammography for breast cancer(102), pregnancy strip test for early detection of pregnancy. Ideally a screening test should detect early onset of disease with 100% sensitivity and specificity. The actual test's sensitivity and specificity vary with the prevalence of the disease

as well as the disease-specific modulation of the molecule(s) being targeted for detection. There are two kinds of screening tests based on a) direct detection of the agent/antigen or b) indirect detection of agent/antigens using the immune responses (such as antibodies or pathogen-specific responses of T-cells or cytokine production). For this review we will focus on screening tests available for direct detection of MBO in animals. These include:

1. Acid-fast staining and microscopy – Bacteria itself visualized
2. Nucleic acid amplification – Bacterial DNA amplified
3. Pathogen specific biomarkers – Bacterial fingerprint detected

Acid-fast staining & microscopy

The cell wall of mycobacteria contains a higher content of complex lipids (>60% as opposed to approximately 5% and 20% in Gram-positive and Gram-negative organisms respectively) including long chain (C_{60}—C_{90}) fatty acids called mycolic acids(103, 104). Mycolic acids make the cell wall extremely hydrophobic and enhance resistance to desiccation, killing by disinfectants, penetration by many of the drugs that are used to treat other infections and staining with basic aniline dyes. Therefore, OIE suggests application of Ziehl-Neelsen method to achieve staining(105). The biggest advantage of this method is direct visualization of the bacteria. Yet this method does not differentiate between Non-MTBC such as *Nocardia*(106) likely leading to false positive results. Paucibacillary infections on the other hand may lead to false negatives. This method is done in parallel to bacterial culturing and acid-fast stain from pure cultures.

The ante-mortem clinical samples such as sputum, bronchial aspirates, bronchoalveolar lavage fluid as well as lymph node biopsies are extremely difficult to obtain in animals. Instead,

blood, urine, milk and feces are a lot easier to collect but the low mycobacterial count in these samples along with technical expertise required to concentrate the bacteria before staining is time consuming and expensive.

Nucleic acid amplification

This technique targets the MTBC specific genes such as IS*6110*, IS*1081* or genes encoding MTBC specific proteins such as MPB70, 38KDa antigen(105). The amplification requires bacterial DNA to remain intact even after disinfection of sample, concentration of the bacteria, rigorous breakdown of the bacterial cell-wall via zirconium bead beating and high-speed centrifugation. If the sample has low bacterial load, or if PCR inhibitors are present, a TB infected animal will pass the screening test undetected as a false negative. WHO DNA amplification on the GeneXpert platform for TB has not been adequately tested for MBO detection. In addition, MBO carries 0-3 copies of IS*6110* making it difficult to get adequate amplification by the GeneXpert probes. To detect the presence of active or latent TB amplification of DNA could be employed, but infection that is clearing out on its own would lead to false positives.

Pathogen specific biomarkers (PSB)

PSB are entities released specifically by live bacteria during the infection cycle. MTBC organisms release highly specific entities including: ESAT-6, CFP-10, MPB70, MPB83, Pks5, Mb2515c, MB1895c and several others as reported by Lamont et al., (2014)(107). PSBs are ideal as screening molecules. Chapter 6 and 7 discusses in detail the PSBs that are being targeted for detection. Using a combination of biomarkers such as CFP-10 and ESAT-6 increases specificity as well as sensitivity. Modulation of PSBs released during pathogen-host interaction suggests that

detection systems based on PSBs can identify different stages of infection as well as serve to monitor response to therapy. This allows for the creation of a screening tool suitable for field diagnosis with high sensitivity as well as specificity and facilitates the early diagnosis of infection in a herd that is critical for control of diseases in animal populations. While specific to MTBC, PSBs do not differentiate between MTB and MBO. Further these molecules are present in minute quantities as compared to albumin levels in serum. Use of concentration techniques along with removal of high abundance proteins in blood can alleviate the problem and allow for the creation of an ideal screening tool for field diagnosis of TB.

Conclusion

Screening tools for TB that can directly detect the causative agent especially under field conditions could prove to be revolutionary in the control of TB around the world. Use of smear microscopy, detection of nucleic acids or PSB in samples requires not just collection of the sample, but also transportation under special conditions, technical expertise as well as expensive lab equipment. This thesis focuses on not just identification of PSB, but their validation as well as application for the development of a screening tool that could be used under field conditions with high sensitivity and specificity. Available tools for screening have multiple drawbacks, but with current progress in technology these can be overcome. By combining the use of highly sophisticated molecular tools of genotyping we are able to enhance our understanding of TB transmission as well as evolution under isolated conditions and communities. Spoligotyping, RFLP, MIRU-VNTR and most importantly WGS are incredible tools that can be put to use to

construct not only the clustering pattern of circulating bacterial strains, but it also facilitates detection of infection sources as well as potential genomic causes of modification in virulence.

REFERENCES

REFERENCES

1. Cambau E, Drancourt M. 2014. Steps towards the discovery of Mycobacterium tuberculosis by Robert Koch, 1882. Clin Microbiol Infect. Blackwell Publishing Ltd.

2. Mulholland C V., Shockey AC, Aung HL, Cursons RT, O'Toole RF, Gautam SS, Brites D, Gagneux S, Roberts SA, Karalus N, Cook GM, Pepperell CS, Arcus VL. 2019. Dispersal of Mycobacterium tuberculosis Driven by Historical European Trade in the South Pacific. Front Microbiol https://doi.org/10.3389/fmicb.2019.02778.

3. Barberis I, Bragazzi NL, Galluzzo L, Martini M. 2017. The history of tuberculosis: From the first historical records to the isolation of Koch's bacillus. J Prev Med Hyg.

4. Daniel TM. 2006. The history of tuberculosis. Respir Med 100.

5. Fellag M, Loukil A, Drancourt M. 2021. The puzzle of the evolutionary natural history of tuberculosis. New Microbes New Infect.

6. The famous people who died of tuberculosis.

7. LoBue PA, Enarson DA, Thoen CO. 2010. Tuberculosis in humans and animals: An overview. Int J Tuberc Lung Dis.

8. Rua-domenech R De, Goodchild AT, Vordermeier HM, Hewinson RG. 2006. Ante mortem diagnosis of tuberculosis in cattle : A review of the tuberculin tests , c -interferon assay and other ancillary diagnostic techniques 81:190–210.

9. Brites D, Loiseau C, Menardo F, Borrell S, Boniotti MB, Warren R, Dippenaar A, Parsons SDC, Beisel C, Behr MA, Fyfe JA, Coscolla M, Gagneux S. 2018. A new phylogenetic framework for the animal-adapted mycobacterium tuberculosis complex. Front Microbiol 9.

10. Kerr TJ, de Waal CR, Buss PE, Hofmeyr J, Lyashchenko KP, Miller MA. 2019. Seroprevalence of mycobacterium tuberculosis complex in free-ranging african elephants (Loxodonta africana) in Kruger national park, South Africa. J Wildl Dis 55.

11. Songthammanuphap S, Puthong S, Pongma C, Buakeaw A, Prammananan T, Warit S, Tipkantha W, Kaewkhunjob E, Yindeeyoungyeon W, Palaga T. 2020. Detection of Mycobacterium tuberculosis complex infection in Asian elephants (Elephas maximus) using an interferon gamma release assay in a captive elephant herd. Sci Rep 10.

12. Larsen RS, Salman MD, Mikota SK, Isaza R, Montali RJ, Triantis J. 2000. Evaluation of a

multiple-antigen enzyme-linked immunosorbent assay for detection of Mycobacterium tuberculosis infection in captive elephants. J Zoo Wildl Med 31:291–302.

13. Mikota SK, Peddie L, Peddie J, Isaza R, Dunker F, West G, Lindsay W, Larsen RS, Salman MD, Chatterjee D, Payeur J, Whipple D, Thoen C, Davis DS, Sedgwick C, Montali RJ, Ziccardi M, Maslow J. 2001. Epidemiology and diagnosis of Mycobacterium tuberculosis in captive Asian elephants (Elephas maximus). J Zoo Wildl Med 32:1–16.

14. Paudel S, Nakajima C, Mikota SK, Gairhe KP, Maharjan B, Subedi S, Poudel A, Sashika M, Shimozuru M, Suzuki Y, Tsubota T. 2019. Mixed mycobacterium tuberculosis lineage infection in 2 elephants, Nepal. Emerg Infect Dis 25:1031–1032.

15. Lyashchenko KP, Greenwald R, Esfandiari J, Mikota S, Miller M, Moller T, Vogelnest L, Gairhe KP, Robbe-Austerman S, Gai J, Waters WR. 2012. Field application of serodiagnostics to identify elephants with tuberculosis prior to case confirmation by culture. Clin Vaccine Immunol 19.

16. Mikota SK, Lyashchenko KP, Lowenstine L, Agnew D, Maslow JN. 2015. Mycobacterial infections in elephants.Tuberculosis, leprosy and mycobacterial diseases of man and animals: the many hosts of mycobacteria.

17. Paudel S, Mikota SK, Thapa J, Lyashchenko KP, Gairhe KP, Dhakal IP, Subedi N, Maharjan B, Subedi S, Kaufman GE, Tsubota T. 2018. Serodiagnosis of elephant tuberculosis: a useful tool for early identification of infected elephants at the captive-wild interface. Eur J Wildl Res 64.

18. Paudel S, Sreevatsan S. 2020. Tuberculosis in elephants: Origins and evidence of interspecies transmission. Tuberculosis.

19. Sarad P, TSUBOTA T. 2016. Tuberculosis in Elephants: A Zoonotic Disease at the Human-Elephant Interface. Japanese J Zoo Wildl Med 21.

20. Lassausaie J, Bret A, Bouapao X, Chanthavong V, Castonguay-Vanier J, Quet F, Mikota SK, Théorêt C, Buisson Y, Bouchard B. 2015. Tuberculosis in Laos, who is at risk: The mahouts or their elephants? Epidemiol Infect 143.

21. Lainé N. 2018. Elephant tuberculosis as a reverse zoonosis. Med Anthropol Theory 5.

22. Lyashchenko KP, Greenwald R, Esfandiari J, Olsen JH, Ball R, Dumonceaux G, Dunker F, Buckley C, Richard M, Murray S, Payeur JB, Andersen P, Pollock JM, Mikota S, Miller M, Sofranko D, Waters WR. 2006. Tuberculosis in elephants: Antibody responses to defined antigens of mycobacterium tuberculosis, potential for early diagnosis, and monitoring of treatment. Clin Vaccine Immunol 13.

23. Miller MA, Finnegan M, Storms T, Garner M, Lyashchenko KP. 2018. Outbreak of mycobacterium tuberculosis in a herd of captive asian elephants (elephas maximus): Antemortem diagnosis, treatment, and lessons learned. J Zoo Wildl Med 49:748–754.

24. Zachariah A, Pandiyan J, Madhavilatha GK, Mundayoor S, Chandramohan B, Sajesh PK, Santhosh S, Mikota SK. 2017. Mtb in wild asian elephants EID 2017 23.

25. Adeogun A, Omobowale O, Owuamanam C, Alaka O, Taiwo V, Van Soolingen D, Cadmus S. 2016. Mycobacterium tuberculosis and Dual M. tuberculosis/M. bovis infection as the cause of tuberculosis in a gorilla and a lioness, respectively, in Ibadan Zoo, Nigeria. Case Reports Vet Med 2016.

26. Choi EW, Lee KW, Kim TM, Park H, Jeon MR, Cho CW, Park JB, Kim S. 2016. Mycobacterium tuberculosis infections in cynomolgus monkey transplant recipients and institution of a screening program for the prevention and control of tuberculosis. BMC Vet Res 12.

27. Pavlik I, Ayele WY, Parmova I, Melicharek I, Hanzlikova M, Svejnochova M, Körmendy B, Nagy G, Cvetnic Z, Katalinic-Jankovic V, Ocepek M, Zolnir-Dovc M, Lipiec M, Havelkova M. 2003. Mycobacterium tuberculosis in animal and human populations in six Central European countries during 1990-1999. Vet Med (Praha) 48.

28. Michel AL, Hlokwe TM, Espie IW, van Zijll Langhout M, Koeppel K, Lane E. 2013. Mycobacterium tuberculosis at the Human/Wildlife interface in a high TB Burden Country. Transbound Emerg Dis 60:46–52.

29. Hlokwe TM, Said H, Gcebe N. 2017. Mycobacterium tuberculosis infection in cattle from the Eastern Cape Province of South Africa. BMC Vet Res 13.

30. Pereira WLA, Bernal MKM, Campos CED, de Arimatéia Freitas J, Rodrigues YC, Lima KVB, Cecim ADA, de Souza AJS. 2018. Tuberculosis caused by mycobacterium tuberculosis complex in a captive tapir (Tapirus terrestris). Acta Sci Vet 46.

31. Miller MA, Greenwald R, Lyashchenko KP. 2015. Potential for serodiagnosis of tuberculosis in black rhinoceros (Diceros bicornis). J Zoo Wildl Med 46:100–104.

32. Alves AS, Araújo SB, Ferreira JS, Medeiros LKG, Oliveira EL, Dantas AFM, Souza AP. 2019. Clinical and therapeutic aspects of infection by Mycobacterium bovis in domestic cat-Case report. Arq Bras Med Vet e Zootec 71.

33. Firmino MO, Frade MTS, Ferreira JS, Alves AS, Ikuta CY, Neto JSF, Souza AP, Dantas AFM. 2018. Mycobacteriae diagnosed in domestic cats in the backland of Paraiba, Brazil. Pesqui Vet Bras 38.

34. Thorel MF, Karoui C, Varnerot A, Fleury C, Vincent V. 1998. Isolation of Mycobacterium

bovis from baboons, leopards and a sea-lion. Vet Res 29.

35. Barandiaran S, Pérez AM, Gioffré AK, Martínez Vivot M, Cataldi AA, Zumárraga MJ. 2015. Tuberculosis in swine co-infected with Mycobacterium avium subsp. Hominissuis and Mycobacterium bovis in a cluster from Argentina. Epidemiol Infect 143.

36. Bernitz N, Kerr TJ, Goosen WJ, Chileshe J, Higgitt RL, Roos EO, Meiring C, Gumbo R, de Waal C, Clarke C, Smith K, Goldswain S, Sylvester TT, Kleynhans L, Dippenaar A, Buss PE, Cooper D V., Lyashchenko KP, Warren RM, van Helden PD, Parsons SDC, Miller MA. 2021. Review of Diagnostic Tests for Detection of Mycobacterium bovis Infection in South African Wildlife. Front Vet Sci. Frontiers Media S.A.

37. Bryan J, den Boon P, McGuirk J, Madigan G, Skuce R, Fogarty U. 2018. Tuberculosis caused by Mycobacterium bovis infection in a donkey. Equine Vet Educ 30.

38. Esple IW, Hlokwe TM, Van Gey Pittius NC, Lane E, Tordiffe ASW, Michel AL, Müller A, Kotze A, Van Helden PD. 2009. Pulmonary infection due to mycobacterium bovis in a black rhinoceros (Diceros bicomis minor) in South Africa. J Wildl Dis 45.

39. Miller M, Michel A, van Helden P, Buss P. 2017. Tuberculosis in Rhinoceros: An Underrecognized Threat? Transbound Emerg Dis.

40. García-Bocanegra I, Barranco I, Rodríguez-Gómez IM, Pérez B, Gómez-Laguna J, Rodríguez S, Ruiz-Villamayor E, Perea A. 2010. Tuberculosis in alpacas (Lama pacos) caused by Mycobacterium bovis. J Clin Microbiol 48.

41. Ku BK, Jeon BY, Kim JM, Jang YB, Jang Y, Yu SY, Kim J, Moon OK, Jung SC, Lee MK, Jeong TN. 2016. Mycobacterium bovis infection in a wild sow (Sus scrofa): first case in Korea. J Vet Sci 17.

42. Lee H, Kim JM, Jang Y, Lee K, Baek K, Lee B, Kim HY, Lee MH, Ryoo S, Bae YC, Choi EJ, So BJ. 2015. Bovine tuberculosis in an Asian small-clawed otter (Aonyx cinerea) in the Republic of Korea. J Vet Diagnostic Investig 27.

43. Hlokwe TM, Sutton D, Page P, Michel AL. 2016. Isolation and molecular characterization of Mycobacterium bovis causing pulmonary tuberculosis and epistaxis in a Thoroughbred horse. BMC Vet Res 12.

44. Di Blasio A, Varello K, Vitale N, Irico L, Bozzetta E, Goria M, Chiavacci L, Zoppi S, Dondo A. 2019. Animal tuberculosis in a free-ranging fallow deer in northwest Italy: a case of “lucky strain survival” or multi-host epidemiological system complexity? Eur J Wildl Res 65.

45. Portillo-Gómez L, Sosa-Iglesias EG. 2011. Molecular identification of Mycobacterium bovis and the importance of zoonotic tuberculosis in Mexican patients. Int J Tuberc Lung Dis

15:1409–1414.

46. JAJOU R, Soolingen DV. 2019. Whole genome sequencing of Mycobacterium tuberculosis complex for species identification, drug susceptibility testing, and epidemiological typingPediatrics.

47. Waters WR, Thacker TC, Nelson JT, DiCarlo DM, Maggioli MF, Greenwald R, Esfandiari J, Lyashchenko KP, Palmer M V. 2014. Virulence of two strains of mycobacterium bovis in cattle following aerosol infection. J Comp Pathol 151:410–419.

48. Rodwell TC, Moore M, Moser KS, Brodine SK, Strathdee SA. 2008. Tuberculosis from Mycobacterium bovis in binational communities, United States. Emerg Infect Dis 14:909–916.

49. Schwartz D. 1980. Epidemiology of human fecundity. Rev Epidemiol Sante Publique 28:7–12.

50. Rodwell TC, Kapasi AJ, Moore M, Milian-Suazo F, Harris B, Guerrero LP, Moser K, Strathdee SA, Garfein RS. 2010. Tracing the origins of Mycobacterium bovis tuberculosis in humans in the USA to cattle in Mexico using spoligotyping. Int J Infect Dis 14:129–135.

51. Tippett SG. 1948. Bovine TuberculosisBritish Medical Journal.

52. Scott C, Cavanaugh JS, Pratt R, Silk BJ, LoBue P, Moonan PK. 2016. Human Tuberculosis Caused by Mycobacterium bovis in the United States, 2006-2013. Clin Infect Dis 63:594–601.

53. Hlavsa MC, Moonan PK, Cowan LS, Navin TR, Kammerer JS, Morlock GP, Crawford JT, LoBue PA. 2008. Human Tuberculosis due to Mycobacterium bovis in the United States, 1995–2005 . Clin Infect Dis 47:168–175.

54. Thoen CO, Steele JH, Gilsdorf MJ. 2008. Mycobacterium Bovis Infection in Animals and Humans: Second EditionMycobacterium Bovis Infection in Animals and Humans: Second Edition.

55. Provider H, Control D, Use I, Robinette C, Saffran L, Ruple A, Deem SLL, Rodriguez-Campos S, Schürch AC, Dale J, Lohan AJ, Cunha M V., Botelho A, Cruz K De, Boschiroli ML, Boniotti MB, Pacciarini M, Garcia-Pelayo MC, Romero B, de Juan L, Domínguez L, Gordon S V., Van Soolingen D, Loftus B, Berg S, Hewinson RG, Aranaz A, Smith NH, Kaneene JB, Miller R, De Kantor IN, Thoen CO, Kaneene JB, Che-Amat A, Ong BL, Kendall EA, Fofana MO, Dowdy DW, Javed T, Lucy A, Aït-Khaled N, Alarcón E, Armengol R, Bissell K, Boillot F, Caminero JA, Chen-Yuan C, Clevenbergh P, Dlodlo R, Enarson DA, Enarson P, Fujiwara PI, Harries AD, Sven H, Hinderaker G, Lienhardt C, Monedero I, Rieder HL, Rusen ID, Trébucq A, Deun A Van, Wilson N, Vogelnest L, Thompson P, Tovey ER, Of PWD, Grzegorzewicz AE, de Sousa-

d'Auria C, McNeil MR, Huc-Claustre E, Jones V, Petit C, Angala SK, Zemanová J, Wang Q, Belardinelli JM, Gao Q, Ishizaki Y, Mikušová K, Brennan PJ, Ronning DR, Chami M, Houssin C, Jackson M, Smith NH, luna, Wu X, Lu W, Shao Y, Liu Q, Shi X, Wang X, Li C, Zhu L, Chen C, Schürch AC, Van Soolingen D, Jackson M, McNeil MR, Brennan PJ, Berg S, Garcia-Pelayo MC, Müller B, Hailu E, Asiimwe B, Kremer K, Dale J, Boniotti MB, Rodriguez S, Hilty M, Rigouts L, Firdessa R, MacHado A, Mucavele C, Ngandolo BNR, Bruchfeld J, Boschiroli L, Müller A, Sahraoui N, Pacciarini M, Cadmus S, Joloba M, Van Soolingen D, Michel AL, Djønne B, Aranaz A, Zinsstag J, Van Helden P, Portaels F, Kazwala R, Källenius G, Hewinson RG, Aseffa A, Gordon S V., Smith NH, Alva A, Aquino F, Gilman RH, Olivares C, Requena D, Gutiérrez AH, Caviedes L, Coronel J, Larson S, Sheen P, Moore DAJ, Zimic M, Lawal JR, Jajere SM, Ibrahim UI, Geidam YA, Gulani IA, Musa G, Ibekwe BU, Collins MT, Madison W, Romha G, Gebru G, Asefa A, Mamo G, Jasenosky LD, Scriba TJ, Hanekom WA, Goldfeld AE, LS BL, 2001 undefined, Interferon- M, Genesusceptibility R, Mycobacterial TO, Krajewska-Wędzina M, Augustynowicz-Kopeć E, Weiner M, Szulowski K, Asante-Poku A, Yeboah-Manu D, Otchere ID, Aboagye SY, Stucki D, Hattendorf J, Borrell S, Feldmann J, Danso E, Gagneux S, Wu Q, Chen M, Buchwald M, Phillips RA, Lainé N, Vitale F, Capra G, Maxia L, Reale S, Vesco G, Caracappa S, Ravimohan S, Kornfeld H, Weissman D, Bisson GP, Reynolds J, Moyes RB, Breakwell DP, Mahara G, Yang K, Chen S, Wang W, Guo X, Claus. D. Buergelt, DVM P, Elliot Williams B, Department of Pathobiology, College of Veterinary Medicine, University of Florida, Gainesville F, Tevere VJ, Hewitt PL, Dare A, Hocknell P, Keen A, Spadoro JP, Young KK, Zink AR, Sola C, Reischl U, Grabner W, Rastogi N, Wolf H, Nerlich AG, Robinette C, Saffran L, Ruple A, Deem SLL, Beaunée G, Vergu E, Joly A, Ezanno P, Maurelli AT, Fernández RE, Bloch CA, Rode CK, Fasano A, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C, Simon S, Harris B, Atkin R, Doggett J, Mayes R, Keating L, Wheeler PR, Parkhill J, Barrell BG, Cole ST, Gordon S V., Hewinson RG, Lieb B, Nuru A, Mamo G, Worku A, Admasu A, Medhin G, Pieper R, Ameni G, Cabibbe AM, Walker TM, Niemann S, Cirillo DM, Snow KJ, Nelson LJ, Sismanidis C, Sawyer SM, Graham SM, American Academy of Family Physicians. AF, Bannon M, Rittenhouse S, Oommen S, Banaji N, Tb H, Zhang YJ, Zhang S, Liu XZ, Wen HA, Wang M, Chen ZW, Cardona P-J, Neelsen F, Zaman K, Hossain S, Arifeen SE, Mahmud A, Begum V, Cook EAJ, de Glanville WA, Thomas LF, Kariuki S, Bronsvoort BM de C, Fèvre EM, Jagielski T, Minias A, van Ingen J, Rastogi N, Brzostek A, Żaczek A, Dziadek J, Cooper R. 2018. Technical Briefing : Tuberculosis as a zoonotic disease. Ann Agric Environ Med 13:1–79.

56. Otchere ID, van Tonder AJ, Asante-Poku A, Sánchez-Busó L, Coscollá M, Osei-Wusu S, Asare P, Aboagye SY, Ekuban SA, Yahayah AI, Forson A, Baddoo A, Laryea C, Parkhill J, Harris SR, Gagneux S, Yeboah-Manu D. 2019. Molecular epidemiology and whole genome sequencing analysis of clinical Mycobacterium bovis from Ghana. PLoS One 14.

57. 2013. The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2011. EFSA J 11.

58. Khattak I, Mushtaq MH, Ahmad M ud D, Khan MS, chaudhry M, Sadique U. 2016. Risk factors associated with Mycobacterium bovis skin positivity in cattle and buffalo in Peshawar, Pakistan. Trop Anim Health Prod 48.

59. Ullah A, Khattak US, Ayaz S, Qureshi MS, Khan I, Jan IU, Khattak I, Taj R, Nigar S, Khan NU, Khan MA, Sohail ML. 2019. Bovine tuberculosis (bTB): Prevalence and associated risk factors in large ruminants in the central zone of Khyber Pakhtunkhwa, Pakistan. Pak J Zool 51.

60. Akhtar F, Javed MT, Aziz-ur-Rehman, Khan MN, Akhtar P, Hussain SM, Aslam MS, Kausar R, Qamar M, Cagiola M. 2015. The use of PCR technique in the identification of Mycobacterium species responsible for bovine tuberculosis in cattle and buffaloes in Pakistan. Trop Anim Health Prod 47.

61. Khattak I, Mushtaq MH, Ahmad MUD, Khan MS, Haider J. 2016. Zoonotic tuberculosis in occupationally exposed groups in Pakistan. Occup Med (Chic Ill) 66.

62. Ullah A, Sadique U, Ayaz S, Qureshi MS, Khan FA. 2020. Bovine tuberculosis (bTB)-isolation and species-specific identification of mycobacterium bovis from bovine raw milk in Pakistan. Sarhad J Agric 36.

63. Wilkins MJ, Meyerson J, Bartlett PC, Spieldenner SL, Berry DE, Mosher LB, Kaneene JB, Robinson-dunn B, Stobierski MG, Boulton ML. 2008. Mycobacterium bovis Infection and Bovine 14:657–660.

64. Sunstrum J, Shoyinka A, Power LE, Maxwell D, Stobierski MG, Signs K, Sidge JL, Brien DJO, Robbe-Austerman S, Davisdson P. 2019. Notes from the Field- Zoonotic Mycobacterium bovis DIsease in Deer Huntees-Michigan, 2002-2017. Centers Dis Control Prev 68:807–808.

65. Torres-gonzalez P, Cervera-hernandez ME, Martinez-gamboa A, Garcia-garcia L, Cruz-hervert LP, Valle MB, Leon AP, Sifuentes-osornio J. 2016. Human tuberculosis caused by Mycobacterium bovis : a retrospective comparison with Mycobacterium tuberculosis in a Mexican tertiary care centre , 2000 – 2015. BMC Infect Dis 1–9.

66. WHO. 2018. Global Tuberculosis ReportGeneva: World Health Organization.

67. De Val BP, Romero B, Tórtola MT, León LH, Pozo P, Mercader I, Sáez JL, Domingo M, Vidal E. 2021. Polyresistant mycobacterium bovis infection in human and sympatric sheep, spain, 2017-2018. Emerg Infect Dis 27.

68. O'Connor CM, Abid M, Walsh AL, Behbod B, Roberts T, Booth L V., Thomas HL, Smith NH, Palkopoulou E, Dale J, Nunez-Garcia J, Morgan D. 2019. Cat-to-human transmission of Mycobacterium bovis, United Kingdom. Emerg Infect Dis 25.

69. Wanzala SI, Nakavuma J, Travis D, Kia P, Ogwang S, Waters WR, Thacker T, Johnson T, Hadi SA, Sreevatsan S. 2019. Retrospective analysis of archived pyrazinamide resistant Mycobacterium tuberculosis complex isolates from Uganda—evidence of interspecies transmission. Microorganisms 7.

70. Davidson JA, Loutet MG, O'Connor C, Kearns C, Smith RMM, Lalor MK, Thomas HL, Abubakar I, Zenner D. 2017. Epidemiology of mycobacterium bovis disease in humans in England, Wales, and Northern Ireland, 2002–2014. Emerg Infect Dis 23.

71. Grandjean-Lapierre S, Thiberville SD, Fellag M, Eghazarian C, Bouzid F, Gavril C, Drancourt M. 2018. Recurrent bilateral Mycobacterium bovis necrotizing epididymitis: A case report. BMC Res Notes 11.

72. 2019. Global Tuberculosis Report, Executive Summary.

73. De La Rua-Domenech R. 2006. Human Mycobacterium bovis infection in the United Kingdom: Incidence, risks, control measures and review of the zoonotic aspects of bovine tuberculosis. Tuberculosis 86:77–109.

74. Khan IA, Khan A, Vocational S, Sciences A, Ali A. 2008. Factors Affecting Prevalence of Bovine Tuberculosis in Nili Ravi Buffaloes. Pak Vet J 28:155–158.

75. Javed MT, Ahmad L, Feliziani F, Pasquali P, Akhtar M, Usman M, Irfan M, Severi G, Cagiola M. 2012. Analysis of some of the epidemiological risk factors affecting the prevalence of tuberculosis in buffalo at seven livestock farms in Punjab Pakistan. Asian Biomed 6:35–42.

76. Tschopp R, Schelling E, Hattendorf J, Aseffa A, Zinsstag J. 2009. Risk factors of bovine tuberculosis in cattle in rural livestock production systems of Ethiopia. Prev Vet Med 89:205–211.

77. Sternberg Lewerin S, Olsson SL, Eld K, Röken B, Ghebremichael S, Koivula T, Källenius G, Bölske G. 2005. Outbreak of Mycobacterium tuberculosis infection among captive Asian elephants in a Swedish zoo. Vet Rec 156.

78. Ghielmetti G, Coscolla M, Ruetten M, Friedel U, Loiseau C, Feldmann J, Steinmetz HW, Stucki D, Gagneux S. 2017. Tuberculosis in Swiss captive Asian elephants: Microevolution of Mycobacterium tuberculosis characterized by multilocus variable-number tandem-repeat analysis and whole-genome sequencing. Sci Rep 7.

79. Seo MG, Ouh IO, Kim M, Lee J, Kim YH, Do JC, Kwak D. 2017. Mycobacterium tuberculosis infection in a domesticated Korean wild boar (sus scrofa coreanus). J Food Prot 80.

80. Moravkova M, Slany M, Trcka I, Havelkova M, Svobodova J, Skoric M, Heinigeova B, Pavlik I. 2011. Human-to-human and human-to-dog Mycobacterium tuberculosis transmission

studied by IS6110 RFLP analysis: A case report. Vet Med (Praha) 56:314–317.

81. Mentula S, Karkamo V, Skrzypczak T, Seppänen J, Hyyryläinen H-L, Haanperä M, Soini H. 2020. Emerging source of infection – Mycobacterium tuberculosis in rescue dogs: a case report. Access Microbiol 2.

82. Szaluś-Jordanow O, Augustynowicz-Kopeć E, Czopowicz M, Olkowski A, Łobaczewski A, Rzewuska M, Sapierzyñski R, Wiatr E, Garncarz M, Frymus T. 2016. Intracardiac tuberculomas caused by Mycobacterium tuberculosis in a dog. BMC Vet Res 12.

83. Vangone L, Cardillo L, Riccardi MG, Borriello G, Cerrone A, Coppa P, Scialla R, Sannino E, Miletti G, Galiero G, Fusco G. 2021. Mycobacterium tuberculosis SIT42 Infection in an Abused Dog in Southern Italy. Front Vet Sci 8.

84. de Sousa DER, Wilson TM, Machado M, Guerra JM, Da Costa DEM, Colling A, de Castro MB. 2018. Disseminated tuberculosis in a dog: A putative zooanthroponosis. Acta Sci Vet 46.

85. Fritsche A, Engel R, Buhl D, Zellweger JP. 2004. Mycobacterium bovis tuberculosis: From animal to man and back. Int J Tuberc Lung Dis 8:903–904.

86. Posthaus H, Bodmer T, Alves L, Oevermann A, Schiller I, Rhodes SG, Zimmerli S. 2011. Accidental infection of veterinary personnel with Mycobacterium tuberculosis at necropsy: A case study. Vet Microbiol 149:374–380.

87. . 2012. Molecular epidemiology of Mycobacterium tuberculosis transmission between cattle and man - A case report. Vet Arh 82.

88. Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, Johnston JC, Gardy J, Lipsitch M, Fortune SM. 2013. Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. Nat Genet 45.

89. Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TEA. 2013. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: A retrospective observational study. Lancet Infect Dis 13.

90. Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, Mohaideen N, Ioerger TR, Sacchettini JC, Lipsitch M, Flynn JL, Fortune SM. 2011. Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection. Nat Genet 43.

91. Comas I, Gagneux S. 2011. A role for systems epidemiology in tuberculosis research. Trends Microbiol.

92. Gagneux S. 2018. Ecology and evolution of Mycobacterium tuberculosis. Nat Rev Microbiol.

93. Cunha EA, Ferrazoli L, Riley LW, Basta PC, Honer MR, Maia R, da Costa IP. 2014. Incidence and transmission patterns of tuberculosis among indigenous populations in Brazil. Mem Inst Oswaldo Cruz https://doi.org/10.1590/0074-0276130082.

94. Malla B, Stucki D, Borrell S, Feldmann J, Maharjan B, Shrestha B, Fenner L, Gagneux S. 2012. First Insights into the Phylogenetic Diversity of Mycobacterium tuberculosis in Nepal. PLoS One 7.

95. Paudel S, Mikota SK, Nakajima C, Gairhe KP, Maharjan B, Thapa J, Poudel A, Shimozuru M, Suzuki Y, Tsubota T. 2014. Molecular characterization of Mycobacterium tuberculosis isolates from elephants of Nepal. Tuberculosis 94.

96. J. Higgins. 2011. Genotyping of Mycobacterium tuberculosis cultured from elephants. Tuberculosis in elephants: science, myths, and beyond. Kansas city, Missouri.

97. Orloski K, Robbe-Austerman S, Stuber T, Hench B, Schoenbaum M. 2018. Whole genome sequencing of Mycobacterium bovis isolated from livestock in the United States, 1989-2018. Front Vet Sci 5.

98. Salvador LCM, O'Brien DJ, Cosgrove MK, Stuber TP, Schooley AM, Crispell J, Church S V., Gröhn YT, Robbe-Austerman S, Kao RR. 2019. Disease management at the wildlife-livestock interface: Using whole-genome sequencing to study the role of elk in Mycobacterium bovis transmission in Michigan, USA. Mol Ecol 28.

99. Maxim LD, Niebo R, Utell MJ. 2014. Screening tests: A review with examples. Inhal Toxicol 26:811–828.

100. Arbyn M, Sankaranarayanan R, Muwonge R, Keita N, Dolo A, Mbalawa CG, Nouhou H, Sakande B, Wesley R, Somanathan T, Sharma A, Shastri S, Basu P. 2008. Pooled analysis of the accuracy of five cervical cancer screening tests assessed in eleven studies in Africa and India. Int J Cancer 123:153–160.

101. Mayrand M-H, Duarte-Franco E, Rodrigues I, Walter SD, Hanley J, Ferenczy A, Ratnam S, Coutlée F, Franco EL. 2007. Human Papillomavirus DNA versus Papanicolaou Screening Tests for Cervical Cancer. N Engl J Med 357:1579–1588.

102. Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS, Hayes MK, Copit DS, Carlson KL, Cink TM, Barke LD, Greer LN, Miller DP, Conant EF. 2014. Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA - J Am Med Assoc 311:2499–2507.

103. 2014. Structure, Biosynthesis, and Genetics of the Mycolic Acid-Arabinogalactan-Peptidoglycan ComplexTuberculosis and the Tubercle Bacillus.

104. Asselineau C, Asselineau J, Lanéelle G, Lanéelle MA. 2002. The biosynthesis of mycolic acids by Mycobacteria: Current and alternative hypotheses. Prog Lipid Res.

105. OIE. 2018. Bovine tuberculosis. In Manual of Diagnostic Tests and Vaccines for Terrestrial AnimalsOIE Terrestrial Manual.

106. Malama S, Munyeme M, Mwanza S, Muma JB. 2014. Isolation and characterization of non tuberculous mycobacteria from humans and animals in Namwala District of Zambia. BMC Res Notes 7.

107. Lamont EA, Janagama HK, Ribeiro-lima J, Vulchanova L, Seth M, Yang M, Kurmi K, Waters WR. 2014. Circulating Mycobacterium bovis Peptides and Host Response Proteins as Biomarkers for Unambiguous Detection of Subclinical Infection 52:536–543.

CHAPTER 3 DISCOVERY OF A PREDOMINANT AND DISTINCT LINEAGE OF MYCOBACTERIUM TUBERCULOSIS IN BRAZILIAN INDIGENOUS POPULATION

Abstract

After nearly a century of vaccination and six decades of drug therapy, TB kills more people annually than any other infectious disease. Substantial challenges to disease eradication remain among vulnerable and underserved populations. The Guarani-Kaiowá people are an indigenous population in Paraguay and the Brazilian state of *Mato Grosso do Sul*. This community, marginalized in Brazilian society, experiences severe poverty. Like other South American indigenous populations, their TB prevalence is high, but the disease has remained largely unstudied in their communities.

Herein, MTB isolates from local clinics were whole genome sequenced, and a population genetic framework was generated. Phylogenetics show MTB isolates in the Guarani-Kaiowá people cluster far from selected reference strains, suggesting divergence and within-population microevolution. Most isolates cluster in a single group, further characterized as MTB sublineage 4.3.3. Closer analysis of SNPs showed numerous variants across the genome, including in drug resistance-associated genes, and with many unique changes fixed in each group.

We report that local MTB strains affecting the Guarani-Kaiowá people have acquired unique polymorphisms, and drug resistance characterization is urgently needed to inform public health to ensure proper care and avoid further evolution and spread of drug-resistant TB.

Introduction

The WHO EndTB strategy focuses on the reduction of incidence and mortality rates and works towards the elimination of TB by 2035, yet the disease remains a substantial public health threat worldwide. Ethnic minorities who face precarious living conditions and lack access to healthcare are at greater risk. TB eradication in aboriginal (indigenous) groups is further complicated by sparse data collection from these communities. The existing evidence indicates that incidence of active TB and prevalence of latent TB infection (LTBI) are significantly higher in indigenous groups compared to non-indigenous populations globally(1). This is a result of socioeconomic inequalities, including limited access to education, lack of employment opportunities and marginalization(2). To achieve TB eradication in these communities, high-resolution surveillance tools are needed. The use of whole-genome sequencing (WGS) in TB research provides the highest possible resolution to study disease transmission. WGS can diagnose infections directly from primary clinical samples, provide results faster, predict antimicrobial agent resistance-associated genotypes, and has applications in outbreak investigations to better define transmission clusters as compared to RFLP, MIRU-VNTR or spoligotyping as discussed in Chapter 1 (3–8).

Despite significant advances in TB control in recent decades, Brazil remains a high-burden country with an estimated 95,000 new cases of TB diagnosed in 2018(9). Indigenous groups in Brazil suffer from incidence rates three times than that of the general population(10). Studies have suggested nearly 50% LTBI prevalence rates in some indigenous territories, along with evidence of recent and ongoing transmission(11),(12). The Guarani-Kaiowá people in the Brazilian

state of *Mato Grosso do Sul* (MS) are Brazil's second largest indigenous group at around 44,000 individuals. For the last decade, MS has borne the country's highest rate of TB among its indigenous population. In 2013, a study of the indigenous people in MS found annual incidence rates of diagnosed active TB to be six times the state average(13). The southern part of the state, Amambai district, stood out with an average annual TB incidence rate of around 400/100,000 in the population of Guarani-Kaiowá ethnicity(13).

The aim of this study was to characterize the genomic patterns of diversification of drug resistance, and assess whether hyperendemic, locally circulating *Mycobacterium tuberculosis* in the indigenous population of the Guarani-Kaiowá is a product of regular introductions from outside the community, or unchecked transmission within the community.

Materials and methods

Study area, population, and design

The population of the Brazilian state of MS is around 2.7 million, including approximately 74,000 indigenous people from nine different groups, of which the Guarani-Kaiowá and the Terena tribes represent the majority. Despite international condemnation of human rights violations against the Guarani-Kaiowá(14), the Guarani-Kaiowá remain confined to overcrowded reservations(15). Most income for the Guarani-Kaiowá in southern MS comes from seasonal work on plantations, where overcrowded workers' barracks have been identified as TB transmission hotspots(16). Poor living conditions of the Guarani-Kaiowá, with high rates of violence(15), low education and income levels(17), and food insecurity(18) favor transmission of TB.

Control of TB in the indigenous populations of Brazil is the responsibility of the Special Secretariat for Indigenous Health of the Ministry of Health (Portuguese: SESAI). SESAI is composed of 34 Special Indigenous Health Districts which organize healthcare centers located in areas with significant indigenous populations. Each center manages smaller healthcare stations inside each indigenous territory as primary points of care. We examined TB cases from the *Amambai* and *Caarapó* healthcare centers during the study period of 2011 until 2016.

TB diagnosis in the study area was carried out at local healthcare centers. Ziehl-Neelsen bacilloscopy was performed until 2014 when this method was replaced by GeneXpert. Ministry of Health's TB protocol(19) states that positive bacilloscopy/GeneXpert results should be confirmed by the culture at the General Laboratory of MS; this confirmation is not always possible due to logistical challenges in getting samples to the laboratory(20) in addition to well-documented difficulties with successfully culturing MTB (21).

Per the Ministry of Health(19), the standard treatment regimen for TB during the study period was two months of rifampin, isoniazid, pyrazinamide, and ethambutol, followed by four months rifampin and isoniazid. Streptomycin and fluoroquinolone are second-line TB drugs and are used only in special regimes of treatment, with the patient under close medical supervision. Testing cultures for susceptibility to first-line drugs was done at the Central Public Health Laboratories of *Mato Grosso do Sul* (*Laboratório Central de Saúde Pública de Mato Grosso do Sul*, Portuguese (acronym LACEN-MS) only with clinical indication (i.e., failure to respond to initial treatment or re-infection.

The indigenous territories included in this study (*Amambai, Guassuty, Jaguari, Kurussu Amba, Limao Verde, Taquaperi,* and *Caarapó*) are assisted by *Amambai* and *Caarapó* health units and are located on the border of Paraguay. Together, these Guarani-Kaiowá communities encompass approximately 18,000 people.

Tuberculosis isolate identification, regrowth, and DNA extraction

Patients diagnosed with TB during the study period were identified in the records of the SESAI. MTB isolates from these patients were identified in the MS Laboratory Environment Management System. All cultures still in storage at LACEN-MS were regrown in the Ogawa-Kudoh medium. DNA from the regrown cultures was extracted and purified at São Paulo's Instituto Adolfo Lutz using the cetyltrimethylammonium bromide (CTAB)-lysozyme method as described by van Embden *et al.* (1993)(22). Purified DNA concentrations were determined by Gene-Quant II (Pharmacia Biotech).

SESAI registered 277 incident cases during the study period. We set a cutoff for sample metadata completeness, which limited our set. Furthermore, not all samples arrived with sufficient quality for sequencing in Michigan, and WGS was ultimately performed on 68/277 samples.

Whole-genome sequencing and barcoding

Sixty-eight samples were submitted for WGS on the MiSeq (Illumina) platform at Michigan State University. Samples were numbered S1-S68. Samples S02 and S36 failed quality checks and were excluded. The remaining 66 isolates was trimmed by Trimmomatic(23) (trimmomatic PE, trimming TruSeq2 adapters, with parameters 2:30:10 LEADING:3 TRAILING:3

SLIDINGWINDOW:4:15 MINLEN:36) and mapped against MTB H37Rv in CLC Genomics Workbench v12.0. Representative global reference lineages included were MTB 96121 (Manila-Lineage 1), HN878 (Lineage 2), HN-024 (Lineage 3), CDC1551 (Lineage 4), H37Rv (Lineage 4), Erdman (Lineage 4), and GM041182 (West African 2-Lineage 6). Multiple representatives of lineage 4 were included due to its historical and modern clinical importance. A whole genome multiple sequence alignment was constructed with default parameters in CLC. To identify the sublineages of MTB circulating in the community, we utilized specific SNPs for "barcoding" of TB, as reported by Coll *et al.* (2018)(24). The whole-genome alignment of Brazilian isolates was searched manually for each of the 414 reported SNP barcodes. If all diagnostic SNPs for a particular sub-lineage were found in an isolate, it was considered part of that group. Summary of results are shown in Table 1.

Core SNP extraction, phylogeny generation, and analysis

After initial analysis, we sought local references to better understand observed diversity. Unfortunately, assembled Brazilian whole genome sequences are few and we reached out to Drs. Brynildsrud and Eldholm, who in 2018 were first and corresponding authors, respectively, on an expansive investigation into *Mycobacterium tuberculosis* lineage 4 isolates. Their dataset includes hundreds of paired-end FASTQ read sets for South American tuberculosis genomes. Reads for nine Brazilian isolates unrelated to our current study were downloaded from NCBI (BioProject PRJEB27366) and processed utilizing the snippy pipeline(25). Briefly, we included CLC-constructed contigs for all 66 of our isolates, assembled whole genomes for global references, and raw paired-end FASTQ reads for nine isolates (referred to in this manuscript as

"local references") from the Brynildsrud et al., 2018, dataset (26). These local references represent sublineages 4.3 (LAM, n=7), 4.4 (S-type, n=1), and 4.7 (Congo, n=1)(27). They were first processed by ABySS(28) (abyss-pe, default parameters, k=96) to construct contigs, and they, along with all contigs from our isolates and assembled genomes for global references, were batch-processed in snippy with the MTB H37Rv Genbank (.gb) file as reference. The snippy output folders for each isolate and local references were processed by snippy core using H37Rv.gb as reference. This output core alignment, comprised solely of variant sites and excluding complex changes like indels, was analyzed in 16 drug resistance-associated genes. To avoid the effects of uninformative, false positive hits in repetitive regions of the genome, snippy-core was run again with the --mask parameter and the MTB H37Rv-based bed mask file included in the snippy package to filter out loci like PE/PPE family proteins. The unmasked and masked core SNP sets were compared, and ~3% of SNPs in both isolates and local references were considered uninformative by this approach. Masked core SNPs were passed into MEGA-X v10.1.8(29). Maximum-likelihood trees were produced with 500 bootstrap replicates and the HKY model with other parameters default, and the tree with highest log-likelihood is shown (Figure 3-1) A quick comparison of ML phylogeny (HKY, default parameters, bs=100) between masked and unmasked core SNPs showed greater bootstrap values for nodes derived from the masked set, suggesting removing uninformative hits yielded a more robust phylogeny. As such,

the masked SNP set was used for final analysis.

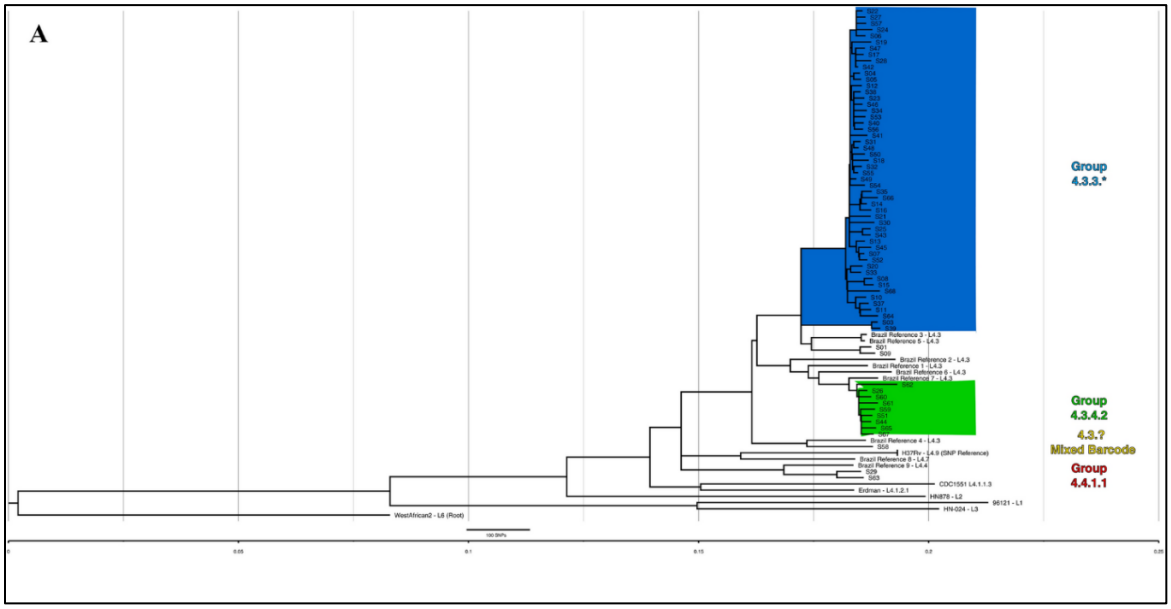

*Figure 3-1: (**A**) Maximum likelihood (ML) tree showing SNP-based phylogenetic analysis of core SNPs extracted by snippy. Analysis by MEGA-X (default parameters, HKY model) with 500 bootstrap replicates. Tree with highest log likelihood is shown (− 53,357.07). Tree was visualized in FigTree, and labeling generated in Inkscape software. The tree is rooted to* M. tuberculosis *GM041182 (West African 2). The tree is drawn to scale, with branch lengths measured in average substitutions per site (7420 sites total). Bootstrap values are robust across the entire tree. Coloring represents groups identified by Coll et al.'s TB barcoding strategy, that groups the isolates based on the presence or absence of lineage-determining SNPs. The tree cleanly groups similarly barcoded TB isolates and suggests divergence of sampled isolates even from recent local reference isolates.*

Trees were visualized in FigTree and MEGA-X. Final labeling of trees was performed with Inkscape v.1.0(30).

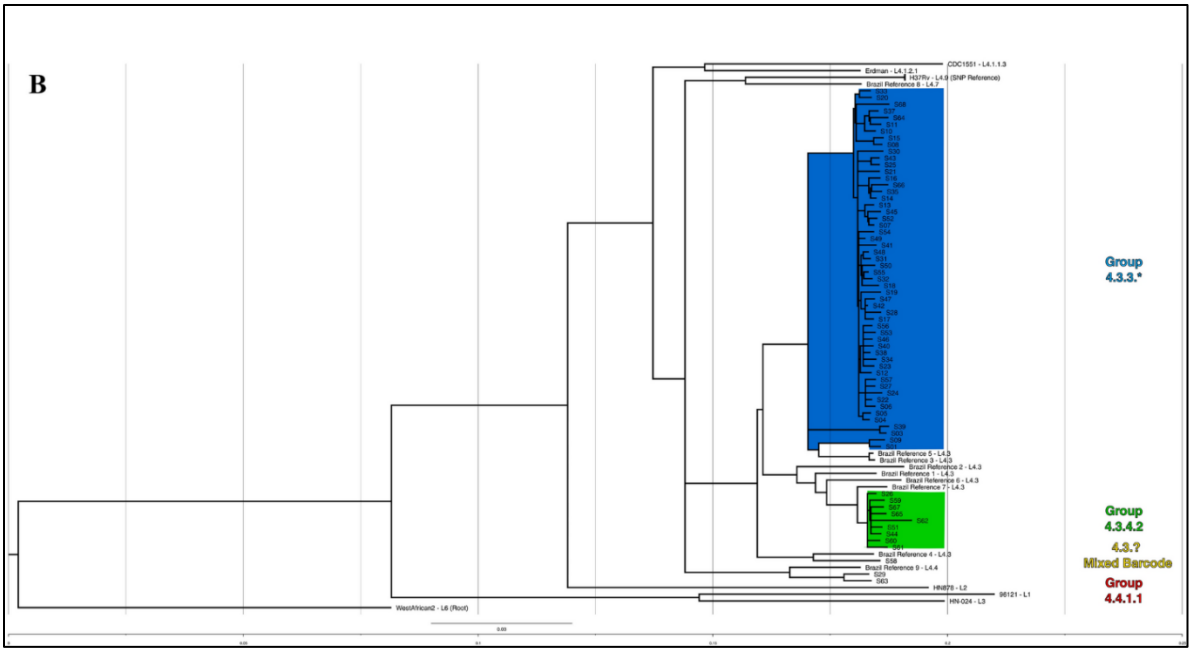

*Figure 3-2: (**B**) Bayesian consensus phylogram showing SNP-based phylogenetic analysis of 7420 core SNPs extracted by snippy. Analysis performed by MrBayes (GTR model, nst = mixed, mcmc = 500,000). Tree was visualized in FigTree, and labeling generated in Inkscape software. The tree is rooted to* M. tuberculosis *GM041182 (West African 2). Coloring represents groups identified by Coll et al.'s TB barcoding strategy. The tree demonstrates nearly identical grouping of all isolates as the maximum likelihood tree in 3-1 (**A**) and lends even more support for barcoding-based grouping.*

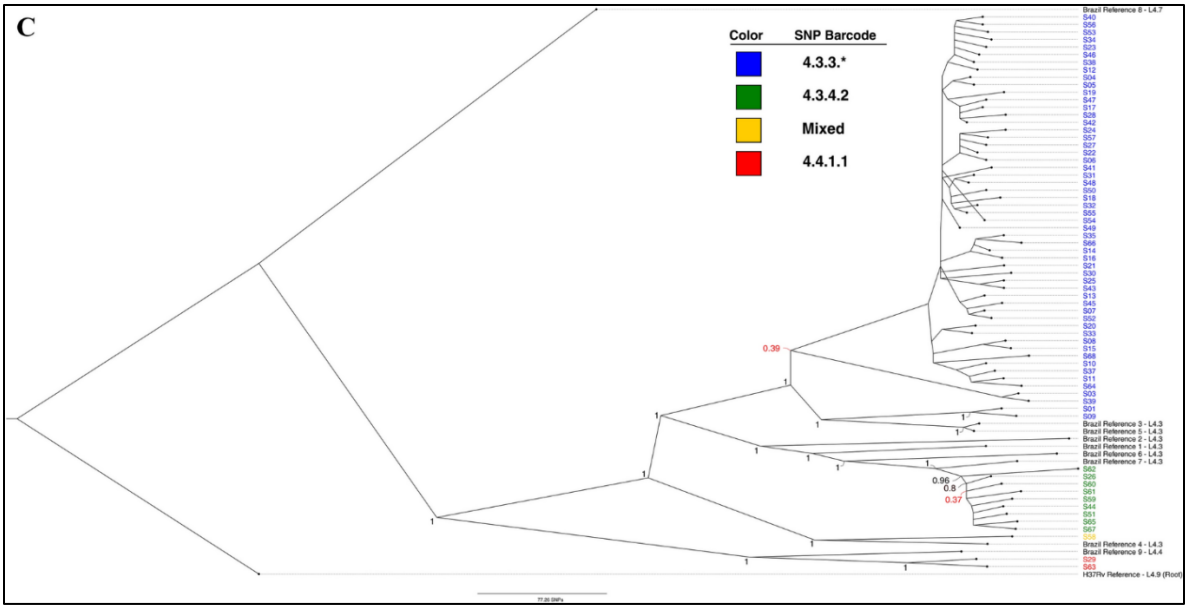

*Figure 3-3: (**C**) Maximum likelihood (ML) tree showing SNP-based phylogenetic analysis of core SNPs extracted by snippy without global reference lineages. Analysis by MEGA-X (default parameters, HKY model) with 1000 bootstrap replicates. Tree with highest log likelihood is shown (− 30,203.04). Tree was visualized in FigTree, and labeling generated in Inkscape software. The tree is rooted to* M. tuberculosis *H37Rv (Lineage 4.9). The tree is drawn to scale, with branch lengths measured in average substitutions per site (3863 sites total). Bootstrap values are labeled at branch points if greater than 70%, and any points where support falls under 70% is labeled in red and further values are not provided for the clade. Coloring represents groups identified by Coll et al.'s TB barcoding strategy. Without global reference lineages, barcoded isolates still group reliably. S09/S01 and S03/S39 (blue clade, bottom four isolates) stand out by branch lengths as divergent from rest of blue clade.*

Results

Characterization of the patient population

Selected samples reflected the characteristics of the overall TB patient population in the study area(31, 32). Diagnosed patients were 62% male (71% of patients from whom we obtained WGS data). TB patients were relatively young, with a median age of 32.5 in the study area (33 years for our subset used WGS analysis). Nearly 50% of patients sought healthcare within a week of symptom onset, but around 20% (in total and WGS subset cases) reported waiting five weeks or longer to seek healthcare. Nine of the 68 patients had been treated for TB one or more times in the past.

Whole-genome sequencing

After genome processing, the 414 lineage-determining single-nucleotide polymorphisms (SNPs) published in Coll *et al.* (2014) were utilized (24). Four main groups emerged from the analysis. The dominant sublineage (n=54/66, (blue)) was 4.3.3. The 4.3.3-determining *Rv1248c* C2526G was always paired with an adjacent G2525T mutation (S842Y). This SNP, among others, led us to tentatively label this sublineage as 4.3.3.*, noting that it and other changes unique to the group may represent a distinct sublineage of 4.3.3 in Brazil. In total, 97% of all MTB analyzed in this study was identified as sublineage 4.3. When analyzing local Brazilian 4.3 references (n=7), only 2 (local references 3 and 5) shared some of the SNPs fixed in our dataset, and 61 SNPs at greater than 80% prevalence in 4.3.3.* were completely absent in all global and local references. Remaining isolates belonged to sub-lineages 4.3.4.2 (n=9/66 (green), 4.4.1.1 (n=2/66 (red)), and

ambiguous S58 (yellow) in either 4.3.2 or 4.3.3, carrying an unusual mixed SNP genotype from multiple lineages (although *Rv1248c* G2525T/C2526G was absent).

A maximum likelihood (ML) phylogenetic tree was generated in MEGA-X to identify genotype clustering patterns. The tree with the highest log likelihood is shown in Figure 3-1. Branch lengths are drawn to scale and represent the number of substitutions. All isolates were well-separated from global MTB references [closest=H37Rv, >600SNPs avg. pairwise distance]. Bootstrap values were well above recommended cutoffs for statistical support(33). The blue clade representing 4.3.3.* contains most of the isolates (n=54/66, ~82%), though four of these isolates (S01/S09, and S03/S39) stand on relatively distant, robust branches and appear distinct. These isolates fell at least ~180 SNPs away from closest local reference (local reference 5 to S01). The green clade, 4.3.4.2, consisted of 9 isolates, grouped next to four local references (local references 1, 2, 6, and 7), but still showed a minimum SNP distance of ~75 SNPs. The red clade (S63/S29) grouped with local reference 9 at ~245 SNPs distance. Finally, S58 (yellow clade) grouped ~200 SNPs away from closest local reference 4. Next, a Bayesian phylogenetic consensus tree was generated in MrBayes and is shown in Figure 3-2. The tree closely replicates the reconstruction by the ML tree, and all results from the previous tree apply. Support for the run and Bayesian tree were excellent. Finally, an additional ML tree built against H37Rv with only local strains is shown in Figure 3-3, providing an at-a-glance visualization of how barcoding aligned with phylogenetic grouping.

Both the barcoding strategy and phylogenetic trees using different methodologies produced nearly identical findings. Over 80% of isolates grouped into a dominant clade typed as 4.3.3.*,

and considerable distance was observed between recent Brazilian strains from another study and the isolates circulating in the studied population.

Single-nucleotide polymorphism analysis and drug resistance-associated genes

Of 68 patients, nine had been treated for tuberculosis at least once before. One of these isolates failed quality control (S02), but of the remaining eight, six were infected by 4.3.3.*, and two by 4.3.4.2. To investigate potential drug resistance in this community, an analysis of 13 drug resistance (DR) associated genes was performed, revealing 24 single nucleotide polymorphisms (SNPs). Of these, 79.2% (19/24) led to point mutations in which a single nucleotide change resulted in a codon that codes for a different aminoacid (missense mutation). Among these missense mutations, 57.9% (11/19) were unique to the Guarani-Kaiowá isolates when compared against the 9 Brazilian isolates from Brynildsrud *et al*. (2018)(26).

Isoniazid resistance is associated with *inhA*, *ahpC*, *kasA,* and *katG*(36, 37). All four genes were analyzed for variants, including 100bp upstream of the start codon for *inhA*. No SNPs were seen in *inhA* or *ahpC*. Two SNPs were identified in *kasA* and six in *katG*. These eight SNPs all led to missense mutations (Figure 3-4). In *kasA*, G805A was present in all 54 isolates in 4.3.3.*. This SNP was shared only by local references 3 and 5. In *KatG*, two mutations were found in three isolates at nucleotide C944, one leading to S315T (S03, S39) and the other to S315N (S17). This SNP was also found to be commonly circulating in all local reference strains except in local reference 1. Only two samples in our dataset were known to be phenotypically resistant to isoniazid – S03 and S17 – and both had SNPs in *kasA* and *katG.*

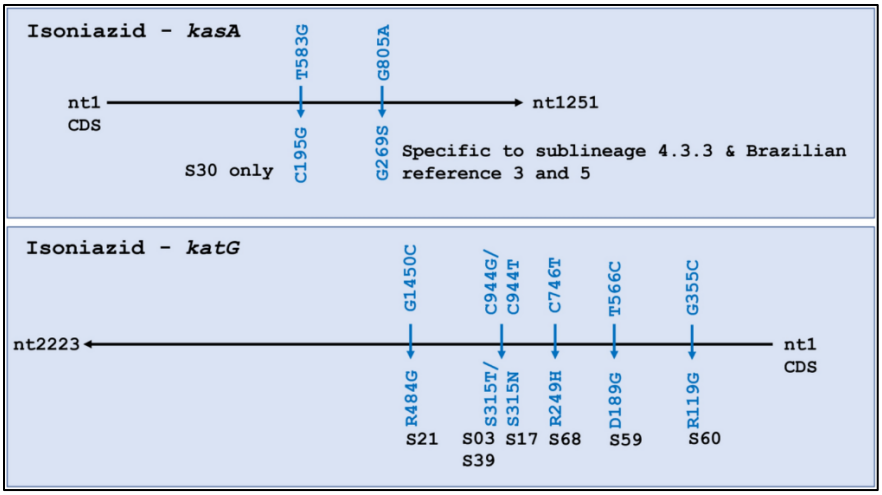

Figure 3-4: Isoniazid resistance-associated genes katG and kasA had multiple single nucleotide polymorphisms, whereas ahpC and inhA were wildtype. Missense mutations are colored blue

The 80bp region in *rpoB* – the rifampicin resistance-determining region (RRDR) – did not carry any mutations However, two missense mutations were seen outside the RRDR in three samples – S03, S68 and S12 (Figure 3-5). S03 was the only sample that was phenotypically rifampin resistant, confirming S03 as multidrug resistant both phenotypically and genotypically due to presence of mutations in *kasA*, *katG,* and *rpoB,* although the *rpoB* SNP fell outside the RRDR. No mutations were observed in *rpoA* in any of the isolates, while S68 showed a missense mutation C3119G (P1040R) in *rpoC*. It is of note that S50 was phenotypically resistant to rifampin, but no SNPs were identified in *rpoB*.

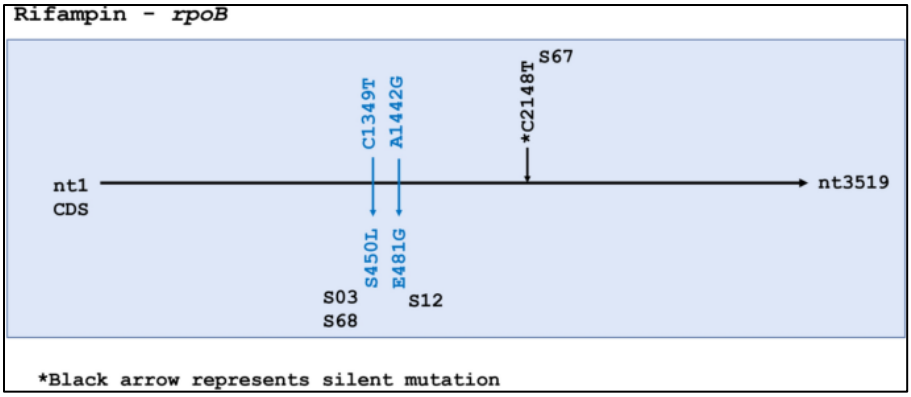

*Figure 3-5: Rifampin resistance-associated gene rpoB had only three single nucleotide polymorphisms (SNPs). Missense mutations are colored blue. S67 had a silent mutation represented by the *black arrow*

Ethambutol resistance-associated mutations occur in three genes identified as *embABC*. No SNPs were detected in *embA*. One silent mutation in *embB*, and 3 silent plus 2 missense mutations in *embC* were identified (Figure 3-6). *embC* missense mutations were seen in S01, S09 and S08, but drug susceptibility data was not available for any of these isolates. The silent mutation at C2781T in *embC* was found in all 66 isolates in our study but was also found in all references except H37Rv.

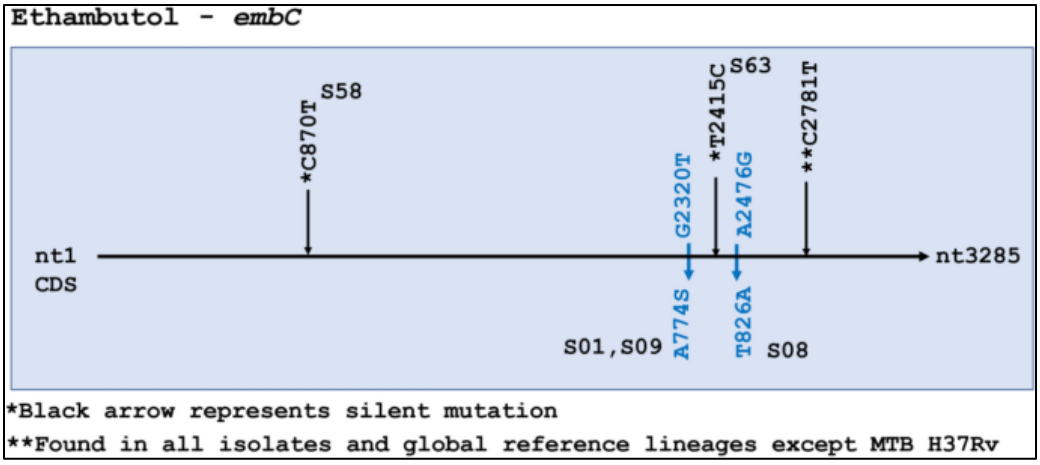

Figure 3-6: Ethambutol resistance-associated gene embC *had 2 missense single nucleotide polymorphisms (SNPs) and one silent SNP (**C27181T) found in all 66 isolates in the data set as well as all reference strains except H37Rv and is not considered particularly informative. All samples were wildtype for* embA. *Missense mutations are colored blue. Silent mutations are represented in *black color*

For the second line anti-tuberculosis drug fluoroquinolone, two genes – *gyrA* and *gyrB* – were analyzed. All SNPs identified were missense – 5 in *gyrA* and 2 in *gyrB*. Out of the 5 SNPs in *gyrA*, G739A (G247S) was fixed in all 54 isolates that fell in sublineage 4.3.3. Three SNPs were common to all 66 isolates as well as our global references except for H37Rv: *gyrA* E21Q, S95T, and G668D. These 3 SNPs were present in almost all Brazilian reference strains, with the exception of S95T & G668D that were absent in local reference 8. The 2 missense SNPs in *gyrB* were found only in two isolates and were absent in Brazilian references (Figure 3-7).

Figure 3-7: Fluoroquinolone resistance-associated genes gyrA *and* gyrB*, had a few yet significant single nucleotide polymorphisms (SNPs). All SNPs were missense (blue). Three SNPs (**) in* gyrA *were fixed in the entire data set and all references except H37Rv, whereas one SNP in* gyrA *was found in 54/66 isolates. The former are not considered particularly informative, while the latter does appear to be. Two SNPs were observed in* gyrB*.*

Discussion

Although TB is hyperendemic among indigenous populations like the Guarani-Kaiowá in Brazil, there has been a lack of investigation into TB transmission in these settings, a red flag for public health policy. This is, to our knowledge, the first study applying WGS to TB samples from an indigenous population in a high-burden, low-and-middle-income country. Our phylogenomic

study indicates that, although cases were diagnosed across six different territories over a study period of six years, the majority of the sequenced cases are caused by MTB clustered into sub-lineage 4.3.3 (Figure 3-1Figure 3-2Figure 3-3, blue clades). The three barcoded sublineages observed showed significant divergence from global references (~600+ SNP average distance), necessitating the inclusion of other recently published Brazilian isolates for comparison. These nine additional isolates (referred to as "local references") and an analysis methodology were graciously shared by Drs. Brynildsrud and Eldholm, and they allowed a more refined comparison between Brazilian TB strains(26). Isolates remain extensively divergent by phylogeny and SNP analysis even when including these local references from 2009-2012. While local reference strains reliably cluster with our isolates, the average SNP distance within 4.3.3.* isolates were ~72 SNPs (SD±45), and the distance from two closest 4.3 local references was ~210 SNPs (SD±10). The strains observed in this study are Latin American-Mediterranean (LAM) type TB, which is itself not rare in the region. However, the extreme prevalence of MTB sublineage 4.3 (64/66, ~97%) is noteworthy. It is far greater than any LAM incidence observed in any prior report the authors could identify in South America or globally, with frequencies typically up to 50%(38–41) and peaking in South America at only ~65% in one study(39). The robust grouping of isolates distant from global references, the divergence from local references, and the unusual prevalence of a single type of TB (LAM) supports a hypothesis that disease in this community is not the result of modern, recurring introductions. Rather, mutations absent in global lineages yet fixed in the isolates from this study suggest earlier, historical introductions of MTB and a subsequent circulation and diversification of local sublineage 4.3 (LAM) within the Guarani-Kaiowá

population. This hypothesis would be expected for an isolated community, but the Guarani-Kaiowá cannot live off their overcrowded territories and are therefore not living in isolation. Our findings rather support the extreme level of marginalization suffered by this indigenous group – to the point where the fingerprints of interaction appear largely undetected by high-resolution genomic data.

In both phylogenetic trees, we observed universal agreement for all barcoded groups. Isolates grouped into the blue clades (Figure 3-1and Figure 3-2) are dominant and responsible for most sampled infections. Both trees and SNP-based classifications support additional uncommon subtypes circulating in this population, like the distant sublineage 4.4.1.1 isolates (red clade) of S29/S63, or the unusual genotype observed in S58 (yellow clade). We find 111 SNPs that are absent in global references but common at 80%+ prevalence in local references and isolates, an additional 63 SNPs present only in 4.3.3.* and the closest related local references 3 and 5, and a further 61 SNPs found at 80%+ prevalence solely in our 4.3.3.* isolates alone. Furthermore, isolates show variability in drug resistance-associated genes, a possible portent of developing resistance. Two variants – *kasA* G805A and *gyrA* G739A – have become dominant and fixed in most sampled infections among the Guarani-Kaiowá population. These fixed variants were specifically identified in isolates of sublineage 4.3.3.*. Additionally, 4 SNPs were identified in all 66 isolates in the fluoroquinolone (*gyrA*) and ethambutol (*embC*) resistance-associated genes. Without careful genome assembly and associated error-checking of the local reference paired-end reads we used, as well as still more local references to compare against, the authors cannot say for certain whether all these SNPs are truly unique, nor can we speculate what effects they

may have, but they do continue to support extensive diversification within the studied population.

Taken together, this is alarming for the Guarani-Kaiowá people from a public health policy perspective, as patients are currently treated as drug-sensitive unless treatment fails, due to insufficient provision of susceptibility kits to the laboratory environment systems (LACENs). In addition to needless suffering, this may have allowed drug-resistant MTB to arise unchecked in Guarani- Kaiowá communities. Drug susceptibility testing should be provided by the Brazilian Ministry of Health as part of the national TB control program, as it is urgently needed in these vulnerable populations; both to lessen the existing high disease burden as well as to prevent the emergence of MDR-TB that can cause further harm and confound control efforts.

Some isoniazid resistance-associated genes, such as *inhA* or *ahpC,* showed no changes*,* while *katG* and *kasA* had multiple mutations. *KatG* normally activates the isoniazid prodrug (INH), while *kasA* contributes to mycolic acid synthesis that isoniazid impairs. Changes to *katG* Ser315 are strongly associated with isoniazid resistance(42). Globally, a single change in *katG* Ser315 is responsible for 64.2% of all phenotypic resistance to isoniazid, the majority of which (95.3%) cause S315T and a few (3.6%) are associated with S315N(36). These substitutions were also observed in a study in China, where single *katG* mutations occurred in majority of the cases, most of which had S315T and two bore S315N(43). In our analysis, S315T was observed in two isolates S03 and S39, while S315N was identified in S17. Though we lack phenotypic data on S39, both S03 and S17 were phenotypically resistant to isoniazid. Jagielski *et al.* 2015 reported on SNPs in *kasA* and found that G269S was present in 22% of MDR clinical cases, and in 10% of isoniazid

mono-resistant (44). Older studies have also reported the same(42, 45–47). In our work, 81.8% of isolates (54/66) showed G269S (G805A), suggesting a much higher prevalence than found elsewhere. Unfortunately, highly limited phenotypic data (n=2/54) hinders further interrogation beyond S03 and S17, which were confirmed isoniazid resistant. These findings merit in-depth study of the population, as literature suggests some of the observed and commonplace SNPs are associated with drug resistance and prompting concern about drug-resistant tuberculosis becoming pervasive in the Guarani-Kaiowá people.

In *rpoB*, drug resistance-associated mutations are mostly confined to the RRDR (codon 507 to 533), coding for the core of the beta-subunit of RNA polymerase. Almost 95% of all resistance-associated mutations exist in this hotspot. In our study, all 3 SNPs were found to be outside the RRDR. Of 66 isolates, S03 was the only one known to be multidrug-resistant, yet both S68 and S03 had the same amino acid substitution S450L that is commonly associated with resistance to rifampicin(48, 49) This mutation falls inside the polymerase's highly conserved catalytic pocket (codon 443–451)(50), suggesting that S68 might also have been multi drug-resistant like S03. Furthermore, S450L has been reported to accumulate significantly more compensatory mutations than other rifampicin resistance-associated SNPs(51). This was observed in our study uniquely in S68, with a missense mutation in rpoC (P1040R) lending further support to possible rifampin-resistance in S68. We lack the phenotypic data to confirm our hypothesis, again calling for intensive drug testing in the community before initiation of treatment (52–54).

Fluoroquinolone, the second-line anti-TB drug, targets a DNA gyrase encoded by *gyrA* and *gyrB*. Like rifampin's *rpoB* hotspot, drug resistance to fluoroquinolone is also associated with

specific mutations in the quinolone resistance-determining region (QRDR) of *gyr*A & *gyrB*(55). In *gyrA*, T378S was a substitution found to be unique to our study (Figure 3-7). Another substitution, G247S, was found 100% of all our 4.3.3.* isolates and was not seen in any other isolates or global references. Only the two Brazilian reference strains (reference 3 and 5) that grouped on the periphery of 4.3.3.* showed this SNP, suggesting these references may be tightly related to our predominant sublineage. This mutation falls outside the QRDR and has not been associated with fluoroquinolone resistance(56). Whether it led to drug resistance in patients is unknown, but it is important to note that while indigenous patients are not treated with fluoroquinolones for TB in Brazil, drugs in this major group are regularly administered for other ailments such as pneumonia, foodborne diseases, and urinary tract infections caused by a wide variety of bacteria such as methicillin-resistant *Staphylococcus aureus*(57), *Enterobacteriaceae*(58), *Pseudomonas aeruginosa*(59), and *Salmonella*(60). It is possible that the use of fluoroquinolones for other conditions in parallel with endemic TB has led to the development of unique circulating TB strains. Fixing of SNPs in the entire data set and the establishment of unique SNPs beyond what are observed in existing sub-lineages support a hypothesis that new TB strains have arisen in the indigenous population(61).

Multiple mutations were identified in nearly all drug resistance-associated genes that were analyzed, with many in isoniazid, ethambutol, and fluoroquinolone drug resistance-associated genes. A 2000-2006 study analyzing 783 TB isolates (82.4% new cases and 17.6% previously treated patients) from the general population in *Mato Grosso do Sul* found evidence of drug resistance in 18.3% (all combinations) of cases. Looking at single-drug primary resistance,

streptomycin showed the highest level of resistance (3.4%), followed by isoniazid (2.9%), ethambutol (1.7%), and rifampin where no resistance was detected. The resistance to streptomycin was believed to be due to endogenous reactivation of TB strains from before the 1980s, as streptomycin was not given to new cases. Acquired resistance was found to isoniazid (7.2%), streptomycin (3.6%), rifampin (2.2%), and ethambutol (1.4%). Although levels of primary resistance were lower than observed in contemporary data from other regions, the levels of acquired resistance were already high in 2000-2006(62). As the relative risk of TB is six times higher for the indigenous versus the non-indigenous population in MS(10) we could speculate that a significant proportion of the samples from the 2000-2006 study could have been from indigenous patients.

Conclusion

In summary, these data provide the first detailed genomic view of hyperendemic TB circulating within the Guarani-Kaiowá people. The vast majority of MTB isolates fell into a dominant clade of the Latin American sub-lineage 4.3.3 by Coll *et al.*'s SNP barcoding strategy(24). This grouping was fully supported by two phylogenetic methods using a core SNP set of 7420 total positions taken from our 66 isolates, 7 global references, and 9 local references. Analysis of SNPs both for drug resistance-associated genes – specifically *gyrA* (fluoroquinolone) and *kasA* (isoniazid) – and for lineage-determination revealed diversification, with unique SNPs fixed in the population. Among 4.3.3.*-classified isolates, there was the universal presence of a two nucleotide *Rv1248c* G2525T/C2526G change. Unlike 4.3.3 sub-lineage determining *Rv1248c* C2526G, these two changes lead to an amino acid substitution (S842Y). When looking at

references, this 4.3.3.*-unique substitution was only observed in 2/7 of Brazilian 4.3 references. This finding as well as the presence of other unique SNPs fixed in the population (*gyrA* and *kasA*) leads us to conclude that most TB in the Guarani- Kaiowá people is the result of a historical introduction of the commonly circulating 4.3.3 lineage, with subsequent diversification in the community. The incidence of sublineage 4.3 is far higher than observed anywhere else in South America, suggesting it is hyperendemic in the population. We have demarcated the most prevalent and distinct clade in this population as 4.3.3.*, and molecular and phenotypic characterization of this microevolution is important to discern whether drug resistance has become fixed in this sub-lineage. This would entail drug resistance testing to perform the phenotypic characterization. Though our study lacked lab-confirmed drug resistance data for all samples, ample literature exists that support that the genotypic mutations seen in our work has the potential for resistance to the standard treatment protocol. We recommend that regular surveillance and drug-susceptibility information be made compulsory before the initiation of treatment. A more targeted public health strategy should be implemented in the region, where TB cases are not screened for drug resistance unless all prior treatments have failed.

REFERENCES

REFERENCES

1. Tollefson D, Bloss E, Fanning A, Redd JT, Barker K, McCray E. 2013. Burden of tuberculosis in indigenous peoples globally: A systematic review. Int J Tuberc Lung Dis https://doi.org/10.5588/ijtld.12.0385.

2. Cormier M, Schwartzman K, N'Diaye DS, Boone CE, dos Santos AM, Gaspar J, Cazabon D, Ghiasi M, Kahn R, Uppal A, Morris M, Oxlade O. 2019. Proximate determinants of tuberculosis in Indigenous peoples worldwide: a systematic review. Lancet Glob Heal https://doi.org/10.1016/S2214-109X(18)30435-2.

3. Crisan A, McKee G, Munzner T, Gardy JL. 2018. Evidence-based design and evaluation of a whole genome sequencing clinical report for the reference microbiology laboratory. PeerJ https://doi.org/10.7717/peerj.4218.

4. Lee RS, Behr MA. 2016. The implications of whole-genome sequencing in the control of tuberculosis. Ther Adv Infect Dis https://doi.org/10.1177/2049936115624630.

5. Walker TM, Kohl TA, Omar S V., Hedge J, Del Ojo Elias C, Bradley P, Iqbal Z, Feuerriegel S, Niehaus KE, Wilson DJ, Clifton DA, Kapatai G, Ip CLC, Bowden R, Drobniewski FA, Allix-Béguec C, Gaudin C, Parkhill J, Diel R, Supply P, Crook DW, Smith EG, Walker AS, Ismail N, Niemann S, Peto TEA, Davies J, Crichton C, Acharya M, Madrid-Marquez L, Eyre D, Wyllie D, Golubchik T, Munang M. 2015. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: A retrospective cohort study. Lancet Infect Dis https://doi.org/10.1016/S1473-3099(15)00062-6.

6. Pankhurst LJ, del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, Lemaitre N, Niemann S, Paul J, Rogers TR, Roycroft E, Smith EG, Supply P, Tang P, Wilcox MH, Wordsworth S, Wyllie D, Xu L, Crook DW. 2016. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: A prospective study. Lancet Respir Med https://doi.org/10.1016/S2213-2600(15)00466-X.

7. van der Werf MJ, Ködmön C. 2019. Whole-genome sequencing as tool for investigating international tuberculosis outbreaks: A systematic review. Front Public Heal.

8. Lalor MK, Casali N, Walker TM, Anderson LF, Davidson JA, Ratna N, Mullarkey C, Gent M, Foster K, Brown T, Magee J, Barrett A, Crook DW, Drobniewski F, Thomas HL, Abubakar I. 2018. The use of whole-genome sequencing in cluster investigation of a multidrugresistant tuberculosis outbreak. Eur Respir J https://doi.org/10.1183/13993003.02313-2017.

9. WHO. 2018. Global Tuberculosis ReportGeneva: World Health Organization.

10. De Viana PVS, Gonçalves MJF, Basta PC. 2016. Ethnic and Racial Inequalities in Notified Cases of Tuberculosis in Brazil. PLoS One https://doi.org/10.1371/journal.pone.0154658.

11. Malacarne J, Rios DPG, da Silva CMFP, Braga JU, Camacho LAB, Basta PC. 2016. Prevalence and factors associated with latent tuberculosis infection in an indigenous population in the Brazilian Amazon. Rev Soc Bras Med Trop https://doi.org/10.1590/0037-8682-0220-2016.

12. Cunha EA, Ferrazoli L, Riley LW, Basta PC, Honer MR, Maia R, da Costa IP. 2014. Incidence and transmission patterns of tuberculosis among indigenous populations in Brazil. Mem Inst Oswaldo Cruz https://doi.org/10.1590/0074-0276130082.

13. Basta PC, Marques M, De Oliveira RL, Cunha EAT, Da Costa Resendes AP, Souza-Santos R. 2013. Social inequalities and tuberculosis: An analysis by race/color in mato grosso do sul, Brazil. Rev Saude Publica https://doi.org/10.1590/S0034-8910.2013047004628.

14. Tauli-Corpuz V, Peoples UHRCSR on R of I, Secretariat UHRC. 2016. Report of the Special Rapporteur on the rights of indigenous peoples on her mission to Brazil. Geneva; United Nations. Geneva.

15. CIMI. 2018. Violence against Indigenous Peoples in Brazil - Data for 2018. Brasil.

16. Sacchi FPC, Croda MG, Estevan AO, Ko AI, Croda J. 2013. Sugar cane manufacturing is associated with tuberculosis in an indigenous population in Brazil. Trans R Soc Trop Med Hyg 107:152–157.

17. Lemos EF, Alves AMDS, Oliveira GDC, Rodrigues MP, Martins NDG, Croda J. 2014. Health-service performance of TB treatment for indigenous and non-indigenous populations in Brazil: A cross-sectional study. BMC Health Serv Res https://doi.org/10.1186/1472-6963-14-237.

18. Brasil F. 2016. O Direito Humano à Alimentação Adequada e à Nutrição do povo Guarani e Kaiowá: um enfoque holístico – Resumo Executivo.

19. 2019. Manual de para o Controle da TuberculoseMinistério da Saúde, 2nd ed. Ministério da Saúde.

20. Malacarne J, Heirich AS, Cunha EAT, Kolte IV, Souza-Santos R, Basta PC. 2019. Performance of diagnostic tests for pulmonary tuberculosis in indigenous populations in Brazil: the contribution of Rapid Molecular Testing. J Bras Pneumol https://doi.org/10.1590/1806-3713/e20180185.

21. Nema V. 2012. Tuberculosis diagnostics: Challenges and opportunities. Lung India 29:259–

266.

22. Van Embden JDAA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM, Small PM. 1993. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: Recommendations for a standardized methodology. J Clin Microbiol.

23. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics https://doi.org/10.1093/bioinformatics/btu170.

24. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, Portugal I, Pain A, Martin N, Clark TG. 2014. A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun https://doi.org/10.1038/ncomms5812.

25. Seemann T. Reads, snippy: fast bacterial variant calling from NGS.

26. Brynildsrud OB, Pepperell CS, Suffys P, Grandjean L, Monteserin J, Debech N, Bohlin J, Alfsnes K, Pettersson JOH, Kirkeleite I, Fandinho F, Da Silva MA, Perdigao J, Portugal I, Viveiros M, Clark T, Caws M, Dunstan S, Thai PVK, Lopez B, Ritacco V, Kitchen A, Brown TS, Van Soolingen D, O'Neill MB, Holt KE, Feil EJ, Mathema B, Balloux F, Eldholm V. 2018. Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation. Sci Adv https://doi.org/10.1126/sciadv.aat5869.

27. Malm S, Ghoma Linguissi LS, Tekwu EM, Vouvoungui JC, Kohl TA, Beckert P, Sidibe A, Rüsch-Gerdes S, Madzou-Laboum IK, Kwedi S, Beng VP, Frank M, Ntoumi F, Niemann S. 2017. New Mycobacterium tuberculosis complex sublineage, Brazzaville, Congo. Emerg Infect Dis https://doi.org/10.3201/eid2303.160679.

28. Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, Birol I. 2017. ABySS 2.0: Resource-efficient assembly of large genomes using a Bloom filter. Genome Res 27:768–777.

29. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol https://doi.org/10.1093/molbev/msy096.

30. 2020. Inkscape v1.0.

31. Malacarne J, Kolte IV, Freitas LP, Orellana JDY, Souza MLP de, Souza-Santos R, Basta PC. 2018. Factors associated with TB in an indigenous population in Brazil: the effect of a cash transfer program. Rev Inst Med Trop Sao Paulo https://doi.org/10.1590/S1678-9946201860063.

32. Freitas LP, Souza-Santos R, Kolte I V, Malacarne J, Basta PC. 2018. Socioeconomic status of

indigenous peoples with active tuberculosis in Brazil: a principal components analysis. bioRxiv https://doi.org/10.1101/290668.

33. Hillis DM, Bull JJ, Hillis, D.M. and Bull JT. 1993. An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. Syst Biol 42:182–192.

34. Sreevatsan S, Pan X, Zhang Y, Kreiswirth BN, Musser JM. 1997. Mutations associated with pyrazinamide resistance in pncA of Mycobacterium tuberculosis complex organisms. Antimicrob Agents Chemother 41:636 LP – 640.

35. Sreevatsan S, Escalante P, Pan X, Gillies DA, Siddiqui S, Khalaf CN, Kreiswirth BN, Bifani P, Adams LG, Ficht T, Perumaalla VS, Cave MD, van Embden JD, Musser JM. 1996. Identification of a polymorphic nucleotide in oxyR specific for Mycobacterium bovis. J Clin Microbiol 34:2007 LP – 2010.

36. Seifert M, Catanzaro D, Catanzaro A, Rodwell TC. 2015. Genetic Mutations Associated with Isoniazid Resistance in Mycobacterium tuberculosis: A Systematic Review. PLoS One 10:e0119628.

37. Desalegn W, Damtie D, Mathewos B, Weldeyohannes D, Mathewos B. 2014. Review on Molecular Mechanism of First Line Antibiotic Resistance in Mycobacterium tuberculosis. Mycobact Dis 04:4–7.

38. Vasconcellos SEG, Acosta CC, Gomes LL, Conceição EC, Lima KV, De Araujo MI, De Lourdes Leite M, Tannure F, De Souza Caldas PC, Gomes HM, Santos AR, Gomgnimbou MK, Sola C, Couvin D, Rastogi N, Boechat N, Suffys PN. 2014. Strain classification of mycobacterium tuberculosis isolates in brazil based on genotypes obtained by spoligotyping, mycobacterial interspersed repetitive unit typing and the presence of large sequence and single nucleotide polymorphism. PLoS One https://doi.org/10.1371/journal.pone.0107747.

39. Mendes NH, Melo FA, Santos AC, Pandolfi JR, Almeida EA, Cardoso RF, Berghs H, David S, Johansen FK, Espanha LG, Leite SR, Leite CQ. 2011. Characterization of the genetic diversity of Mycobacterium tuberculosis in São Paulo city, Brazil. BMC Res Notes https://doi.org/10.1186/1756-0500-4-269.

40. Dantas NGT, Suffys PN, Carvalho W da S, Gomes HM, de Almeida IN, de Assis LJ, Augusto CJ, Gomgnimbou MK, Refregier G, Sola C, de Miranda SS. 2015. Genetic diversity and molecular epidemiology of multidrug-resistant Mycobacterium tuberculosis in Minas Gerais State, Brazil. BMC Infect Dis https://doi.org/10.1186/s12879-015-1057-y.

41. Verza M, Scheffer MC, Salvato RS, Schorner MA, Barazzetti FH, Machado H de M, Medeiros TF, Rovaris DB, Portugal I, Viveiros M, Perdigão J, Kritski A, Bazzo ML. 2020. Genomic

epidemiology of Mycobacterium tuberculosis in Santa Catarina, Southern Brazil. Sci Rep https://doi.org/10.1038/s41598-020-69755-9.

42. Hazbon MH, Brimacombe M, Bobadilla M, Cavatore M, Varma-basil M, Billman-jacobe H, Lavender C, Fyfe J, Garcı L, Ine C, Bose M, Chaves F, Murray M, Eisenach KD, Cave MD, Leo AP De, Alland D. 2006. Population Genetics Study of Isoniazid Resistance Mutations and Evolution of Multidrug-Resistant Mycobacterium tuberculosis † 50:2640–2649.

43. Luo D, Chen Q, Xiong G, Peng Y, Liu T, Chen X, Zeng L, Chen K. 2019. Prevalence and molecular characterization of multidrug-resistant M. tuberculosis in Jiangxi province, China. Sci Rep 9:1–8.

44. Jagielski T, Bakula Z, Roeske K, Kamiski M, Napio´rkowska A, Augustynowicz-kopec E, Zwolska Z, Bielecki J. 2015. Mutation profiling for detection of isoniazid resistance in Mycobacterium tuberculosis clinical isolates ´. J Antimicrob Chemother 70:3214–3221.

45. Ramaswamy S V, Reich R, Dou S, Jasperse L, Pan X, Wanger A, Quitugua T, Graviss EA. 2003. Single Nucleotide Polymorphisms in Genes Associated with Isoniazid Resistance in Mycobacterium tuberculosis 47:1241–1250.

46. Sun Y, Lee ASG, Wong S, Paton NI. 2007. Analysis of the role of Mycobacterium tuberculosis kasA gene mutations in isoniazid resistance. Clin Microbiol Infect 13:833–835.

47. Nusrath Unissa A, Subbian S, Elizabeth L, Selvakumar N. 2016. Infection , Genetics and Evolution Overview on mechanisms of isoniazid action and resistance in Mycobacterium tuberculosis 45:474–492.

48. Datta G, Nieto LM, Davidson RM, Mehaffy C, Pederson C, Dobos KM, Strong M. 2016. Longitudinal whole genome analysis of pre and post drug treatment Mycobacterium tuberculosis isolates reveals progressive steps to drug resistance. Tuberculosis 98:50–55.

49. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S, Gagneux S. 2012. Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nat Genet https://doi.org/10.1038/ng.1038.

50. Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, Darst SA. 2001. Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase 104:901–912.

51. Zuo T, Liu Q, Gan M, Gao Q. 2018. Comprehensive identification of compensatory mutations in rifampicin-resistant Mycobacterium tuberculosis strains. Chinese J Tuberc Respir Med 41:207–212.

52. Zaw MT, Emran NA, Lin Z. 2018. Mutations inside rifampicin-resistance determining region

of rpoB gene associated with rifampicin-resistance in Mycobacterium tuberculosis. J Infect Public Health 11:605–610.

53. Prim RI, Schörner MA, Senna SG, Nogueira CL, Carolina A, Figueiredo C, Oliveira JG De, Rovaris DB, Bazzo ML. 2015. Molecular profiling of drug resistant isolates of Mycobacterium tuberculosis in the state of Santa Catarina , southern Brazil. Mem Inst Oswaldo Cruz 110:618–623.

54. Nebenzahl-guimaraes H, Jacobson KR, Farhat MR, Murray MB. 2014. Systematic review of allelic exchange experiments aimed at identifying mutations that confer drug resistance in Mycobacterium tuberculosis. J Antimicrob Chemother 331–342.

55. Johnson R, Elizabeth MS, Louw GE, Warren RM, van Helden PD, Victor TC. 2006. Drug resistance in Mycobacterium tuberculosis. Curr Issues Mol Biol 8:97–112.

56. Malik S, Willby M, Sikes D, Tsodikov O V., Posey JE. 2012. New Insights into Fluoroquinolone Resistance in Mycobacterium tuberculosis: Functional Genetic Analysis of gyrA and gyrB Mutations. PLoS One 7:e39754.

57. de Oliveira TLR, Cavalcante FS, Chamon RC, Ferreira RBR, dos Santos KRN. 2019. Genetic mutations in the quinolone resistance-determining region are related to changes in the epidemiological profile of methicillin-resistant Staphylococcus aureus isolates. J Glob Antimicrob Resist 19:236–240.

58. Minarini LAR, Darini ALC. 2012. Mutations in the quinolone resistance-determining regions of gyrA and parC in enterobacteriaceae isolates from Brazil. Brazilian J Microbiol https://doi.org/10.1590/S1517-83822012000400010.

59. de Oliveira Santos IC, Pereira de Andrade NF, da Conceição Neto OC, da Costa BS, de Andrade Marques E, Rocha-de-Souza CM, Asensi MD, D'Alincourt Carvalho-Assef AP. 2019. Epidemiology and antibiotic resistance trends in clinical isolates of Pseudomonas aeruginosa from Rio de janeiro - Brazil: Importance of mutational mechanisms over the years (1995–2015). Infect Genet Evol https://doi.org/10.1016/j.meegid.2019.05.015.

60. Almeida F, Seribelli AA, Cazentini Medeiros MI, Rodrigues D dos P, De MelloVarani A, Luo Y, Allard MW, Falcão JP. 2018. Phylogenetic and antimicrobial resistance gene analysis of Salmonella Typhimurium strains isolated in Brazil by whole genome sequencing. PLoS One https://doi.org/10.1371/journal.pone.0201882.

61. Avalos E, Catanzaro D, Catanzaro A, Ganiats T, Brodine S, Alcaraz J, Rodwell T. 2015. Frequency and geographic distribution of gyrA and gyrB mutations associated with fluoroquinolone resistance in clinical Mycobacterium tuberculosis isolates: A systematic review. PLoS One 10:1–24.

62. Marques M, Cunha EAT, Ruffino-Netto A, Andrade SM de O. 2010. Drug resistance profile of mycobacterium tuberculosis in the state of mato grosso do sul, Brazil, 2000-2006. J Bras Pneumol.

CHAPTER 4 DISCOVERY OF GENOME-WIDE SINGLE NUCLEOTIDE POLYMORPHISMS IN MYCOBACTERIUM TUBERCULOSIS VARIANT BOVIS RAVENEL AS POTENTIAL CONTRIBUTORS OF ATTENUATION

Abstract

Recent experimental calf infection studies show that *Mycobacterium tuberculosis* variant *bovis* strain Ravenel (herein MBO Ravenel), a strain used for pathogenesis studies, has become attenuated in the cattle host. The present study was undertaken to identify genomic features possibly associated with attenuation of the strain. Whole genome sequencing was performed to identify regions of differences (RD) as well as single nucleotide polymorphisms (SNP) to seek an explanation for the attenuation observed in experimental studies. The analysis involved comparison of MBO Ravenel against the genomes of three pathogenic strains of MBO (AF2122-97, 10-7428 and 95-1315). Out of 173 SNPs in MBO Ravenel, nine were unique, missense mutations in genes with annotated putative functions, impacting two major pathways of MBO: 1) Cell wall synthesis & transport [*espH* (A103T), *mmpL8* (V888I), *aftB* (H484Y), *eccC*$_5$ (T507M), *rpfB* (E263G)] and 2) Lipid metabolism & respiration [*mycP*$_1$(T125I), *pks5* (G455S), *fadD*29 (N231S), *fadE29* (V360G)]. While MBO Ravenel is similar to classic MBO in terms of RDs, it is lacking any of the deletions characteristic of the attenuated MBO Bacillus Calmette Guérin strain. It does contain multiple SNPs not seen in other strains across genes for several virulence-associated pathways. Understanding these subtle changes and their role in pathogenesis will heighten our understanding of pathogenesis in this major zoonotic threat.

Introduction

Mycobacterium tuberculosis variant *tuberculosis* (MTB) causes tuberculosis (TB) predominantly in humans (1–4) whereas *Mycobacterium tuberculosis* variant *bovis* (MBO) is the primary cause of TB in cattle and other warm-blooded animals (5–7). It is zoonotic (6, 8), and causes ~150,000 cases of illness in humans every year (9). Over the course of evolution, *Mycobacterium cannetti* underwent deletions (identified as regions of deletions or RDs) and polymorphisms that led to its adaptation in different hosts, giving rise to the MTBC. MTB and MBO are both members of the complex yet primarily infect different hosts. MBO eventually lost a region of its genome (RD1) that further led to its attenuation and formation of MBO Bacillus Calmette Guerrin (MBO BCG) (10–14). Attenuated strain MBO BCG has been used as a vaccine to prevent childhood TB and pulmonary TB (15–17) with variable efficacy (17–20). In recent times, another lab-attenuated strain of MBO has emerged. This strain is identified as MBO-Ravenel and has been maintained at the Trudeau Institute in rabbits – a non-traditional host. MBO Ravenel provokes cellular responses to its antigens in experimentally infected cattle yet leads to no pathological lesions. Because of this deficiency, pathogenesis studies no longer employ MBO Ravenel. A separate series of studies (UNPUBLISHED) performed experimental infections with MBO Ravenel in cattle to define modulation to the host's cellular immune response as well as any associated pathological lesions. The current study expands on these earlier experimental findings and is focused on the identification of genomic factors that could explain the observed clinical attenuation.

Methods

Whole genome sequencing of MBO Ravenel

MBO Ravenel was obtained from National Animal Disease Center (APHIS, USDA) and cultured on Middlebrook 7H10 slants (Hardy Diagnostics, Santa Maria, CA) for 14 days. Colonies were harvested for DNA extraction. Whole genome sequencing entailed paired-end (2 x 150bp) library preparation followed by NovaSeq Illumina sequencing (Novogene). All reads were quality-checked, and adapters trimmed by Novogene's in-house custom software (v1.0). Sequences were then checked for contamination using Kraken 2 with default parameters and the author-provided "Standard" database (21). All reads were used to assemble the genome *de novo* irrespective of Kraken's taxonomy assignment. ABySS v2.1.5 (k-value=96) was used to assemble the genome (22). RagTag v1.1.0 was used to perform MBO-AF2122/97 (GenBank accession number LT708304) reference-based assembly correction followed by scaffolding (23). QUAST v5.0.2 was used to analyze scaffolds (24), and generate Circos plots (25) and Icarus views. Uninformative contigs (<200 bp) were removed before submission to the Prokaryotic Genome Annotation Pipeline (PGAP) v5.1 (26). The draft genome was submitted to both PATRIC (27) and NCBI.

Genome comparisons

Identification of single nucleotide polymorphisms (SNPs) in MBO Ravenel

MBO Ravenel's genomic comparisons were performed against the reference strain MBO AF2122/97, MBO 95-1315 (isolated from deer, PATRIC ID 1765.15) and MBO 10-7428 (isolated from cattle). MBO 10-7428 was submitted to Novogene for resequencing as described above, assembled as a draft genome (PATRIC ID 1765.618; GenBank: JAGEUC000000000.1), and used for

comparative genomics and SNP extraction. Snippy (28) was used to call SNPs from MBO Ravenel, MBO 10-7428 and MBO 95-1315 using MBO AF2122/97 as the reference. MBO Ravenel specific SNPs were derived in an iterative fashion. The snippy-core script was used to extract core SNPs from the four genomes, and SNPs not unique to MBO Ravenel were removed from the analysis. The SNPs that remained at this stage were unique to MBO Ravenel. Among these, only the missense mutations with a putative function were crosschecked with the annotated genome on PATRIC.

Region of difference (RD) analysis

RD-Analyzer (29) was used to identify *in silico* regions of difference for MBO Ravenel when compared to MBO AF2122-97 using raw reads of the available genome.

Results

Whole genome sequencing

The MBO Ravenel genome has been assembled and deposited in publicly accessible databases (GenBank: JAGEUB000000000.1; PATRIC: 1765.617). The sequencing yielded 9,074,522 spots, with 2 150bp reads per spot. Draft assembly of the genome after removal of uninformative contigs (<200bp) yielded 18 final contigs. Total length of the genome was 4,377,551bp with a GC percentage of 65.6%. Length of the longest contig (N50) was 4,371,545bp with a coverage of 625.8x. The NCBI-based Prokaryotic Genome Annotation Pipeline (PGAP) was able to identify 4,058 coding sequences (CDSs), 3 rRNAs, 45 tRNAs, 3 noncoding RNAs, and 192 pseudogenes.

Genome comparisons

MBO-Ravenel SNPs versus MBO-pathogenic strains

A total of 974 single nucleotide polymorphisms (SNPs) were shared between MBO AF2122/97, MBO 95-1315, MBO 10-7428 and MBO Ravenel. One-hundred and seventy-three out of 974 SNPs were uniquely identified in MBO Ravenel. Of these unique SNPs, 95 were missense, but only 54/95 had putative functions assigned. We further selected only SNPs that fell in regions without potential for misassembly as determined by QUAST, leaving a subset of 32 SNPs. Nine of these 32 unique-missense SNPs were identifiable as 'specialty genes' in the functionally annotated genome available on PATRIC (1765.617). According to PATRIC features such as 'Specialty Genes' are identified via "BLASTP-based sequence similarity mapping to reference genes are collected from reputed external databases or manually curated by the PATRIC team. These features (genes) include Antibiotic Resistance: Mapped from CARD, NDARO, and PATRIC-curated AMR genes. Drug Targets: Mapped from DrugBank and TTD". These nine SNPs were cross-checked in Mycobrowser for their functional categorization (30). Two main functional categories were identified (i) cell wall and cell processes (*espH*, *mmpL*8, *aftB*, *eccC*$_5$, *rpf*B) and (ii) respiration or lipid metabolism (*mycP*$_1$, *pks*5, *fadD*29, *fadE*29)

MBO Ravenel's Region of difference

Region of difference (RD) analysis of MBO Ravenel assigned it as classic MBO type. RD9, RD4, RD7, RD8, RD10, RD11 and RD12 were absent. Strain 10-7428 was also classified as classic MBO with same RD regions as that of MBO Ravenel.

Discussion

With the increase in prevalence of extremely drug resistant TB in humans as well as inability

	POS	Gene Name	Gene Identifier	Mycobrowser classification	PRODUCT	DNA sequence change	Amino acid change	Amino acid nomenclature used in this study (one-letter code)
1	4305161	*mycP 1*	MB3913C	Intermediary metabolism and respiration	membrane-anchored mycosin (serine protease) (subtilisin-like protease) (subtilase-like)(mycosin-1)	c.374C>T	p.Thr125Ile	T125I
2	4283808	*espH*	MB3897	Cell wall and cell processes	esx-1 secretion-associated protein	c.307G>A	p.Ala103Thr	A103T
3	4229553	*mmpL8*	MB3853C	Cell wall and cell processes	conserved integral membrane transport protein	c.2662G>A	p.Val888Ile	V888I
4	4208072	*aftB*	MB3835C	Cell wall and cell processes	possible arabinofuranosyltransferase	c.1450C>T	p.His484Tyr	H484Y
5	2015643	*eccC 5*	MB1812	Cell wall and cell processes	esx conserved component esx-5 type vii secretion system protein	c.1520C>T	p.Thr507Met	T507M
6	1129357	*rpfB*	MB1036	Cell wall and cell processes	Probable resuscitation-promoting factor	c.788A>G	p.Glu263Gly	E263G
7	1715469	*pks5*	MB1554C	Lipid metabolism	Probable polyketide synthase	c.1363G>A	p.Gly455Ser	G455S
8	3262864	*fadD29*	MB2974C	Lipid metabolism	fatty-acid-amp ligase (fatty-acid-amp synthetase) (fatty-acid-amp synthase)	c.692A>G	p.Asn231Ser	N231S
9	3929690	*fadE29*	MB3573C	Lipid metabolism	Probable Acyl-CoA dehydrogenase	c.1079T>G	p.Val360Gly	V360G

Table 4-1: Nine MBO Ravenel specific single nucleotide polymorphisms (SNPs) hypothesized to contribute to attenuation of MBO Ravenel strain. These missense SNPs were extracted in comparison to 3 pathogenic MBO strains- AF2122-97, 10-7428 and 95-1315. These SNPs were from regions without potential missassembly points as flagged by QUAST of the draft genome. c. = coding DNA reference sequence, p. = protein-level amino acid sequence (31).

of MBO BCG to provide adequate immunity against all form of TB among all age groups, there is an urgent need to develop better control methodologies. Herein we focus on one of the primary preventive care strategies (vaccination) where we propose that the study of MBO Ravenel could lead to a potential new vaccine candidate. The clinical studies have indicated that MBO Ravenel is attenuated and is able to elicite immune response against infection with MBO in cattle. It is now possible to identify the potential genomic factors associated with MBO Ravenel's clinical

attenuation (32). The associations identified here could provide a better understanding of the bacterium along with facilitation in targeting alternative modes for vaccine development.

The present study utilized the virulent strains MBO AF2122/97, MBO 95-1315 and MBO 10-7428 to draw comparisons against MBO Ravenel. The analysis led to identification of 32 MBO Ravenel specific non-synonymous SNPs. Among these, a subset of 9 SNPs were selected based on affected gene functional annotation. These SNPs may contribute to reduced virulence and an attenuated phenotype as they affect not just cell wall synthesis and transportation associated genes, but also those critical for the metabolism and survival of the mycobacterium.

Cell wall and cell processes associated SNPs- $eccC_5$ (T507M), espH (A103T), mmpL8 (V888I), aftB (H484Y), rpfB (E263G)

Members of MTBC possess unique Type VII secretion systems (ESX systems). This secretion system contributes to virulence (ESX-1, -3, -5), nutrient uptake (ESX-5), metal homeostasis (ESX-3), and export of PE/PPE family proteins (ESX-5)(33). Disruption of these systems is associated with attenuation (34). In MBO Ravenel's ESX-5 system, we identified the polymorphism C1530T (T507M) in *eccC*$_5$. The ATPase encoded by *eccC*$_5$ has 3 nucleotide binding domains (NBD-1, NBD-2, NBD-3). These 3 domains are hypothesized to be essential for substrate recognition (35). Ates *et al.* (2015) demonstrated that NBD mutations impaired bacterial growth. The mutation in MBO Ravenel's *eccC*$_5$ (T507M) falls directly adjacent to NBD-1 (K506), which may destabilize binding and function of this virulence-associated system(36). MBO Ravenel also has mutation G307A (A103T) in *espH,* a gene associated with the ESX-1 virulence system (33).

The mycobacterial membrane protein [large] (*mmpL*) genes encode a broad family of transmembrane-transport proteins believed to be involved in fatty acid transportation (37–40). MBO Ravenel's *mmpL8* gene had the missense mutation G2662A (V888I). The mutation might affect transportation of virulence-associated compounds.

The arabinofuranosyltransferase enzyme is encoded by *aftB*, and is involved in biosynthesis of arabinogalactan that forms a fundamental component of the mycobacterial cell wall (41, 42). Raad *et al.* (2010) developed a Δ*aftB* mutant (42). The mutant had a perturbed outer membrane that sloughed off in substantial quantity into culture media, demonstrating that *aftB* function prevented destabilization of the outer membrane. Jankute *et al.* (2017) demonstrated similar results in *M. smegmatis*(43). In MBO Ravenel, we see one missense SNP C1450T (H484Y) in *aftb.* Whether this mutation results in deficient biosynthesis of arabinogalactan remains to be explored.

Resuscitation promoting factor (RPF), in partnership with RPF-interacting protein A (RipA), facilitates mycobacterial cells in their transition from dormancy to active infection(44–48). It is encoded by the *rpfB* gene. The resuscitation promoting factor cleaves peptidoglycan, employing E292 present in its catalytic pocket (49, 50). In MBO Ravenel, we observe *rpfB* A788G (E263G). Even though this polymorphism is outside Squeglia *et al*.'s observed catalytic pocket (50), the amino acid change might affect the folding of the protein.

Respiration or lipid metabolism - fadD29 (N231S)), fadE29 (V360G), pks5 (G455S), mycP$_1$ (T125I)

Lipids are used as the primary source of carbon by mycobacteria (3, 51, 52). The *fadD* and *fadE* family of genes are involved in long chain fatty-acid synthesis in mycobacteria, encoding

ligase, synthetase and dehydrogenase. The enzyme encoded by *fadD29* converts long chain fatty acids to acyl adenylates (53). These acyl adenylates are required for phenol glycolipid (PGL) production, which in turn are required for the synthesis of outer membrane of MTB (53). In the present study, *fadD29* had SNP A692G (N231S). Synthesis of PGLs is required for mycobacterial viability, so non-synonymous changes like this warrant scrutiny (54, 55). On the other hand, *fadE29* permits MTB to import host cholesterol. The carbon supply is essential for survival during chronic infection. Knockout studies by Thomas *et al.*, 2011 and Gilbert *et al*., 2018, determined that *fadE28* and *fadE29* are essential for degradation of cholesterol-metabolites (56, 57). *fadE29* has missense mutation T1079G (V360G) in MBO-R. It is possible that the combined downstream effect of non-synonymous mutations in both *fadE29* and *fadD29* would reduce bacterial viability.

Polyketide synthase 5, encoded by *pks5*, has been identified as a virulence-associated biomarker of MTBC infection in cattle in our lab (58). The product of *pks5* is thought to be involved in multimethyl-branched fatty acid synthesis required for lipooligosaccharides (LOS) biosynthesis (59). Loss of *pks5* leads to severe MTB growth defects in animal models (60). In MBO Ravenel, we observed a G1363A mutation (G455S) in *pks5*. Whether the missense mutation has an effect on the function of the resulting enzyme requires experimental validation.

The serine protease mycosin ($mycP_1$) is a conserved membrane component of ESX-1 and ESX-5 systems (33). Ohol *et al.* (2010) found that the inhibition of $mycP_1$ protease activity leads to increased ESX-1 substrate secretion as well (61). Hence C374T (T125I) in $mycP_1$ is another polymorphism of interest in MBO Ravenel, as destabilization of this protease is known to lead to dysregulation of the tightly controlled ESX-1 machinery. Considering the ESX-1-containing RD-1

is believed to be a major contributor to attenuation of MBO BCG(10, 11, 62, 63), potential disruption of this system by alternate means could explain similar virulence deficiencies despite RD-1's presence.

Out of 32 total genes with unique missense SNPs, nine with functional annotations have been described here. We hypothesized that a single mutation alone cannot be associated with the attenuation observed in MBO Ravenel. Though a cumulative effect of all the SNPs might have led to impairment of pathogenesis. However, it would be far-fetched to claim these 9 SNPs are the sole cause of MBO Ravenel's attenuation. The remaining 23 unique missense SNPs, as well as those in unannotated genomic locations, must be investigated further.

Conclusion

Akin to MBO BCG that is used as vaccine in humans, our aim was to discover the pathways impacted by mutations in MBO Ravenel. The set of SNPs explored in the work are identified in genes associated with pathogenesis of MBO. Whether they lead to functional changes or not is yet to be explored. Continued studies using MBO Ravenel will deepen our understanding of MBO's mechanisms of infection that could potentially lead to development of a candidate strain for vaccination and prevention of disease.

REFERENCES

REFERENCES

1. Koch; R. 1882. Die Äetiologie der Tuberkulose. Berliner Klin Wochenschrift 15:221–230.

2. Gordon S V., Parish T. 2018. Microbe profile: Mycobacterium tuberculosis: Humanity's deadly microbial foe. Microbiol (United Kingdom) 10.

3. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S V., Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajahdream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulsten JE, Taylor K, Whitehead S, Bartell BG. 1998. Erratum: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence (Nature (1998) 393 (537-544)). Nature.

4. Chai Q, Zhang Y, Liu CH. 2018. Mycobacterium tuberculosis: An adaptable pathogen associated with multiple human diseases. Front Cell Infect Microbiol. Frontiers Media S.A.

5. Tippett SG. 1948. Bovine TuberculosisBritish Medical Journal.

6. Wilkins MJ, Meyerson J, Bartlett PC, Spieldenner SL, Berry DE, Mosher LB, Kaneene JB, Robinson-dunn B, Stobierski MG, Boulton ML. 2008. Mycobacterium bovis Infection and Bovine 14:657–660.

7. Mathews F, Macdonald DW, Taylor GM, Gelling M, Norman RA, Honess PE, Foster R, Gower CM, Varley S, Harris A, Palmer S, Hewinson G, Webster JP. 2006. Bovine tuberculosis (Mycobacterium bovis) in British farmland wildlife: The importance to agriculture. Proc R Soc B Biol Sci 273:357–365.

8. Kleeberg HH. 1984. Human tuberculosis of bovine origin in relation to public health*. Rev sci tech Off int Epiz 3:23–27.

9. WHO. 2019. Global Tuberculosis Report.

10. Heijmenberg I, Husain A, Sathkumara HD, Muruganandah V, Seifert J, Miranda-Hernandez S, Kashyap RS, Field MA, Krishnamoorthy G, Kupz A. 2021. ESX-5-targeted export of ESAT-6 in BCG combines enhanced immunogenicity & efficacy against murine tuberculosis with low virulence and reduced persistence. Vaccine https://doi.org/10.1016/j.vaccine.2021.08.030.

11. Waters WR, Palmer M V., Nonnecke BJ, Thacker TC, Scherer CFC, Estes DM, Hewinson RG,

Vordermeier HM, Barnes SW, Federe GC, Walker JR, Glynne RJ, Hsu T, Weinrick B, Biermann K, Larsen MH, Jacobs WR. 2009. Efficacy and immunogenicity of Mycobacterium bovis ΔRD1 against aerosol M. bovis infection in neonatal calves. Vaccine 27:1201–1209.

12. Pym AS, Brodin P, Majlessi L, Brosch R, Demangel C, Williams A, Griffiths KE, Marchal G, Leclerc C, Cole ST. 2003. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. Nat Med 9:533–539.

13. Copin R, Coscollá M, Efstathiadis E, Gagneux S, Ernst JD. 2014. Impact of in vitro evolution on antigenic diversity of Mycobacterium bovis bacillus Calmette-Guerin (BCG). Vaccine 32:5998–6004.

14. Bespiatykh D, Bespyatykh J, Mokrousov I, Shitikov E. 2021. A Comprehensive Map of Mycobacterium tuberculosis Complex Regions of Difference. mSphere 6.

15. Chretien J. 1992. Tubercle and lung disease: Introduction. Tuber Lung Dis.

16. McShane H. 2011. Tuberculosis vaccines: Beyond bacilli Calmette-Guérin. Philos Trans R Soc B Biol Sci.

17. Damjanovic D, Khera A, Afkhami S, Lai R, Zganiacz A, Jeyanathan M, Xing Z. 2015. Age at mycobacterium bovis BCG priming has limited impact on anti-tuberculosis immunity boosted by respiratory mucosal AdHu5Ag85A immunization in a murine model. PLoS One 10.

18. Lule SA, Mawa PA, Nkurunungi G, Nampijja M, Kizito D, Akello F, Muhangi L, Elliott AM, Webb EL. 2015. Factors associated with tuberculosis infection, and with anti-mycobacterial immune responses, among five year olds BCG-immunised at birth in Entebbe, Uganda. Vaccine 33.

19. Nieuwenhuizen NE, Kaufmann SHE. 2018. Next-generation vaccines based on Bacille Calmette-Guérin. Front Immunol.

20. Moliva JI, Turner J, Torrelles JB. 2015. Prospects in Mycobacterium bovis Bacille Calmette et Guérin (BCG) vaccine diversity and delivery: Why does BCG fail to protect against tuberculosis? Vaccine.

21. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biol 20:257.

22. Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, Birol I. 2017. ABySS 2.0: Resource-efficient assembly of large genomes using a Bloom filter. Genome Res 27:768–777.

23. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. 2019. RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. Genome Biol 20.

24. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075.

25. Krzywinski M, Schein J, Birol II, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA, STeven JJ, Marra MA, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645.

26. Tatusova T, Dicuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624.

27. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, MacHi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJC, Yoo HS, Zhang C, Zhang Y, Sobral BW. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res 42.

28. Seemann T. 2015. Snippy: Rapid haploid variant calling and core SNP phylogeny. GitHub.

29. RD-analyzer.

30. Kapopoulou A, Lew JM, Cole ST. 2011. The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. Tuberculosis 91:8–13.

31. S O, ML G, JT den D, RB W. 2007. Standard mutation nomenclature in molecular diagnostics: practical and educational challenges. J Mol Diagn 9:1–6.

32. Khare S, Hondalus MK, Nunes J, Bloom BR, Garry Adams L. 2007. Mycobacterium bovis ΔleuD auxotroph-induced protective immunity against tissue colonization, burden and distribution in cattle intranasally challenged with Mycobacterium bovis Ravenel S. Vaccine 25:1743–1755.

33. Vaziri F, Brosch R. 2019. ESX/Type VII Secretion Systems—An Important Way Out for Mycobacterial Proteins. Microbiol Spectr 7.

34. Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK. 1996. Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. J Bacteriol 178:1274–1282.

35. DiGiuseppe Champion PA, Stanley SA, Champion MM, Brown EJ, Cox JS. 2006. C-terminal

signal sequence promotes virulence factor secretion in Mycobacterium tuberculosis. Science (80-) 313:1632–1636.

36. Ates LS, Ummels R, Commandeur S, van der Weerd R, Sparrius M, Weerdenburg E, Alber M, Kalscheuer R, Piersma SR, Abdallah AM, Abd El Ghany M, Abdel-Haleem AM, Pain A, Jiménez CR, Bitter W, Houben ENG. 2015. Essential Role of the ESX-5 Secretion System in Outer Membrane Permeability of Pathogenic Mycobacteria. PLoS Genet 11:1–30.

37. Converse SESE, Mougous JDJD, Leavell MDMD, Leary JAJA, Bertozzi CR, Cox JSJS. 2003. MmpL8 is required for sulfolipid-1 biosynthesis and Mycobacterium tuberculosis virulence. Proc Natl Acad Sci USA 100:6121– 6126.

38. Seeliger JC, Holsclaw CM, Schelle MW, Botyanszki Z, Gilmore SA, Tully SE. 2012. Elucidation and chemical modulation of sulfolipid-1 biosynthesis in Mycobacterium tuberculosis. J Biol Chem 287:7990– 8000.

39. Lamichhane G, Tyagi S, Bishai WR. 2005. Designer arrays for defined mutant analysis to detect genes essential for survival of Mycobacterium tuberculosis in mouse lungs. Infect Immun 73:2533– 2540.

40. Domenech P, Reed MB, Barry CE. 2005. Contribution of the Mycobacterium tuberculosis MmpL protein family to virulence and drug resistance. Infect Immun 3492– 3501.

41. Seidel M, Alderwick LJ, Birch HL, Sahm H, Eggeling L, Besra GS. 2007. Identification of a novel arabinofuranosyltransferase AftB involved in a terminal step of cell wall arabinan biosynthesis in Corynebacterianeae, such as Corynebacterium glutamicum and Mycobacterium tuberculosis. J Biol Chem 282.

42. Raad RB, Méniche X, De Sousa-D'Auria C, Chami M, Salmeron C, Tropis M, Labarre C, Daffé M, Houssin C, Bayan N. 2010. A deficiency in arabinogalactan biosynthesis affects Corynebacterium glutamicum mycolate outer membrane stability. J Bacteriol 192.

43. Jankute M, Alderwick LJ, Noack S, Veerapen N, Nigou J, Besra GS. 2017. Disruption of mycobacterial aftB results in complete loss of terminal β(1 → 2) arabinofuranose residues of Lipoarabinomannan. ACS Chem Biol 12.

44. Kana BD, Gordhan BG, Downing KJ, Sung N, Vostroktunova G, Machowski EE, Tsenova L, Young M, Kaprelyants A, Kaplan G, Mizrahi V. 2008. The resuscitation-promoting factors of Mycobacterium tuberculosis are required for virulence and resuscitation from dormancy but are collectively dispensable for growth in vitro. Mol Microbiol 67.

45. Kana BD, Mizrahi V. 2010. Resuscitation-promoting factors as lytic enzymes for bacterial growth and signalingFEMS Immunology and Medical Microbiology.

46. Hett EC, Chao MC, Steyn AJ, Fortune SM, Deng LL, Rubin EJ. 2007. A partner for the resuscitation-promoting factors of Mycobacterium tuberculosis. Mol Microbiol 66.

47. Hett EC, Chao MC, Deng LL, Rubin EJ. 2008. A mycobacterial enzyme essential for cell division synergizes with resuscitation-promoting factor. PLoS Pathog 4.

48. Russell-Goldman E, Xu J, Wang X, Chan J, Tufariello JAM. 2008. A Mycobacterium tuberculosis Rpf double-knockout strain exhibits profound defects in reactivation from chronic tuberculosis and innate immunity phenotypes. Infect Immun 76.

49. Cohen-Gonsaud M, Barthe P, Bagnéris C, Henderson B, Ward J, Roumestand C, Keep NH. 2005. The structure of a resuscitation-promoting factor domain from Mycobacterium tuberculosis shows homology to lysozymes. Nat Struct Mol Biol 12.

50. Squeglia F, Romano M, Ruggiero A, Vitagliano L, De Simone A, Berisio R. 2013. Carbohydrate recognition by RpfB from mycobacterium tuberculosis unveiled by crystallographic and molecular dynamics analyses. Biophys J 104.

51. Lima P, Sidders B, Morici L, Reader R, Senaratne R, Casali N, Riley LW. 2007. Enhanced mortality despite control of lung infection in mice aerogenically infected with a Mycobacterium tuberculosis mce1 operon mutant. Microbes Infect 9:1285–1290.

52. Tekaia F, Gordon S., Garnier T, Brosch R, Barrell B., Cole S. 1999. Analysis of the proteome of Mycobacterium tuberculosis in silico. Tuber Lung Dis 79:329–42.

53. Siméone R, Léger M, Constant P, Malaga W, Marrakchi H, Daffé M, Guilhot C, Chalut C. 2010. Delineation of the roles of FadD22, FadD26 and FadD29 in the biosynthesis of phthiocerol dimycocerosates and related compounds in mycobacterium tuberculosis. FEBS J 277:2715–2725.

54. Trivedi OA, Arora P, Sridharan V, Tickoo R, Mohanty D, Gokhale RS. 2004. Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. Nature 428.

55. Siméone R, Léger M, Constant P, Malaga W, Marrakchi H, Daffé M, Guilhot C, Chalut C. 2010. Delineation of the roles of FadD22, FadD26 and FadD29 in the biosynthesis of phthiocerol dimycocerosates and related compounds in Mycobacterium tuberculosis. FEBS J 277:2715–2725.

56. Gilbert S, Hood LC, Seah SYK. 2018. Characterization of an aldolase involved in cholesterol side chain degradation in Mycobacterium tuberculosis. J Bacteriol 200.

57. Thomas ST, VanderVen BC, Sherman DR, Russell DG, Sampson NS. 2011. Pathway profiling in Mycobacterium tuberculosis: Elucidation of cholesterol-derived catabolite and enzymes that catalyze its metabolism. J Biol Chem 286.

58. Lamont EA, Janagama HK, Ribeiro-lima J, Vulchanova L, Seth M, Yang M, Kurmi K, Waters WR. 2014. Circulating Mycobacterium bovis Peptides and Host Response Proteins as Biomarkers for Unambiguous Detection of Subclinical Infection 52:536–543.

59. Etienne G, Malaga W, Laval F, Lemassu A, Guilhot C, Daffé M. 2009. Identification of the polyketide synthase involved in the biosynthesis of the surface-exposed lipooligosaccharides in mycobacteria. J Bacteriol 191:2613–21.

60. Rousseau C, Sirakova T, Dubey V, Bordat Y, Kolattukudy P, Gicquel B, Jackson M. Virulence attenuation of two Mas-like polyketide synthase mutants of Mycobacterium tuberculosis. Microbiol 149:1837–1847.

61. Ohol YM, Goetz DH, Chan K, Shiloh MU, Craik CS, Cox JS. 2010. Mycobacterium tuberculosis MycP1 Protease Plays a Dual Role in Regulation of ESX-1 Secretion and Virulence. Cell Host Microbe 7:210–220.

62. Ganguly N, Siddiqui I, Sharma P. 2008. Role of M. tuberculosis RD-1 region encoded secretory proteins in protective response and virulence. Tuberculosis.

63. Pym AS, Brodin P, Brosch R, Huerre M, Cole ST. 2002. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti. Mol Microbiol 46.

CHAPTER 5 VALIDITY OF MTBC-SPECIFIC BIOMARKERS IN EXPERIMENTALLY INFECTED NON-HUMAN PRIMATES

Abstract

Pathogen specific biomarkers (PSBs) are highly specific targets for disease detection. A panel of 16 PSBs were discovered in our laboratory of which three - Pks5, MB2515c and MB1895c, were validated in experimental as well as naturally infected cattle and deer sera. Using indirect ELISA developed for cattle and deer, we tested for the presence of these biomarkers in experimentally infected non-human primates. We also tested for their ability to differentiate between infected from uninfected primate samples. The results prove the applicability of the PSBs in detection of not just ruminants, but also primates. This provides an opportunity for enhanced diagnostics to be applied for surveillance and monitoring of TB not just in primate sanctuaries around the world but also in humans in the future.

Introduction

According to WHO *Mycobacterium tuberculosis* is among the world's top 10 disease causing agents in humans(1). It is an infectious disease estimated to infect one-fourth of the world's population (1). The End TB strategy and UN's Sustainable Development Goals (SDGs) have set an ambitious target to end the epidemic of TB by 2030. To achieve this target breakthroughs in vaccines, diagnostics and treatment are required by the year 2025. Studies have indicated the use of pathogen-specific-biomarkers as a more specific diagnostic target than those released by the host themselves (2–4). Increased test specificity would allow enhanced screening in high

burden areas, and also overcome the problem of missing out immunocompromised people (e.g those co-infected with HIV/AIDS or have chronic TB) and facilitates detection of latent TB.

MTBC is comprised of organisms with highly similar genomes. MTB and MBO are members of the complex with genome similarity of 99.95% (5, 6). The genomic similarity and zoonotic nature of MBO is of importance because many of the same diagnostic tools are used since it overlaps with the human and animal diagnostics for TB. Tuberculin skin test, interferon-gamma assay, as well as antibody tests used for animal TB screening are also used for human and primate TB testing. These tests have proved to be useful since these are reliable host-biomarkers (7, 8). Using pathogen-specific biomarker would enhance the specificity of the diagnostics that would save millions of dollars in low-prevalence countries both for human and bovine TB by preventing false positives. This idea has been explored by Lamont et al., (2014) where they identified a unique panel of 16 PSBs in cattle infected with *Mycobacterium bovis*. Of these three PSB biomarkers Pks5, MB2515c and MB1895c were employed successfully to detect bovine TB in naturally as well as experimentally infected cattle sera. Wanzala et al. (2016) took the work one step further and validated these biomarkers in white-tailed deer infected with MBO in both experimental and naturally infected animals with contemporaneous controls. The primary goal of this project was to develop rapid, simple and point-of-care diagnostics for tuberculosis across multiple species, including humans, in alignment with the WHO recommended 'ASSURED' (Affordable. Specific. Sensitive. User-friendly. Rapid and Robust. Equipment free. Delivered to end users.) guideline (9). The objective of the current body of work was to test the three PSBs

(Pks5, MB2515c and MB895c) in non-human-primates infected with the MTB, after the successful validation work performed in cattle and deer.

Methods

Biomarkers & Antibodies

The three biomarkers Pks5, MB2515c and MB1895c previously identified and validated in cattle and white-tailed deer were evaluated in a primate experimental-infection model (10, 11). The antibodies previously generated and used by Lamont et al. 2014 to detect the biomarkers in cattle, were optimized for the purpose of indirect ELISA for each biomarker in primates(10).

Sample

A longitudinal experimental infection model of tuberculosis developed at Tulane University Primate Center was garnered for this study. Eight non-human primates (NHP) were infected with MTB CDC1551 with (n=5) or without (n=3) a pre-infection with simian immunodeficiency virus (SIVmac239 300 TCID50 IV). Plasma was collected from animals infected with SIVmac239 at baseline (pre-infection), and then week 3, 15, 18 and 21 post MTB infection. The MTB infected animals were tuberculin skin tested at three time points during the study: pre-infection (26th Sept 2011), week 5 (02nd Nov 2011), and week 10 (14th Dec 2011).

Plasma samples of these 8 animals was tested for the presence of MTBC-specific biomarkers. Serum collected at baseline was used as an internal-negative control. Further, a set of 21 uninfected non-human primate serum samples from Yerkes National Primate Research Center also served as negative controls for the study.

Indirect ELISA

To test the presence of the biomarkers Pks5, MB2515c and MB1895c, indirect ELISA was performed as described (10). Serum was diluted (Pks5 1:100 dilution, MB2515c and MB1895c 1:50 dilution) using carbonate-bi-carbonate buffer (pH 9.6, 0.05M). A 96 well plate was coated with 50µL of diluted serum and incubated overnight at 4°C. The wells were dry blotted followed by washing thrice with 200µL of 1X PBS-tween 20 buffer (10mM sodium phosphate, 0.15M NaCl, 0.05% Tween-20, pH 7.5). The wells were blocked with 200µL of 5% dry milk powder in 1X TBS buffer for two hours at 37°C. They were washed again thrice with 200µL of 1X PBS-Tween. The wells were incubated with 100µL of primary monoclonal antibody (Pks5-1:10,000, MB2515c and MB1895c 1:5000) diluted in 1% dry milk in 1X TBS-tween, for one hour at room-temperature. After the incubation, the wells were washed thrice with 1X PBS-tween. 100µL of secondary antibody (neutravidin-HRP, 1:5000 dilution with 1% dry milk in 1X TBS-tween) was added to the wells and incubated for one hour at room temperature. The wells were then washed as described before. An aliquote of 100-µL of TMB was added to the wells and incubated at room temperature in the dark. After 30 minutes the reaction was quenched with 100µL of 0.05M sulfuric acid. The optical densities were recorded at 450nm.

Data analysis

Every serum sample was assayed in triplicates. The optical density ($O.D._{450nm}$) of the three wells was averaged. O.D. of the blank wells was deducted from the averaged O.D of each negative sample. The resulting O.D was graphed for each biomarker over time.

Results & Discussions

The indirect ELISA results are clearly indicative of increase in all three TB-specific biomarkers as infection progressed in the NHPs. Pks5 (Rv1527c) and MB2515c (Rv2488c; LuxR family) show consistent rise in abundance as infection progressed. While MB1895c (Rv1864c; hypothetical conserved protein) showed an increasing trend, it was not diagnostic in early stages of infection in the primate model. One-month post-infection $O.D._{450nm}$ for Pks5 was ≤ 0.3, for MB2515c ≤ 0.5, for MB1895c ≤ 0.2. By the fifth month post-infection. $O.D._{450nm}$ for Pks5 was ≤ 0.7, for MB2515c ≤ 1.2, for MB1895c ≤ 0.3. The twenty-one uninfected primates had $O.D._{450nm}$ for Pks5 ≤ 0.02, for MB2515c ≤ 0.03, for MB1895c ≤ 0.04. This suggests that MB2515c is the most sensitive early-stage biomarker for NHP TB diagnostics, whereas MB1895c was the least detectable.

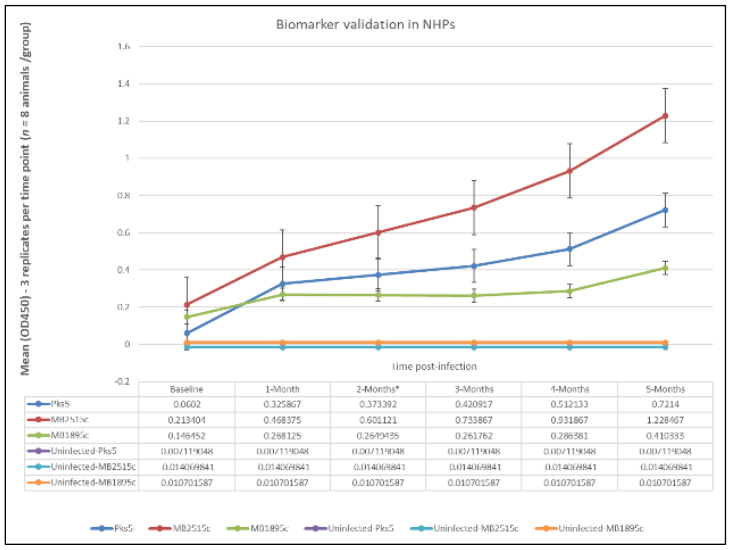

Figure 5-1: The top 3 trendlines correspond to Mycobacterium tuberculosis complex specific biomarkers in infected non-human-primates (NHPs): Pks5 (blue), MB2515c (red), MB1895c (green). Of the 3 biomarkers MB2515c was the most effective detector of infection.

The study was performed on five experimentally infected primate sera. While the sample size is relatively small, it was powered for infection studies. Most all vaccine-challenge studies for TB using NHP model use similar sample sizes. That said, biomarker diagnostics work would be strengthened further by increasing the sample sizes. Since NHP studies are expensive, few primate centers are performing experimental infection studies and sample availability is scarce. A larger sample size of naturally infected primate population along with associated disease signalment history as well as concordance with other diagnostic tests would enable us to make more generalizable recommendations for biomarker's use in primate sanctuaries and likely for human TB as well. Evaluating these biomarkers in different primate species would strengthen our observations of using the trio for TB detection in a wide range of animals with species-specific optimizations.

The findings described above support the work done earlier by Lamont et al. (2014) in cattle and Wanzala et al. (2016) in deer. The pathogen specific biomarkers are now confirmed to be detectable in three different animal species. In cattle Pks5 performed better than MB2515c or MB1895c, though all three of them were able to differentiate between uninfected and infected (and exposed) cases. In deer, Pks5 was the best diagnostic marker. Signal-to-noise ratios of both Pks5 and MB2515c increased gradually and reached a peak at day 60 (2 months post-infection). Yet unlike cattle and deer where Pks5 had performed better than MB2515c and MB1895c, in primates MB2515c appeared to be the best biomarker as it was detectable much earlier post-infection as well as at a higher O.D. than other two biomarkers. Though like deer, $O.D._{450nm}$ of

both Pks5 and MB2515c gradually rises over the course of infection in primates, the level of MB1895c stabilizes or drops after 1-month post-infection.

The work convincingly shows the detectability of the MTC-specific biomarkers across varied species. The levels of all three biomarkers varies in different animal species at different time points. But combined they can prove to be an especially useful trio-target for diagnostics. Future work in primate sanctuaries in Africa is currently underway and should help validate these biomarkers for routine field use.

REFERENCES

REFERENCES

1. WHO. 2020. Global TUBERCULOSIS report 2020.

2. Aagaard C, Govaerts M, Okkels LM, Andersen P, Pollock JM. 2003. Genomic approach to identification of Mycobacterium bovis diagnostic antigens in cattle. J Clin Microbiol 41:3719–3728.

3. Russell DG. 2007. Who puts the tubercle in tuberculosis? Nat Rev Microbiol 5:39–47.

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

5. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S V., Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajahdream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulsten JE, Taylor K, Whitehead S, Bartell BG. 1998. Erratum: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence (Nature (1998) 393 (537-544)). Nature.

6. Garnier T, Eiglmeier K, Camus J, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C, Simon S, Harris B, Atkin R, Doggett J, Mayes R, Keating L, Wheeler P, Parkhill J, Barrell B, Cole S, Gordon S, Hewinson R. 2003. The complete genome sequence of Mycobacterium bovis. Proc Natl Acad Sci U S A 100:7877–82.

7. Metwally MA, Hamouda HM, Yassin AS, Amin MA, Riad emad M. 2020. Interferon Gamma Release Assay, A Powerful Tool for the Detection of Human and Bovine Tuberculosis in the Greater Cairo Area Compared to Other Diagnostic Tools. Sains Malaysiana 49:785–792.

8. Waters WR, Vordermeier HM, Rhodes S, Khatri B, Palmer M V, Maggioli MF, Thacker TC, Nelson JT, Thomsen B V, Robbe-austerman S, Garcia DMB, Schoenbaum MA, Camacho MS, Ray JS, Esfandiari J, Lambotte P, Greenwald R, Grandison A, Sikar-gang A, Lyashchenko KP. 2017. Potential for rapid antibody detection to identify tuberculous cattle with non-reactive tuberculin skin test results. BMC Vet Res 1–7.

9. Mabey D, Peeling RW, Ustianowski A, Perkins MD. 2004. Diagnostics for the developing world. Nat Rev Microbiol.

10. Lamont EA, Janagama HK, Ribeiro-lima J, Vulchanova L, Seth M, Yang M, Kurmi K, Waters

WR. 2014. Circulating Mycobacterium bovis Peptides and Host Response Proteins as Biomarkers for Unambiguous Detection of Subclinical Infection 52:536–543.

11. Wanzala SI, Palmer M V., Waters WR, Thacker TC, Carstensen M, Sreevatsan S. 2017. Evaluation of pathogen-specific biomarkers for the diagnosis of tuberculosis in white-tailed deer (78:729–734.

CHAPTER 6 DEVELOPMENT OF A MULTIDIMENSIONAL PROTEOMIC APPROACH TO DETECT CIRCULATING IMMUNE COMPLEXES IN CATTLE EXPERIMENTALLY INFECTED WITH *MYCOBACTERIUM BOVIS*

Abstract

The objective of the study was to evaluate a high-resolution method to identify pathogen-specific biomarkers in serum of calves infected with *Mycobacterium bovis*. Serum samples from four calves infected with MBO were collected before and after infection at weeks 9, 14, 15, 31, and 36. Immune-complex-associated mycobacterial antigens in the serum were enriched using an immunochromatography method termed, dual path platform (DPP). All regions of antigen capture zones, that consisted of monospecific rabbit polyclonal antibodies raised against MTB lysates, on DPP strips were excised and analyzed by multidimensional proteomics. The resulting proteins were then passed through 4 rigorous peptide quality filters. False-hits, decoys, non-MTBC proteins were all removed followed by individual quality check of those remaining. Peptides were then checked on NCBI's BLASTp for MTBC specificity.

Proteins in two of the animals passed the multipronged-highly stringent peptide quality analysis. Animal#54 had seven unique MTBC proteins at week 14 post-infection, while animal#56 had four at week 36 post-infection along with one immunoglobulin.

MTBC -specific peptides identified in this study were identified in 2 animals and at 2 separate time points post infection. Further studies with better enrichment protocols and using larger sample sizes and replications are required to develop a TB-specific diagnostic tool for bovine tuberculosis.

Introduction

Mycobacterium bovis causes tuberculosis primarily in cattle but it is also zoonotic. Transmission to humans occurs through close contact with infected animals or via consumption of contaminated animal products (e.g., unpasteurized milk or dairy products) (1–3). The primary screening test used in the field is tuberculin-based skin test which is time-consuming, labor intensive and associated with low sensitivity and variable specificity. Variability in specificity is caused by species differences and technique being used(4, 5). Ultimately a false-positive can lead to a considerable financial burden on farmers deterring control measures. Thus, there is a need for highly specific rapid field tests that are cost effective.

Immune complexes are formed by the non-covalent binding of antigens with antibody molecules circulating real-time(6). Lyashchenko et al.(7), reported the presence of *Mycobacterium* specific immune complexes in cattle experimentally infected with MBO detectable by the dual-path platform (DPP) assay that utilizes polyclonal antibodies against MTB whole-cell antigens. This provided an unprecedented opportunity to interrogate MTBC-specific antigens enriched by polyclonal tuberculosis-specific antibodies using high resolution technique of liquid chromatography followed by dual mass-spectrometry (LC-MS/MS). LC-MS/MS can detect proteins at abundances as low as 10^{-15} moles, thereby enabling discovery of circulating in infected animals. In the present study, high-resolution multidimensional mass spectrometry analysis of the DPP-captured immune complexes was evaluated for its ability to identify the captured MBO-specific peptides that may aid in the development of a highly accurate tuberculosis diagnostics for animals and humans.

Materials & Methods

Seven Holstein calves obtained from a TB-free herd in IA and housed in a biosafety level 3 (BSL-3) facility at the National Animal Disease Center, Ames, IA were infected at 11 months of age with 8×10^3 CFU of virulent MBO (95-1315; USDA Animal Plant and Health Inspection Service [APHIS] designation) by aerosol. This strain of MBO was isolated from a white-tailed deer in Michigan, USA. Serum was sampled for serum at multiple time points pre- and post-infection over the next 11.5 months at which point the animals were euthanized (7, 8). Necropsy of all the calves revealed presence of granulomas in multiple organs specific to bovine tuberculosis and bacterial culturing from infected tissues confirmed the presence of MBO in all 7 animals infected.

In this pilot study we focused on 4 out of the 7 calves present in the original study(7), since they had the highest levels of circulating immune complexes to increase the probability of biomarker discovery. Pre- and post-inoculation samples collected at weeks 9, 14, 15, 31, and 36 were used to identify mycobacterial specific peptides. To characterize the circulating immune complexes-associated with MTBC, a rapid DPP-Ag assay was performed (Figure 6-1). The DPP antigen capture zone (test line) was coated with rabbit polyclonal antibodies raised against MTB whole-cell lysate to enable capture of mycobacterial antigen-antibody complexes(7, 8). Pre-infection (baseline) sera from these four animals served as negative controls. Triplicates of each time point from every animal were made pre and post-infection (which summed up to 27 DPP-Ag assay strips for analysis) for each week 0, 9, 14, 31 and 36. A 50 µL aliquot of serum sample was placed on three independent DPP-Ag strips for each time-point, to allow for antigen enrichment of molecules on the capture zone, which were then processed as one single sample

to allow for maximum enrichment, enhanced sensitivity, efficient use of the LC-MS/MS and improved proteomics profile generation.

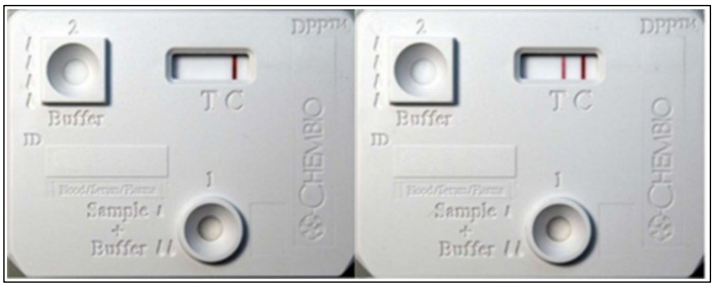

Figure 6-1: Dual-path platform assay kit showing positive and negative controls. Dual-path platform assay was used to detect circulating antigen-antibody complexes in calves infected with Mycobacterium bovis. The rabbit polyclonal antibodies immobilized on the test line(T) acted as the capture reagent for the circulating immune-complexes in the infected animal's serum as well as signal detector when coated onto nano-gold-particles. DPP strip case: ***Left****: Negative Control (serum from uninfected animals),* ***Right****: Positive control.*

The immune-complex capture zone of 2 mm width was excised and analyzed by LC-MS/MS analysis performed at University of Minnesota's Center for Mass Spectrometry and Proteomics (CMSP). Triplicates of DPP-assays for every animal were pooled for analysis. A region 2 mm upstream of the antigen capture zone (or the DPP test line) was also analyzed by LC-MS/MS. The enormous volume of peptide data generated by LC-MS/MS was passed through a series of stringent filters before the final candidates were considered.

First, PEAKS (Bioinformatics Solutions Inc.) software was used to query peptides generated in each triplicate-pooled-sample through LC/MS-MS against a database that included all documented peptides from MTBC, cattle and rabbit proteins. These results were then analyzed

by Scaffold (version Scaffold_4.7.5, Proteome Software Inc., Portland, OR) to validate all MS/MS based peptide identifications and to allow combined visualization of all sample results. All identified peptides were compared against a decoy database (generated in Scaffold_4.7.5), consisting of randomized peptide sequences, to remove any spurious hits. Second, any protein that matched against the decoy database, was removed from further analysis. We focused only on the MTBC proteins because they offer highest possible specificity for bovine tuberculosis diagnostics. The third filter was based on an individual quality check of the proteins within Scaffold. Peptide identifications were accepted in Scaffold if they could be established at greater than 95.0% probability by the Peptide Prophet algorithm(9) with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least 2 identified peptides. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Peptides and proteins that were selected in the third filter had percent probabilities varying from 74 to 100%. The fourth and the last filter was the identification of MTBC specificity using the National Center for Biotechnology Information (NCBI)'s non-redundant database BLASTp (basic local alignment search tool for proteins) analysis where two aspects were investigated: (1) *E*-value ($<1e^{-10}$) and (2) the species match of the peptides. If the proteins matched with any bacteria other than MTBC, they were excluded from further consideration. Additionally, if any peptides had an *E*-value higher than $1e^{-10}$, which suggested that the species match was likely non-random, they were also removed from further consideration. This last filter was excessively stringent as it eliminated most of peptide hits

discovered after decoy database search. Some of the peptides eliminated may still be useful in a future validation study.

The same pipeline was followed for identifying cattle specific immunoglobulins, where immunoglobulins were passed through all the filters described for MTBC proteins. Additionally, the proteins that overlapped between pre-infection and post-infection test-lines were excluded as it suggested that they were not associated with the MBO infection, rather existed in the background.

Results & Discussion

The peptides generated from LC-MS/MS analysis resulted in identification of 26,945 proteins. Forty-nine percent of these were eliminated after the decoy database search. Of these, 3.73% were identified with the MTBC repertoire, 26.02% proteins were of host (bovine) origin and 21.35% were of leporine origin. DPP strips of all post-infection samples, except at week 31, had MBO proteins. After analysis, 11 MTBC-specific proteins were identified in two MBO-infected animals (Table 6-1).

Animal ID	Pre-infection			*Mycobacterium bovis* infection			
	DPPAg result[a])	*Mycobacterium bovis* proteins identified[b]	Cattle specific immunoglobulins	Week post-inoculation	DPP-Ag result[a]	*Mycobacterium bovis* proteins identified	Cattle specific immunoglobulins
51	0	0	0	9	124	0	0
54	0	0	0	14	788	7	0
				15	772	0	0
56	0	0	0	36	485	4	1
57	0	0	0	31	447	0	0

Table 6-1: Enumeration of pathogen-derived proteins detected by mass spectrometry from DPP-Ag assay strips processed with serum samples from cattle experimentally infected with Mycobacterium bovis.

[a]DPP reader data (reflectance) in relative light units obtained as described (7)

[b]Pooled DPP-Ag strips processed with pre-infection sera from four calves.

At week 14 (post-infection) serum from animal #54 showed 7 proteins that corresponded to peptides in MTBC with a BLAST *E*-value lower than $1e^{-10}$ (Table 6-2 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6028705/table/T2/). At week 36 post-infection, serum of animal #56 had 4 proteins that corresponded to MTBC with an E-value lower than $1e^{-10}$ (Table 6-2).

Mycobacterium tuberculosis complex proteins at wk14	Peptide sequences used for BLASTp analysis
Acyltransferase	QDGSASYDAAVR-MLKAGELVGVYPEATISR
Esterase	VFGAADPR-FACVVRAFASMFPGR
LLM class F420-dependent oxidoreductase	QKDYDEYGYR-FGTAGSRLDDLAAPLPR
Transposase, partial	MDPTEDQARALAR-VTGIGTVKPSLRVLR
Transcriptional regulatory protein embr2	FGILGPLEISAGFRSLPLGTPK-SPLGRLPLR
Hypothetical protein Mb3478	GASPATAAR-LPPALNPDDADALPTTDRLTTR
Polyketide synthase	DGDRVLAIVR-LVDAPLPSWTHRTLMLSR-MFNSLGIQYGPAFSGLVAVHTAR-LFVVTRSAASVLPSDLANLEQAGMR
***Mycobacterium tuberculosis* complex proteins at wk36**	**Peptide sequences used for BLASTp analysis**
Helicase helz	VYAHHGGARLHGEALRDHLER-RGNVLAAMAKLK-IDEMIEEKKALADLVVTDGEGWLTELST
Hypothetical protein Mb1791	FGVTINDVVVALCAGALRR-VPSQISDPAQR
Hypothetical protein Mb2390c	HGHGRDVAAHR-TGGHRQASSRIK-HQKPGDVPRDPRC
Chromosome partition protein Smc	LDTMAANLARLTDLTTELR-LAVRTAEER
Cattle Immunoglobulin at week 36	**Peptide sequences used for BLASTp analysis**
PREDICTED: killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 2 isoform X2	GEMLTSGHAPADFVIGPMTLASAGTYR

Table 6-2: List of Mycobacterium tuberculosis complex-specific high confidence proteins at week 14 and week 36 and cattle immunoglobulin at week 36 that passed exclusion criteria.

At week 14 post-infection in animal #54 polyketide synthase was detected, which plays a role in the growth of the bacteria and is considered a potential virulence factor(10). The detection of this protein at such an early stage in MBO infection agrees with other studies(11, 12) where polyketide synthase was detected through different techniques but at similar time points. Lamont et al.(11), showed that polyketide synthase can be used as a useful marker for detecting MBO infection in a multi-cut off fashion, based on the prevalence of the disease.

Killer cell immunoglobulin-like receptor ($1e^{-18}$) in animal #56 at week 36 corresponded to cattle (*Bos taurus*) specific immunoglobulin and alone passed all analysis filters. Even though pre-infection DPP-assays from all 4 animals were pooled together to enhance the probability of capturing all mycobacterial circulating immune complexes at baseline to compare them with proteins detected post-infection, no immune-complexes were detected at baseline.

The panel of mycobacterial proteins and cattle specific immunoglobulin reported in the present study may be specific to the infection stage at which they were detected, as the proteins

seen at week 14 did not overlap with those detected at week 36 post infection. Alternatively, since these distinct proteins sets were found in two different animals, they could be a result of animal-to-animal variation in host response to the infection.

A major limitation of our study was sample size. Since the LC/MS-MS analysis itself was expensive and limited amounts of infected animal sera were available, multiple replications on the same animals were not possible. Additionally, multiple logistical and financial issues precluded us working with larger sample sizes: (1) working with agricultural animals for experimental infection with a BSL-3 pathogen and (2) Expenses associated with a BSL-3 cost, animal costs as well as personnel. Thus, to compensate for this limitation triplication of every animal's sample was performed.

Furthermore, the use of an antiserum derived against MTB may have compromised specificity of our approach to detect MBO specific antigens, although these organisms are genetically very closely related. Future analysis though should include multiple replications of experimental infections followed by DPP assay and LC-MS/MS to discover MBO specific peptides in a reproducible and accurate fashion. Furthermore, a field validation on multiple exposure levels in outbreaks would be necessary for this technology to be applicable in routine bovine TB diagnostics.

Conclusion

In conclusion, the panel of 11 proteins reported in this study are specific to MBO. Further studies with more robust enrichment methods and larger sample sizes would be required to confirm these findings. Further validation of the identified circulating immune-complexes in

naturally infected cattle would enable us to effectively and broadly apply the DPP technology in field.

REFERENCES

REFERENCES

1. Hlavsa MC, Moonan PK, Cowan LS, Navin TR, Kammerer JS, Morlock GP, Crawford JT, LoBue PA. 2008. Human Tuberculosis due to Mycobacterium bovis in the United States, 1995–2005 . Clin Infect Dis 47:168–175.

2. OIE. 2018. Bovine tuberculosis. In Manual of Diagnostic Tests and Vaccines for Terrestrial AnimalsOIE Terrestrial Manual.

3. Wilkins MJ, Meyerson J, Bartlett PC, Spieldenner SL, Berry DE, Mosher LB, Kaneene JB, Robinson-dunn B, Stobierski MG, Boulton ML. 2008. Mycobacterium bovis Infection and Bovine 14:657–660.

4. Cousins D V., Florisson N. 2005. A review of tests available for use in the diagnosis of tuberculosis in non-bovine species. OIE Rev Sci Tech 24:1039–1059.

5. De La Rua-Domenech R. 2006. Human Mycobacterium bovis infection in the United Kingdom: Incidence, risks, control measures and review of the zoonotic aspects of bovine tuberculosis. Tuberculosis 86:77–109.

6. De Rycke L, Peene I, Hoffman IEA, E Kruithof A, Union A, Meheus L, Lebeer K, Wyns B, Vincent C, Mielants H, Boullart L, Serre G, Veys EM, Keyser F De. 2004. Rheumatoid factor and anticitrullinated protein antibodies in rheumatoid arthritis: diagnostic value, associations with radiological progression rate, and extra-articular manifestations 1587–1593.

7. Lyashchenko KP, Greenwald R, Sikar-gang A, Sridhara AA, Johnathan A, Lambotte P, Esfandiari J, Maggioli MF, Thacker TC, Palmer M V, Ray W. 2017. Early Detection of Circulating Antigen and IgM-Associated Immune complexes during Mycobacterium bovis Infection in Cattle 24:1–11.

8. Waters WR, Vordermeier HM, Rhodes S, Khatri B, Palmer M V, Maggioli MF, Thacker TC, Nelson JT, Thomsen B V, Robbe-austerman S, Garcia DMB, Schoenbaum MA, Camacho MS, Ray JS, Esfandiari J, Lambotte P, Greenwald R, Grandison A, Sikar-gang A, Lyashchenko KP. 2017. Potential for rapid antibody detection to identify tuberculous cattle with non-reactive tuberculin skin test results. BMC Vet Res 1–7.

9. Keller AD, Nesvizhskii AI, Kolker E, Aebersold R. 2002. Empirical statistical model to estimate the accuracy of protein identifications made by MS/MS and database search. Proc 50th ASMS Conf Mass Spectrom Allied Top 74:37–38.

10. Forrellad MA, Klepp LI, Gioffré A, García JS, Morbidoni HR, de la Paz Santangelo M, Cataldi AA, Bigi F. 2013. Virulence factors of the mycobacterium tuberculosis complex. Virulence. Taylor and Francis Inc.

11. Lamont EA, Janagama HK, Ribeiro-lima J, Vulchanova L, Seth M, Yang M, Kurmi K, Waters WR. 2014. Circulating Mycobacterium bovis Peptides and Host Response Proteins as Biomarkers for Unambiguous Detection of Subclinical Infection 52:536–543.

12. Wanzala SI, Palmer M V., Waters WR, Thacker TC, Carstensen M, Sreevatsan S. 2017. Evaluation of pathogen-specific biomarkers for the diagnosis of tuberculosis in white-tailed deer 78:729–734.

CHAPTER 7 DNA APTAMERS DEVELOPED AGAINST THE VALIDATED MTBC BIOMARKERS AS TOOLS FOR FIELD DIAGNOSTICS

Abstract

Discovery and validation of *Mycobacterium tuberculosis Complex*-specific-biomarkers (Pks5, MB2515c and MB1895c) offers an unprecedented opportunity to develop rapid and specific diagnostic platforms for field detection of bovine tuberculosis in multiple animal species including humans. This chapter focuses on utilization of these biomarkers as a target for development of DNA aptamers that can enable development of inexpensive diagnostics for tuberculosis. Short segments of the biomarkers were recombinantly produced, confirmed by mass spectrometry, and used as the target for one-step DNA aptamer selection from a randomized combinatorial DNA library via sequential salt elution. The selected aptamer pool was amplified, TA cloned, and sequenced to identify redundancy of candidate aptamers. Selected candidate sequences were checked for structure stability using mFold software and specificity against each of the three biomarkers Pks5, MB2515c and MB1895c, using a SouthWestern blot. A total of five aptamer sequences qualified for specificity testing, and one anti-MB2515c aptamer sequence was confirmed to bind our target. Selected DNA aptamers have the potential to be applied to develop a point-of-care diagnostic test development and validation with infected serum samples. The assay will significantly support the disease control program through enhanced detection of infected animals in the field.

Introduction

BTB is a chronic debilitating zoonotic disease caused by MBO, which primarily infects cattle and other warm-blooded animals. In the US, white-tailed deer serve as a wildlife reservoir for this disease. Worldwide, MBO causes 1.4% of all human pulmonary TB cases according to WHO(1). This prevalence is an underestimate as currently applied diagnostics do not differentiate MTB from MBO. Furthermore, prevalence in specific geographical locations and communities such as in Mexico, Southern California, etc., are quite high. For example, an outbreak caused by MBO was reported in the Hispanic community in New York due to consumption of unpasteurized dairy products imported from Mexico (2–4). Another study reported that 10% of all human TB cases in California were caused by MBO (5).

Critically, controlling the disease in animals is a primary approach to preventing/mitigating spread of MBO into human populations. It is estimated that it costs $1.5 Million, if BTB is detected in a herd of 1000 animals (6). With 12,000 plus farms at risk in Michigan alone, risk of infection is an economic threat for the farmers in addition to the emotional and psychological damage associated with financial and animal losses. BTB is endemic in Michigan and is maintained in white-tailed deer since first identified in 1975. Historically BTB circulating in the deer has been transmitted to those who come in close contact with the deer like hunters, venison-farmers, and consumers. Three cases of direct transmission from deer to humans were confirmed from 1994 till 2007 (7, 8).

The rise in BTB infected cattle herds has created a state of urgency in the last few years to Michigan agriculture, causing its status to be changed by the US Department of Agriculture from

TB accredited free zone (AFZ) to modified accredited zone (MAZ) that includes Alcona, Alpena, Montmorency ans Oscoda counties since October 13th, 2014. This change has led to restrictions imposed on domesticated animal movement from the endemic region of MI to other areas in the state or across state borders, unless the cattle are being moved directly to slaughter or have a BTB-free-certification. As of March 2021, 81 herds had been identified to be infected with BTB. The situation has become even more concerning since the 81st herd was identified outside the MAZ.

An 'ideal' point-of-care test, per the World Health Organization would meet the ASSURED criteria (affordable, sensitive, specific, user-friendly, robust and rapid, equipment-free, deliverable to those who need them) and allow expediated diagnostic decision-making. The discovery of pathogen-specific biomarkers (PSB) Pks5, MB2515c, MB1895c by Lamont et al., (2014) provides an unprecedented opportunity to develop and deploy a field test for BTB in congruence with WHO recommended criteria (9).

Aptamers are single stranded oligonucleotides discovered in the 1990's by Tuerk and Gold (10). Aptamers are either RNA, DNA or peptides that provide a high level of selectivity and specificity to a receptor of interest. This distinctiveness arises from the combinatorial library that is synthesized via randomization of nucleotide sequences. Aptamers are generally between 25 and 60 nucleotides long. An aptamer of length 40 nucleotides would provide 10^{24} unique sequences at synthesis in a one-molar solution. Each sequence provides a 3-dimensional structure that has a potential to bind with the target molecules via hydrogen bonds and Van der Waal's forces. Aptamers have the advantage of not only binding with the target molecule with

affinity like that provided by an antibody, but also cheaper to mass produce with minimal batch-to-batch variation. We and others have demonstrated aptamer success with a wide array of target molecules to select aptamers against, such as viruses, bacteria, small molecules, and prions (11–20).

The ultimate objective of this research is to combine the discovery of PSBs with aptamer technology in order to develop a field-based diagnostic tool. The first part of this project focuses on in-lab selection and specificity confirmation of anti-PSB aptamers.

Materials & Methods

Target peptide selection

Small peptides used by Lamont et al. (2014) for antibody production and those identified by Hadi et al. 2019 through LC-MS/MS were used for target protein synthesis (Table 7-1)(9, 21). A third set of "short peptides" were also identified using a stepwise computational approach. First, peptide regions of each biomarker with high probability and predicted avidity of interactions with bovine MHC class 1 (BoLA) molecules on NetBoLApan Server were selected. Second, those peptide regions that fall in the conserved regions of the protein were chosen. Third on these that passed the second filter, a BLASTp (Basic Local Alignment Search Tool-proteins) search on National Center for Biotechnology Information (NCBI) was performed. Fourth and last test was based on the conservation of these peptides in the published whole genome sequences available on PATRIC (Pathosystems Resource Integration Center).

Biomarker	Peptide number	Peptide sequence from			Longer peptide used for aptamer selection
		Lamont, et al. 2014	Hadi, et al. 2018	BoLa and MTBC specific by NCBI	
MB2515c	1.	MDRRPRDFEQSRRRC	-	VRLFISPRTVHSHLT	**MDRRPRDFEQSRRRC**RCNALRAGSMLASMSKIHPGVDVVPVDWSADGVSELVPTGTVTLLLADIEGATHLPGSQLDTTAIAKLDRTLTELVRE
	2.				MGDSEFDDAWAEGTALSIDEAIAYAQRGRGARKRPTSGWGALTPTELEVALLVGEGLSNKEIG**VRLFISPRTVHSHLT**HVYTKLGLSSRLQLAQQAARRGESERGPS
MB1895c	3.	ARVRANPDPRAQSKLTG-C	-	TFTRAAKPGAYLRVI	**MARVRANPDPRAQSKLTG**IDKVAASEAVMVRAPGSMHAGVGSGLVGDTVGNPKLHGGDDQAVYAYAREDLDAWETQLHRTLHNGMFGENLTTSGVDVTYARIGERWRIGSDGLVLEVSAPRIPCRTFAAFLDLRYWIK**TFTRAAKPGAYLRVI**
Pks5	4.	MGKERTKTVDRTRVTP-C	VQYPDGRLVDAPLPSWTHRTLMLSR	-	**MGKERTKTVDRTRVTP**VAVIGMGCRLPGGIDSPDRLWEALLRGDDLVTEIPADRWDIDEYYDPEPGVPGRTDCKWGAYLDNVGDFDPEFFGIGEKEAIAIDP
	5.				MPIATLAAMRRGEQLPFGLRGFVADVHNAGAKVDFS**VQYPDGRLVDAPLPSWTHRTLMLSR**EDSHRSHTGAVQAVHPLLGAHVHLLEEPERHVWQAGVG

Table 7-1: Peptide motifs selected for each Biomarker for expression.

These short sequences along with neighboring amino acids on both ends of peptides conforming to a size between 10 and 20 kDa, were selected for cloning and recombinant protein expression. These “longer peptide” sequences (Table 7-2) were confirmed to be specific to Mycobacterium tuberculosis complex (MTBC) by performing a BLASTp. For the recombinant expression of these “longer peptides”, codon optimized nucleotide sequences were synthesized (GeneScript) with a c-terminal histag on a pET24a plasmid.

Peptide name	MTBC or E. coli	Nucleotide seq	Peptide seq encoded	Estimated Weight (Kilo	Size (bp)
MB2515c Lamont	MTBC	**ATG GAC CGG CGG CCA AGA GAT TTC GAG CAA TCT AGG CGC AGG TGC** CGA TGT AAT GCG CTA CGC GCT GGC TCG ATG CTG GCT AGC ATG AGC AAG ATT CAC CCC GGC GTC GAC GTG GTA CCT GTG GAC TGG AGC GCA GAT GGT GTG AGC GAG CTT GTA CCG ACG GGC ACG GTG ACG TTG CTG CTG GCC GAT ATC GAG GGC GCG ACC CAC CTG CCG GGG TCT CAG CTT GAC ACG ACG GCG ATC GCC AAA CTG GAC CGA ACG CTG ACC GAA TTG GTT CGC GAG	**M D R R P R D F E Q S R R R C** R C N A L R A G S M L A S M S K I H P G V D V V P V D W S A D G V S E L V P T G T V T L L L A D I E G A T H L P G S Q L D T T A I A K L D R T L T E L V R E		
	E. coli	G GAT CCG **ATG GAC CGT CGT CCG CGT GAT TTC GAG CAG AGC CGT CGT CGT TGC** CGT TGC AAC GCG CTG CGT GCG GGT AGC ATG CTG GCG AGC ATG AGC AAG ATC CAC CCG GGC GTG GAC GTG GTT CCG GTT GAC TGG AGC GCG GAT GGT GTG AGC GAG CTG GTT CCG ACC GGT ACC GTG ACC CTG CTG CTG GCG GAT ATT GAA GGT GCG ACC CAC CTG CCG GGC AGC CAA CTG GAC ACC ACC GCG ATT GCG AAA CTG GAT CGT ACC CTG ACC GAG CTG GTT CGT GAA CTC GAG	D P **M D R R P R D F E Q S R R R C** R C N A L R A G S M L A S M S K I H P G V D V V P V D W S A D G V S E L V P T G T V T L L L A D I E G A T H L P G S Q L D T T A I A K L D R T L T E L V R E L E	10.01	279
MB2515c Hadi's	MTBC	ATG GGT GAC AGC GAG TTC GAC GAT GCC TGG GCC GAA GGA ACT GCG TTA TCG ATC GAT GAG GCG ATC GCG TAC GCG CAA CGA GGC CGC GGT GCA CGG AAA CGC CCG ACA AGT GGC TGG GGC GCG TTG ACA CCG ACA GAG CTT GAG GTG GCT CTA CTG GTC GGT GAG GGA CTC AGC AAC AAG GAG ATC GGC **GTC CGC CTA TTC ATC TCG CCG CGG ACC GTG CAT TCC CAC TTG ACC** CAC GTC TAC ACC AAG CTC GGC CTG TCC TCC CGT CTT CAA CTC GCG CAG CAA GCT GCG CGC CGG GGC GAA TCA GAA CGA GGG CCG TCG	M G D S E F D D A W A E G T A L S I D E A I A Y A Q R G R G A R K R P T S G W G A L T P T E L E V A L L V G E G L S N K E I G **V R L F I S P R T V H S H L T** H V Y T K L G L S S R L Q L A Q Q A A R R G E S E R G P S	11.56	321
	E. coli	G GAT CCG ATG GGT GAC AGC GAG TTT GAT GAT GCG TGG GCG GAA GGT ACC GCG CTG AGC ATC GAT GAG GCG ATT GCG TAT GCG CAA CGT GGT CGT GGC GCG CGT AAG CGT CCG ACC AGC GGT TGG GGC GCG CTG ACC CCG ACC GAG CTG GAA GTG GCG CTG CTG GTT GGT GAA GGC CTG AGC AAC AAG GAA ATC GGT **GTG CGT CTG TTT ATT AGC CCG CGT ACC GTG CAC AGC CAC CTG ACC** CAC GTT TAT ACC AAA CTG GGC CTG AGC AGC CGT CTG CAG CTG GCG CAG CAA GCG GCG CGT CGT GGT GAG AGC GAA CGT GGT CCG AGC CTC GAG	D P M G D S E F D D A W A E G T A L S I D E A I A Y A Q R G R G A R K R P T S G W G A L T P T E L E V A L L V G E G L S N K E I G **V R L F I S P R T V H S H L T** H V Y T K L G L S S R L Q L A Q Q A A R R G E S E R G P S L E		
MB1895c Lamont's and Hadi's combined	MTBC	**ATG GCT CGT GTT CGG GCG AAC CCT GAT CCG CGC GCG CAG TCG AAG TTG ACC GGA** ATC GAC AAA GTG GCG GCA TCT GAG GCG GTC ATG GTG CGG GCA CCC GGG TCG ATG CAT GCG GGT GTC GGC AGC GGC CTG GTT GGC GAC ACC GTC GGC AAC CCG AAG CTC CAT GGC GGT GAT GAT CAG GCC GTC TAT GCG TAT GCG CGA GAA GAC CTC GAC GCG TGG GAA ACC CAG CTT CAC CGC ACC CTT CAC AAC GGA ATG TTC GGC GAG AAT CTG ACC ACC TCG GGC GTC GAC GTG ACG TAC GCG CGG ATT GGT GAA CGC TGG CGC ATC GGC TCC GAC GGA TTG GTG TTG GAA GTC TCG GCG CCC CGG ATC CCG TGC CGG ACC TTC GCG GCC TTC CTG GAT TTG CGT TAT TGG ATC AAG **ACC TTC ACC CGG GCC GCC AAA CCA GGC GCC TAC CTA CGG GTG ATC**	**M A R V R A N P D P R A Q S K L T G** I D K V A A S E A V M V R A P G S M H A G V G S G L V G D T V G N P K L H G G D D Q A V Y A Y A R E D L D A W E T Q L H R T L H N G M F G E N L T T S G V D V T Y A R I G E R W R I G S D G L V L E V S A P R I P C R T F A A F L D L R Y W I K **T F T R A A K P G A Y L R V I**	16.5	459
	E. coli	G GAT CCG **ATG GCG CGT GTG CGT GCG AAC CCG GAC CCG CGT GCG CAG AGC AAG CTG ACC GGC** ATC GAT AAA GTT GCG GCG AGC GAG GCG GTG ATG GTT CGT GCG CCG GGT AGC ATG CAT GCG GGT GTG GGC AGC GGT CTG GTG GGC GAC ACC GTT GGT AAC CCG AAG CTG CAC GGT GGC GAC GAT CAG GCG GTT TAC GCG TAT GCG CGT GAG GAC CTG GAT GCG TGG GAA ACC CAA CTG CAC CGT ACC CTG CAC AAC GGC ATG TTC GGT GAA AAC CTG ACC ACC AGC GGC GTG GAC GTT ACC TAC GCG CGT ATC GGC GAG CGT TGG CGT ATT GGC AGC GAT GGT CTG GTG CTG GAA GTT AGC GCG CCG CGT ATC CCG TGC CGT ACC TTC GCG GCG TTT CTG GAT CTG CGT TAC TGG ATT AAG **ACC TTT ACC CGT GCG GCG AAA CCG GGT GCG TAT CTG CGT GTG ATT** CTC GAG	D P **M A R V R A N P D P R A Q S K L T G** I D K V A A S E A V M V R A P G S M H A G V G S G L V G D T V G N P K L H G G D D Q A V Y A Y A R E D L D A W E T Q L H R T L H N G M F G E N L T T S G V D V T Y A R I G E R W R I G S D G L V L E V S A P R I P C R T F A A F L D L R Y W I K **T F T R A A K P G A Y L R V I** L E		
PKS5 Lamont	MTBC	**ATG GGT AAG GAG AGA ACA AAG ACC GTG GAT AGG ACA CGT GTT ACG CCG** GTA GCT GTC ATT GGT ATG GGG TGC CGG CTG CCG GGG GGC ATC GAC TCA CCC GAT CGG TTG TGG GAG GCG TTG CTG CGG GGC GAC GAT CTG GTC ACC GAG ATC CCC GCC GAC CGC TGG GAC ATC GAC GAG TAC TAC GAC CCC GAA CCC GGC GTG CCC GGA CGC ACC GAC TGC AAA TGG GGC GCG TAC CTC GAT AAC GTC GGC GAC TTT GAT CCC GAG TTC TTC GGG ATC GGG GAG AAA GAA GCG ATA GCG ATC GAT CCG	**M G K E R T K T V D R T R V T P** V A V I G M G C R L P G G I D S P D R L W E A L L R G D D L V T E I P A D R W D I D E Y Y D P E P G V P G R T D C K W G A Y L D N V G D F D P E F F G I G E K E A I A I D P	11.34	319
	E. coli	G GAT CCG **ATG GGC AAG GAG CGT ACC AAA ACC GTG GAC CGT ACC CGT GTT ACC CCG** GTG GCG GTT ATC GGT ATG GGC TGC CGT CTG CCG GGT GGC ATT GAC AGC CCG GAT CGT CTG TGG GAG GCG CTG CTG CGT GGC GAC GAT CTG GTG ACC GAA ATC CCG GCG GAC CGT TGG GAC ATT GAT GAG TAC TAT GAT CCG GAA CCG GGT GTG CCG GGT CGT ACC GAC TGC AAG TGG GGT GCG TAC CTG GAT AAC GTT GGC GAC TTC GAT CCG GAA TTC TTT GGT ATC GGC GAG AAA GAA GCG ATC GCG ATT GAT CCG CTC GAG	D P **M G K E R T K T V D R T R V T P** V A V I G M G C R L P G G I D S P D R L W E A L L R G D D L V T E I P A D R W D I D E Y Y D P E P G V P G R T D C K W G A Y L D N V G D F D P E F F G I G E K E A I A I D P L E		
PKS5 Hadi	MTBC	ATG CCG ATC GCA ACG CTT GCC GCG ATG CGG CGC GGG GAA CAG CTG CCG TTC GGG TTG CGC GGC TTC GTC GCC GAC GTG CAC AAC GCC GGC GCC AAG GTG GAC TTC TCT **GTC CAG**	M P I A T L A A M R R G E Q L P F G L R G F V A D V H N A G A K V D F S **V Q Y P D G R L V D A P L P S W T H R T L M L S R** E D S H R S H T G A V Q A V H	11.02	297
	E. coli	G GAT CCG ATG CCG ATT GCG ACC CTG GCG GCG ATG CGT CGT GGT GAG CAG CTG CCG TTC GGT CTG CGT GGC TTT GTG GCG GAC GTT CAC AAC GCG GGT GCG AAG GTG GAT TTC AGC	D P M P I A T L A A M R R G E Q L P F G L R G F V A D V H N A G A K V D F S **V Q Y P D G R L V D A P L P S W T H R T L M L S R** E D S H R S H T G A V Q A		

Table 7-2: The longer peptides expressed for aptamer selection are tabulated below. The first column represents the peptide names based on publication source. The 3rd column represents the nucleotide sequences that are translated into the selected peptide sequence in MTBC*. These nucleotide sequences were codon optimized for expression in E. coli. The nucleotide sequence was confirmed by the Translate tool on ExPASy-Swiss Institute of Bioinformatics (Bioinformatics Resource Portal). The colored part of nucleotide sequences shows the peptide sequences chosen earlier in Table 7-1.*

In-lab protein expression, purification, confirmation

Cloning

The pE24a plasmid system was used to transform *E. coli* (BL21 Competent Cells, Expression Strains, MilliporeSigma, BL21(DE3) Competent Cells, 2×20 µl. Catalog# 69450-3) per manufacturer's recommendations. Briefly, each synthesized plasmid was dissolve in 40µL TE buffer to have a working concentration of 100ng/µl. One-µl (10ng/ µl) of the plasmid was added to 20-µl of *E. coli* BL21 cells and incubated for 5 minutes on ice. The cells were heat shocked for 30 seconds at 42°C followed by quick chill on ice for 5 minutes. Eighty-µl of S.O.C media was added to the transformed cells. These cells were incubated at 37°C while shaking at 250 rpm for 60 minutes. A 33.3-µL volume of the transformed cells were plated on selective media (LB-agar medium with kanamycin at a final concentration of 30µg/ml) and incubated for 15-18 hours at 37°C. Colonies for each of the 5 different plasmids were picked and master plated. The master colonies were confirmed for the presence of the plasmids by T7 PCR..

The amplicons were purified using QIAquick mini-elute kit [Cat. Nos. 28104 and 28106]. The purified amplicons were quantified using Nanodrop and confirmed by sanger sequencing .

Biomarker peptide expression

Protein expression for all five plasmids was performed by individually testing different temperatures (37°C, 21°C) and isopropylthio-β-galactoside (IPTG) concentrations (100µM, 500µM, 1000µM) in 25 ml Luria broth with kanamycin at 30ug/ml (LB-Kan30) concentration.

Biomarker peptide confirmation

Recombinant proteins from soluble fraction of cell pellets were purified using immobilized metal ion affinity chromatography (IMAC). HisPur™ Cobalt Resin (Catalog # 89965) was used according to manufacturer's protocol. The purified protein elutes were quantified by Qubit™ Protein BR Assay Kit (Invitrogen™, Fisher Scientific, Catalog # Q33212) and Qubit 3 Fluorometer (Catalog # Q33216).

Expressed proteins were further confirmed by Western blotting. Peptides were prepared for electrophoresis under de-naturing conditions using 2x Laemmli sample buffer (Bio-Rad. Catalog # 161-0737) and β-mercaptoethanol (Bio-Rad. Catalog # 161-0710) at 1:1 ratio followed by heating at 95°C for 15 minutes and quick chill. Samples were electrophoresed at 100V for 1.5 hours on a 4-20% polyacrylamide gel (4–20% Mini-PROTEAN® TGX™ Precast Gel 15 well size #4561096EDU). The gels were then electroblotted onto NCM at 60V for 120 minutes at 4°C using a western blot transfer buffer (Pierce™ 10X Western Blot Transfer Buffer, Methanol-free. Catalog # 35040). The membrane was washed with 1X PBS (phosphate buffered saline) once and then blocked overnight at 4°C with 2% bovine serum albumin (BSA) dissolved in 1X PBS. Membranes were washed thrice after blocking with 1XPBS-T (10 mM 1xPBS, pH7.2, 0.05%Tween-20). Anti-Hist Tag HRP-conjugated antibody (R&D Systems, Cataloge # MAB050H) was used at a concentration of 1:4000 for the detection of his-tagged peptides. It was diluted using 1X PBS-T and the membranes were incubated with this primary antibody for 1 hour at room temperature. It was followed by washing as before. Last step involved the addition of luminol peroxidase substrate in 1:1 ratio (Amersham ECL Prime Western blotting detection reagent. GE Healthcare.

Catalog # RPN2236) to the membrane. NCM was incubated for five minutes in dark, then quickly dipped in 1XPBS-T and visualized. A second round of Western blot on purified protein elutes was also performed.

Finally, all peptides were confirmed by mass-spectrometry analysis of excised bands from acrylamide gels, at the Proteomics Core facility at MSU.

Aptamer selection

One-step aptamer selection

Candidate aptamer selection against Pks5 and MB2515c was performed separately using a one-step aptamer selection protocol, using sequential salt elution method as described by Hmila, et al., 2017 (13). A DNA aptamer library with a randomized central region (40-mer) and flanked on either end by constant primer binding regions (20-bp each) was manufactured (Integrated DNA Technologies, Inc., Coralville, IA). The purified peptides (1mg/well) were immobilized onto the 96-well ELISA plate in triplicate (Corning 96 well standard strip well microplates clear high binding). The target peptides were mixed with 100-µL sodium-carbonate-bi-carbonate buffer to enhance binding with the solid-phase. After overnight incubation at 4°C, the peptide solution was removed. This was followed by an hour-long incubation with 10µM of 100µL aptamer library. The wells were dry blotted, followed by washing for 5 min with 100µL of 1X PBS. The washing solution was collected and pooled from each triplicate. This was repeated with 0.5M, 0.75M, 1.0M, 1.25M and 1.5M solution of sodium chloride. The increasing salt concentration was used to elute the aptamers with the sequence with highest affinity appearing at the highest salt concentration. The

aptamers in 1.5 M pooled sample were amplified using the flanking primer sequences (WP20R1 and primer WP20F1) (Table 7-3)

Reagent	Details
Aptamer Library	**5'**-AGT-GCA-AGC-AGT-ATT-CGG-TC-**N40**-GGC-ATC-ACG-CAT-CAG-CTT-TA-**3**'
Forward primer (WP20F1)	5'-AGT-GCA-AGC-AGT-ATT-CGG-TC-3'
Reverse primer (WP20R1)	5'-TAA-AGC-TGA-TGC-GTG-ATG-CC-3’
Biotinylated WP20F1	5'-Bio/-AGT-GCA-AGC-AGT-ATT-CGG-TC-3'

Table 7-3: Sequences of Aptamer library and primers used in the work.

Specificity Analysis of selected aptamers- Dot blot

The selected aptamers from 1.5M salt elute were amplified from the original pool using the same PCR protocol and reagents as before, except for WP20F1 which was biotinylated to create 5’biotinylated DNA aptamers (Table 7-3).

Dot blot on a NCM was performed to test for specificity. One-thousand-ng per 2µl of each peptide was blotted in duplicate on the NCM and allowed to dry at room temperature for 1 hour. The membrane was blocked with 2% BSA in 1XPBS overnight (shaking) at 4°C. Washing (with shaking) was performed once with 1XPBS-T for 5 minutes. In the meantime, 100 µL of the biotinylated aptamer library was heat denatured at 94°C for 15 minutes to separate the biotinylated strand from the un-biotinylated strand. The denatured aptamer amplicon was mixed with 900 µL 1XPBS-T and incubated for 1-hour in a closed container to prevent drying. It was then washed thrice with 1XPBS-T for 5 minutes each and incubated with Neutravidin-HRP (Avidin,

NeutrAvidin™, Horseradish Peroxidase conjugate. Thermo Fisher Scientific. Catalog # A2664) in 1XPBS-T at 1:4000 dilution for 1 hour. This was followed by washing thrice as described above. Finally, luminol peroxidase substrate in 1:1 ratio was added to the membrane. NCM was incubated for 5 minutes in dark, then quickly dipped in 1XPBS-T and visualized.

Aptamer sequence identification

Aptamers that eluted at 1.5M salt concentrations, were amplified from the original elute pool, purified, and cloned (TA Cloning™ Kit, with pCR™2.1 Vector. Thermo Fischer Invitrogen. Catalog # k204040). Blue-white screening for transformed colonies was followed by amplification of cloned aptamer candidates followed by a clean-up and sequencing (Genewiz). One-hundred and twenty-five white (transformed) colonies were picked for anti-Pks5 aptamers, and 119 white colonies were picked for anti-MB2515c aptamers. Sequences were analyzed for redundancy using MEGA 7(22). The aptamer sequences that appeared more than once (redundant), had the highest G-C content, and had stable secondary structure predictions by *mFold* (23), were selected for further testing (Table 7-4).

Specificity testing of Selected Aptamers

The selected aptamers were tested for specificity before synthesis. Dot blot test was used as described earlier to test the selected aptamer amplicons from the clones. The aptamers (along with the primer flanking regions) that showed binding were then synthesized at Integrated DNA Technologies with a biotin molecule on the 5'-end.

Biotinylated aptamers were subjected to specificity analysis using a South-Western blot. The peptides were prepared for electrophoresis under denaturing conditions as described for

Western blot. For every aptamer, duplicates of both Pks5 (1µg/per well) and MB2515c (1µg/per well) were loaded on a precast 4-20% gradient polyacrylamide gel. They were electrophoresed at 100V for 90 minutes. The gels were then electroblotted onto NCM at 60V c for 120 minutes at 4°C using a Western blot transfer buffer. The membranes were blocked with 2% BSA in 1XPBS-T, followed by washing the membranes thrice with 1XPBS-T for 5 min each and probed with biotinylated DNA aptamer (1µM/lane in 1 ml of nuclease free water). The membranes were washed thrice again. The aptamer binding was detected using 1:4000 dilution of Neutravidin-HRP (Avidin, NeutrAvidin™, Horseradish Peroxidase conjugate. Thermo Fisher Scientific. Catalog # A2664) and luminol 1:1 ratio (Amersham ECL Prime Western blotting detection reagent. GE Healthcare. Catalog # RPN2236) as a chemiluminescent substrate.

	#ID	Aptamer sequence	GC %
Anti-Pks5 aptamers	**54** & 67	ATCCGTTAGGAAGTGACGGTTAGTTGGTCTATTGAGTATT	35
	1 & 15	ACATTCTAACATCATTCATAAAGTCAAGGCATCCAACACA	35
	55 & 64	GGCATTAATATTAGCTAGAATAACATGACAACGCTTTAGC	35
Anti-MB2515c aptamer	29 & 30	ACTAGCTCTCGAGCCTACTGCCAATTGCTCAAATTCTCATG	47.5

Table 7-4: Redundant aptamer sequences identified through TA cloning and sanger sequencing, against both Pks5 and MB2515c. None of them showed binding on South-Western blot. The bold ID # are the ID used in this paper.

Results

Aptamer Selection and Specificity Analysis

Aptamer selection was first attempted using the “short peptides”. The small size of the peptides led to the selection of aptamers with non-reproducible results when evaluated by specificity testing. This is likely due to the relatively small sized peptides failing to bind strongly to solid phases (microtiter wells) during aptamer selection. This led to selection of five ‘long peptide” sequences that were cloned and expressed for a second round of aptamer selection.

Two out of the five peptides were expressed successfully- Pks5-4 (aka Pks5) and MB2515c-2 (aka MB2515c). Pks5 was expressed at 4 hours post-induction at 100µM IPTG, whereas MB2515c was expressed at 2 hours post induction with 100µM IPTG. Both were identified in and extracted from the soluble fraction of the bacterial cell pellet and confirmed using Western blot (Figure 7-1 and Figure 7-2) and mass-spectrometry analysis (Figure 7-3 and Figure 7-4). Fifty-six percent amino acids in the peptide sequence of Pks5 were identified in MS while 70.1% amino acids in the peptide sequence of MB2515c were detected.

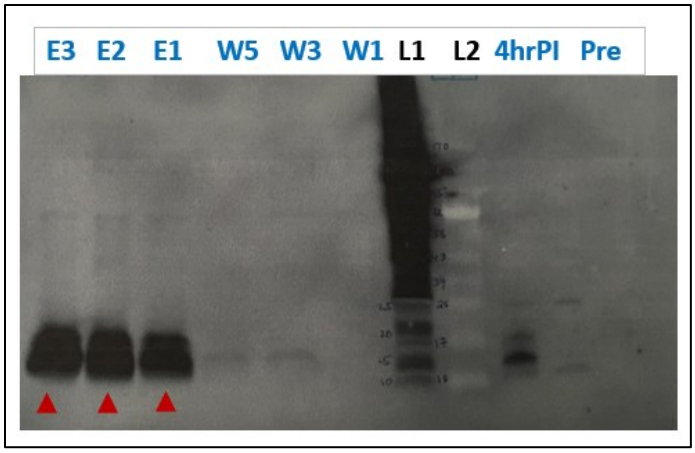

Figure 7-1: Western blot confirmed that Pks5 was present in unpurified sample 4 hours post-induction (4hrPI) and was then successfully purified (E1, E2, E3) by IMAC. Small quantities could be identified in pre-induction (Pre) sample as well as in the washes (W1, W2, W3).

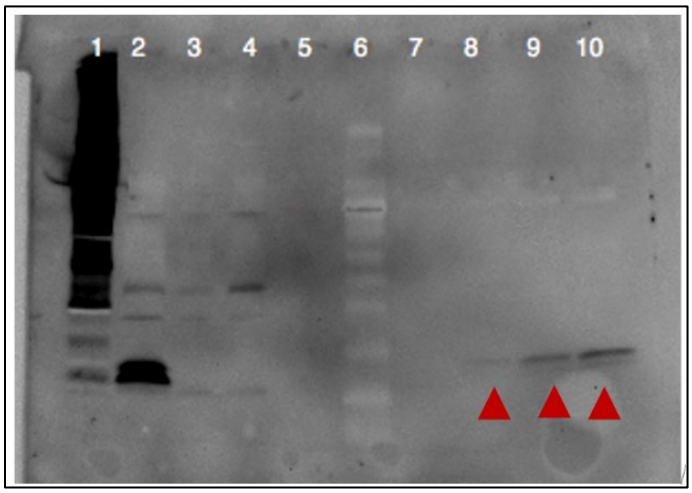

Figure 7-2: Western blot confirmed that MB2515c was present in unpurified sample 2 hours post-induction (well 2) and was then successfully purified (E1, E2, E3= well 8, 9 and 10) by IMAC.

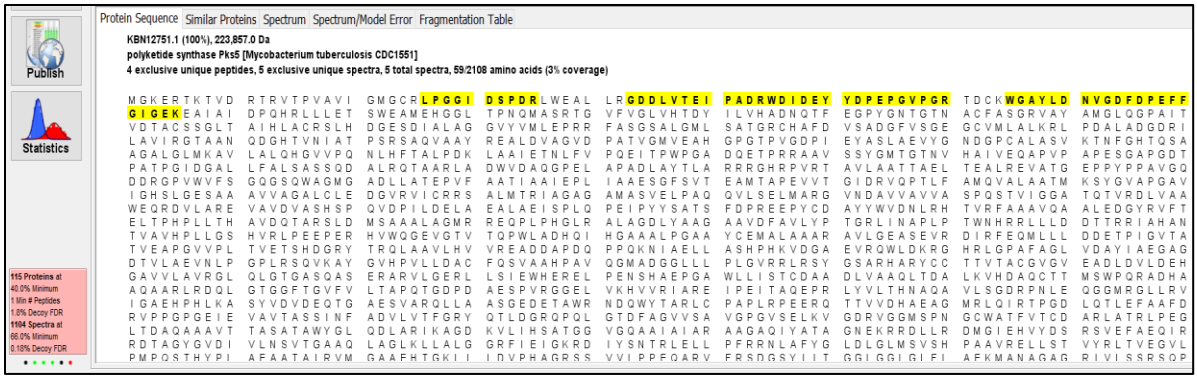

Figure 7-3: The scaffold results showed that 70% of MB2515c-2's sequence was detected. The sequences detected were highlighted in yellow in the scaffold program.

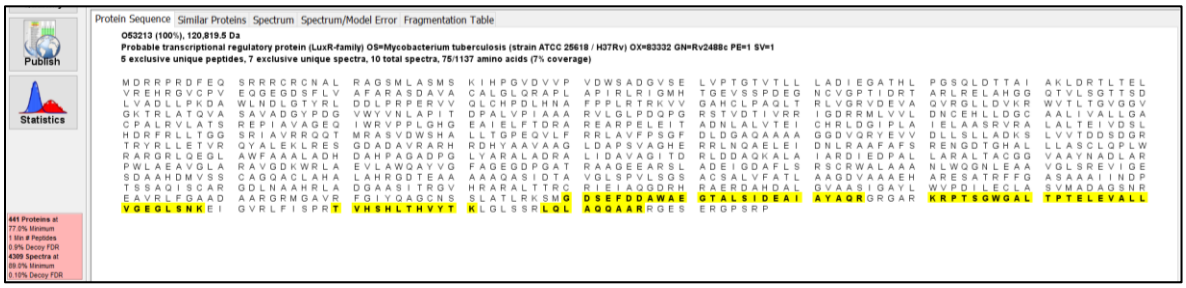

Figure 7-4: Out of 102 AAs in Pks5's sequence, 56.86% were detected. The sequences detected were highlighted in yellow in the scaffold program.

Aptamer selection was performed using sequential salt elution method. Both anti-biomarker-aptamers amplified from 1.5M salt solution showed binding in a dot blot assay (Figure 7-5). Sanger sequencing of these amplicons led to identification of three redundant sequences for Pks5 and one redundant sequence for MB2515c (Table 7-4).The redundancy is suggestive of a nucleotide sequence that folds and binds more strongly with the target peptide, thus getting amplified more than other sequences. Apart from the redundant sequences, a few more were selected due to their high GC content, since GC content affects the stability of secondary structures (Table 7-5).

	#ID	Aptamer sequence	GC content%
Anti-Pks5 aptamers	#22	GTGGGGATGGGTGGGGCGAGGTGAGGCCAAGGTGCCCCGA	70
	#25	GGGGGGTTAAGGAGGGCCGCCCTGCGCAGCGAGTACCGGG	75
	#69	TTTGGACAATAACCGGCCGAGTAGCCCGGGCGACATCTTA	55
	#110	GTCGGGAGTTGGCTAGTGAACAAAACGGGGAGGTGTCCGG	57.5
Anti-MB2515c aptamer	#44	ACAATTATAGCCGCGGTACCGAACCATCCGGGTTGCCCCG	60
	#63	CAGCAGGTGCACACTTGCTTAAGGGCTGCTCGCATCTCTA	55
	#64	GTGTGCCCAAGGTATGTCAGCCATGGGCTCCGGGCCCAGG	67.5

Table 7-5: Additional aptamer sequences with higher GC % and stable secondary structure that were tested by dot blot for specific binding with the target peptide. Only anti-MB2515c-aptamer # 44 showed binding on Southwestern blot.

Therefore, based on redundancy, GC content as well as thermal stability as predicted by *mFold*, a total of 11 aptamer sequences were selected for further specificity testing. Four of these 11 aptamers demonstrated binding with the respective peptides on dot blot assay. This was followed by Southwestern blotting for verification of these four aptamers.

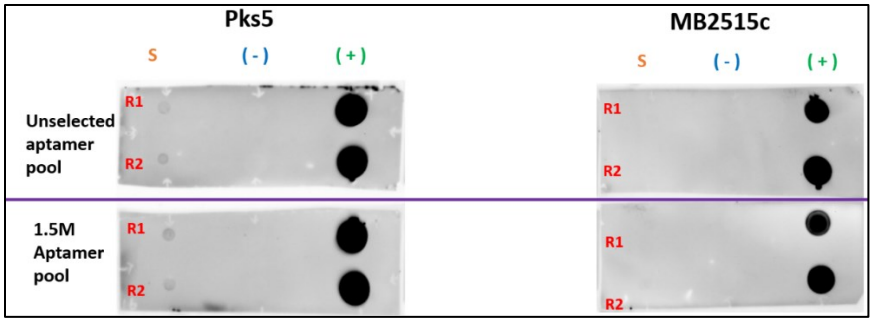

Figure 7-5: Dot blot assay for both Pks5 and MB2515c. The assay was performed in duplicates (R1, R2) using selected 1.5M aptamer pool as compared to the unselected aptamer pool. Faint bands can be seen for both sample dots (S). No band was seen for negative control (-). A dark bright dot was seen for the biotinylated primer used as a positive control (+).

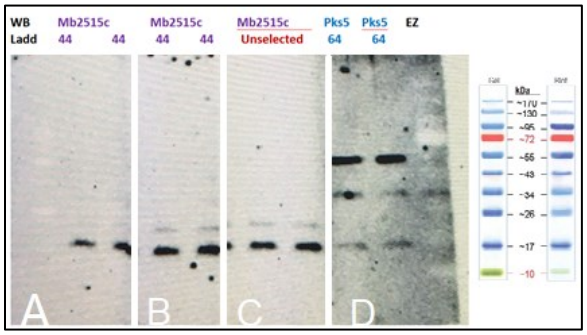

Figure 7-6: Specificity analysis of the selected aptamers was done using Southwestern blot. Strip A-C was electroblotted with MB2515c (11.56KDa) and D with Pks5 (11.34KDa). Each strip was probed with the aptamer specifically selected against the electroblotted peptide. Strip A- Anti-MB2515c aptamer (#44), Strip B- Anti-MB2515c aptamer (#44), Strip C- Unselected aptamer library as control for MB2515c, Strip D- Anti-Pks5 aptamer (#64). Both strip A and B reproducibly showed that the selected anti-MB2515c aptamer binds specifically. Whereas strip D shows non-specific binding of the selected anti-Pks5 aptamer.

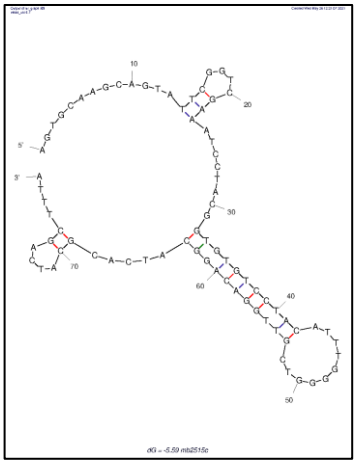

Figure 7-7: Anti-MB2515c aptamer (ID # 44), folded at 37˚C with primer flanking regions.

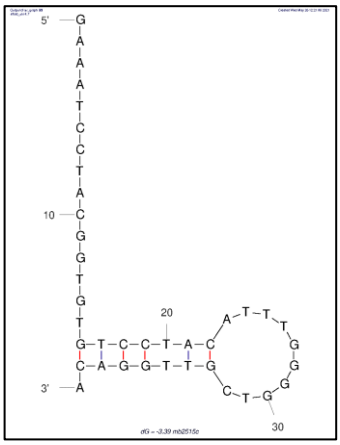

Figure 7-8: Anti-MB2515c aptamer (ID # 44), folded at 37˚C without primer flanking regions. The folded region does not change its structure with or without the primer folding region.

Out of the four Anti-MB2515c-Aptamer#44 showed reproducible binding on Southwestern blotting (Figure 7-6). *mFold* analysis of that sequence indicated formation of a stable secondary structure both with and without the primer flanking regions (Figure 7-7 and Figure 7-8). The stem-loop structure contains a G-tetrad suggesting potential for strong binding and stability of this aptamer.

Discussion

Numerous host-specific-biomarkers have been identified for detection of infection caused by tuberculosis in humans (24–27). Though effective in one species, the tests cannot be applied without modification to other species. Significant optimizations along with reselection of detection molecules is necessary. Instead of re-inventing the wheel every time a new host species has to be tested for TB, we applied the peptidome hypothesis. We posited that the anti-inflammatory conditions induced by the host pathogen interaction at the site of the infection may results in the release of several classes of low molecular weight cellular and bacterial proteins or products into the blood. Therefore, the blood peptidome likely contain these molecules that can serve as biomarkers of subclinical infection. Follow-up on studies confirmed that the *Mycobacterium* specific proteins released in circulation provide significant specificity to detect BTB (28, 29). Studies by Lamont et al. (2014) and Wanzala et al. (2016) validated the use of Mycobacterial-specific biomarkers in not just one but two different animal species. To extend the validation beyond ruminants in lieu of extending it to humans, we explored the presence of these biomarkers in non-human primates (NHP) as described in Chapter 5. With successful validation of biomarkers in cattle, deer and NHP, we proceeded to selecting DNA aptamers against the pathogen specific antigens Pks5, MB2515c and MB1895c.

In the past, aptamers have been developed against highly immunogenic TB antigens such as malate synthase (30), 23 kDa MPT64 (31–40), 10 kDa culture filtrate protein (CFP-10) (41–43), 6 kDa early secreted antigenic targets ESAT-6 (42, 43), HspX (44–47) and even whole bacteria (48, 49). Since it is approximated that one-third of all TB human patients are seronegative for any

single antigen (50), a combination of these antigens have been employed to identify aptamers with improved sensitivity and specificity (42, 43). Even on combination of certain biomarkers like CFP-10 and ESAT-6 that have been designed to be detected in laboratory set-up, the pre-requisites of a point-of-care device for screening TB in multiple animal species includes a) biomarkers pre-validated in multiple animal species, b) biomarkers abundantly secreted, c) detectable during latent infection as well as in early stages of infection with high specificity and sensitivity, and d) biomarkers that can differentiate *Mycobacterium tuberculosis* Complex infection from those caused by Mycobacteria other than TB (MOTT) as well as other environmental mycobacteria such as *Mycobacterium smegmatis*. Pks5, MB2515c as well as MB1895c identified in our prior work fulfill the above criteria and became the ideal target for a new TB screening tool with aptamers as detection molecules.

The functional sites of biomarkers Pks5, Mb2515c and MB1895c have not been experimentally confirmed nor have their tertiary structure been determined. Since studies performed with RNA aptamers suggest that the binding of aptamers occurs to functional part of protein target (51–54), we targeted the sequence of protein predicted to be the most antigenic using NetBoLApan Server, conserved regions of protein as well as BLASTp analysis. The approach allowed Lamont et al., (2014) to select antibodies that were highly specific for the biomarker. Therefore, a similar approach for peptide sequence identification was taken for aptamer selection that proved to be equally effective.

For expression of Pks5, MB2515c and MB1895c peptides, multiple temperature as well as IPTG induction concentrations were explored. Reducing temperature of expression led to

success, whereas the concentration of IPTG did not affect expression as much. Our results agreed with prior studies (43, 55, 56) where MTBC protein ESAT-6's expression was achieved at lower temperatures (25°C). Even though only two of the five peptides were successfully expressed, we suggest testing expression at an even lower temperature for the three unexpressed peptides.

The pool of aptamer selected after Sanger sequencing for affinity testing all had GC content (above 50%). They even had G-rich regions at the 3'-terminus of aptamer as well as motifs "TGGGG", "GTGG" or "CTGG" as reported earlier by Mozioglu et al., (2016) for the TB aptamers they had selected(48). Interestingly the aptamer (Anti-MB2515c-#44) that passed the affinity testing by both dot blot and Southwestern assay, did not display G-rich region at 3'-terminus, but it had the motif "TGGGG" in its loop structure along with 60% GC content. This is suggestive of a strong binding and a very specific aptamer. These characteristics were complemented by same folding pattern at 37°C as well as 25°C, with or without flanking primer sequences.

Moving forward the aptamer will be characterized further to determine its limit of detection, strength of detection, functional region as well as specificity of binding in the Prescence of other proteins. If the aptamer passes all these hurdles, it will then be tested to detect TB in serum samples of infected animals. A comprehensive validation work would then proceed ultimately leading the path for a point-of-care device capable of BTB detection in multiple animal species under field conditions.

Conclusion

This work is the very first description of selection of aptamers against pathogen-specific-biomarkers Pks5, MB2515c and MB1895c for detection of bovine tuberculosis in multiple animal

species. We expressed the biomarkers Pks5 as well as MB2515c at room temperatures. It was then proceeded by selection of aptamers that consistently showed binding with the target peptide. Anti-MB2515c aptamer # 41 was found to be the best candidate in our work so far. With re-expression of remaining peptides at very low temperature, followed by aptamer selection and validation of aptamer # 44, we expect to be a lot closer to realizing the very first TB detection tool that can be applied to all species- animals and humans alike.

REFERENCES

REFERENCES

1. 2019. Global Tuberculosis Report, Executive Summary.

2. Hlavsa MC, Moonan PK, Cowan LS, Navin TR, Kammerer JS, Morlock GP, Crawford JT, LoBue PA. 2008. Human Tuberculosis due to Mycobacterium bovis in the United States, 1995–2005 . Clin Infect Dis 47:168–175.

3. Scott C, Cavanaugh JS, Pratt R, Silk BJ, LoBue P, Moonan PK. 2016. Human Tuberculosis Caused by Mycobacterium bovis in the United States, 2006-2013. Clin Infect Dis 63:594–601.

4. Rodwell TC, Kapasi AJ, Moore M, Milian-Suazo F, Harris B, Guerrero LP, Moser K, Strathdee SA, Garfein RS. 2010. Tracing the origins of Mycobacterium bovis tuberculosis in humans in the USA to cattle in Mexico using spoligotyping. Int J Infect Dis 14:129–135.

5. Rodwell TC, Moore M, Moser KS, Brodine SK, Strathdee SA. 2008. Tuberculosis from Mycobacterium bovis in binational communities, United States. Emerg Infect Dis 14:909–916.

6. Chiu LJV, Tauer LW, Smith RL, Grohn YT, Verteramo Chiu LJ, Tauer LW, Smith RL, Grohn YT. 2019. Assessment of the bovine tuberculosis elimination protocol in the United States. J Dairy Sci 102:2384–2400.

7. Sunstrum J, Shoyinka A, Power LE, Maxwell D, Stobierski MG, Signs K, Sidge JL, Brien DJO, Robbe-Austerman S, Davisdson P. 2019. Notes from the Field- Zoonotic Mycobacterium bovis DIsease in Deer Huntees-Michigan, 2002-2017. Centers Dis Control Prev 68:807–808.

8. Wilkins MJ, Meyerson J, Bartlett PC, Spieldenner SL, Berry DE, Mosher LB, Kaneene JB, Robinson-dunn B, Stobierski MG, Boulton ML. 2008. Mycobacterium bovis Infection and Bovine 14:657–660.

9. Lamont EA, Janagama HK, Ribeiro-lima J, Vulchanova L, Seth M, Yang M, Kurmi K, Waters WR. 2014. Circulating Mycobacterium bovis Peptides and Host Response Proteins as Biomarkers for Unambiguous Detection of Subclinical Infection 52:536–543.

10. Tuerk C, Gold L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science (80-) 249:505–510.

11. Wang P, Hatcher KL, Bartz JC, Chen SG, Skinner P, Richt J, Liu H, Sreevatsan S. 2011. Selection and characterization of DNA aptamers against PrPSc. Exp Biol Med 236:466–476.

12. Takemura K, Wang P, Vorberg I, Surewicz W, Priola SA, Kanthasamy A, Pottathil R, Chen SG, Sreevatsan S. 2006. DNA aptamers that bind to PrPCand not PrPScshow sequence and structure specificity. Exp Biol Med 231:204–214.

13. Hmila I, Wongphatcharachai M, Laamiri N, Aouini R, Marnissi B, Arbi M, Sreevatsan S, Ghram A. 2017. A novel method for detection of H9N2 influenza viruses by an aptamer-real time-PCR 243:83–91.

14. Wongphatcharachai M, Wang P, Enomoto S, Webby RJ, Gramer MR, Amonsin A, Sreevatsan S. 2013. Neutralizing DNA aptamers against swine influenza H3N2 viruses. J Clin Microbiol 51:46–54.

15. Lamont EA, Wang P, Enomoto S, Borewicz K, Abdallah A, Isaacson RE, Sreevatsan S. 2014. A Combined Enrichment and Aptamer Pulldown Assay for Francisella tularensis Detection in Food and Environmental Matrices 1–19.

16. Joshi R, Janagama H, Dwivedi HP, Senthil Kumar TMA, Jaykus LA, Schefers J, Sreevatsan S. 2009. Selection, characterization, and application of DNA aptamers for the capture and detection of Salmonella enterica serovars. Mol Cell Probes 23:20–28.

17. Lamont EA, He L, Warriner K, Labuza TP, Sreevatsan S. 2011. A single DNA aptamer functions as a biosensor for ricin. Analyst 136:3884–3895.

18. Bannantine JP, Radosevich TJ, Stabel JR, Sreevatsan S, Kapur V, Paustian ML. 2007. Development and characterization of monoclonal antibodies and aptamers against major antigens of Mycobacterium avium subsp. paratuberculosis. Clin Vaccine Immunol 14:518–526.

19. Waybrant B, Pearce TR, Wang P, Sreevatsan S, Kokkoli E. 2012. Development and characterization of an aptamer binding ligand of fractalkine using domain targeted SELEX. Chem Commun 48:10043–10045.

20. White SP, Sreevatsan S, Frisbie CD, Dorfman KD. 2016. Rapid, Selective, Label-Free Aptameric Capture and Detection of Ricin in Potable Liquids Using a Printed Floating Gate Transistor. ACS Sensors 1:1213–1216.

21. Hadi SA, Waters WR, Palmer M, Lyashchenko KP, Sreevatsan S. 2018. Development of a multidimensional proteomic approach to detect circulating immune complexes in cattle experimentally infected with Mycobacterium bovis. Front Vet Sci 5.

22. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol https://doi.org/10.1093/molbev/msw054.

23. 2003. mfold-DNA folding form. Nucleic Acids Res 31:3406–15.

24. Liu C, Xiang G, Jiang D, Liu L, Liu F, Luo F, Pu X. 2015. An electrochemical aptasensor for detection of IFN-γ using graphene and a dual signal amplification strategy based on the exonuclease-mediated surface-initiated enzymatic polymerization. Analyst 140:7784–7791.

25. Chang CC, Lin S, Lee CH, Chuang TL, Hsueh PR, Lai HC, Lin CW. 2012. Amplified surface plasmon resonance immunosensor for interferon-Gamma based on a streptavidin-incorporated aptamer. Biosens Bioelectron 37:68–74.

26. Farid S, Meshik X, Choi M, Mukherjee S, Lan Y, Parikh D, Poduri S, Baterdene U, Huang CE, Wang YY, Burke P, Dutta M, Stroscio MA. 2015. Detection of Interferon gamma using graphene and aptamer based FET-like electrochemical biosensor. Biosens Bioelectron 71:294–299.

27. Li H, Song S, Wen M, Bao T, Wu Z, Xiong H, Zhang X, Wen W, Wang S. 2019. A novel label-free electrochemical impedance aptasensor for highly sensitive detection of human interferon-gamma based on target-induced exonuclease inhibition. Biosens Bioelectron 142.

28. Aagaard C, Govaerts M, Okkels LM, Andersen P, Pollock JM. 2003. Genomic approach to identification of Mycobacterium bovis diagnostic antigens in cattle. J Clin Microbiol 41:3719–3728.

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

30. Dhiman A, Kumar C, Mishra SK, Sikri K, Datta I, Sharma P, Singh TP, Haldar S, Sharma N, Bansal A, Ahmad Y, Kumar A, Sharma TK, Tyagi JS. 2019. Theranostic Application of a Novel G-Quadruplex-Forming DNA Aptamer Targeting Malate Synthase of Mycobacterium tuberculosis. Mol Ther - Nucleic Acids 18:661–672.

31. Sypabekova M, Bekmurzayeva A, Wang R, Li Y, Nogues C, Kanayeva D. 2017. Selection, characterization, and application of DNA aptamers for detection of Mycobacterium tuberculosis secreted protein MPT64. Tuberculosis 104:70–78.

32. Sypabekova M, Dukenbayev K, Tsepke A, Akisheva A, Oralbayev N, Kanayeva D. 2019. An aptasensor for the detection of Mycobacterium tuberculosis secreted immunogenic protein MPT64 in clinical samples towards tuberculosis detection. Sci Rep 9.

33. Chen H, Liu K, Li Z, Wang P. 2019. Point of care testing for infectious diseases. Clin Chim Acta 493:138–147.

34. Cheon HJ, Lee SMP, Kim S-R, Shin HY, Seo YH, Cho YK, Lee SMP, Kim M Il. 2018. Colorimetric

Detection of MPT64 Antibody Based on an Aptamer Adsorbed Magnetic Nanoparticles for Diagnosis of Tuberculosis. J Nanosci Nanotechnol 19:622–626.

35. Gou D, Xie G, Li Y, Zhang X, Chen H. 2018. Voltammetric immunoassay for Mycobacterium tuberculosis secretory protein MPT64 based on a synergistic amplification strategy using rolling circle amplification and a gold electrode modified with graphene oxide, Fe3O4 and Pt nanoparticles. Microchim Acta 185.

36. Zhu C, Liu J, Ling Y, Yang H, Liu Z, Zheng R, Qin L, Hu Z. 2012. Evaluation of the clinical value of ELISA based on MPT64 antibody aptamer for serological diagnosis of pulmonary tuberculosis. BMC Infect Dis 12.

37. Thakur H, Kaur N, Sareen D, Prabhakar N. 2017. Electrochemical determination of M. tuberculosis antigen based on Poly(3,4-ethylenedioxythiophene) and functionalized carbon nanotubes hybrid platform. Talanta 171:115–123.

38. Thakur H, Kaur N, Sabherwal P, Sareen D, Prabhakar N. 2017. Aptamer based voltammetric biosensor for the detection of Mycobacterium tuberculosis antigen MPT64. Microchim Acta 184:1915–1922.

39. Qin LH, Ma ZZ, Zheng RJ, Liu ZH, Lu JM, Hu ZY. 2008. Selecting and affinity analysis of DNA aptamers to MPT64 protein from Mycobacterium tuberculosis. Chinese J Tuberc Respir Dis 31:453–458.

40. Qin LH, Zheng RJ, Yang H, Feng YH, Wang J, Hu ZY. 2008. The methodology for selecting DNA aptamer to MPT64 protein from Mycobacterium tuberculosis with SELEX technology. Chinese J Microbiol Immunol 28:176–180.

41. Li J, Hu K, Zhang Z, Teng X, Zhang X. 2019. Click DNA cycling in combination with gold nanoparticles loaded with quadruplex DNA motifs enable sensitive electrochemical quantitation of the tuberculosis-associated biomarker CFP-10 in sputum. Microchim Acta 186.

42. He F, Xiong Y, Liu J, Tong F, Yan D. 2016. Construction of Au-IDE/CFP10-ESAT6 aptamer/DNA-AuNPs MSPQC for rapid detection of Mycobacterium tuberculosis. Biosens Bioelectron 77:799–804.

43. Rotherham LS, Maserumule C, Dheda K, Theron J, Khati M. 2012. Selection and Application of ssDNA Aptamers to Detect Active TB from Sputum Samples. PLoS One 7.

44. Lavania S, Das R, Dhiman A, Myneedu VP, Verma A, Singh N, Sharma TK, Tyagi JS. 2018. Aptamer-Based TB Antigen Tests for the Rapid Diagnosis of Pulmonary Tuberculosis: Potential Utility in Screening for Tuberculosis. ACS Infect Dis 4:1718–1726.

45. Dhiman A, Haldar S, Mishra SK, Sharma N, Bansal A, Ahmad Y, Kumar A, Sharma TK, Tyagi JS. 2018. Generation and application of DNA aptamers against HspX for accurate diagnosis of tuberculous meningitis. Tuberculosis 112:27–36.

46. Das R, Dhiman A, Mishra SK, Haldar S, Sharma N, Bansal A, Ahmad Y, Kumar A, Tyagi JS, Sharma TK. 2019. Structural switching electrochemical DNA aptasensor for the rapid diagnosis of tuberculous meningitis. Int J Nanomedicine 14:2103–2113.

47. Kumari P, Lavania S, Tyagi S, Dhiman A, Rath D, Anthwal D, Gupta RK, Sharma N, Gadpayle AK, Taneja RS, Sharma L, Ahmad Y, Sharma TK, Haldar S, Tyagi JS. 2019. A novel aptamer-based test for the rapid and accurate diagnosis of pleural tuberculosis. Anal Biochem 564–565:80–87.

48. Mozioglu E, Gokmen O, Tamerler C, Kocagoz ZT, Akgoz M. 2016. Selection of Nucleic Acid Aptamers Specific for Mycobacterium tuberculosis. Appl Biochem Biotechnol 178:849–864.

49. Zhang X, Feng Y, Duan S, Su L, Zhang J, He F. 2019. Mycobacterium tuberculosis strain H37Rv Electrochemical Sensor Mediated by Aptamer and AuNPs-DNA. ACS Sensors 4:849–855.

50. Bekmurzayeva A, Sypabekova M, Kanayeva D. 2013. Tuberculosis diagnosis using immunodominant, secreted antigens of Mycobacterium tuberculosis. Tuberculosis.

51. Huang D Bin, Vu D, Cassiday LA, Zimmerman JM, Maher LJ, Ghosh G. 2003. Crystal structure of NF-κb (p50)2 complexed to a high-affinity RNA aptamer. Proc Natl Acad Sci U S A 100.

52. Jaeger J, Restle T, Steitz TA. 1998. The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor. EMBO J 17.

53. Kettenberger H, Eisenführ A, Brueckner F, Theis M, Famulok M, Cramer P. 2006. Structure of an RNA polymerase II-RNA inhibitor complex elucidates transcription regulation by noncoding RNAs. Nat Struct Mol Biol 13.

54. Storz G, Altuvia S, Wassarman KM. 2005. An abundance of RNA regulators. Annu Rev Biochem.

55. Chaudhary VK, Kulshreshta A, Gupta G, Verma N, Kumari S, Sharma SK, Gupta A, Tyagi AK. 2005. Expression and purification of recombinant 38-kDa and Mtb81 antigens of Mycobacterium tuberculosis for application in serodiagnosis. Protein Expr Purif 40.

56. Kulshrestha AA, Gupta AA, Verma NN, Sharma SK, Tyagi AKA, Chaudhry V, Chaudhary VK. 2005. Expression and purification of recombinant antigens of Mycobacterium tuberculosis

for application in serodiagnosis. Protein Expr Purif 44:75–85.

CHAPTER 8 CONCLUSION & RECOMMENDATIONS

The work performed herein enhanced our knowledge about one of the world's oldest disease-causing group of agents, MTC. Three unique aspects of TB were explored: disease transmission pattern in a marginalized indigenous community; genomic exploration of the evolution of a pathogenic bacteria to an attenuated strain; validation, and application of mycobacterial specific biomarkers (Pks5, MB2515c, MB1895c) in developing point-of-care diagnostics. The research led to noteworthy discoveries that further pave the way for integrating One Health in the TB control programs around the world.

For specific aim 1 we explored hyperendemicity of MTB within the Guarani-Kaiowá people in Chapter 2. The genome wide SNP analysis was performed for drug-resistance-associated genes as well as for lineage identification which led to the definition of multiple fixed and unique SNPs in the community. This led us to conclude that most TB in the Guarani- Kaiowá people is the result of a historical introduction of the commonly circulating 4.3.3 lineage, with subsequent diversification in the community. The major limitation of our study was the lack of lab-confirmed drug resistance data available for all samples. We suggest that the government improve upon its TB surveillance in the community along with introducing drug-resistance testing as a policy before the initiation of treatment. The study showed the importance of regular surveillance that would allow a community specific TB-control policy to be developed.

Specific aim 2 allowed us to investigate the genomic mutations that could have led to attenuation of MBO. Unlike MBO-BCG, MBO-Ravenel had no large deletions. Instead, multiple missense SNPs were identified in genes critical for the metabolism and survival of the

Mycobacteria. The experimental infection of cattle, pathological findings (performed but not described in this thesis), and the genomic evidence suggests that MBO Ravenel has evolved into an attenuated strain via accumulation of SNPs in genes associated with pathogenicity rather than loss of major genomic regions that have been described for BCG.

For specific aim 3, we successfully validated the presence of Pks5, MB2515c and MB1895c in experimentally infected non-human primates. Even though we had serum of only five infected primates, validation in experimentally infected animals provided us with the groundwork that was needed to extend the testing in naturally infected animal sera. Through our alliances in Africa multiple primate species are already being tested for the presence of these PSBs under natural settings. This validation work in naturally infected animals from the field would provide more support to our results and potentially extension of the project to humans.

Continuing with specific aim 3, the above work was followed by detection of circulating immune complexes captured on the DPP device. Mass-spectrometry analysis showed that the pathogen specific biomarker Pks5 could be extracted from the serum samples of infected animals. Thus, providing confidence to move towards development of a lateral flow device for detection of these PSBs. The major limitation of this work was the relatively small sample size. The limitation was addressed by having replicates for each time point that was tested.

For the last part of specific aim 3, we succeeded in selecting an aptamer against one of the biomarkers (MB2515c). We faced multiple limitations in our work. Initially a small peptide fragment (20-25AA) of each biomarker was used for aptamer selection. Unfortunately, the small size of the peptide hindered our ability to have consistent results for the specificity testing for

the selected aptamers. To overcome this problem, we expressed longer peptides (100 AA long). Expression of MTBC peptides using *E. coli* is challenging. We had to test multiple expression conditions before we were able to express two (Pks5 and MB2515c) of the three selected biomarkers. For the expression of MB1895c we suggest using a temperature even lower than room temperature. If that does not help, then changing expression host to *M. smegmatis* could provide the solution to our problem. The last limitation of this work was the fact that we were able to only select one aptamer against one biomarker. The detection specificity relies on our ability to detect a panel of biomarkers rather than just one. Performing another round of aptamer selection would likely provide the much-needed panel of aptamers that would allow the development of the first prototype of a point-of-care device for TB detection.

Taken together, this body of research has laid down the groundwork for an improved diagnostic test for TB, explored the unique evolution patterns of MTB as well as MBO that would facilitate in improved public health policy and potentially improved vaccines. The findings in this thesis would help the scientific community to develop robust scientific foundations to inform policy to control and eradicate TB.