ASSESSING SINGLE NUCLEOTIDE POLYMORPHISMS IN CLINICAL ISOLATES IN A SOCIAL NETWORKING TUBERCULOSIS TRANSMISSION STUDY IN KAMPALA, UGANDA AND THE CHARACTERIZATION OF *MYCOBACTERIUM TUBERCULOSIS* α-CRYSTALLIN (ACR) PROTEIN

by

EDRISS YASSINE

(Under the Direction of FREDERICK QUINN)

ABSTRACT

Tuberculosis (TB) in humans is the leading cause of death by a single infectious agent in the world today with approximately 2 billion individuals worldwide latently infected with the bacterium *Mycobacterium tuberculosis* (*Mtb*). Although the mortality rate for TB has been decreasing slowly in the last two decades, it still has not reached the levels set forth by the WHO End TB strategy. Currently, the most accepted method for eradication of disease is the Tuberculosis Directly Observed Treatment Shortcourse (TB-DOTS) program established by the WHO. Although this program has been finding and treating cases of active TB around the world, it is not a preventative method. Therefore, individuals with active disease may potentially pass on the disease before they have been diagnosed. Using Whole Genome Sequencing, Single Nucleotide Polymorphism analysis and examining transmission dynamics of TB in social networks outside of the home, we analyze the genome of clinical *Mtb* isolates to determine transmission between individuals. We show that using a threshold of $\leq$ 12 SNPs as indicative of person-to-person transmission, clusters of transmission between individuals can be inferred.

Incorporating this data with epidemiological data may provide a blueprint for the integration of public health interventions to prevent future TB transmission. In addition to transmission, TB latency is another major problem in the fight for eradication. During latency, individuals harboring *Mtb* show no clinical symptoms and before entering dormancy, the bacterium produces a protein called α-crystallin (Acr) which aids in survival while it's in a metabolically inactive state. Acr is the dominant detectable cellular protein during hypoxia encompassing 25% of total proteins produced, however, not much is still understood about the mechanisms of the protein. The process by which Acr is transported outside of the cell is a major missing piece of the puzzle. Using constructed plasmids encoding different regions of the *acr* operon as well as determining the phosphorylation state of Acr, we further the understanding of the transport process. Altogether, this research will inform future research for the development of effective vaccines and antibiotics.

INDEX WORDS: *Mycobacterium tuberculosis*, Whole Genome Sequencing, Single Nucleotide Polymorphism, Mycobacteria, α-crystallin, Phosphorylation

ASSESSING SINGLE NUCLEOTIDE POLYMORPHISMS IN CLINICAL ISOLATES IN A SOCIAL NETWORKING TUBERCULOSIS TRANSMISSION STUDY IN KAMPALA, UGANDA AND THE CHARACTERIZATION OF *MYCOBACTERIUM TUBERCULOSIS* α-CRYSTALLIN (ACR) PROTEIN

by

EDRISS YASSINE

BS, San Jose State University, 2006

MPH, Boston University, 2014

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2020

ASSESSING SINGLE NUCLEOTIDE POLYMORPHISMS IN CLINICAL ISOLATES IN A SOCIAL NETWORKING TUBERCULOSIS TRANSMISSION STUDY IN KAMPALA, UGANDA AND THE CHARACTERIZATION OF *MYCOBACTERIUM TUBERCULOSIS* α-CRYSTALLIN (ACR) PROTEIN

by

EDRISS YASSINE

Major Professor:	Frederick Quinn
Committee:	Russell Karls
	Lisa Shollenberger
	M. Stephen Trent
	David Peterson

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
May 2020

DEDICATION

This work is dedicated to my mom and dad, my wife Brianne, and to my entire family, friends and co-workers who have supported me throughout this long journey. Thank you for always reminding me that it would be worth it at the end.

ACKNOWLEDGEMENTS

The people to acknowledge and thank for encouraging me throughout this process could take a while, but I will try to keep it as short as I can. First, my advisor, Dr. Fred Quinn has been an amazing source of support during my PhD journey. He was there every step of the way and with his wealth of knowledge addressed every concern that I had. Next, I would like to thank Dr. Lisa Shollenberger for being a constant source of encouragement, advice and friendship. She was there for me from day one. From my initial rotation as a new PhD student at the University of Georgia to passing on her knowledge of everything PhD-related along the way. I am so happy you now have your own lab. You will be an amazing P.I. Dr. Karls has been there for me throughout this process with my protein work. Thank you for teaching me plasmid maps, everything cloning and sharing your extensive knowledge of Mycobacteria. Drs. Trent and Peterson. Your amazing advice and suggestions during my committee meetings consistently pointed me toward the right direction and brought me back into my lane when I was veering off course. Finally, Dr. Whalen, who allowed me to be included in his Ugandan social network study. This enabled me to learn many aspects of bioinformatics that I was not anticipating learning when I first came to UGA. This work also allowed me to travel to Uganda to expand my research capabilities and meet a new network of colleagues with similar research interests.

Next, I wanted to acknowledge my parents, Marzia and Nassim, without whom, attaining a higher education would not have been possible. Immigrating to this country with nothing but the clothes on our backs, working their fingers to the bone with multiple jobs just to make sure their kids could have a better life. Now that I am older I am able to look back at our lives and

realize all of the personal sacrifices that you made so Tahmina and I would not feel left out of anything in life. All of the knowledge and life experiences bestowed on us have made us the people who we are today, and I wouldn't trade it for anything in the world. My sister Tahmina for always being there and listening when I needed someone to talk to and encouraging me to pursue life. My niece Delilah and my nephew Zayne for always making my heart smile when I see you both. Although I have missed a lot of your growing up because I am far away, that will hopefully change soon. Erika and Billy Senese for giving us a home away from home, being our travel buddies and allowing us to come over more times than we probably should have. You were my family when mine was far away. Finally, Chris Slade. What can I say? Been together since the beginning of this program. Started together and now finishing together. I am so grateful that we became friends and I hope you achieve everything you hope for.

Last, but not least, my amazing wife, Brianne. Together through thick and thin. Thank you for encouraging me through the hard times and never letting me quit at my low moments. I couldn't have made it through this without your love and support. Thank you for reading all of my drafts and I'm sorry that you had to drive to Atlanta every week for work. I hope I can make it up to you some day. You are the greatest adventure partner anyone can ask for.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Tuberculosis (TB) is a disease that has coexisted with humans for over 70,000 years and despite its long relationship, it is still an inherent part of overall disease burden today [1]. According to the World Health Organization's (WHO) Global Tuberculosis Report 2019, in 2018, it is estimated that there were more than 1.2 million deaths among HIV negative individuals attributed to TB with 10 million overall new TB cases [2]. The WHO also estimates that there are over two billion individuals currently infected with *Mycobacterium tuberculosis* (*Mtb*) living throughout the world. This global pathogen has surpassed HIV/AIDS as the leading cause of mortality by a single etiological agent [3]. Due to its endemic and persistent prevalence in the world, the fight to eradicate this disease has been a public health problem for many years.

One of the most successful programs used to combat TB disease burden has been the WHO's Tuberculosis Directly Observed Treatment Shortcourse (TB-DOTS). Introduced worldwide in the 1990's, the TB-DOTS program aims to identify those with TB disease and treat them with a course of the most effective antibiotics to try to reduce the overall disease burden in hopes of control and ultimate eradication [4]. The term "directly observed" is used to describe an intervention in which patients are watched as the medicine is taken during the first three months of treatment, thereby ensuring the antibiotics are taken on time and in correct combination to help treat the disease and prevent the development of antibiotic resistance [4–6]. In conjunction with TB-DOTS, the WHO implemented The End TB Strategy with a goal to end the TB pandemic [7]. This program has set a target to have a 95% reduction in TB deaths and a 90% reduction in TB incidence rate by the year 2035. Despite these goals and treatments, the TB

eradication program has not seen as much of a decline in disease burden as hoped. This is evidenced by the aforementioned disease burden today; thus, new measures are needed to help control the process. One reason that TB-DOTS has not been as effective as hoped is that the program is based on finding active TB disease cases and treating them rather than employing a preventative model. Relying on a treatment model puts the program's success mostly on the reporting of cases by individuals and governments which can result in major delays in diagnosis and treatment, increasing the probability of diseased individuals transmitting the pathogen. For example, many people in the areas of the world where TB is endemic do not have easily accessible healthcare and by the time a person is diagnosed with active disease and treatment is implemented, they could have potentially infected many other individuals in their household or social network, thereby perpetuating disease [8].

Switching to a preventative method first requires a new system to identify individuals and locations within an area where transmission may occur most frequently and then using that information, implement targeted prevention methods. One new method used to confirm transmission among individuals is Whole Genome Sequencing (WGS) and Single Nucleotide Polymorphism (SNP) analysis. WGS has given epidemiologists and public health investigators the ability to decode an organism's DNA and look at each nucleotide individually [9]. A SNP is a nucleotide base variation at a single position in a DNA sequence. Generally, a nucleotide difference is considered a SNP when more than 1% of the population does not carry that specific nucleotide at the position [10]. Using the bioinformatics data gathered from the bacterial genomes, it can be determined, based on the number and location of SNPs between two isolates, if the isolates potentially came from the same source; the smaller the SNP amount, the more related the isolates. The current gold standard used in the literature for relatedness of isolates based on SNP

difference is set at a threshold of ≤12 SNPs per genome [11]. If enough transmission events are identified in a particular location, then prevention efforts can be specifically designed for that location to help prevent disease transmission.

Another major barrier in the control of TB within the population is the latency stage of disease. Latent tuberculosis infection (LTBI) is defined as having evidence of *Mtb* infection via standard diagnostic tests, but not having clinical symptoms of active disease [12–14]. During latency, the immune system cannot eliminate the infecting mycobacteria, so instead it attempts to control the spread of the organism by walling off the bacteria in a granuloma. Granulomas are aggregate structures, primarily in the lungs and lymph nodes, that form to contain the bacteria within a core surrounded by innate and adaptive immune cells [15–17]. These cores can become hypoxic thus forcing the mycobacteria to shift their metabolism into a slowly replicative or dormant mode that can persist for decades or for the duration of a person's lifetime [18]. In hypoxic conditions one of the major proteins produced by *Mtb* is the 16 kDa, α-Crystallin (Acr, HspX, Rv2031c). Acr belongs to the family of heat shock proteins and it is hypothesized that its primary job is to function as a chaperone to assist in the refolding of proteins and prevent auto-aggregation [19–21]. When the bacterium is in stationary phase, Acr is the dominant detectable cellular protein and can account for approximately 25% of total proteins present in culture [22]. Studies have shown that Acr is essential for the survival of *Mtb* during LTBI. Deletion of the *acr* gene has been shown to produce a two-fold increase in the elimination of the bacteria using standard anti-tuberculosis antibiotics and a lower relapse rate than wild-type *Mtb* suggesting that treatments to eliminate the disease could be focused on Acr [23]. The development of targeted therapeutics toward Acr, may help alleviate disease burden, especially those with LTBI.

This thesis will focus on two specific aims. The first will explore the issue of TB transmission between individuals with the use of WGS SNP analysis and the second will further our understanding of the function of Acr in LTBI.

Specific Aim 1: Define molecular changes in *Mtb* bacilli transmitted among contacts in a Ugandan social network. In addition to host factors, it is likely that there are differences present in the genomes of some strains of *Mtb* that allow them to propagate disease more efficiently and transmit between hosts. Our working hypothesis is that there are sequence differences present in the genomes of *Mtb* that make the bacteria more infectious and enable more effective spread to more contacts. Using WGS and analysis of SNPs, we will help identify genetic changes in *Mtb* isolates causing active TB disease within a Ugandan social network. More specifically, we will determine if there are specific genetic variations occurring that allow the bacteria to propagate disease more effectively in humans. To test our hypothesis, we will perform two sub-aims. First, we will use WGS and SNP analysis to obtain interpretable data for comparison purposes of the active infections within the social network. Next, we will identify any SNPs between the genetically-related active TB cases and determine if the SNPs are correlated with a gene(s) needed for transmission of the bacteria or genes correlated with active disease and LTBI, including Acr.

Specific Aim 2: Characterizing the mechanism of secretion of Acr from *Mtb* to the extracellular environment. Our working hypothesis is that when phosphorylated, *Mtb* Acr protein does not require co-transcribed gene products for secretion outside of the cell. Using bacterial plasmid cloning and site directed mutagenesis, we will further the understanding of the secretion

mechanism of Acr from the bacterium to the outside environment. To test our hypothesis, we will perform four sub-aims. First, we will clone the *acr* operon region in smaller segments into an *M. tuberculosis* genome-integrating backbone, starting with *acr* alone and subsequently adding upstream and downstream genes within the operon, resulting in a series of *acr* regional plasmids. Next, we will electroporate the constructed plasmids into *M. smegmatis* (a rapid growing commensal organism) to determine which regions of the operon are necessary for Acr secretion. Furthermore, we will determine the phosphorylation states of Acr from both a recombinant form of Acr isolated from *E. coli* and a native form from *Mtb* using mass spectrometry. Lastly, we will determine if the introduction of a Ser → Ala mutation in the determined phosphorylated serine residue effects the extracellular secretion of Acr.

CHAPTER 2

LITERATURE REVIEW

2.1 History of *Mycobacterium tuberculosis*

Mycobacterium tuberculosis (*Mtb*) is a facultative, intracellular pathogen that is the primary agent responsible for tuberculosis (TB) in humans. The bacterium is classified under the phylum *Actinobacteria* and family *Mycobacteriaceae*. Phylogeographical analyses have determined that *Mtb* originated in Africa approximately 70,000 years ago, co-evolving with humans during their migration out of Africa and becoming an established pathogen due to the increase of the human population as time progressed [1]. Evidence of TB has been found in mummified remains from ancient Egypt and depicted in art from that era of more than 5000 years ago, both showing evidence of characteristics of Pott's disease, a form of TB that occurs outside the lungs causing fusion of two adjacent vertebrae [24]. The writings of Hippocrates from ancient Greece also referenced the deadly disease referring to it as "Phthisis" [25]. The infectivity of the agent was demonstrated in 1865 by Jean-Antione Villemin, a surgeon that worked with TB patients, when he inoculated a rabbit with liquified material from the tubercle lesion of a dead patient [25]. Several months after inoculation, Villemin found that the rabbit had developed human-like TB. Although Villemin had been able to show that material from tubercle lesions could produce disease, the organism itself was not identified as the etiological agent until 1882 by the pioneering microbiologist Robert Koch for which he won the Nobel Prize in Medicine in 1905 [25]. Several years after identifying *Mtb* as the causative agent, Koch presented the idea that he had

identified a substance, called tuberculin, from the bacteria that could help combat the disease. Subsequently, two other investigators, Clemens Freiherr von Pirquet and Charles Mantoux, developed a diluted form of tuberculin that could be used to diagnose the presence of latent TB. The current form of purified protein derivative (PPD) skin test used to diagnose potential TB today was developed by Florence Seibert in the 1930's [25].

Despite its long history, TB is still one of the leading causes of death in the world today. According to the World Health Organization (WHO) global tuberculosis report, in 2018, it was estimated that there were more than 1.2 million deaths among HIV-negative individuals attributed to TB and 10 million overall new cases of TB disease [2]. Of the estimated 2 billion individuals infected with *Mtb* globally, it is estimated that 5-10% will develop active TB disease in their lifetime. There also are many factors that compound the occurrence of disease and complicate its treatment, such as the emergence of drug resistance and comorbid diseases like HIV/AIDS. Of the new cases of TB in 2017, 558,000 were multi-drug resistant (MDR-TB) and 1.2 million individuals were co-infected with HIV/AIDS [3]. The problem has reached such great proportions that tuberculosis has surpassed HIV/AIDS in global prevalence and as the leading cause of death by a single infectious agent in the world. The overall outlook of disease burden seems to be improving as the WHO reports that the incidence and mortality rate worldwide has been declining since 2000 by 2% and 3% per year, respectively. However, the ever-expanding burden to world public health, compounded with the threat of continuing emergence of drug resistance, compels us to propagate our knowledge to further advance the development of new treatments, vaccines and diagnostics.

Mycobacterium tuberculosis is the causative agent of the majority of human TB disease, however, the *Mycobacterium tuberculosis* complex (MTBC) group is composed of 9

mycobacterial species that cause a form of tuberculous or tuberculous-like disease in humans and/or other animals. The bacteria that cause disease in mainly humans are *M. tuberculosis*, *M. africanum* and *M. canettii*, *M. microti* is a pathogen of rodents, *M. bovis*, *M. caprae*, *M. mungi*, and *M. orygis* primarily target bovine species and *M. pinnipedii* mainly infect pinnipeds, such as seals and walruses [26]. Zoonotic transmission of *M. bovis* from cattle to humans occurs in approximately 1.4% of U.S. cases and likely significantly higher in animal TB endemic areas. [27] with numbers as high as 76% of human TB shown to be *M. bovis* in areas of Tunisia [28]. Zoonotic TB transmission can occur through close contact with infected cattle and other animals by aerosol inhalation or through the consumption of contaminated animal products such as unpasteurized milk. Non-tuberculous mycobacterial species have also been identified, many of which are environmental saprophytes or cause other forms of disease including lung infections in immunocompromised, elderly or those with genetic disorders that change the standard commensal-host relationship. Examples of non-tuberculous species include *M. avium*, *M. abscessus* and *M. kansasii* [29].

The first, and only, approved vaccine against TB, the attenuated derivative of *M. bovis* is strain Bacille-Calmette-Guérin (BCG), which was developed in 1921. Albert Calmette and Camille Guerin in 1908 at the Pasteur Institute in Lille grew a virulent strain of an *M. bovis* isolate on bile, glycerin and potato medium. After passaging the bacteria 230 times over the next decade, they generated an attenuated form of the original isolate that no longer caused disease and thus could be used as a vaccine [30]. The vaccination efforts were almost stopped in 1930 in Lubeck, Germany, when company producing the BCG vaccine accidentally contaminated the stocks with a virulent *M. tuberculosis* strain resulting in the death of 73 children [30]. The BCG

vaccine continued to be used, however, the route of vaccination was changed from oral to parenteral injection. This route is safer yet likely a less efficacious vaccination route, however, the injection route remains today. Despite its success, the BCG vaccine is most effective against preventing disseminated forms of disease in children and generally ineffective against pulmonary form in adults.

Treatment of TB disease took great strides in the mid 1900's. One of the first drugs used to combat TB was streptomycin in 1945. H. Corwin Hinshaw and Hugh Feldman used streptomycin in their Guinea pig model to show that the disease status of the infected animals improved when given the drug [31]. Concurrently, Jorgen Lehmann, developed para-aminosalicyclic acid (PAS), that also had the ability to kill *M. tuberculosis* bacilli during infection and improve disease outcome. Unfortunately, both streptomycin and PAS were only moderately effective, were prone to development of resistance and both had major side effects. Of the currently used drugs, isoniazid, was discovered in 1951 [31] and rifampin, a derivative of rifamycin, was FDA approved in 1971 [32]. These latter two drugs, along with pyrazinamide and ethambutol, are components of the four drug, first line anti-*Mtb* cocktail given to treat active infection today.

2.2 Tuberculosis Disease Control

In 1993, the WHO declared tuberculosis a global public health emergency and from then on have put forth efforts to eradicate the disease. To try to accomplish this, the WHO established the Tuberculosis Directly Observed Treatment Shortcourse (TB-DOTS) program [4] and The End TB Strategy [7]. Introduced worldwide in the 1990's, the TB-DOTS program is a "diagnose and treat" program aimed to identify those with TB disease and treat them with a course of the most

effective antibiotics to try to reduce the disease burden. TB-DOTS operates on five main principles: 1) Commitment from governments to sustain TB control activities 2) Detection of TB cases by sputum sample 3) A standardized treatment with antibiotics for those identified as having a positive smear test 4) A regular supply of antibiotics and 5) A reporting and documentation system for everyone in the program [4]. One of the main components for success is the directly observed aspect in which patients are watched as they take their antibiotics. Despite success in some parts of the world, TB-DOTS has been deemed a failure in sub-Saharan Africa due to the lack of implementation, below average public health systems, poverty and the prevalence of HIV [33].

In conjunction with TB-DOTS, the WHO's End TB Strategy is a global program with set goals and targets to achieve on the road to eradication. The program has been adopted by the members of the World Health Assembly and by the year 2025, the program has set a goal of a 75% reduction in TB deaths and a 50% reduction in TB incidence worldwide. Furthermore, by the year 2035, the program would like to increase the previous numbers to a 95% reduction in TB deaths and a 90% reduction in incidence rate [7]. The program is based upon three main pillars: 1) Integrated, patient-centered care and prevention: Diagnosis and treatment of all individuals with TB disease 2) Bold policies and supportive systems: Political involvement and universal coverage of the program and 3) Intensified research and innovation: Continuing the research for new drugs and strategies to end disease. These programs, working together, have only had moderate success in the fight against this deadly disease, however, with an increased effort, the tides will change for the positive in the near future.

2.3 Transmission and Pathogenesis of Disease

Mycobacterium tuberculosis is primarily transmitted through the aerosol route when a person with an active infection coughs, sneezes or speaks forcefully, thus producing bacteria-containing droplets that can be inhaled by another person. Once inhaled, the infection that follows is typically established in the lungs, however, the bacteria can spread to other organs such as the kidneys, spine and brain. In the airways the bacteria come in contact with dendritic cells (DCs), alveolar macrophages and lung epithelial cells. Although bacteria interact with all of these cell types, they are primarily internalized by the lung alveolar macrophages via phagocytosis. There are two ways that the bacteria are recognized for phagocytosis: either by activation of complement receptors on macrophages or recognition of bacterial mannose residues by macrophage receptors [34]. Once the bacteria are internalized they may remain as intracellular pathogens and replicate there while evading the host immune system [35]. The bacteria circumvent killing by the macrophages by possessing mechanisms that prevent the acidification of the phagosome, reduce damage from reactive oxygen species (ROS) and block recognition of infected macrophages by CD4+ T cells [36]. The bacterium prevents the acidification of the phagosome by blocking the fusion of the phagosome and lysosome by interfering with normal phagosomal biogenesis. This may be done by inhibiting the function of phagosomal proteins Rab5, Rab7 and EEA1, all of which play a role in the maturation and recruitment of the lysosome [36].The prevention of damage by ROS is not yet well understood, however, it is believed that *M. tuberculosis* encodes a gene called *ahpC* whose translated protein catabolizes the produced oxidant, peroxynitrite ($ONOO^-$), and prevents it from damaging the bacteria [34,36] Once the innate immune system becomes inadequate and the adaptive immune system is needed, the bacteria further evade immune system recognition by interfering with antigen presentation

that is needed to activate an effective response. Studies have shown that the bacteria have the ability to reduce MHC-II presentation leading to a decreased ability of T-cells to recognize those foreign antigens and destroy the infected cells [37].

2.4 Cell Wall Structure

The mycobacterial cell wall structure is complex and contributes to bacterial resistance to initial host innate defense systems, facilitating survival in inhospitable conditions in which other pathogens would not (Figure 1). Over 60% of the composition of the cell wall is made of lipids including mycolic acids [38,39]. Immediately outside the cytoplasmic membrane are covalently-attached layers of peptidoglycan, arabinogalactan and mycolic acids [38]. The mycolic acids are long, branched fatty acids that contain between 70-90 carbons. These lipids are linked to arabinogalactan that together form the waxy coating of the *M. tuberculosis* bacilli [40]. The outer cell wall layer is composed of free lipids, fatty acids and glycoproteins. The combination of proteins, sugars and lipids make the bacteria in this genus difficult to degrade in addition to preventing the entry of small charged molecules. This coating has been shown to be associated with the virulence of the pathogenic species by providing resistance to antibiotics, killing by acidic or alkaline substances and resistance to oxidation, allowing *Mtb* to survive in macrophages and establish infection [39].

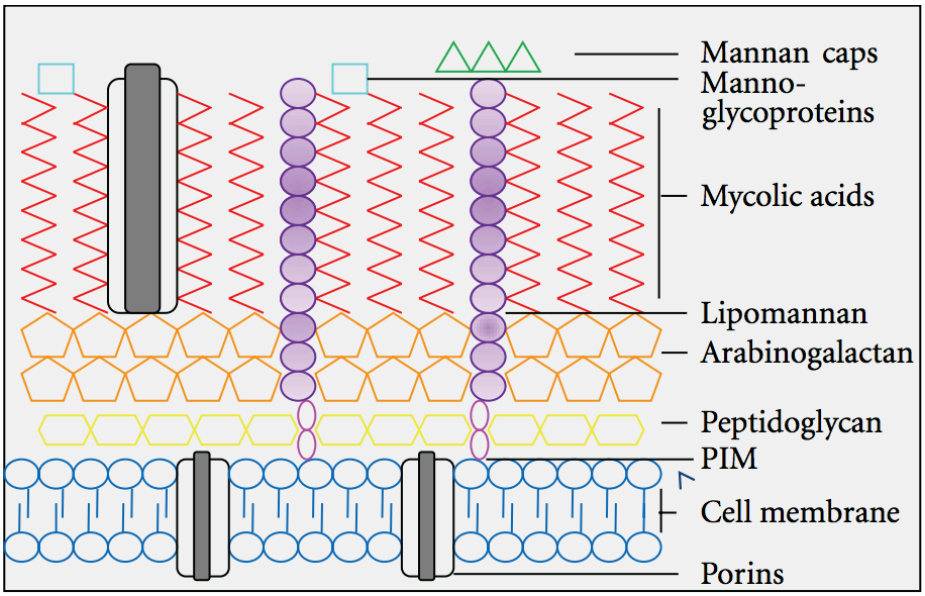

Figure 1: Illustration of a mycobacterial cell wall composition [41]

2.5 Tuberculosis Disease

It is estimated that in approximately 90% of *Mycobacterium tuberculosis* infection cases, the bacteria do not cause infection and instead are completely cleared by the lung innate immune response. In the remaining 5-10% of exposed individuals, the bacteria are dormant or slowly replicating and establish a latent TB infection (LTBI). Once the bacteria have established infection in the lungs, it is hypothesized that the host's immune system forms chronic inflammatory lesions, called granulomas, around the site of infection. In the early stages of granuloma formation, the cells that compose these structures are the innate immune cells recruited to the site of infection, including epithelial macrophages, neutrophils, multinucleated giant cells and foamy macrophages. However, as the adaptive immune system becomes involved, more specialized cells, including different variations of T and B cells, including Th1, T_{reg}, B_1 and B_2 cells, aggregate with the macrophages and form a more defined structure (Figure 2) [15]. With time, the centers of the granulomas liquefy and can become necrotic with material

referred to as caseum [16]. During the latency period, an individual harboring the bacteria is asymptomatic and hypothesized to no longer be infectious. An individual with LTBI can remain in this state for decades; however, there is a 10% lifetime risk of disease reactivation usually associated with the decline of the individual's immune status [18]. The risk of reactivation of disease can also be increased due to factors like co-infection with other immune modifying diseases such as HIV/AIDs, or by non-infectious means including the use of steroids, malnutrition, smoking or diabetes [18].

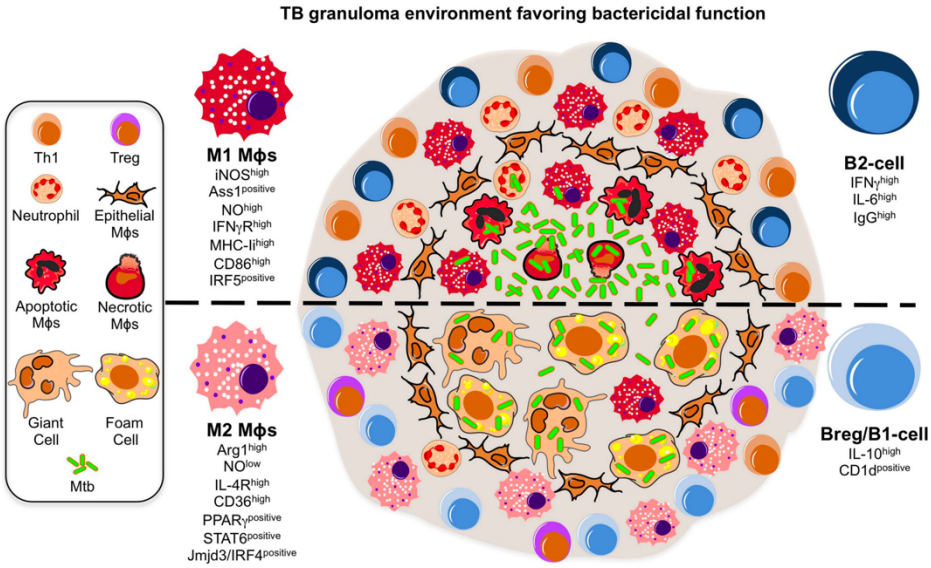

Figure 2: Depiction of the cellular structure of a granuloma [15]

The approximately 100 genes identified as being involved in *M. tuberculosis* survival within the granuloma are part of the DosR regulon. The genes in the DosR regulon are activated when the bacilli are exposed to nitric oxide or to hypoxic conditions. The regulon is comprised of 48 co-regulated genes, including alternative electron transport pathways (*fdxA*) and nitrate metabolism (*narX*) [42,43]. The DosR regulon is regulated through a three-component regulatory system composed of two sensor kinases, DosS and DosT, and a regulator, DosR. When the

bacteria sense their environment shifting toward a less oxygenated state, the DosR regulon genes are up-regulated. Studies have shown that *Mtb* deficient in DosR have a 10,000-fold decrease in survival rate during anerobic conditions [43]. Furthermore, genes of the DosR regulon may be associated more with sensing environmental signals rather than virulence due to the fact that many of the genes also are conserved in non-tuberculous mycobacteria [42]. The system is activated when DosT senses hypoxic conditions or elevated nitric oxide and activates DosR via transfer of a phosphate group to its aspartate residue. Following this event, DosS is activated, which fully initiates the regulon [44]. Both DosS and DosT are needed for full activation of the regulon.

The study of *Mtb* pathogenesis, especially the latent stages, can be challenging due the difficulty in using physiologically appropriate animal models and the differences in environmental interactions with the bacteria when grown *in vitro* and *in vivo*. One method used to mimic hypoxic stress on the mycobacteria is the Wayne culture. The Wayne culture allows the study of hypoxic conditions without the use of long-term non-rodent mammalian animal models. In this method, the *Mtb* cultures are grown in a sealed flask containing a limited 0.5 head-space to liquid medium volume ratio while being slowly agitated during which time the oxygen levels are slowly depleted as the bacteria replicate thus providing a hypoxic niche [45]. Another method for the study of *Mtb* in late stage disease is growth as a pellicle [46]. The pellicles differ from the Wayne model as they are grown in a stationary phase, without agitation, forming a thick bacterial layer at the air-liquid interface of the medium. Pellicles have been shown to produce higher amounts of mycolic acids and several pellicle-specific proteins, including Rv0097 and FabG4 [47,48]. Furthermore, the pellicle-grown bacteria show less antibiotic sensitivity than wild type cultures indicating a protective factor of the biofilm for the bacteria [48]. Some animal models

for studying LTBI exist, however, they are not completely representative of disease in humans. For example, the mouse model has been used due to the ease of manipulation, but the mouse model can be problematic because the formation of granulomas in mice and in humans differ due to the mice granulomas not becoming hypoxic, calcified, or caseous [49]. One of the more widely used models for LTBI has been the Guinea pig model. Although these animals do not develop human-like LTBI, they do develop granulomas with hypoxic cores. The development of these types of cultures and models may be essential for the study of *Mtb* survival during LTBI and for the development of new vaccine and drug targets against other stages of disease.

2.6 *Mycobacterium tuberculosis* Genome

The *Mycobacterium tuberculosis* H37Rv genome is comprised of approximately 4.4 million base pairs that potentially encode over 4,000 genes [50,51]. The genome contains a relatively high G+C content of 65.6%. Furthermore, it has been shown to contain 50 genes encoding functional RNAs and over 3,900 open reading frames (ORFs) [50]. Of the genes identified, the functions of approximately 40% are known or suspected. *M. tuberculosis* inherently contains some antibiotic resistance due to its outer cell wall layer previously described. However, some resistance is also found encoded in the genome of the bacterium. For example, resistance to β-lactams is one of several resistances encoded in the genome [50]. Furthermore, the *M. tuberculosis* genome encodes for a wide array of genes involved in lipid metabolism. It has been shown that there are over 250 known enzymes encoded by *M. tuberculosis* for the purpose of fatty acid metabolism. This accounts for five times as many enzymes as encoded in the genomes of other bacterial species such as *E. coli* [50]. The *Mtb* genome also encodes several virulence factors that the bacterium uses to establish itself in the

host. These include catalase-peroxidase to protect against reactive oxygen species, *mce*, which encodes macrophage colonizing factor and *sigA*, encoding the primary transcription sigma factor, which is needed intracellular and extracellular growth [52].

The genome of *M. tuberculosis* encodes the virulence factors, but it also provides a useful gauge for determining species-specific diversity. This latter trait is important because it provides another level of classification that can extend beyond host and environmental factors. Currently, seven global *M. tuberculosis* lineages have been identified: 1-Indo-Oceanic, 2-East Asian (Beijing), 3-East African Indian, 4-Euro-American, 5-West Africa I, 6-West Africa II and 7-Ethiopia-Horn of Africa (Figure 3) [53]. Lineages are important for implementing control measures because it has been shown that different lineages may correlate with different epidemiology and disease outcomes [54,55]. One method of using genomics to determine lineage is by the detection of single nucleotide polymorphisms (SNPs). A SNP is a nucleotide base variation at a single position in a DNA sequence. Generally, a nucleotide difference is considered a SNP when approximately more than 1% of the population does not carry that specific nucleotide at the position[10]. SNPs can be found in both coding and non-coding regions of sequences and may or may not change the amino acid sequence depending on the nucleotide substitution. If a SNP occurs in a coding region, it can present in two forms, synonymous or non-synonymous. Synonymous SNPs are those that do not change the coding sequences due to the change in nucleotide; they are referred to as silent because they do not change the function of the protein that they encode [56]. In contrast, non-synonymous SNPs are substitutions that do alter the amino acid sequence. Non-synonymous mutations can be further subdivided into two groups, missense and nonsense. A missense mutation occurs when a single nucleotide change encodes a different amino acid and a nonsense mutation occurs when the change encodes a stop codon. The

encoding of a stop codon usually renders the protein inactive. SNPs can alter the susceptibility to certain diseases. For example, it has been shown that certain SNPs in the human genome are the cause of cystic fibrosis [57]. Furthermore, at the bacterial level, it has been shown that SNPs in certain *M. tuberculosis* genes like *katG, mabA* and *Rv1772* encode drug resistance to one of the aforementioned primary TB drugs, isoniazid [58]. Through genomic analyses, specific SNPs can be identified as unique to a particular lineage, giving one the ability to distinguish between lineages based on the presence or absence of the SNP. SNP analysis can be further used to identify sub-lineages. Therefore, understanding SNPs and being able to identify differences between strains can help increase knowledge of how SNPs alter functions of the organism. This may potentially lead to early genomic identification systems for diseases like TB, during asymptomatic periods in order to help prevent or treat the disease and prevent transmission.

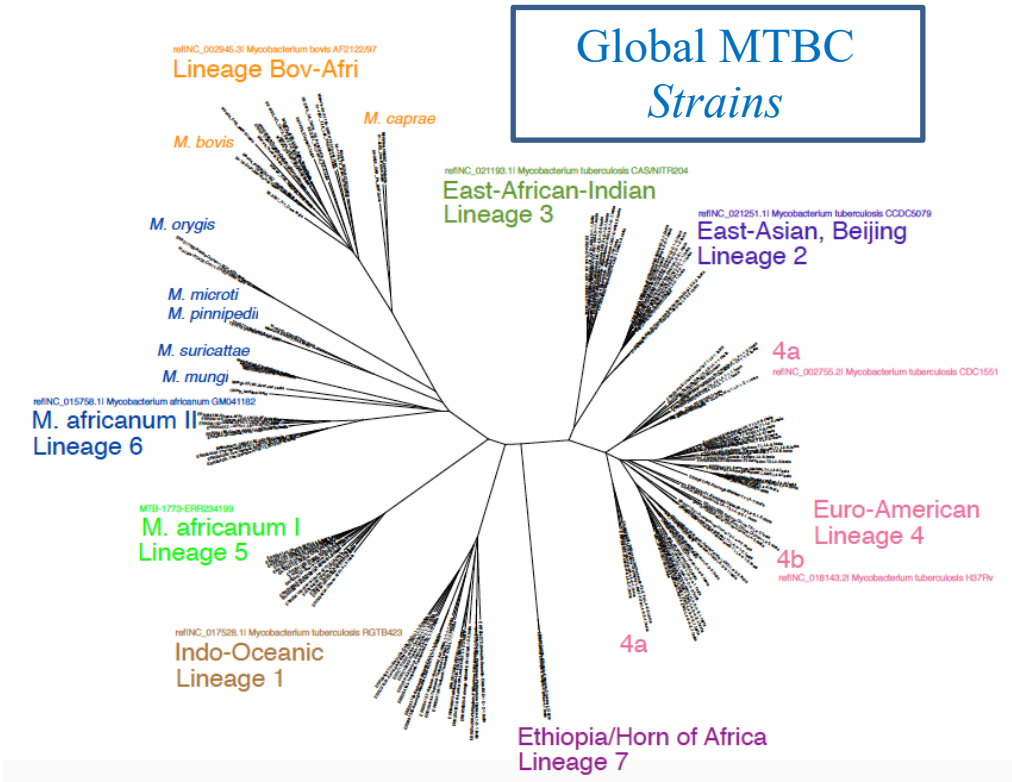

Figure 3. *M. tuberculosis* Phylogenetic Tree*. Phylogenetic tree separating *Mtb* isolates into its 7 lineages using SNP analysis. *Adapted from T. Stuber, USDA, Ames, IA

2.7 Whole Genome Sequencing

Whole genome sequencing (WGS) has given us the ability to examine an organism's genetic structure down to the single nucleotide. This ability has opened up a world of knowledge that humans did not have access to previous to this discovery. The pioneer of sequencing was two-time Nobel prize winner, Dr. Frederick Sanger. In 1977, Sanger developed his dideoxy method of sequencing that used radiolabeled dideoxynucleotides (ddNTPs) incorporated into DNA strands polymerized from plasmid, phage, or genomic sequences to determine the position of specific nucleotide bases in a genome [9]. This method of sequencing is known as the Sanger method. One of the next great achievements in sequencing was developed in 1986 by Leroy Hood with the introduction of automated DNA sequencing [59]. Hood showed that DNA sequences could be determined with a computer and without the use of gels. Fluorescently-labeled ddNTPs were used to detect the unknown base as it passed by a florescent detector in a machine and the information stored in a computer. This is still the basis of the current array of sequencers in use today. It was not until 1995 when the genome of the first living organism was sequenced. Craig Venter and colleagues used genomic sequencing to obtain the complete nucleotide sequence of the bacterium *Haemophilus influenzae* [60]. This led to one of the biggest milestones in sequencing history which was achieved in 2001 when the Human Genome Project published its first draft of the complete 14.8 billion base pair human genome sequence [61]. Capabilities for storing and analyzing the sequences also have been developing and improving culminating in 1982 with Walter Goad initiating the database called GenBank where all publicly available genomic sequences are stored and available for use. As of June, 2019, GenBank contains over 213 million sequences (NCBI, 2019). Sequencing genomes also has become less expensive as technology has

advanced. The human genome project had a budget of $3 billion dollars to sequence the human genome; this is in stark contrast to the estimated costs in 2017 of $0.014 per megabase, which would render most genome sequencing to be very inexpensive [63].

2.8 α-crystallin (Acr)

The *Mycobacterium tuberculosis* α-crystallin (Acr, HspX, Rv2031c) protein is 16kDa in size and encoded by the *acr*/*hspX* gene in *Mtb*. The mammalian homolog of α-crystallin was first identified as a chaperone protein acting as a lens structural protein in the eyes of vertebrates [19]. Its function in the eye has been identified to be part of the refractive element needed to produce the refractive index in the lens of the eye and maintaining lens clarity. When performing an alignment comparison, it is observed that human Acr and *Mtb* Acr contain only a 18% identity to each other [64]. Both forms of the protein contain the core α-crystallin domain shared by all members of the heat shock-family of proteins. Upon further analysis, the mammalian homolog was also found in other tissues, including the heart and the brain [19]. The *Mtb* Acr belongs to the family of heat shock proteins (HSP) whose primary job is to function as a chaperone; to assist in the refolding of proteins and to prevent auto-aggregation. Furthermore, HSPs also prevent the thermal denaturation of proteins as well as increase resistance to hydrogen peroxide [65]. In *Mtb*, Acr is known to be expressed during LTBI. The *acr* gene is part of the aforementioned DosR regulon that is up-regulated following low oxygen stress or the presence of nitric oxide. The protein is highly up-regulated during stationary phase and is undetectable during logarithmic phase of the bacterium. The hypoxic environment induced during LTBI serves as a signal for the induction of the protein. When the bacterium is in stationary phase, Acr is the dominant detectable cellular protein and can account for approximately 25% of total proteins [22]. Studies

have shown that Acr is essential for the survival of *Mtb* during LTBI. Deletion of the *acr* gene has been shown to eliminate bacterial burden 2 weeks faster than wild-type using standard anti-tuberculosis antibiotics and a lower relapse rate than infections with wild-type bacilli suggesting that treatments to eliminate the disease should be focused on *acr* [23].

The precise role for Acr in LTBI is not well understood, however, some preliminary studies have shown that Acr may contribute to latency by inhibiting several essential functions of antigen-presenting cells including dendritic cells (DC). First, a majority of cells of the adaptive immune system, including DCs, need both primary and co-stimulatory factors to activate the cell. For DCs, these factors include the molecules CD80, CD86 and CD40. It has been shown that Acr can stop the maturation of DCs by down-regulating the expression of these three molecules and thereby blocking DC activation of T cells [66]. Next, DCs treated with Acr show an increase in the expression of PD-L1 and Tim-3, which are both associated with the DC's ability to induce tolerance. Tolerance may allow the bacilli to hide from the host immune system and survive for longer periods of time [66]. Furthermore, Acr has the ability to modulate the expression of suppressor of cytokine signaling (SOCS) genes. SOCS-1, SOCS-2 and SOCS-3 are negative regulators of cytokine signaling in the JAK/STAT pathway. The JAK/STAT pathway is used by the cell to transmit extracellular signals to target gene promoters in the nucleus to exert transcriptional regulation [67]. Acr alters the function of the pathway by regulating the expression the SOCS genes, preventing signaling and weakening DC function [66]. Finally, Acr has been reported to suppress DC maturation by inhibiting the activity of NF-kB (nuclear factor kappa light chain enhancer of activated B cells). NF-kB is a protein factor that controls several processes in the cell including transcription of DNA and cytokine production. The activity of NF-kB is regulated by the inhibitor IkBα. The IkBα/NF-kB complex together inhibits function of

NF-kB; however, when activation signals are received, IkBα is phosphorylated and degraded allowing NF-kB to enter the nucleus [68]. To attenuate DC function, Acr down-regulates the activity of NF-kB, followed by a decrease in the amount of cytokine production and loss of activity for the DC [66]. All of these factors together allow the bacteria to circumvent the host immune system and persist in the human body.

Phosphorylation is a type of post-translational modification in which addition of a phosphate group to a protein serine, threonine or tyrosine residues induces a conformational change that can either activate or deactivate the protein [69]. Autophosphorylation occurs when the protein has a kinase function that phosphorylates its own amino acid and is most commonly used to convert extracellular signals into cellular functions. Acr has been shown to be a phosphoprotein that contains internal protein kinase activity that allows the autophosphorylation of the protein exclusively at a serine residue contained in the Acr genetic sequence [70]. Autophosphorylation activity is conserved within the genus *Mycobacterium* and has been shown in *M. smegmatis* as well as in other proteins in *Mtb*. Studies have shown that when autophosphorylation is inhibited in proteins, like *Mtb* PknB, there is a decreased growth of the bacteria [71]. From these results, it can be inferred that autophosphorylation may play a significant role in the physiology and metabolism of *Mtb* and may potentially be a target for future therapeutics.

CHAPTER 3

Assessing Transmission of *Mycobacterium tuberculosis* in a Defined Social Network Using Single Nucleotide Polymorphism Threshold Analysis

[1] Edriss Yassine, Lauren Cowan, Ronald Galiwango, James Posey, Willy Ssengooba, Fred Ashaba, Moses L. Joloba, Sarah Zalwango, Christopher Whalen and Frederick Quinn. To be submitted to *The Journal of Infectious Disease*

Abstract

Tuberculosis (TB) in humans is the leading cause of death by a single infectious agent in the world today with approximately 2 billion individuals worldwide latently infected with the etiological agent *Mycobacterium tuberculosis* (*Mtb*). Although the mortality rate for TB has been decreasing slowly in the last two decades, it still has not reached the goals set forth by the World Health Organization (WHO) End TB strategy. Currently, the most accepted method for controlling this disease is the Tuberculosis Directly Observed Treatment Shortcourse (TB-DOTS) program established by the WHO. Although this program has been finding and treating cases of active TB around the world, it is not a preventative method, and therefore, individuals with TB may potentially transmit disease before diagnosis. Uganda is one of the 30 high TB burden countries. Using whole genome sequencing, single nucleotide polymorphism (SNP) analysis and examining transmission dynamics of TB in social networks outside of the home, we analyzed the genomes of 113 *Mtb* clinical isolates and established that in the Rubaga Division of Kampala, Uganda, these isolates grouped into *Mtb* complex (MTBC) lineages 1, 2, 3 and 4, with the most isolates grouping into lineage 4. Possible transmission pairs containing $\leq$ 12 SNPs between genomes were identified in lineages 1, 3 and 4 with dominant transmitters identified in lineages 3 and 4. Incorporating this analysis with epidemiological data may provide a blueprint for the integration of public health interventions to decrease future TB transmission in this area of Uganda.

Introduction

Tuberculosis (TB) in humans is caused primarily by infection with *Mycobacterium tuberculosis* (*Mtb*). Most TB disease is generated when the bacilli transmit person-to-person via the aerosol route from an individual with an active infection coughing, sneezing, or speaking forcefully. Once the mycobacteria-containing droplets are inhaled by an individual in close proximity, the infection that follows is typically established in the lungs; however, the bacteria can disseminate to other organs such as the kidneys, spine and brain [34,72].

The World Health Organization (WHO) estimates that in 2018, there were 10 million new TB cases and 1.5 million deaths [2]. TB is again the leading infectious cause of death in the world today due to a single agent. An estimated two billion individuals may be latently infected with approximately 5-10% being at risk for reactivation TB in their lifetime [2]. Although the overall outlook for disease control has been reported to be trending positively, with incidence and mortality rates declining by 2% and 3%, respectively, since the year 2000, we are still below the goals set forth by the WHO End TB Strategy [2,7].

In an eradication effort begun in the 1990's, the WHO established the Tuberculosis Directly Observed Treatment Shortcourse (TB-DOTS) program [4] aiming to identify individuals worldwide with active disease and administer a course of antibiotics for treatment. However, TB-DOTS is based on a passive, find-and-treat approach, meaning there is opportunity for new infections to occur before a positive confirmation of a diagnostic test is obtained. A more preventative approach would identify infected individuals early in the disease course, and thus preventing subsequent transmission and preventing disease [73]. Despite success in some parts of the world, TB-DOTS has been significantly less effective in sub-Saharan Africa due to the lack of available rapid and sensitive diagnostic tests, below average public health systems and

manpower, systemic poverty and the high prevalence of HIV [33]. In 2018, 24% of all new TB cases occurred on the African continent with 8.6% among individuals with HIV [2]. Additionally, the lack of an effective vaccine is impeding global TB control efforts. The only globally approved vaccine is Bacille-Calmette-Guérin (BCG) [30] derived from *Mycobacterium bovis* almost a century ago. However, it is ineffective against pulmonary disease in adults and it is not recommended for general use in the United States due to invalidation of the Mantoux TB skin test in those receiving the vaccine. There are several new TB vaccines in development and clinical testing, but currently none have proven to more protective than BCG in clinical trials.

Uganda is one of the 30 high TB burden countries identified by the WHO with 86,000 new TB cases and an incidence rate of 200/100,000 in 2018 [2,74]. The disease burden is compounded by comorbidity with HIV as 34,000 of the new cases were individuals who were HIV-positive[74]. Previous studies have demonstrated *M. tuberculosis* transmission between household contacts [75,76], which was long thought to be the primary means of dissemination. More recent epidemiological studies show however, that *M. tuberculosis* transmission is more likely to occur outside of the household [72,77]. From outbreak investigations, research shows that transmission of *M. tuberculosis* bacilli can occur in social settings [78,79] and at other events in the community [80,81]. However, the frequency of occurrence in these settings outside of the household is not known.

Whole genome sequencing (WGS) has given researchers the ability to examine an organism's genetic structure down to the single nucleotide. In *M. tuberculosis*, WGS also has allowed investigators to determine genetic diversity within the species and identify genomic variances potentially involved in pathogenesis [82]. Another interwoven use of WGS is to highlight transmission patterns based on the detection of single nucleotide polymorphisms (SNPs). A SNP

is a nucleotide base variation at a single position in a common DNA sequence. Generally, a SNP is considered authentic when more than 1% of the population does not carry that specific nucleotide at the position [10]. SNPs can be found in both coding and non-coding regions of sequences and may or may not change the amino acid sequence depending on the nucleotide substitution. In *M. tuberculosis*, it has been shown that single SNPs in certain genes like *katG, mabA* and *Rv1772* can enable drug resistance to one of the primary TB drugs, isoniazid [58]. In our study, using WGS and SNP analysis of *M. tuberculosis* isolates collected from active TB cases within a Ugandan social network study [83], we assessed transmission of disease by comparing the number of SNPs among the isolates using the SNP threshold method. The transmission data presented can be used to determine possible transmission hotspots within Ugandan social networks that can then be the target of public health intervention efforts in the future.

Materials and Methods:

Growth and DNA Isolation of Clinical Isolates

Culturing and manipulation of *M. tuberculosis* isolates were performed in the College of American Pathologist (CAP)-accredited, Mycobacteriology (BSL-3) Laboratory in the Department of Medical Microbiology, Makerere University College of Health Sciences, Kampala, Uganda. Clinical isolates were sub-cultured on Middlebrook 7H10 agar (Becton and Dickinson, New Jersey), incubated at 37°C in a CO_2 incubator (ThermoFisher Scientific, Massachusetts) set at 5%. Growth was observed daily for 28 days. The bacteria were harvested and suspended in absolute ethanol (Sigma Aldrich, Missouri) for inactivation by suffocation. After inactivation, chromosomal DNA was extracted using the protocol outlined in the ZR Fungal/Bacterial DNA Microprep kit (Zymo Research, California) with a slight modification.

Because a Bead-beater instrument was not available, bacterial cells in ZR BashingBead Lysis tubes were attached to a vortexer and shaken for 5 minutes for lysis. After elution of each sample, the DNA concentrations were measured using a Nanodrop spectrophotometer. The DNA extracts were then shipped at ambient temperature to the Department of Infectious Diseases, University of Georgia, College of Veterinary Medicine, Athens, GA.

Sterility Testing

Sterility testing of the DNA samples was performed prior to WGS. Middlebrook 7H10 agar (Becton and Dickinson, New Jersey) in 150 mm Petri dishes was spotted with 1 μL of each sample. One microliter of *Mycobacterium bovis* BCG was used as a positive growth control. Plates were incubated at 37°C/5% CO_2 for six weeks and observed for growth. After the DNA samples were confirmed negative for growth, the remainder of the DNA samples were transferred to 96-well plates and stored at -20°C until processed for DNA sequencing.

Whole Genome Sequencing (WGS) and Single Nucleotide Polymorphism (SNP) Analyses

Sequencing libraries were prepared using Nugen Ultralow V2 or Nextera XT V2 following manufacturer's recommended protocol. The libraries were sequenced on a NextSeq 500 using mid output V2 chemistry (2 x 150bp) or on a Miseq using V2 chemistry (2 x 250bp). SNP analysis was conducted using BioNumerics 7.6.3 (Applied Maths NV). Reference guided assemblies were created using BioNumerics Reference Mapper 1.2.3 [84] with *M. tuberculosis* H37Rv (NCBI NC_00962.3) used as the reference genome for alignment. The settings for base calling were set as follows: minimum total coverage =3, minimum forward coverage =1, minimum reverse coverage=1, Single base threshold = 0.75, double base threshold=0.85, triple base threshold=0.95, and gap threshold = 0.5. Isolates found with an average coverage of the genome of less than 50 were re-sequenced. Reference guided assemblies are compared using

Bionumerics 7.6.3 SNP analysis filters. For a SNP to be retained in the analysis, it had to meet the following criteria: Have a total coverage of 5 reads, not contain ambiguous bases, not contain unreliable bases, not contain gaps and not be within 12 base pairs of adjoining called SNPs. Non-informative SNPs were also excluded from further analysis. The number of high-quality SNPs determined to be present between two isolates were recorded as the SNP distance. Using the SNP threshold method, we used the Walker et al. limit of ≤ 12 SNPs being the determinant of relatedness between two isolates [11].

Transmission Network Analysis

Transmission networks were created using R statistical software (Vienna, Austria) and data visualization package qgraph. SNP distance matrices outputted by the BioNumerics pipeline were supplied into qgraph and desired output settings (color and SNP ranges) were selected to create the transmission network.

***Mycobacterium tuberculosis* gene SNP Search**

SNPs present in *M. tuberculosis* isolates were searched using UNIX command line tools. When the position of each SNP was attained, specific codon mutations were visualized using Integrative Genomics Viewer (IGV) (Broad Institute, MA).

Results:

Study Design

This transmission study was conducted in the Rubaga Division of Kampala, Uganda, located in the western division of the city. According to the Uganda Bureau of Statistics' National Population and Housing Census 2014, Rubaga has a population of approximately 380,000 individuals [85]. Tuberculosis is a growing problem in this area of the city with the

prevalence of positive TB smear tests estimated to be 1,025 per 100,000 individuals, with a third of cases also being HIV-positive [86]. Adults, 15 years of age and older, presenting with TB symptoms and residing in the Rubaga Division were given a clinical test and acid-fast staining was performed on two sputum samples. Individuals were included in the study if they showed clinical symptoms of pulmonary TB in addition to two positive sputum smears.

SNP Analysis

The genome of *M. tuberculosis* provides a useful means of determining species-specific diversity. Currently, seven global *M. tuberculosis* complex (MTBC) lineages have been identified: 1-Indo-Oceanic, 2-East Asian (Beijing), 3-East African Indian, 4-Euro-American, 5-West Africa I, 6-West Africa II and 7-Ethiopia-Horn of Africa [53]. Lineages are important for implementing control measures because it has been shown that different lineages may correlate with different epidemiology and potential disease outcomes [54,55].

Sequences from a total of 143 isolates were analyzed using the BioNumerics pipeline. Isolates were excluded from further analysis if they did not meet the inclusion criteria of >50X average read coverage. If a sample was initially found to not meet the inclusion criteria, it was analyzed a second time for thoroughness. Samples were fully excluded after failing to meet the criteria during both analyses. Of the 143 sequences analyzed, 25 sequences were excluded for not meeting inclusion criteria, one for failing the *de novo* assembly process, 3 were mixed samples and one for general sequencing failure. After exclusion, a total of 113 isolates were included in the final SNP analysis (Figure 4).

Lineage 1

Of the 113 isolates analyzed, 2 isolates, 17918 and 20850, grouped into MTBC lineage 1, Indo-Oceanic. SNP analysis determined that the two isolates were identical with 0 SNPs occurring between them; indicating a possible transmission pair transmitted from the same individual.

Lineage 2

A single isolate, 28272, grouped into MTBC lineage 2, East Asian (Beijing). A second isolate forming a transmission pair was not identified thus indicating the single isolate was an isolated strain within the sampled population.

Lineage 3

A total of 23 isolates grouped into MTBC lineage 3, East African Indian, separating into 2 transmission clusters of interest (Figure 5). Isolates 16294, 20695, 20839, 19621, 20918 and 22199 formed cluster 1 and isolates 20060, 20061 and 18346 formed cluster 2 (Table 1). The minimum spanning tree (MST) for lineage 3, cluster 1 (Figure 6) shows two sets of isolates are identical with 0 SNPs between isolates: 16294 and 20695 as the first pair, and 19621 and 20918 as the second. Furthermore, each isolate from within each of these groups differs by a single SNP from each other. Finally, in regard to cluster 1, isolates 16294 and 20695 differ by one SNP from isolates 20839 and 22199. The MST for lineage 3 cluster 2 (Figure 7) shows that isolates 20060 and 20061 are identical with 0 SNPs between each other and that those two isolates differ by 2 SNPs from isolate 18346. All samples from the two clusters contain $\leq$ 12 SNPs that may indicate that isolates from cluster 1 have been transmitted from a single individual and all isolates from cluster 2 may have been transmitted from another single individual.

Table 1: Table of MTBC lineage 3 clusters showing possible transmission pairs with the number of SNPs between isolates.

Cluster Number	Isolate Pair IDs	Number of SNPs Between Pairs
1	16294 20695	0
	19621 20918	0
	16294 19621	1
	16294 20918	1
	20695 19621	1
	20695 20918	1
	16294 20839	1
	20695 20839	1
	16294 22199	1
	20695 22199	1
2	20060 20061	0
	20060 18346	2
	20061 18346	2

Lineage 4

There was a total of 87 isolates grouped into MTBC lineage 4 Euro-American, separating into 19 clusters (Figure 8). Of the 19 clusters, we found transmission pairs containing $\leq$ 12 SNPs in 17 of the 19 clusters; all clusters with the exception of clusters 2 and 9. Clusters with two or more, non-identical, isolates can be represented as a minimum spanning tree (MST) or a neighbor joining tree (NJT) (Figures 10-17). The number of SNPs between isolates in each cluster can be found in Table 2. The MST for cluster 1 shows isolates 19891 and 27889

containing 4 SNPs. Cluster 3 is defined by isolate 16607 differing by 6 SNPs from 21779 and by 2 SNPs from 16608. Cluster 4 shows 7 SNPs between isolates 19034 and 26720. Isolates 14956 and 19895 from cluster 6 differ by 2 SNPs. The greatest number of isolate pairs in lineage 4 were found in cluster 13. 19077 and 20606 are identical containing no SNPs. These two isolates each differ from 16732 by 9 SNPs and, finally, 16732 differs from 15634 by 8 SNPs. In the last several clusters of Table 2, we see that in cluster 15, isolates 17778 and 17782 have a 3 SNPs; in cluster 17, 18673 and 20148 are on the threshold of 12 SNPs and in cluster 18, 17085 and 14774 are almost identical with 1 SNP. Clusters 5, 7, 8, 10, 11, 12, 14, 16 and 19 contain identical pairs with 0 SNPs.

Table 2: Table of MTBC lineage 4 clusters showing possible transmission pairs with the number of SNPs between isolates.

Cluster Number	Isolate Pair IDs	Number of SNPs Between Pairs
1	19891 27889	4
3	16607 21779	6
	16607 16608	2
4	19034 26720	7
5	22466 22468	0
6	14956 19895	2
7	15545 15547	0
8	23229 26963	0
10	13577 13578	0
	13577 13579	0
	13578 13579	0
11	20574 20603	0
12	19595 19801	0
	19595 19832	0
	19801 19832	0

Table 2 (continued): Table of MTBC lineage 4 clusters showing possible transmission pairs with the number of SNPs between isolates.

Cluster Number	Isolate Pair IDs	Number of SNPs Between Pairs
13	19077 20606	0
	19077 16732	9
	20606 16732	9
	16732 15634	8
14	14158 14159	0
15	17778 17782	3
16	17549 17551	0
17	18673 20148	12
18	17085 14774	1
19	20253 20634	0

Isolates Visualized as a Transmission Network

Using the data generated by this study, we see that the isolates from each lineage form distinct transmission networks connected to each other based on the number of SNPs between isolate pairs. The isolates in lineage 3 (Figure 9A) and lineage 4 (Figure 9B) both form identifiable networks and can be visualized based on the number of SNPs separating the isolates. Each node is connected to another if they can be associated with each other within the network. Possible transmission pairs containing $\leq$ 12 SNPs are highlighted in red to indicate where they fit in the transmission network. Sixteen of the 23 lineage 3 isolates and 67 of the 87 lineage 4 isolates included in the analysis form their respective transmission networks.

SNPs Present in *M. tuberculosis* Virulence Genes

Virulence-associated genes of *M. tuberculosis* containing SNPs among the isolates in this study have been listed in Table 3. One hundred percent of isolates contained at least one SNP in genes *htrA2*, *ctpV*, *pks12,* and *pstA1* when compared to the *M. tuberculosis* H37Rv reference genome. Furthermore, greater than half of all isolates also contained SNPs in the genes *mce1*, *plcA*, *plcB*, *pks7*, *dosT* and *pks5.* These top 10 genes were further evaluated to determine the specific SNP(s) present in each gene and to assess any potential problems that could be caused by the mutation(s) (Table 4). Of the SNPs found, the large number, and type of mutations present in the genes *plcA* and *plcB* showed the need for further evaluation. Both of these genes are translated in the reverse orientation in the *M. tuberculosis* genome, and therefore, the SNP positions occur early in the protein-coding sequence. In the sequence of *plcB*, position 2630173 shows a nonsense mutation from a serine to a stop codon. This is potentially problematic as a stop codon at the early stages of the coding sequence may indicate the presence of a truncated/nonfunctional protein. Of the 8 SNPs present in *plcA*, half were found to be synonymous; however, some of the non-synonymous mutations can potentially cause protein-folding and side-chain interaction problems.

Table 3: Table of 30 *M. tuberculosis* virulence genes containing SNPs among study isolates relative to reference strain H37Rv

Gene name	Rv Number	Description	Number of Isolates Containing SNP	Percentage of Isolates Containing SNP
htrA2	Rv0983	Serine protease and chaperone	113	100
ctpV	Rv0969	Copper efflux p-type ATPase	113	100
pks12	Rv2048c	Polyketide synthase	113	100
pstA1	Rv0930	Inorganic phosphate ABC transporter	113	100
mce1	Rv0166	Mammalian cell entry protein	105	93
plcA	Rv2351c	Phospholipase C	92	81
plcB	Rv2350c	Phospholipase C	87	77
pks7	Rv1661	Polyketide synthase	66	58
dosT	Rv2027c	Transcriptional regulator	63	56
pks5	Rv1527c	Polyketide synthase	62	55
fadD26	Rv2930	Fatty acid CoA synthase	29	26
RD1	Rv3868	Esx1 component	26	23
dosR	Rv3133c	Transcriptional regulator	17	15
pknD	Rv0931c	Protein kinase D	11	10
pknE	Rv1743	Serine/Threonine kinase E	11	10
sigC	Rv2069	Sigma factor C	8	7
erp	Rv3810	Exported repetitive protein	4	4
esxB	Rv3874	esx1 component	4	4
mce2	Rv0586	Mammalian cell entry protein	4	4
esxD	Rv3874	Esx1 component	4	4
sodC	Rv0432	Superoxide dismutase C	3	3
acg	Rv2032	unknown	3	3
ahpC	Rv2428	Alkyl hydroperoxide reductase C	2	2
mce4	Rv3501c	Mammalian cell entry protein	2	2
pcaA	Rv0470c	Mycolic acid synthase	1	1
hspX	Rv2031c	Alpha Crystallin protein	1	1
mce3	Rv1964	Mammalian cell entry protein	1	1
hbhA	Rv0475	Heparin binding hemagglutinin protein	0	0
esxA	Rv3875	Esx1 component	0	0
katG	Rv1908c	Catalase peroxidase enzyme	0	0

Table 4: Top 10 *M. tuberculosis* virulence genes containing SNPs from study isolates showing SNP codon-specific changes.

Gene	Rv#	SNP	Position	Codon Change	AA Change
htrA2	Rv0983	T → C	1100234	CCT →CCC	Synonymous
ctpV	Rv0969	C → A	1079927	ACC →ACA	Synonymous
pks12	Rv2048c	G → C	2296042	GGT →CGT	Gly →Arg
		G → T	2297287	TGC →TTC	Cys →Phe
		A → G	2300237	CCA →CCG	Synonymous
		A → T	2300546	CGA →CGT	Synonymous
		T → G	2300552	TGT →TGG	Synonymous
psta1	Rv0930	C → T	1037911	GCG →GTG	Ala →Val
		T → C	1037012	AAT →AAC	Synonymous
mce1	Rv0166	C → T	196642	ACC →ATC	Thr →Ile
plcA	Rv2351c	C → G	2631556	CCG →CGG	Pro →Arg
		T → C	2631565	ATG →ACG	Met →Thr
		T → C	2631574	GTG →GCG	Val →Ala
		G → A	2631583	AGC →AAC	Ser →Asn
		G → A	2631599	GGG →GGA	Synonymous
		A → G	2631620	TAA →TAG	Synonymous
		A → G	2631971	CAA →CAG	Synonymous
		G → C	2631977	CCG →CCC	Synonymous
plcB	Rv2350c	C →G	2630158	ACC →AGC	Thr →Ser
		A → G	2630161	GAT →GGT	Asp →Gly
		C → G	2630173	TCA →TGA	Ser →Stop
		C → G	2630176	ACA →AGA	Thr →Arg
		G →A	2630182	CGA →CAA	Arg →Gln
		T →A	2630184	TGT →AGT	Cys →Ser
		C →T	2630188	GCT →GTT	Ala →Val
		T →G	2630206	GTC →GGC	Val →Gly
		G →A	2630211	GGC →AGC	Gly →Ser
		A →G	2630215	AAG →AGG	Lys →Arg
pks7	Rv1661	T →G	1875544	GTT →GGT	Val →Gly
dosT	Rv2027c	C →T	2273627	CCC →CCT	Synonymous
pks5	Rv1527c	G →A	1724120	AGG →AGA	Synonymous

Discussion:

In this study we showed that possible transmission relationships do exist between numerous *M. tuberculosis* isolates collected from patients presenting with pulmonary TB symptoms in the Rubaga Division of Kampala, Uganda based on genome sequence comparisons. A total of 143 clinical isolates were processed using the BioNumerics pipeline; after implementation of the inclusion criteria, 113 isolates were included in the SNP analysis. Two isolates grouped into MTBC lineage 1, Indo-Oceanic, a single isolate grouped into lineage 2, East Asian (Beijing), 23 isolates grouped into lineage 3, East African Indian, and 87 grouped into lineage 4, Euro-American. According to the SNP analysis, using a threshold of $\leq$ 12 SNPs as indicative of a transmission pair, we found transmission pairs in all lineages containing at least 2 isolates, with lineage 4 having the highest frequency. Lineage 4 contained the most transmission pairs including the most isolates. This should be expected as lineage 4 is the dominant lineage present in Uganda.

Among the patients with pulmonary TB in this region of Uganda, the predominant grouping of *M. tuberculosis* was lineage 4, Euro-American, followed by lineage 3, East African Indian. Furthermore, isolates were shown to be related as possible transmission pairs by containing $\leq$ 12 SNPs between their genomes when sequenced. One possible transmission pair was identified from lineage 1, 13 possible transmission pairs were identified from a total of 23 isolates grouping in lineage 3, and 25 possible transmission pairs were identified from a total of 87 isolates grouping in lineage 4. Furthermore, the most-robust transmitter from lineage 3 was the identical pair of isolates 16294 and 20095 from cluster 1, which were potential sources of transmission for isolates 20839, 22199, 19621 and 20918. The best transmitter from lineage 4 was shown to be isolate 16732 from cluster 13, which was shown to potentially be the

transmission source for isolates 15634, 19077, 20606 and 20824. The only caveat being that isolate 20824 contained 13 SNPs compared to isolate 16732, and our threshold is ≤ 12 SNPs. It has yet to be determined how significant 1 SNP is in this type of analysis; subsequent overlapping epidemiological data may provide assistance in determining this answer. Analyzing the number of SNPs between isolate pairs also demonstrates that the pairs from lineage 3 have fewer numbers of SNPs as a whole with the majority of pairs containing 0 or 1 SNPs, with 2 SNPs being the highest. This is in contrast to lineage 4 where isolate pairs contain SNPs up to, and including, the threshold of ≤ 12 SNPs.

When a pairwise SNP matrix is generated to visualize the isolates as a network, clear relationships can be seen by the connection of the isolates to each other with differing amounts of SNPs colored for easy visualization. These data not only show possible transmission of *M. tuberculosis* isolates between individuals, but the transmission networks identified, once combined with epidemiological data, will allow public health interventions to be implemented in the social gathering establishments that are frequented by the human TB transmitters.

Multiple mutations in the genes *plcA* and *plcB*, especially a nonsense mutation in the latter, bring about the question of if there is a survival advantage to the bacteria by having certain mutations present given that the mutations occur early in the protein coding sequence. The *plcABCD* family of genes encode a phospholipase C, playing a role in pathogenesis by cleaving phospholipids during intracellular infection [87]. Alteration and/or inactivation of those genes in our study isolates can potentially be advantageous for transmission or the establishment of infection in the host in this cohort of individuals. Further investigation needs to be performed on these mutations to obtain any correlations with these observations.

The success of these public health interventions can possibly aid in meeting the future goals and guidelines established by the WHO and introduce a more proactive method of early intervention prior to disease onset. This would possibly prevent TB far more effectively compared to the current passive programs like TB-DOTS.

There are several future studies that can be performed based on the data generated in this study. For example, the project protocol required patients to give a minimum of two sputum samples; however, the relationship between the isolates found in samples from the same person were not analyzed. Therefore, future studies should consider analyzing SNP differences between isolates collected from the same patient to determine if there are any genetic differences between samples originating from the same donor. This would possibly determine if a person carries more than one strain of *M. tuberculosis* during infection in that region or if transmission occurred from multiple individuals. One limitation of this study is that it was conducted in one division of Kampala, Rubaga: therefore, this type of analysis should be expanded beyond the Rubaga Division to determine more transmission networks where interventions can be incorporated and to make the data more generalizable to the entire country.

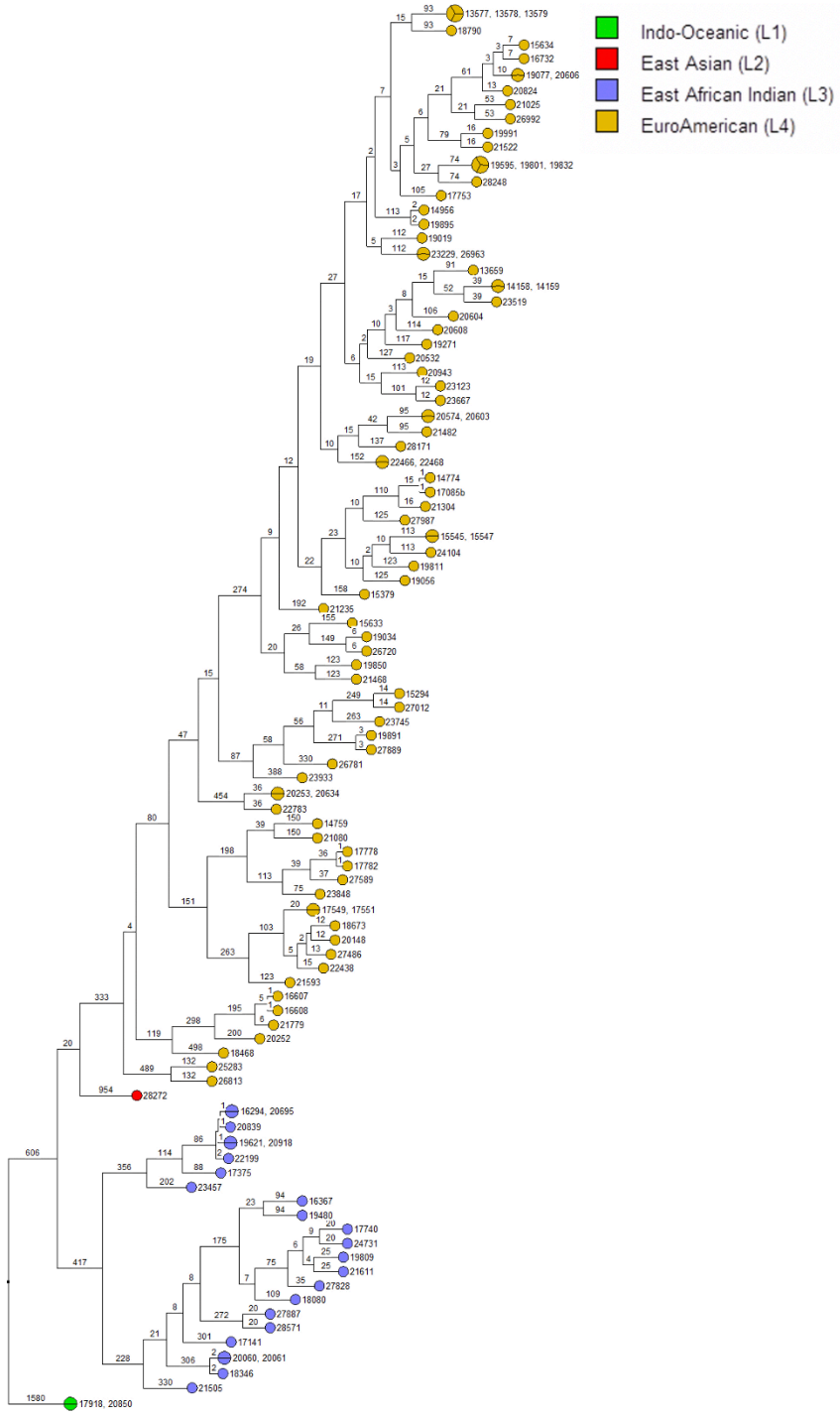

Figure 4. UPGMA tree of all isolates separated into color-coded MTBC lineages using SNP analysis after WGS. Branch numbers indicate SNP distance between isolates.

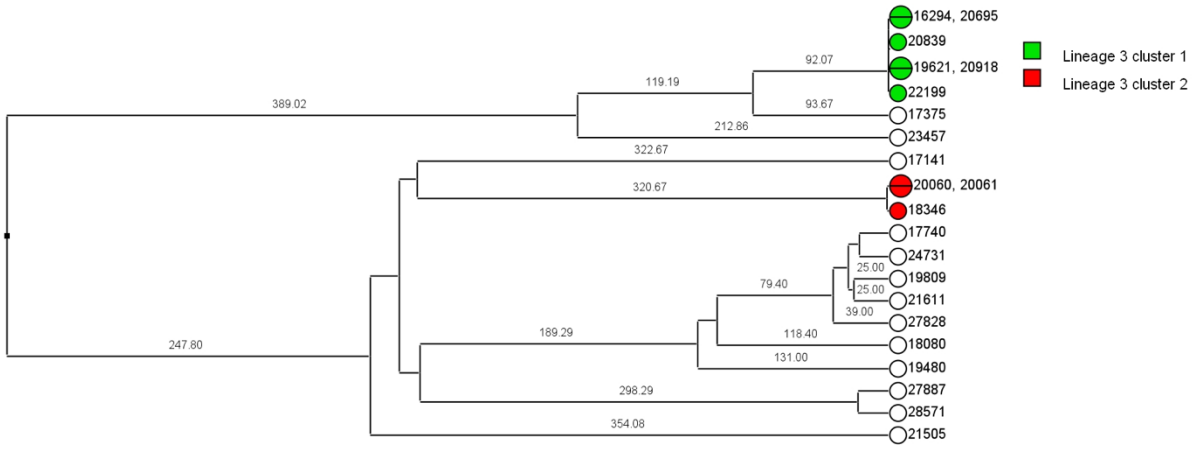

Figure 5: UPGMA tree of isolates grouped into MTBC lineage 3 and separated into possible transmission clusters. Branch numbers indicate SNP distance between isolates.

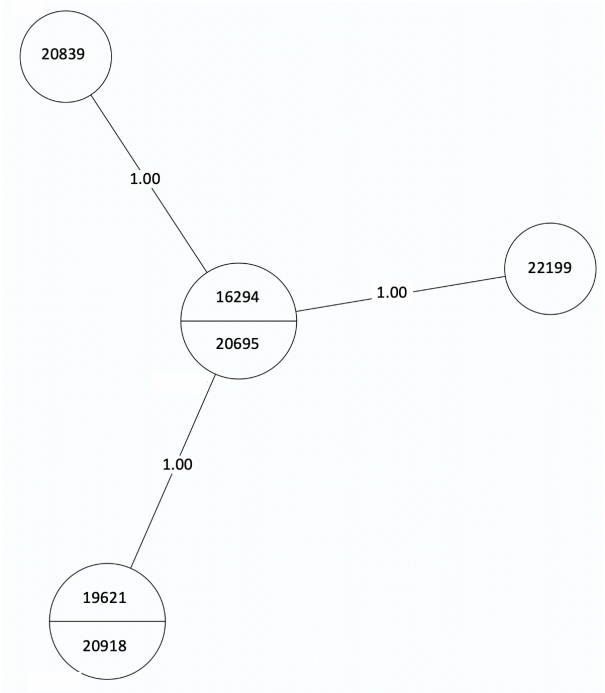

Figure 6: MST of MTBC lineage 3, cluster 1. Numbers between branches indicate SNP distance. Two samples in one circle indicate identical isolates with 0 SNPs.

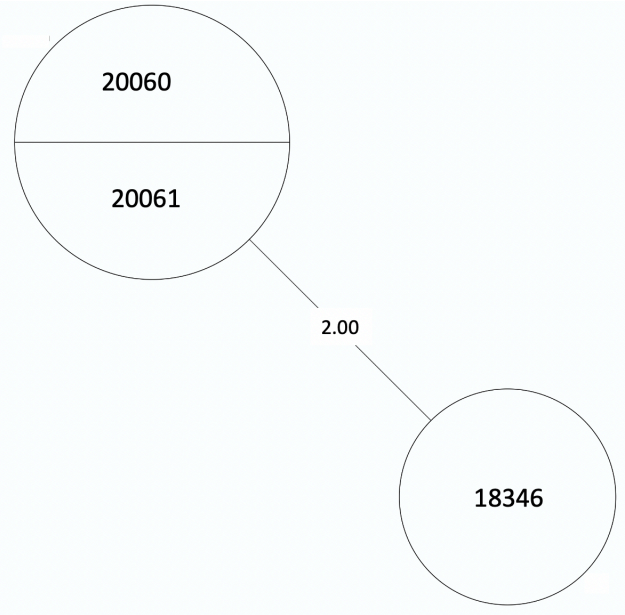

Figure 7: MST of MTBC lineage 3, cluster 2. Numbers between branches indicate SNP distance. Two samples in one circle indicate identical isolates with 0 SNPs.

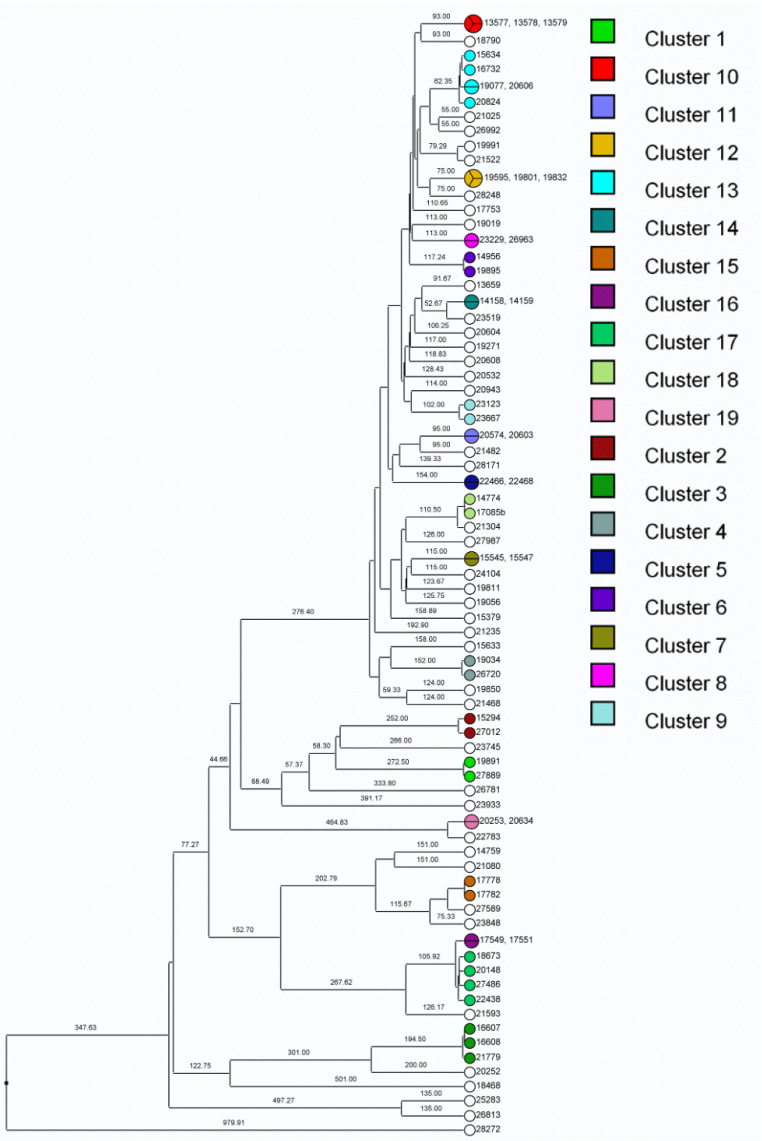

Figure 8: UPGMA tree of isolates grouped into MTBC lineage 4 and separated into possible transmission clusters. Branch numbers indicate SNP distance between isolates.

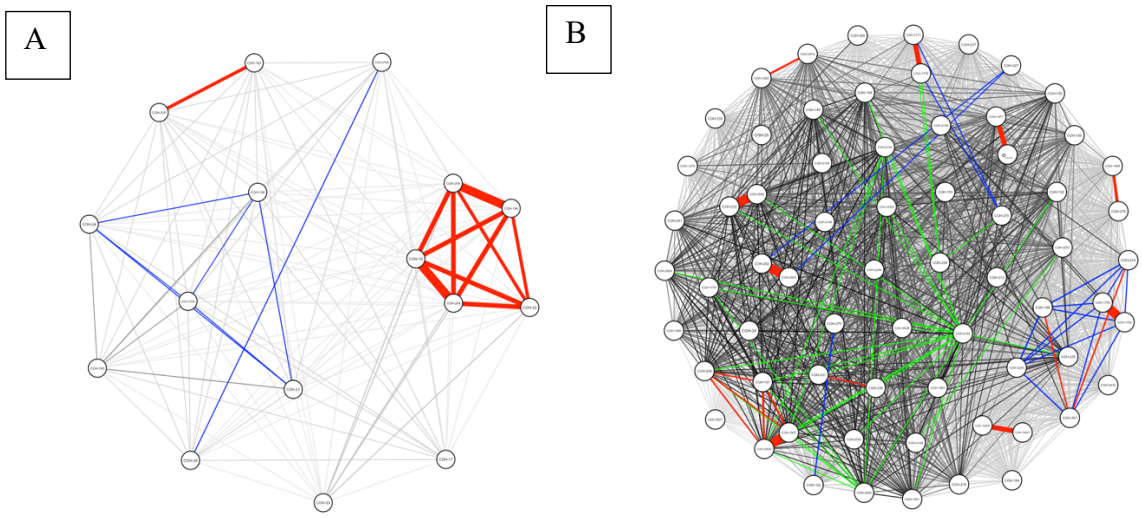

Figure 9: Pairwise SNP matrix visualized as a network colored by number of SNPs: 0-12 SNPs (red), 13-50 SNPs (blue), 51-100 SNPs (green) and >100 SNPs (black). Each node represents an individual isolate **A.** Isolates grouping in MTBC lineage 3 forming a transmission network **B.** Isolates grouping in MTBC lineage 4 forming a transmission network.

Supplemental Figures:

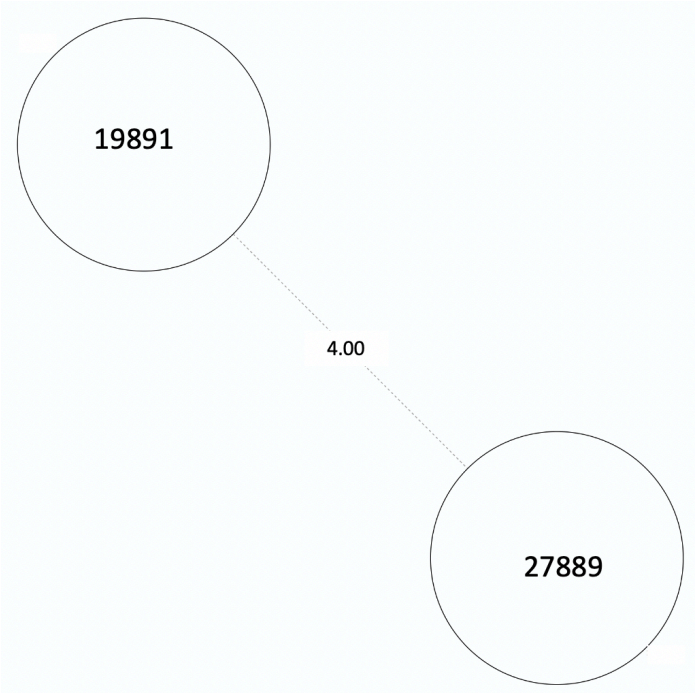

Figure 10: MST of MTBC lineage 4, cluster 1. Numbers between branches indicate SNP distance.

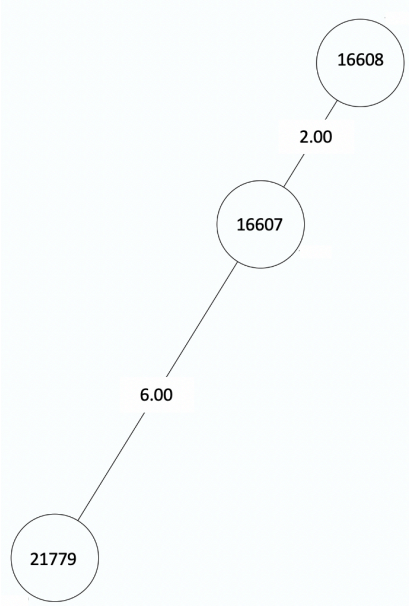

Figure 11: MST of MTBC lineage 4, cluster 3. Numbers between branches indicate SNP distance.

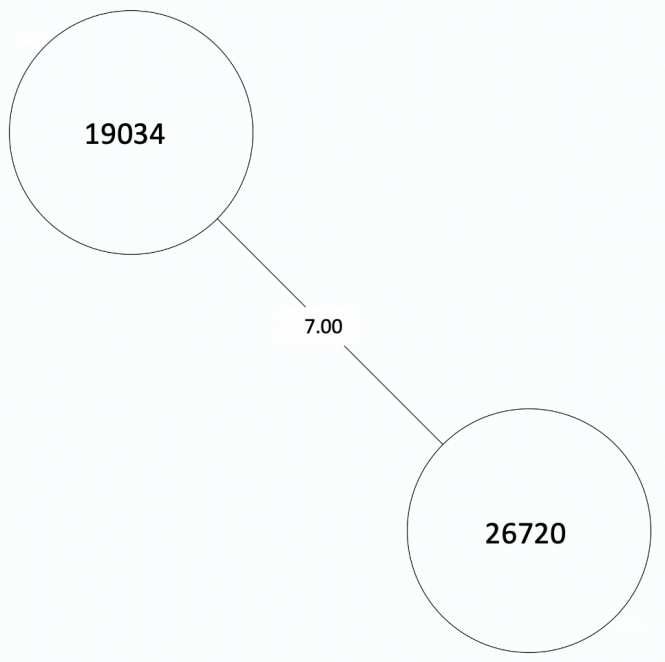

Figure 12: MST of MTBC lineage 4, cluster 4. Numbers between branches indicate SNP distance.

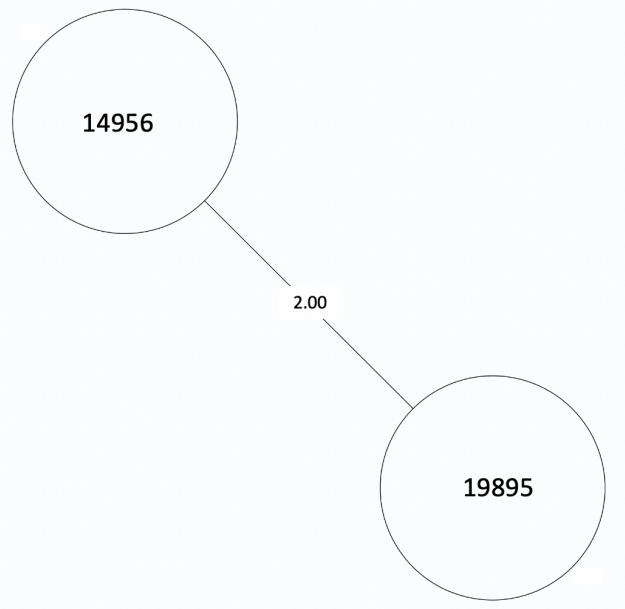

Figure 13: MST of MTBC lineage 4, cluster 6. Numbers between branches indicate SNP distance.

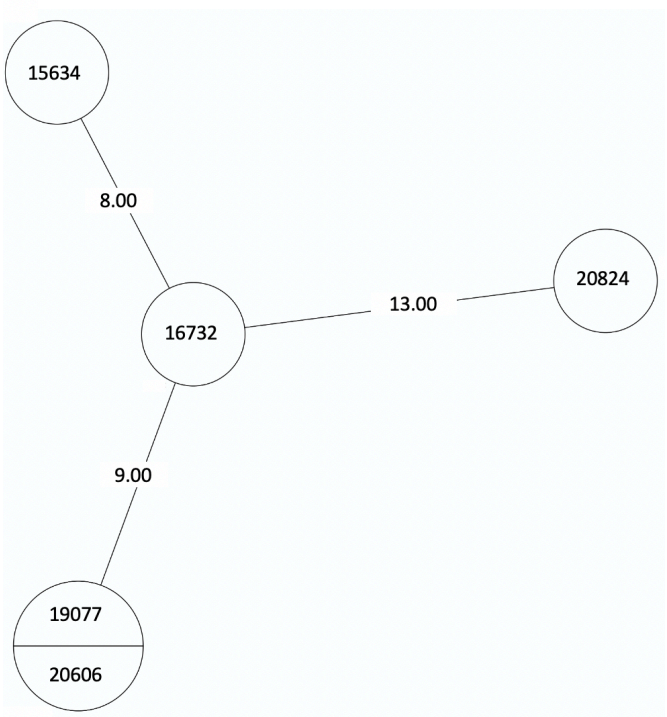

Figure 14: MST of MTBC lineage 4, cluster 13. Numbers between branches indicate SNP distance. Two samples in one circle indicate identical isolates with 0 SNPs.

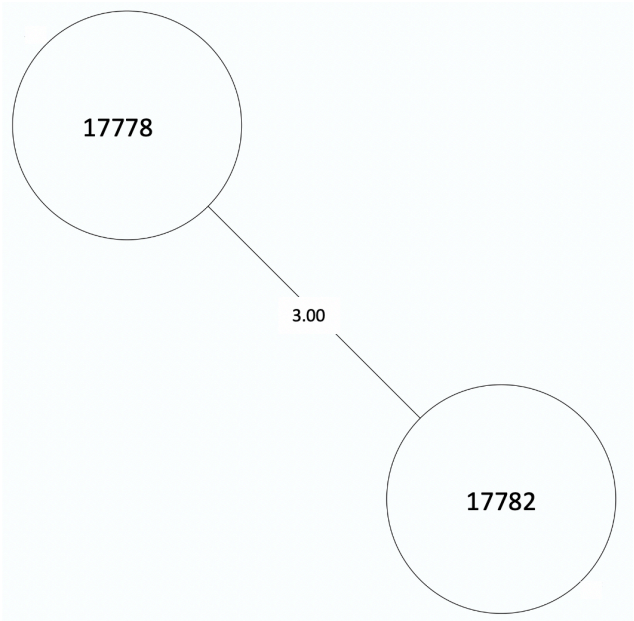

Figure 15: MST of MTBC lineage 4, cluster 15. Numbers between branches indicate SNP distance.

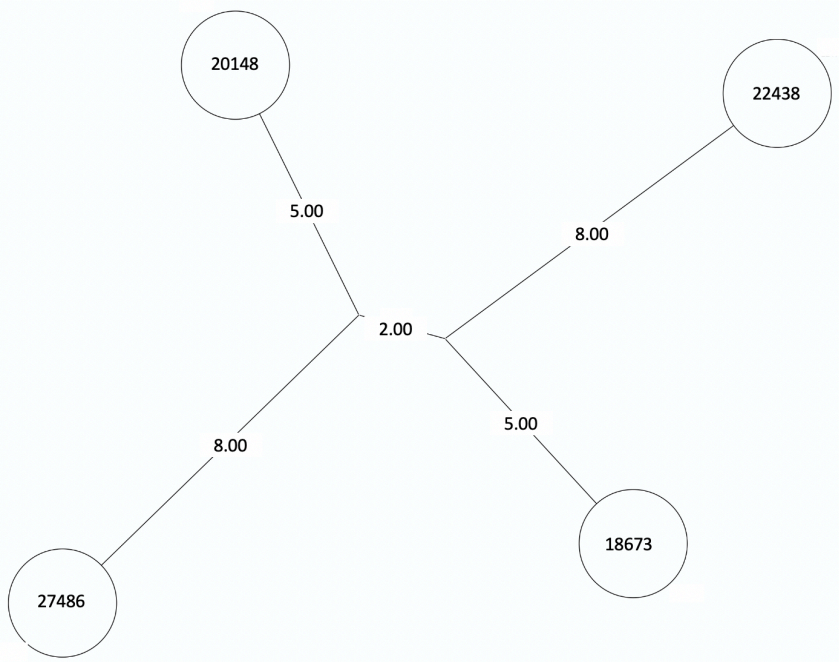

Figure 16: Neighbor Joining Tree (NJT) of MTBC lineage 4, cluster 17. Numbers between branches indicate SNP distance.

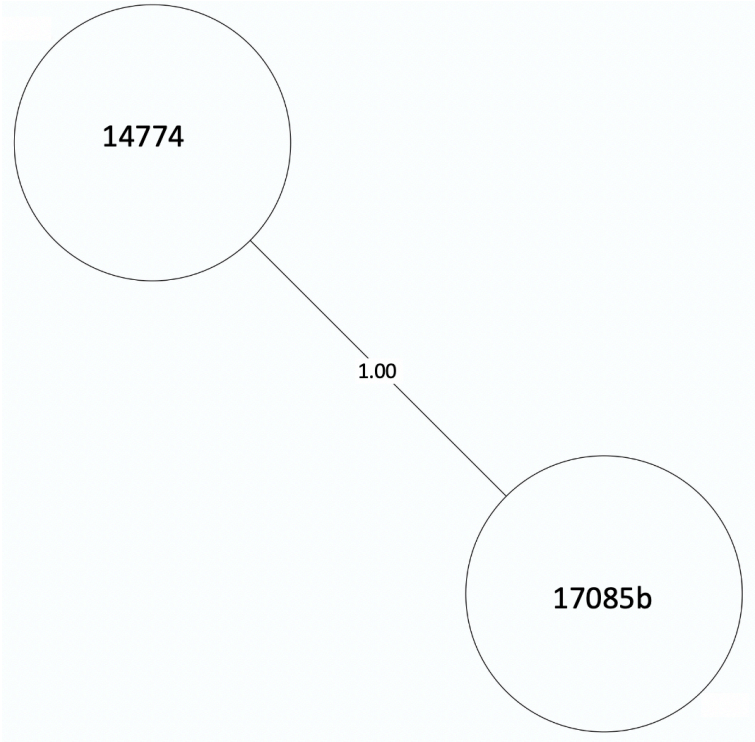

Figure 17: MST of MTBC lineage 4, cluster 18. Numbers between branches indicate SNP distance.

References

1. Gupta A, Kaul A, Tsolaki AG, Kishore U, Bhakta S. Mycobacterium tuberculosis: Immune evasion, latency and reactivation. Immunobiology. 2011;217:363–74.

2. Yates TA, Khan PY, Knight GM, Taylor JG, Mchugh TD, Lipman M, et al. The transmission of Mycobacterium tuberculosis in high burden settings. Lancet Infect Dis [Internet]. 2016 [cited 2019 Dec 9];16(2):227–38. Available from: www.thelancet.com/infectionVol

3. WHO. World Tuberculosis Report 2019. 2019.

4. WHO. The End TB Strategy, World Health Organisation, Geneva. 2015 [cited 2019 Aug 28]; Available from: https://www.who.int/tb/strategy/End_TB_Strategy.pdf?ua=1

5. Maher D, Mikulencak M. What is DOTS? A Guide to Understanding the WHO-recommended TB Control Strategy Known as DOTS. 1999.

6. Meertens RM, Mj Van De Gaar V, Spronken M, De Vries NK. Prevention praised, cure preferred: results of between-subjects experimental studies comparing (monetary) appreciation for preventive and curative interventions. BMC Med Inform Decis Mak [Internet]. 2013 [cited 2019 Dec 16];13(136). Available from: http://www.biomedcentral.com/1472-6947/13/136

7. Whalen CC. Failure of Directly Observed Treatment for Tuberculosis in Africa: A Call for New Approaches. Clin Infect Dis. 2006;42(7):1048–50.

8. Luca S, Mihaescu T. History of BCG Vaccine. Maedica (Buchar) [Internet]. 2013 [cited 2019 Aug 14];8(1):53–8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3749764/pdf/maed-08-53.pdf

9. WHO. Uganda Tuberculosis Profile 2018. 2019.

10. WHO. An Investigation of Household Contacts of Open Cases of Pulmonary Tuberculosis amongst the Kikuyu in Kiambu, Kenya. Bull World Heal Organ. 1961;25(6):831–

50.

11. Guwatudde D, Nakakeeto M, Jones-Lopez EC, Maganda A, Chiunda A, Mugerwa RD, et al. Tuberculosis in Household Contacts of Infectious Cases in Kampala, Uganda. Am J Epidemiol. 2003;158(9):887–98.

12. Buu TN, Van Soolingen D, Huyen MNT, Lan NNT, Quy HT, Tiemersma EW, et al. Tuberculosis Acquired Outside of Households, Rural Vietnam. Emerg Infect Dis [Internet]. 2010 [cited 2019 Dec 9];16(9):1466–8. Available from: http://phil.cdc.gov/phil.

13. Classen CN, Warren R, Richardson M, Hauman JH, Gie RP, Ellis JHP, et al. Impact of social interactions in the community on the transmission of tuberculosis in a high incidence area. Thorax. 1999;54:136–40.

14. Yaganehdoost A, Graviss EA, Ross MW, Adams GJ, Ramaswamy S, Wanger A, et al. Complex Transmission Dynamics of Clonally Related Virulent Mycobacterium tuberculosis Associated with Barhopping by Predominantly Human Immunodeficiency Virus-Positive Gay Men. J Infect Dis. 1999;180(4):1245–51.

15. Verver S, Warren R, Munch Z, Richardson M, Van der spuy G, Borgdorff M, et al. Proportion of tuberculosis transmission that takes place in households in a high-incidence area. Lancet [Internet]. 2004 [cited 2019 Dec 9];363(9404):212–4. Available from: www.thelancet.com

16. Cavalcante SC, Durovni B, Barnes GL, A Souza FB, Silva RF, Barroso PF, et al. Community-randomized trial of enhanced DOTS for tuberculosis control in Rio de Janeiro, Brazil. Int J Tuberc Lung Dis. 2010;14(2):203–9.

17. Sharma K, Verma R, Advani J, Chatterjee O, Solanki HS, Sharma A, et al. Whole Genome Sequencing of Mycobacterium tuberculosis Isolates From Extrapulmonary Sites. Omi A

J Integr Biol [Internet]. 2017 [cited 2019 Dec 9];21(7):412–25. Available from: http://phaster.ca

18. Ahmadian A, Gharizadeh B, Gustafsson A, Sterky F, Nyren P, Uhlen M, et al. Single-Nucleotide Polymorphism Analysis by Pyrosequencing. Anal Biochechemistry [Internet]. 2000 [cited 2019 Aug 19];280:103–10. Available from: https://pdf.sciencedirectassets.com/272363/1-s2.0-S0003269700X00425/1-s2.0-S0003269700944932/main.pdf?X-Amz-Security-Token=AgoJb3JpZ2luX2VjEPb%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJIMEYCIQCuAFouCWDTpt6nLViuBhOpBZ8XYVQau8OruxotpZKMYQIhAKhcY60aVfWn

19. Ramaswamy S V, Reich R, Dou S-J, Jasperse L, Pan X, Wanger A, et al. Single nucleotide polymorphisms in genes associated with isoniazid resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother [Internet]. 2003;47(4):1241–50. Available from: papers2://publication/uuid/F59DC034-E0BA-4A8D-8362-FA595F71A069

20. Sekandi JN, Zalwango S, Martinez L, Handel A, Kakaire R, Nkwata AK, et al. Four Degrees of Separation: Social Contacts and Health Providers Influence the Steps to Final Diagnosis of Active Tuberculosis Patients in Urban Uganda. BMC Infect Dis. 2015;

21. Pouseele H, Supply P. Accurate Whole Genome Sequencing Based Epidemiological Surveillance of Mycobacterium tuberculosis. Methods Microbiol. 2015;42:359–94.

22. Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis [Internet]. 2013 Feb [cited 2017 Oct 27];13(2):137–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23158499

23. UBOS. National Population and Housing Census 2014. National Population and Housing Census. 2017.

24. Sekandi JN, List J, Luzze H, Yin XP, Dobbin K, Corso PS, et al. Yield of undetected tuberculosis and human immunodeficiency virus coinfection from active case finding in urban Uganda. Int J Tuberc Lung Dis. 2014;18(6):754.

25. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, et al. A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun [Internet]. 2014;5:4812. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4166679&tool=pmcentrez&rendertype=abstract

26. Ford CB, Shah RR, Kato Maeda M, Gagneux S, Murray MB, Cohen T, et al. Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug resistant tuberculosis. Nat Genet. 2013;45(7):784–90.

27. Hernández-Pando R, López B, Aguilar D, Orozco H, Burger M, Espitia C, et al. A marked difference in pathogenesis and immune response induced by different Mycobacterium tuberculosis genotypes. Clin Exp Immunol. 2003;133:30–7.

CHAPTER 4

CHARACTERIZATION OF *MYCOBACTERIUM TUBERCULOSIS* ALPHA-CRYSTALLIN PROTEIN PHOSPHORYLATION AND TRANSPORT

Background:

The *acr/hspX* operon region of *M. tuberculosis* (Figure 18) is induced under hypoxic conditions or the presence of high levels of nitric oxide. These conditions not only induce the *acr* operon itself, but also the DosR regulon. After induction, the Acr protein must be exported to assist in its predicted dormancy functions and to modify the antigen presentation capabilities of innate immune components such as Dendritic cells. The process by which Acr is transported outside of the cell is still not well understood. Using plasmids encoding different regions of the *acr* operon as well as determining the phosphorylation state of Acr, we will try to further the understanding of the transport process. Acr has been reported to be a phosphoprotein that contains internal protein kinase activity that allows the autophosphorylation of the protein exclusively at a serine residue contained in the *acr* amino acid sequence [70]. However, since this publication, no studies have determined which of the nine serine residues present in Acr is phosphorylated. As phosphorylation is a key factor in the activity of biological molecules, it is important that knowledge be advanced in the interaction between phosphorylation and protein transport.

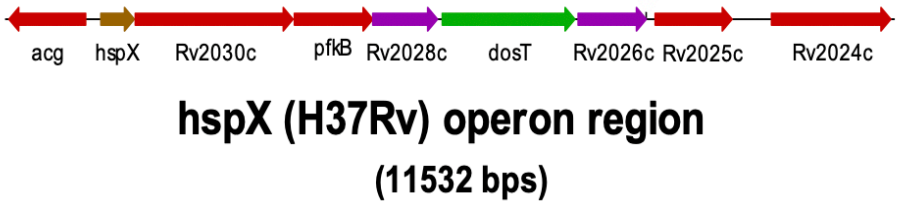

Figure 18: Gene map of the *hspX* operon region including the upstream gene *acg* and downstream genes adjacent to the operon. Figure was generated using Cone Manager Software (Sci Ed, Colorado).

Methods:

Generating pMV306 Derivatives Encoding *M. tuberculosis hspX* Operon Genes

Plasmid pRK249 (Figure 19) was generated by incorporating the *hspX* gene into a pMV306 backbone encoding the mycobacteriophage L5 att/int region for chromosomal integration at the L5:attB site [88]. A 649-bp *hspX* region was amplified by PCR using *M. tuberculosis* H37Rv genomic DNA and primers HindIII-up-hspx-193 and HindIII-down-acr (see Table 6 detailing primer sequences used in this study). Plasmid pMV306 and the PCR product were prepared for ligation by digestion with *Hin*dIII and subsequent dephosphorylation of the cut ends. The PCR product and the digested pMV306 plasmid were ligated and used to transform *E. coli* TAM1 cells. Plasmid DNA from transformants were digested with *Bgl*II, to determine insert orientation. Plasmids yielding bands of 2747 bp and 1881 bp were expected to have the *hspX* gene in the same orientation of other genes on the plasmid. A candidate with this banding pattern was confirmed by DNA sequencing using primer P1473 to have *hspX* inserted without mutations; this plasmid was named pRK129.

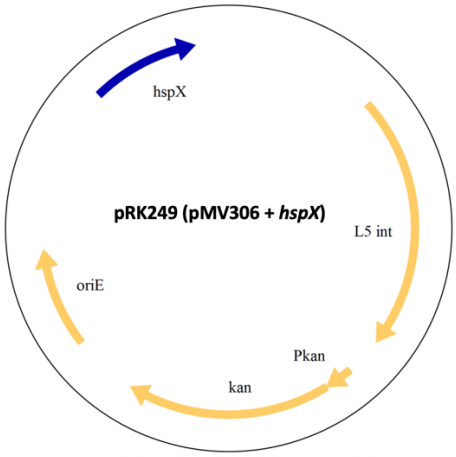

Figure 19. Plasmid map for pRK249 showing each individual plasmid component, including the kanamycin resistance gene, origin of replication of *E. coli* and the L5 chromosomal integration region. The blue arrow indicates the inserted *hspX* gene.

Plasmid pRK251 (Figure 20) was generated by incorporating the *hspX* gene together with the divergently transcribed, adjacent gene *acg* into a pMV306 plasmid backbone. A 1648-bp *acg-hspX* region was amplified by PCR using *M. tuberculosis* H37Rv genomic DNA and primers HindIII-down-acg+1 and HindIII-down-acr. Plasmid pMV306 and the PCR product were prepared for ligation by digestion with *Hin*dIII and subsequent dephosphorylation of the cut ends. The *Hin*dIII-digested PCR product and the pMV306 vector were ligated to create plasmid pRK251. Correct orientation of the ligated PCR region was verified by digestion with *Eco*RI, yielding cut bands of 1479-bp and 4149-bp. pRK251 was confirmed for the presence of the inserted genes without mutations by DNA sequencing using primers P1473, P2022 and P1247.

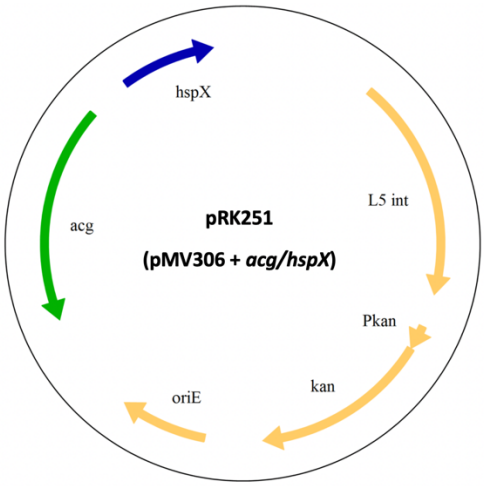

Figure 20. Plasmid map for pRK251 showing each individual plasmid component, including kanamycin resistance gene, origin of replication of *E. coli* and the L5 chromosomal integration region. The blue and green arrows indicate the inserted *hspX* and *acg* genes, respectively.

Plasmid pRK253 (Figure 21) was generated by incorporating the *hspX* gene together with upstream gene *acg* and downstream genes *Rv2030c*, *pfkB* and *Rv2028c* into a pMV306 plasmid backbone. A5592-bp *acg-hspX-Rv2030c-pfkB-Rv2028c* region was amplified by PCR using *M. tuberculosis* H37Rv genomic DNA and primers HindIII-down-acg+1 and HindIII-downRv2028c#2. Plasmid pMV306 and the PCR product were prepared for ligation with the by digestion with *Hin*dIII and subsequent dephosphorylation of the cut ends. The PCR product and the digested pMV306 plasmid were ligated to create plasmid pRK253. Correct orientation of the ligated PCR region was verified by digestion of pRK253 with *Age*I, yielding cut bands of 3670-bp and 5902-bp. pRK253 was confirmed for the presence of the inserted genes without mutations by DNA sequencing using primers P1473, P2022, P1247, P1243, 166-Down-Acr-F, 950-Down-Acr-F, 322uppfkB-F, P1797, P1557 and P1561.

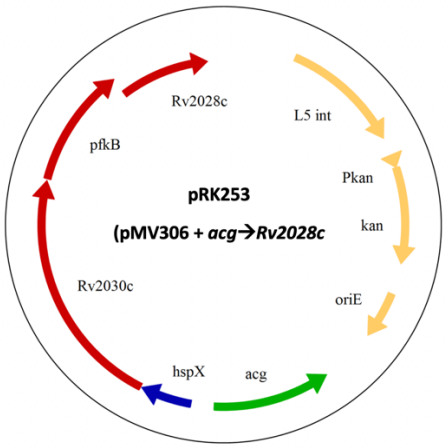

Figure 21. Plasmid map for pRK253 showing each individual plasmid component, including kanamycin resistance gene, origin of replication of *E. coli* and the L5 chromosomal integration region. The blue, green and red arrows indicate the inserted *hspX*, *acg* and *Rv2030c* to *Rv2028c* genes, respectively.

Plasmid pRK255 (Figure 22) was generated by incorporating the *hspX* gene together with upstream gene *acg* and downstream genes *Rv2030c*, *pfkB*, *Rv2028c*, *dosT* and *Rv2026c* into a pMV306 plasmid backbone. The 8285-bp *acg-hspX-Rv2030c-pfkB-Rv2028c-dosT-Rv2026c* region was amplified by PCR using *M. tuberculosis* H37Rv genomic DNA and primers HindIII-down-acg+1 and HindIII-downRv2026c+2. Plasmid pMV306 and the PCR product were prepared for ligation by digestion with *Hin*dIII and subsequent dephosphorylation of the cut ends. The PCR product and the digested pMV306 plasmid were ligated to create plasmid pRK255. Correct orientation of the ligated PCR region was verified by digestion of pRK255 with *Acl*I and *Nhe*I, yielding cut bands of 7718-bp and 4547-bp. pRK255 was confirmed for the presence of the inserted genes without mutations by DNA sequencing using primers P1473, P2022, P1247, P1243, 166-Down-Acr-F, 950-Down-Acr-F, 322uppfkB-F, P1797, P1557, P1561, P1346, P1567, P1565, P1571 and P1569.

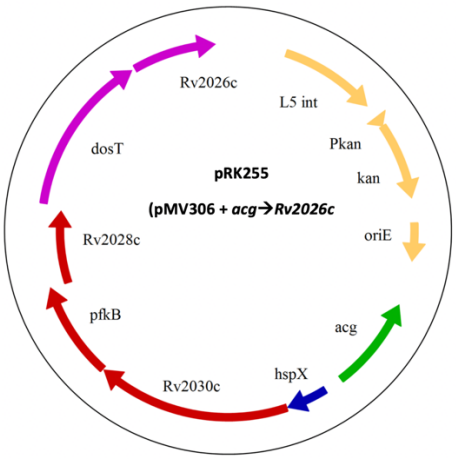

Figure 22. Plasmid map for pRK255 showing each individual plasmid component, including kanamycin resistance gene, origin of replication of *E. coli* and the L5 chromosomal integration region. The blue, green, red and magenta arrows indicate the inserted *hspX*, *acg*, *Rv2030c* to *Rv2028c* and *dosT* to *Rv2026c* genes, respectively.

Electroporation of Plasmids into *M. smegmatis*

Five mL overnight cultures of *E. coli* recombinant clones containing each plasmid were grown in Luria-Bertani medium containing 50 µg/mL kanamycin. Plasmids were extracted using a miniprep kit (Zymo Research, CA). DNA concentration was measured using a NanoDrop spectrophotometer (ThermoFisher Scientific, CA). A 50 mL culture of *M. smegmatis* was grown in Middlebrook 7H9 medium (BD/Difco, NJ) supplemented with 0.5% glycerol, 0.05% Tween 80 and 10% albumen-dextrose-saline (ADS) [89] until reaching OD_{600} = 0.8. Cells were centrifuged (3200 x g) for 10 minutes and washed twice with 10% glycerol solution. Cells were suspended in 5 mL 10% glycerol. For electroporation, 200 µL of cells were mixed with 100 ng of plasmid and transferred to an electroporation cuvette. Electroporation was performed using a BioRad GenePulser (BioRad, CA) with settings at 2500 V, 25 µF, 1000 Ohms with a 2 mm gap cuvette.

Cells were rescued in 1 mL 7H9 medium and incubated 3 hours before plating on Middlebrook 7H10 agar plates (BD/Difco, NJ) supplemented with 0.5% glycerol, 0.05% Tween 80, 10% ADS and 50 μg/mL kanamycin for candidate selection. Colonies were subsequently picked, bacteria lysed by boiling at 100°C for 10 minutes and plasmid content in the transformed colonies was verified via PCR with the following primer pairs: pRK249: P1420 and P1870, pRK251: P1473 and P2377, pRK253: P1473 and P2377, P1419 and P1841, P1553 and P1554, P1557 and P1558, pRK255: P1473 and P2377, P1419 and P1841, P1553 and P1554, P1557 and P1558, P1561 and P1562, P1565 and P1566.

Electroporation of Plasmids into *M. marinum*

A 50 mL culture of *M. marinum* was grown in Middlebrook 7H9 medium (BD/Difco, NJ) supplemented with 0.5% glycerol, 0.05% Tween 80 and 10% albumen-dextrose-saline (ADS) [89] until reaching OD_{600} = 0.8. Cells were centrifuged (3200 x g) for 10 minutes and washed twice with 10% glycerol solution. Cells were suspended in 5 mL 10% glycerol. For electroporation, 200 μL of cells were mixed with 100 ng of plasmid and transferred to an electroporation cuvette. Electroporation was performed using a BioRad GenePulser (BioRad, CA) with settings at 2500 V, 25 μF, 1000 Ohms with a 2 mm gap cuvette.

Cells were rescued in 1 mL 7H9 medium and incubated overnight before plating on Middlebrook 7H10 agar plates (BD/Difco, NJ) supplemented with 0.5% glycerol, 0.05% Tween 80, 10% ADS and 50 μg/mL kanamycin for candidate selection. Colonies were subsequently picked, bacteria lysed by boiling at 100°C for 10 minutes and plasmid content in the transformed colonies was verified via PCR with the following primer pairs: *M. marinum* pRK249: P1420 and P1870, pRK251: P1473 and P2377, pRK253: P1473 and P2377, P1419 and P1841, P1553 and

P1554, P1557 and P1558, pRK255: P1473 and P2377, P1419 and P1841, P1553 and P1554, P1557 and P1558, P1561 and P1562, P1565 and P1566.

hspX Serine to Alanine Site Directed Mutagenesis Plasmids

Plasmids pEY1 – pEY9 (Table 5, Figure 23B) were generated by introducing specific mutations into the serine amino acid sequence and incorporating the generated sequence into the pET23bhspx plasmid (Figure 23A) from *E. coli*, strain BL21(DE3)/pLys5/pET23bhspx, encoding for the Acr protein with a C-terminal 6-His tag. The regions were amplified by PCR using *M. tuberculosis* H37Rv genomic DNA and specified primers. To prepare the substrate and plasmid backbone for ligation, both were enzymatically digested with specified enzymes and the plasmid subsequently were dephosphorylated. The PCR product and the digested plasmid were ligated together to generate each plasmid encoding the desired mutation. Candidates were screened for the *hspX* region by PCR using primers P2305 and P2306 yielding a band of 500-bp. Each plasmid was sequenced to confirm no unwanted mutations were present in the sequence.

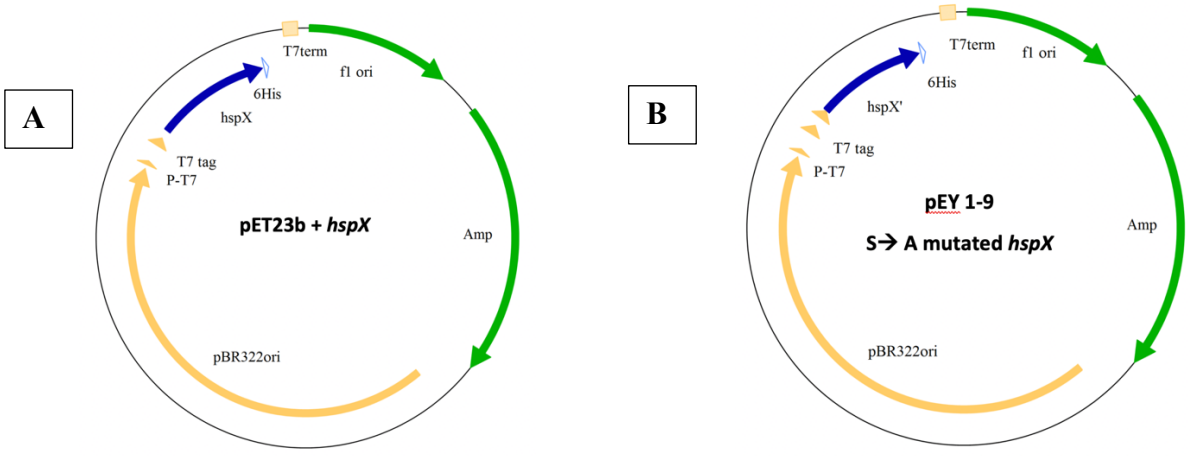

Figure 23. A. Plasmid map for pET23bhspX used as the backbone for the pEY1-pEY9 plasmids. The blue arrow indicates the *hspX* region **B.** Plasmid map for pEY-pEY9 with the blue arrow indicating the *hspX* region that has been mutated. Each plasmid contains a mutation at a unique serine residue.

Table 5: Serine to alanine site directed mutagenesis plasmids showing codon change at specific serine residues in *hspX*

Plasmid	Serine #	Mutation	Primers used for mutation and amplification (F/R)	Enzymes used for ligation preparation
pEY1	1	TCC → GCC	P2391/P2383	*Sac*II/*Xho*I
pEY2	2	TCT → GCT	P2392/P2383	*Sac*II/*Xho*I
pEY3	3	TCA → GCA	P2393/P2383	*Sac*II/*Xho*I
pEY4	4	TCG → GCG	P2384/P2377	*Bsi*WI/*Hin*dIII
pEY5	5	TCC → GCC	P2378/P2383	*Bsi*WI /*Xho*I
pEY6	6	TCG → GCG	P2379/P2383	*Bsi*WI /*Xho*I
pEY7	7	TCG → GCG	P2380/P2383	*Bsm*I/*Xho*I
pEY8	8	TCG→ GCG	P2381/P2383	*Bsm*I/*Xho*I
pEY9	9	TCC → GCC	P2384/P2382	*Xho*I/*Hin*dIII

Purification of Serine to Alanine Mutated Recombinant Acr Protein

Using methodology described earlier, each of the nine generated plasmids were introduced into *E. coli* BL21-Gold(DE3) expression strain via electroporation using a BioRad GenePulser electroporator (BioRad, CA) with settings at 2500 V, 25 μF, 1000 Ohms and a 2 mm gap cuvette (Agilent Technologies, CA). Ten mL overnight cultures of each transformant was grown in Luria-Bertani medium containing 50 μg/mL carbenicillin. Cultures were diluted 1:100 in 2 L flasks and incubated at 37°C to reach OD_{600} = 0.5. Next, cultures were induced with 0.2 mM IPTG and incubated an additional 2 hours. Cultures were centrifuged at 7000 x g for 10 minutes to pellet bacteria and supernatant was removed. Bacterial pellets were re-suspended in Buffer C [90] without urea (1M Tris HCl pH 7.9, 2M KCl, 0.5M EDTA, 50% glycerol, 30mM $ZnSO_4$, 0.035% BME), PMSF was added to prevent protease activity and the cells were lysed using a bacterial cell disruptor (Constant Systems, UK). After lysis, samples were centrifuged at 21000 x g for 15 minutes and supernatants were collected for protein purification.

Purification of recombinant protein was performed by Nickel affinity chromatography. Briefly, 5 mL Nickel columns were prepared per manufacturer instructions (ThermoFisher

Scientific, CA) containing 2 mL Ni-NTA His-Bind resin (Novagen, MA). Each column was equilibrated with Buffer C without urea and supernatant was then added to the column. Resin was washed with Buffer C without urea containing 10 mM imidazole to reduce non-specific binding and subsequently eluted using Buffer C without urea containing 250 mM imidazole. Elution fractions were analyzed for protein using SDS-PAGE and fractions containing protein were combined to maximize yield. Protein was dialyzed in 1x PBS and concentration was determined using Pierce BCA protein assay kit (ThermoFisher Scientific, CA).

Demonstration of rAcr Function

To determine if the rAcr protein can induce some of the same cytokines as the native Acr protein, we measured the production of two cytokines, CCL-2 (MCP-1) and TNF-α, that have been reported to be induced by native Acr in bone marrow-derived and J774 macrophages [91]. MCP-1 is a class of chemoattractant cytokine (chemokine) that recruits monocytes, T –cells and dendritic cells to sites of inflammation and TNF-α is one of the key signaling factors in the inflammation process. RAW 247 and J774 macrophage monolayers were cultured in DMEM medium containing 10% FBS at 37ºC/5% CO_2 until approximately 90% confluent. Approximately $2x10^5$ cells were transferred into wells of a 48-well plate. The cells were allowed to adhere overnight and subsequently rAcr was added to final concentrations of 1, 5 ,10, 15 and 25 μg/ml to triplicate wells. Plates were incubated overnight. After 24 hours, supernatants were harvested, and sandwich ELISA performed to detect CCL2 and TNF-α levels produced.

Detection of Phosphorylation using pIMAGO Western Blot

Ten μg of protein was combined with kinase buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 2 mM DTT, 0.01% Brij35 at pH = 7.5), 10 mM DTT and 1mM ATP. The mixture was incubated for 2 hours at 30°C followed by 2 hours at 37°C at a final volume of 500μL. After incubation, volume was reduced to 20 μL using a Speedvac concentrator and samples subsequently boiled in 1X NuPage-LDS sample buffer (ThermoFisher Scientific, CA) for 10 minutes. Samples were run on a 12% Tris-Glycine-SDS gel along with phosphorylation positive control provided in the pIMAGO biotin phosphoprotein detection kit (Tymora Analytical, Indiana). Protein was transferred to a nitrocellulose membrane (Bio-Rad, California) using Tris-Glycine buffer without SDS. Western blotting was performed according to the protocol in the pIMAGO kit. Bands were visualized using Luminata western HRP substrate (Millipore, MA)..

Table 6. Primers used in this study

Primer	Sequence 5' to 3'
HindIII-up-hspx-193	GAG ACA AGC TTC AAC CTC CGC TGT TCG ATA
HindIII-down-acr	GCA TGA AGC TTC AGT TGG TGG ACC GGA TC
HindIII-down-acg+1	GAT GAA GCT TGC TAC CGG TGA TCC TTA GC
HindIII-downRv2028c#2	GTC TGA AGC TTA CCC AGC GGC ACA GTC AC
HindIII-downRv2026c+2	CTC TGA AGC TTA CTA GCT GGG ACG GAC GA
166-down-acr-F	GGT CGC ATG GGA GGT TGC
322-uppfkB-F	CCA ACG CAA AGC GGT TCG G
950-down-acr-f	GGT CAA TCG GTA CGT TCG C
P1243	aat CTC GAG aac tga cca ctg ggt ccg tg
P1247	cgc GGA TCC aAT Ggc cac cac cct tcc cgt t
P1346	TAG ATC CCG CCG CCG AG
P1419	ggt ggt tGC TCT TCc aac gac gtc gac att atg gtc cgc
P1420	ggt ggt tGC TCT TCc aac ggt cag ctg acc atc aag gc
P1473	CTC ACA TGT TCT TTC CTG CGT TAT CCC
P1496	GAT CCG GAA CAT AAT GGT GCA GG
P1553	gaa cct ctc gag gtg acg agc c
P1554	ggA CTA GTc cgc taa ccg cgg tca tcg
P1557	ggg cag gat caa tac gtt tgg cac
P1558	ggACTAGTcag ctc gcc gtg cac tg
P1561	gac cag gag cat gtg gag caa c
P1562	ggACTAGTccg tga ttc ctc gaa gag acg
P1565	gtc gag aac atg ccg acc gg
P1566	ggACTAGT cgg ata ctg ctc ttg cca gcc
P1567	gaAGATCT agc gga acg tcg atc ggg
P1569	cg gtg atc gtc gtc cgt ccc
P1571	gaAGATCTc gtt ggc gtc gac gga tca g
P1797	gtgAAGCTTaCAG CTC ATC CTT acc tcc cca ttt ggc ggc ttc ggc cg
P1841	cgGGATCC aaa acg act cac cga c
P1870	gag cct ttc gtt tta ttt gat gcc tgg cag
P2305	g atc cgc atg ctt aat cgt cag gag gtc gac tgat ggc cac cac cct tcc cgt tca gcg
P2306	ccaATGCAT cag tgg tgg tgg tgg tgg tgc tcg
P2377	gaa ccg tac gcg aat tcT gCA cga ccg tcg aag tcc ttc tg
P2378	tt cgc gta cgg tGc ctt cgt tcg cac ggt gtc g
P2379	tt cgc gta cgg ttc ctt cgt tcg cac ggt gGc gct gcc ggt agg tgc tga cg
P2380	Tga caa ggg cat tct tac tgt gGc ggt ggc ggt ttc gga agg
P2381	Tga caa ggg cat tct tac tgt gtc ggt ggc ggt tGc gga agg gaa gcc aac cga aaa gc
P2382	cactCTCGAGg ttg gtg gCc cgg atc tga atg tgc ttt tcg
P2383	cactCTCGAGg ttg gtg gac cgg atc tg
P2384	gtgAA GCT TAT GGC CAC CAC CCT TCC C
P2391	cca ccc gcg gGc cct ctt ccc Aga gtt ttc tga gct gtt cg
P2392	cca ccc gcg gtc cct ctt ccc Aga gtt tGc tga gct gtt cgc ggc
P2393	cca ccc gcg gtc cct ctt ccc Aga gtt ttc tga gct gtt cgc Agc ctt cccAGc att cgc cgg act ccg g

Table 7. Plasmids used in this study

Plasmid	Relevant Features	Source/Reference
pMV306	*oriE*,*aph*, mycobacteriophage L5 att/int	Kong & Kunimoto, 1995
pOSN20	pSR173 with *lacZ* promoter cloned upstream of *Ptb21* to enhance expression *tetR*(B)	Shey-Njila et al. 2019
pRK247	pOSN20 with 6-His tagged *Mtb hspX*	This study
pRK249	pMV306 with *Mtb hspX gene*	This study
pRK251	pMV306 with *Mtb hspX*, downstream gene *acg*	This study
pRK253	pMV306 with *Mtb hspX*, downstream gene *acg* and upstream genes *Rv2030c*, *pfkB* and *Rv2028c*	This study
pRK255	pMV306 with *Mtb hspX*, downstream gene *acg* and upstream genes *Rv2030c*, *pfkB*, *Rv2028c*, *dosT* and *Rv2026c*	This study
pEY1	pET23bhspx containing TCC → GCC mutation in 1st Serine of *Mtb hspX* gene	This study
pEY2	pET23bhspx containing TCT → GCT mutation in 2nd Serine of *Mtb hspX* gene	This study
pEY3	pET23bhspx containing TCA → GCA mutation in 3rd Serine of *Mtb hspX* gene	This study
pEY4	pET23bhspx containing TCG → GCG mutation in 4th Serine of *Mtb hspX* gene	This study
pEY5	pET23bhspx containing TCC → GCC mutation in 5th Serine of *Mtb hspX* gene	This study
pEY6	pET23bhspx containing TCG → GCG mutation in 6th Serine of *Mtb hspX* gene	This study
pEY7	pET23bhspx containing TCG → GCG mutation in 7th Serine of *Mtb hspX* gene	This study
pEY8	pET23bhspx containing TCG→ GCG mutation in 8th Serine of *Mtb hspX* gene	This study
pEY9	pET23bhspx containing TCC → GCC mutation in 9th Serine of *Mtb hspX* gene	This study

Wayne Cultures

Mycobacterium tuberculosis strain MC^2-6206 defective for leucine and pantothenic acid biosynthesis [92] was cultured in 20 mL of Middlebrook 7H9 medium (BD/Difco, NJ) supplemented with 0.5% glycerol, 0.05% Tween 80 and 10% albumen-dextrose-saline (ADS) [89],

plus 50 μM leucine and 50 μM calcium pantothenate in rubber-stoppered crimp-top vials. Methylene blue was added to the culture at a concentration of 1.5 μg/mL as an indicator of hypoxia and a rubber stopper was crimped on before incubation at 37ºC for 30 days.

Mycobacterium smegmatis mc2- 155 was cultured in 20 mL of Middlebrook 7H9 medium (BD/Difco, NJ) supplemented with 0.5% glycerol, 0.05% Tween 80 and 10% albumen-dextrose-saline (ADS) [89] in a rubber-stoppered crimp-top vial. Methylene blue was added to the culture at a concentration of 1.5 μg/mL as an indicator of hypoxia and a rubber stopper was crimped on before incubation at 37ºC for 30 days.

Mycobacterium smegmatis strains harboring plasmid pRK249, pRK251, pRK253 or pRK255 were grown in 20 mL of Middlebrook 7H9 medium (BD/Difco, NJ) supplemented with 0.5% glycerol, 0.05% Tween 80 and 10% albumen-dextrose-saline (ADS) [89], plus 50 μg/mL kanamycin in a rubber-stoppered crimp-top vial. Methylene blue was added to the culture at a concentration of 1.5 μg/mL as an indicator of hypoxia and a rubber stopper was crimped on before incubation at 37ºC for 30 days.

Post incubation, cultures were centrifuged at 3500 x g for 15 minutes to pellet the bacteria and supernatant was harvested for western blot analysis. Supernatant was concentrated 6X using 10 kDa cutoff centrifugal filtered tubes (Millipore, MA). Whole cell lysate from pelleted bacteria was harvested using Quick DNA fungal/bacterial kit (Zymo Research, CA).

Express 6His-Tagged Acr from Tetracycline Inducible Promotor

Plasmid pRK247 (Figure 24B) was generated by incorporating the *hspX* gene containing a 6-His tag into a pOSN20 (Figure 24A) backbone encoding a tetracycline inducible promotor and origin of replication to be expressed in *Mycobacterium spp*. A 649-bp 6-His-*hspX* region

was amplified by PCR using DNA from *E. coli*, strain BL21(DE3)/pLys5/pET23bhspx and primers P2305 and P2306. The resulting product was enzymatically digested with enzymes *Sph*I and *Nsi*I. Plasmid pOSN20 was prepared for ligation with the PCR product by digestion with *Sph*I and *Nsi*I and subsequent dephosphorylation of the cut ends. Both digested products were ligated together and transformed into competent *E. coli* TAM-1 cells via electroporation as described previously. After selection of candidates, bacteria were lysed, and analytical restriction digestion was performed to confirm correct orientation of the ligated product. Candidate DNA was digested with enzyme *Aat*II revealing expected bands of 4707-bp and 2533-bp. Candidates displaying correct band size were DNA sequenced to determine the presence of possible mutations using primer P1196 and -80ºC freezer stocks of final candidate were created for future use.

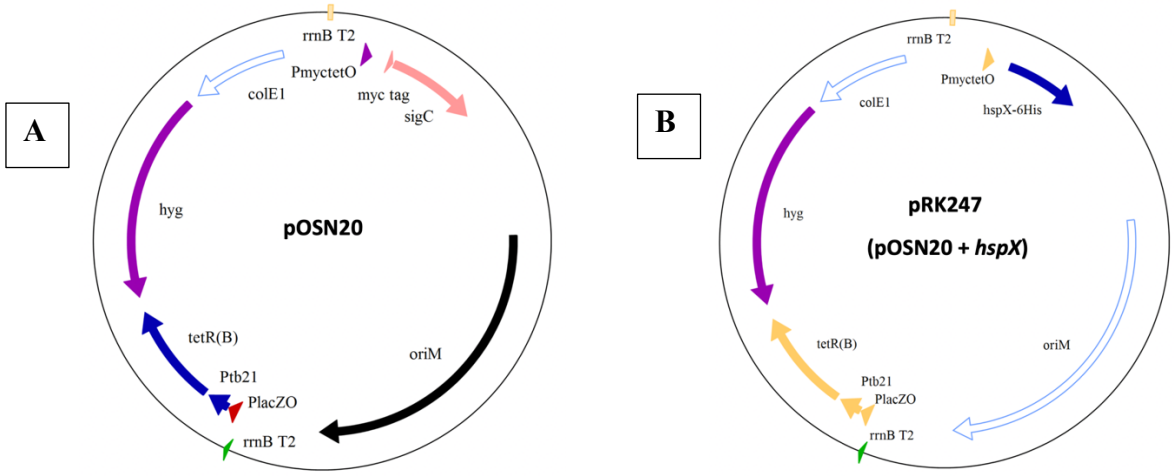

Figure 24. A. Plasmid map for pOSN20 used as the backbone to construct pRK247. The pink arrow indicates the region that was replaced with *hspX* to construct pRK247 **B.** Plasmid map for pRK247 showing replacement of *myc* tag and *sigC* gene with *hspX* indicated by a blue arrow.

Electroporation of pRK247 into *M. tuberculosis*, strain MC^2 6206

Two 50mL cultures of *M. tuberculosis*, strain MC^2 6206 defective for leucine and pantothenic acid biosynthesis [92] were grown to $O.D_{600}$ = 0.6 in Middlebrook 7H9 medium (BD/Difco, NJ) supplemented with 0.5% glycerol, 0.05% Tween 80 and 10% albumen-dextrose-saline (ADS) [89] , plus 50 μM leucine, 50 μM calcium pantothenate and 50 μg/mL hygromycin B to $O.D_{600}$ = 0.6. Cultures were pelleted at 3500 x g for 4 minutes and pellets were subsequently washed 4x with room temperature wash buffer containing 10% glycerol and 0.05% Tween 80. Pellets were washed again with 10% glycerol and resuspended in a final volume of 300μL of 10% glycerol. One hundred μL of cells were mixed with 100 ng of plasmid and transferred to electroporation cuvette. Electroporation was performed using BioRad GenePulser (BioRad, CA) with settings at 2500 V, 25 μF, 1000 Ohms and 2 mm gap cuvette.

Cells were rescued in 1 mL 7H9 medium supplemented with 0.5% glycerol, 0.05% Tween 80 and 10% ADS. Cells were then incubated for 24 hours before being plated on Middlebrook 7H10 agar plates (BD/Difco, NJ) supplemented with 0.5% glycerol, 0.05% Tween 80, 10% ADS and 50 μg/mL hygromycin B for candidate selection. Colonies were picked, bacteria lysed and screened by PCR using primers P1196 and P1543. Freezer stocks (-80^{o}C) were made with candidates containing the 583-bp band.

Detection of 6His-Acr from pRK247 by Western Blot

A 1-mL freezer stock culture of pRK247 was inoculated into Middlebrook 7H9 medium (BD/Difco, NJ) supplemented with 0.5% glycerol, 0.05% Tween 80 and 0.025% tyloxapol, plus 50 μM leucine, 50 μM calcium pantothenate and 50 μg/mL hygromycin B. After 2 week incubation at 37^{o}C to OD_{600} = 0.6, the culture was induced with 100 ng/mL anhydrotetracycline

for 48 hours. Culture was centrifuged at 3200 x g for 10 minutes to pellet the cells and supernatant harvested for western blot analysis. Whole cell lysate was extracted using a Bead beater and lysis buffer from Quick DNA fungal/bacterial kit (Zymo Research, CA). PMSF (0.5 mM) was added to prevent protein degradation. Western blot was performed using HRP conjugated anti-6His antibody (Abcam, MA).

Results:

Tested rAcr Functions are Comparable to Native Acr

Results indicated that the rAcr induced RAW247 cells to produce MCP-1 at levels similar to those reported when using the native Acr protein [91]. The level of MCP-1 produced in cultures increased in a dose-dependent manner with increasing amounts of rAcr (Figure 25). J774 cells were not shown as they produced large amounts of MCP-1 in untreated cells beyond the limits of assay detection.

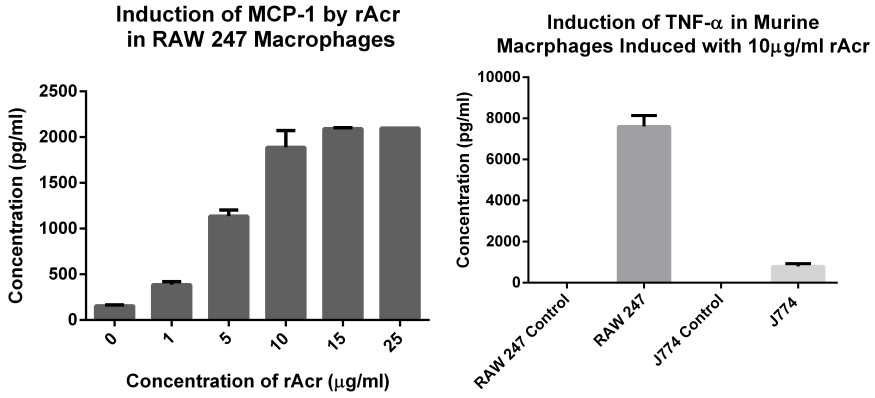

Figure 25. CCL-2 (MCP-1) and TNF-α cytokine panels of rAcr-treated macrophages. Left: RAW 247 macrophages treated with rAcr show increasing levels of MCP-1 as rAcr concentration increases. **Right:** RAW 247 and J774 macrophages treated with 10 μg/ml of rAcr show increased levels of TNF-α compared to untreated control.

Serine Phosphorylation Detection using pIMAGO

Mycobacterium tuberculosis Acr contains nine serine residues. As discussed in the literature review section, it has been previously shown that Acr is phosphorylated at a serine residue; however, it has not been established which of the nine serines are phosphorylated. To determine the location of the phosphorylated residue, each serine present in Acr was individually mutated to an alanine via site-directed mutagenesis. Western blot analysis using pIMAGO biotin phosphoprotein detection kit showed that phosphorylation of Acr was present in all constructed mutants except for pEY6 which contains a mutation at the sixth serine residue (Figure 26). From this preliminary information, we can infer that the sixth serine residue present in Acr is the potential site of phosphorylation for the protein.

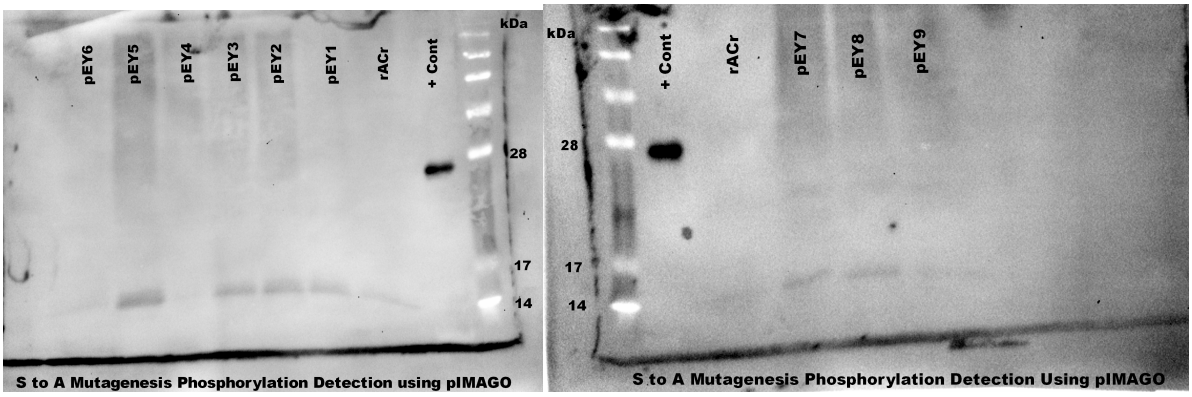

Figure 26. Western blot detection of the serine to alanine mutated residues of Acr using non-radioactive pIMAGO biotin phosphoprotein detection kit. Phosphorylation bands can be seen in all lanes except for pEY6, suggesting a lack of phosphorylation at the that site. Phosphorylated beta-casein protein used as the positive control.

Detection of Recombinant 6-His Acr from a Tetracycline-Inducible Promoter in *M. tuberculosis* strain MC^2 6206

Transport of Acr after synthesis from the cytoplasm of the bacterium to the outside environment is still not well understood. To try to further elucidate the transport mechanism, we assessed the production of Acr via a tetracycline-inducible promoter using transformed plasmid

pRK247 and *M. tuberculosis* strain MC^2 6206 and compared this to protein produced by the wild-type strain of MC^2 6206. After performing western blot analysis on both the recombinant and wild-type protein we observed that both are almost exclusively located in the bacterial whole cell lysate and not secreted to the supernatant (Figure 27). Furthermore, the wild-type protein can be detected as both a monomer and a dimer, whereas the recombinant is only present as a monomer.

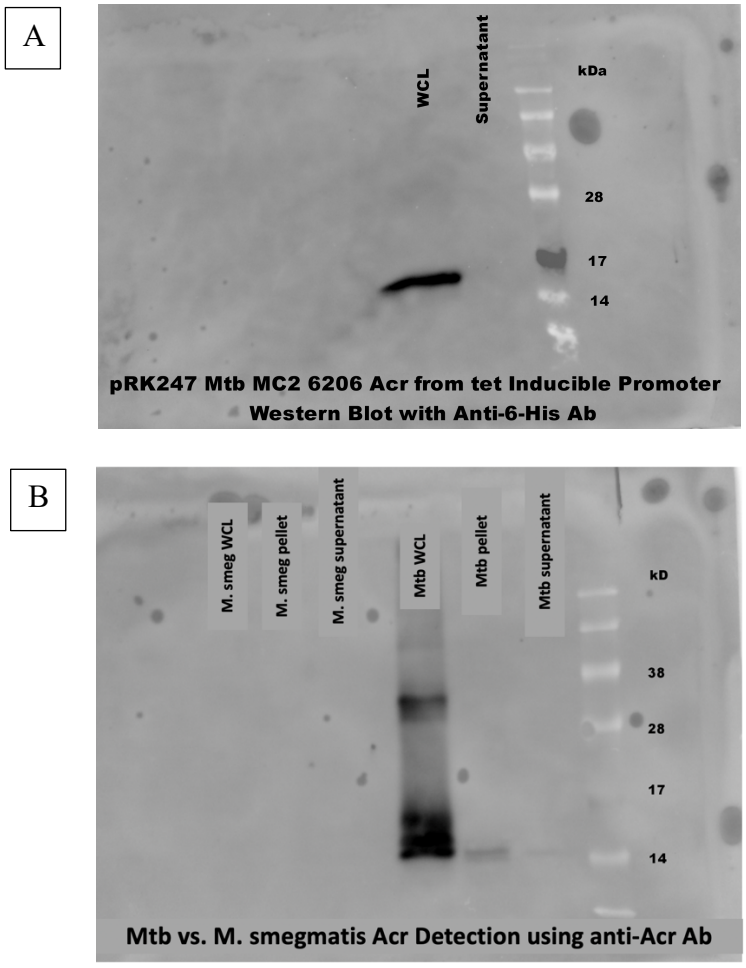

Figure 27. A: Western blot analysis of tetracyline induced pRK247 plasmid producing a recombinant form of Acr. Protein is exlusivley found in the bacterial whole cell lysate fraction as a monomer and is not found in the supernatant. **B:** Western blot analysis on wild-type *M.tuberculosis* MC^2 6206 and wild-type *M. smegmatis*. Protein is almost exclusivley located in the bacterial whole cell lysate as both a monomer and a dimer of *M. tuberculosis*. No protien is detected for *M. smegmatis* in any fraction.

Necessity of Adjacent Genes in *hspX* Operon for the Production of Acr Protein

Little is known regarding regulation and control within the *hspX* operon itself and whether all or some of the adjacent genes to *hspX* are needed for the protein to be produced and exported out of the cell during hypoxic growth conditions. To assess the roles of the adjacent genes, four plasmids were generated to encompass the operon, starting with *hspX* by itself and adding adjacent genes to *hspX* in each subsequent plasmid construct. Each plasmid was electroporated into *M. smegmatis*, a species that inherently lacks the *hspX* operon, and grown in hypoxic Wayne cultures to mimic protein production within *in vivo* granulomas [45] . After a one-month incubation period, the supernatant and whole cell lysates were harvested for western blot analysis using anti-Acr antibody. From the western blot we did not find any evidence of Acr in the supernatant or whole cell lysate fractions (Figure 28).

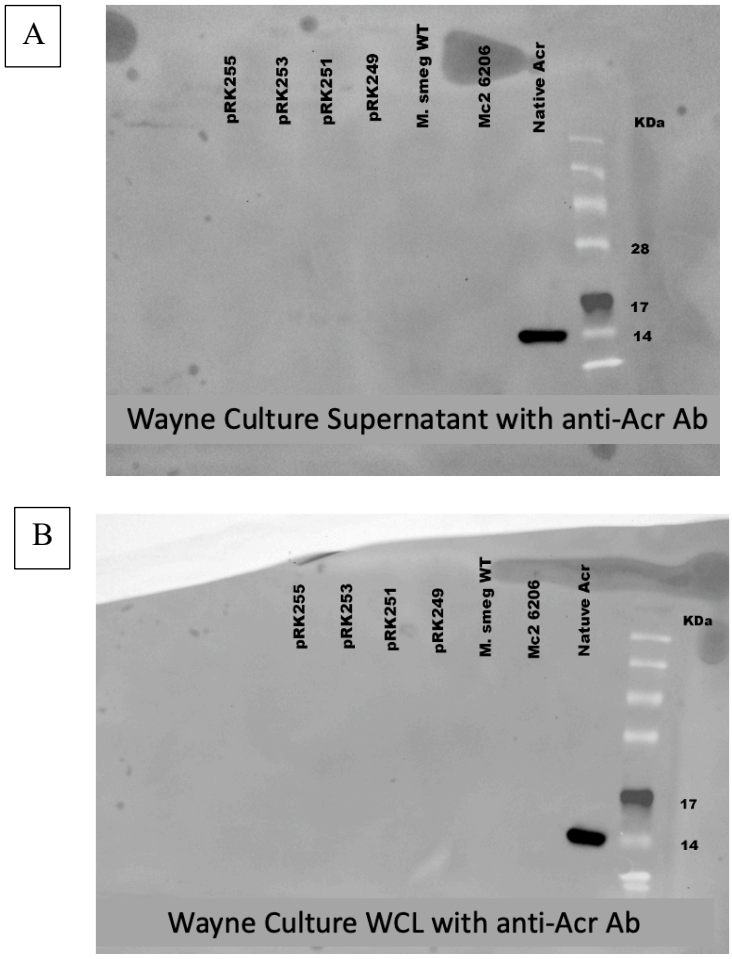

Figure 28. A: Western blot analysis of the Wayne culture supernatants of *M. smegmatis* containing pMV306 derivative plasmids after one-month incubation period. Samples were probed with anti-Acr antibody. Native Acr was used as a positive control. No Acr can be detected in any Wayne culture supernatant. **B:** Western blot analysis of Wayne culture bacterial whole cell lysate after one-month incubation period probed with anti-Acr antibody. Native Acr was used as a positive control. No Acr can be detected in any Wayne culture whole cell lysate.

Because we did not detect any protein from *M. smegmatis* containing tet-inducible *hspX* plasmids, we sought to determine if inserting the same plasmids into *M. marinum* may yield positive results as *M. marinum* has been shown to express similar dormancy and granuloma causing characteristics as *M. tuberculosis* [79,93]. After electroporation of the plasmids, a western blot will be performed on both aerobically-grown and Wayne-cultured *M. marinum* cultures to determine the presence of Acr. At this time this experiment remains ongoing.

Acr Secretion Pathway Predictions

To try to further elucidate the pathway of Acr secretion from the bacterium to the outside environment, we attempted to use online prediction software to determine if the Acr genetic sequence contained any motifs or secretion systems that have historically been present in other secreted proteins. The first secretion method investigated was the presence of a Sec system. The Sec translocase, present in all bacteria, uses a chaperone protein to guide the protein of interest to a membrane channel which ultimately transports the protein across the bacterial outer membrane at the expense of ATP [94]. The second method, the twin-arginine translocation (Tat) pathway, uses a specific signal motif, S/TRRxFLK, with x being a polar amino acid in the sequence, which interacts with a Tat docking system. Once in the interaction takes place, the docking system transports the protein across the membrane through a second pore forming protein complex using proton energy [95]. The last method we investigated was the possibility of Acr being an autotransporter containing a Type V secretion system. Type V secretion systems consist of a protein to be secreted and a ß-barrel domain in the outer membrane that is used to transport the protein [96]. The difference between this system and the Sec or Tet pathway is that the Type V does not use any exogenous energy source to move protein across the membrane. When the nucleotide sequence of *M. tuberculosis* Acr was uploaded into the two prediction software programs, Pred-Tat [97] showed that the sequence did not contain either a Sec signal peptide nor a Tat signal peptide (Figure 29A). Furthermore, the transmembrane helices predicting program TMHMM (DTU Bioinformatics, Denmark) also showed that there were no transmembrane helices predicted in the nucleotide sequence of Acr (Figure 29B). Taken together, these results

show that Acr is not secreted from the bacterium via a Sec, Tat or Type V secretion pathway and must use another less predominant pathway.

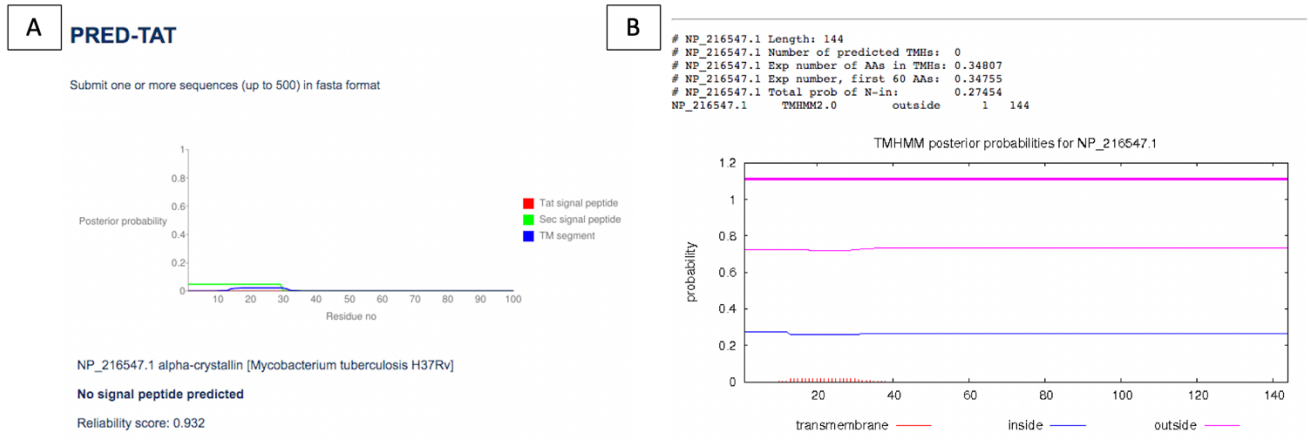

Figure 29. A. Results from Pred-Tat program indicating the absence of both a Sec and a Tat sequence from *M. tuberculosis* Acr nucleotide sequence. **B.** Results from TMHMM program indicating the absence of transmembrane helices in the Acr nucleotide sequence.

M. tuberculosis Phosphorylation via Mass Spectrometry

Native Acr purified from wild-type *M. tuberculosis* H37Rv and a recombinant form of Acr purified from plasmid induced *E. coli* were sent for analysis via top-down, whole protein, mass spectrometry to determine the phosphorylation states of each type of Acr and to confirm the earlier site directed mutagenesis findings. Results from the analysis for the native form, showed no serine phosphorylation, however, the analysis did find a phosphorylated threonine residue at the T3 site. I addition to threonine phosphorylation, the analysis also showed methionine oxidation at sites M2, M5 and M15. Results for the recombinant Acr also showed no serine phosphorylation, however, it did show the same methionine oxidation sites as the native form.

CHAPTER 5

CONCLUSIONS

Tuberculosis (TB) is currently responsible for more human deaths annually than any other single infectious agent [2]. It's ability to be successful in a host for decades without detection is attributable to the bacterium's ability to evade the host immune system and remain undetected. The ability of *Mycobacterium tuberculosis* (*Mtb*) to survive in many host environments including granulomas and macrophages make it difficult to target treatments that can have adequate impact on all potential hiding areas at the same time. Another factor that contributes to the difficultly of detection is the fact that most individuals do not express any symptoms during the latent infection stage and do not know if they harbor the bacterium. It is because metabolically dormant bacteria in the latent stage are antibiotic resistant, it would be impractical to administer antibiotics to an individual. Antibiotics are given to individuals whom show signs of active disease and have a positive TB skin test, however, if the bacteria are encased in a tubercle, then it is possible that the drugs cannot reach the bacteria to induce their effects or if they would have any effect at all due to the metabolic inactivity of the bacteria itself. During active infection, individuals transmit the bacterium to others via the aerosol route when they cough, sneeze or speak forcefully. Delays in diagnosis of TB, especially in areas of the world with poor healthcare, may contribute to millions of new cases per year. A compounding effect to this issue is also the presence of HIV co-infection, as those co-infected have an increased chance of developing active TB disease due to a declining immune system. The current WHO eradication and treatment programs, The End TB Strategy [7] and TB-DOTS [4], are not meeting the set goals

and expectations of their respective programs due to their passive, find and treat, approach. If successful eradication is to be achieved, then the transmission of *M. tuberculosis* needs to be further studied to have an impact on future public health interventions.

Whole Genome Sequencing (WGS) has given researchers the ability to look at an organism at the DNA level. WGS has allowed the translation of countless species' genomes to a level where each base pair in their DNA sequence can be looked at individually. When combined with Single Nucleotide Polymorphism (SNP) analysis, detecting mutations present at an individual base pair site in the genome, WGS can be a powerful tool that has the ability to decipher ancestry as well as person-to-person transmission and help elucidate who and where are responsible for transmission. In chapter 3 of this dissertation, we used the modified BioNumerics pipeline co-developed with the Centers for Disease Control and Prevention to determine the number of SNPs between isolates gathered from patients presenting with pulmonary TB symptoms in the Rubaga Division of Kampala, Uganda. DNA was extracted from each of the samples given by the patients and sequenced using an Illumina sequencer and the Illumina sequencing method. The sequenced genomes were then processed by the aforementioned BioNumerics pipeline to determine the number of SNPs in each genome as compared to *M. tuberculosis* H37Rv as a reference. Each genome was grouped into its proper MTBC lineage and the number of SNPs were then compared to others in the same MTBC lineage. Using a SNP threshold of $\leq$ 12 SNPs as indicative of a possible transmission pair [98], we were able to determine which of the 113 isolates fit this criteria. The analysis resulted in 2 isolates grouping into MTBC lineage 1, Indo-Oceanic, 1 isolate in lineage 2, East Asian (Beijing), 23 isolates grouping into MTBC lineage 3, East African Indian, and 87 isolates grouping into MTBC lineage 4, Euro-American. Among the patients with pulmonary tuberculosis, the predominant

sub-strain of *M. tuberculosis* was lineage 4, followed by lineage 3. Furthermore, lineage 4 contained the highest frequency of transmission pairs along with having the most isolates altogether. Lineage 4 having the higher frequency of transmitters may be indicative of that lineage possibly being more virulent in that region of the world. This was not surprising as previously mentioned studies have shown that the predominant lineage in Uganda is lineage 4. Isolates grouped into lineage 1 and 2 could not be included in frequency analysis as more samples falling into those lineages are needed.

The SNP analysis also showed which isolate(s) were likely transmitting to others in their specific lineages. In lineage 3, the identical pair of isolates 16294 and 20095, from cluster 1, were potential transmitters to isolates 20839, 22199, 19621 and 20918. The most robust transmitter from lineage 4 was shown to be isolate 16732, from cluster 13, which was shown to potentially transmit to isolates 15634, 19077, 20606 and 20824. Furthermore, a pairwise SNP difference matrix showing bacterial transmission between isolates as a network was also generated. From these matrices, we could see how the isolates grouped as a network which allows a visual representation of the data.

The presence of SNPs in virulence associated genes of *M. tuberculosis* can reveal how some strains of the bacteria are more prone to cause disease than others. Numerous mutations were found in the *plcA* and *plcB* phospholipase C genes of the isolates in this study. These SNPs were of interest because they all occurred at the beginning of the protein coding sequence and because one mutation in *plcB* leading to a nonsense mutation. In our subset of isolates, the bacteria are able to establish infection, as evidenced by the patients in the study, meaning that the SNPs in those genes may be advantageous to the survival and pathogenesis of the bacteria. This is in agreement with a study by Talarico et al. who also found mutations in the *plc* group of

genes, in particular *plcD* in the Beijing strain of *M. tuberculosis*, a strain designated as one of the most virulent [87]. Grasping a better understanding of transmission networks will allow public health interventions to be implemented in the social gathering establishments that are frequented by transmitters. The success of these public health interventions can possibly be more effective compared to the current passive programs like TB-DOTS.

Chapter 4 of this dissertation focused on furthering the understanding of the *M. tuberculosis* dormancy heat shock protein alpha-crystallin (Acr, HspX). This 16 kDa protein has been studied for many years, however, it remains unclear how the protein is exported. Acr has homology to the mammalian Acr protein found in the eye with the proteins having approximately 18% identity [64]. Both belong to the family of heat shock proteins (HSP) whose primary job is to function as a chaperone and help re-fold proteins. Acr is the primary protein expressed during hypoxic conditions accounting for 25% of detectible protein [22]. Although Acr has been hypothesized to contribute to dormancy, how or if the protein gets out of the bacterium after production is still not well understood. Phosphorylation is a type of post-translational modification that can induce conformational changes in a protein that can ultimately alter its function. Although there is limited evidence demonstrating the role of phosphorylation in Acr, there is no evidence to its effects on protein function. A previous study had shown that Acr was exclusively phosphorylated at a serine residue, however, it did not elucidate which of the nine serine residues in Acr was the phosphorylation site [70].

To try to further the understanding of Acr phosphorylation, we generated 9 *hspX* mutants that encoded a serine to alanine mutation at each of the individual Acr serine residues. Using pIMAGO western blot phosphorylation detection kit, we observed that Acr was phosphorylated at the 6th serine residue (amino acid 103 of 144) as that was the only mutant to not produce a

band during detection. To confirm this finding, the native form of Acr and a recombinant form were analyzed by mass spectrometry. Mass spectrometry results did not confirm the presence of serine phosphorylation; however, it did reveal multiple phosphorylated threonine residues in both the native and recombinant Acr. Looking further into threonine phosphorylation, research has shown that in other organisms such as *Pseudomonas aeruginosa*, threonine phosphorylation of the gene products of *ppkA* and *pppA* were associated with export of virulence factors via a Type 6 Secretion System (T6SS)[99]. Since our aim was to elucidate Acr export, we wanted to determine if there were any homologues of *ppkA* and *pppA* present in *M. tuberculosis* whose gene products may potentially act in the same manner. A BLASTP search indicated a 39% identity between the gene products of *ppkA* and *M. tuberculosis pknB*, a serine/threonine kinase, and a 36% identity between the gene products of *pppA* and *M. tuberculosis pstB*, a serine/threonine phosphatase. PknB regulates aspects of cell growth and survival, while PstB is part of an ABC transporter complex. The role of these proteins in association with Acr is an area of further study that is beyond the scope of this dissertation.

Furthermore, to determine the effect of the adjacent genes in the *hspX* operon in the secretion of Acr, we generated four plasmids, each containing *hspX* and a range of adjacent genes as previously described in the methods section. After electroporating the plasmids into *M. smegmatis*, we did not find evidence of Acr production in aerobically-grown cultures nor in hypoxic Wayne cultures in 7H9 medium. The antibacterial resistance cassette incorporated into the plasmid was expressed, allowing the bacteria to replicate on agar with antibiotics, however, further experiments are needed to determine the reason why we did not detect Acr.

In conclusion, this dissertation, and the work presented, highlights the needs in two aspects of *Mycobacterium tuberculosis* pathogenesis, person-to-person transmission of the

bacteria and mycobacterial dormancy. Chapter 3 showed that using WGS and SNP analysis, we have the ability to detect possible transmission between pairs of clinical TB samples by looking at the number of SNPs present between the two isolates. All isolate pairs that contained ≤ 12 SNPs should be considered possible transmission pairs. Once these isolates have been identified, they can be clustered and a bacterial transmission network can be generated to visualize the transmission patterns present. When the transmission data is combined with the epidemiological data, the results will give researchers and public health officials the ability to direct preventative public health efforts toward eradication of this deadly disease.

Chapter 4 showed that Acr still remains a protein that needs to be further studied with regard to export and post-translational modification. Due to the protein's predicted role in dormancy, it is critical to the study of *M. tuberculosis* and its eradication, that scientists better understand how virulence factors are exported and the roles that proteins like Acr impact metabolic dormancy, especially since 1/3 of the world's population harbors the bacterium as a latent infection. The question of whether or not phosphorylation plays any role in the export process also needs to be elucidated. If TB transmission and the export of Acr can be better understood, then targeted eradication efforts and potentially powerful drugs could be developed to combat the ability of *M. tuberculosis* to be transmitted or hide from a host immune system. These answers could facilitate development of faster and more efficient detection methods for active disease or for more-effective treatments for individuals, or prophylactic measures to aid in the struggle to eradicate TB.

CHAPTER 6

FUTURE DIRECTIONS

When attempting to answer research questions in the course of a study, often times more questions and future experiments arise. This may be due to interesting findings in our data or other new developments within the literature. There are several future directions and studies that can be conducted based on the research presented in this dissertation.

Aim 1 of this dissertation focused on the transmission of *Mycobacterium tuberculosis* (*Mtb*) using whole genome sequencing (WGS) and single nucleotide polymorphism (SNP) threshold analysis. Based on the findings of the study, there are several future studies that can be pursued. First, the study protocol called for participants to submit at least two sputum samples which were cultured for growth and all bacteria recovered were sub-cultured for potential future analysis. One proposed future study would be to analyze the SNPs between isolates of the same person to determine if the isolates are genetically identical or if a person has been infected with multiple *Mtb* strains. This would help identify if a person has come into contact and transmission has occurred with multiple individuals. This information would contribute to the epidemiological side of the study and the aspect of social networks by contributing to the social mixing patterns of the individuals in the study. Second, this current study is limited to the Rubaga Division of Kampala, Uganda. To make the study findings more generalizable, the study should be extended to other parts of the country to incorporate more study individuals and increase the chances of finding potential transmission hotspots outside of the current geographical limitation. Lastly,

analysis of SNPs in genes present in the study isolates can potentially be expanded beyond virulence genes. The presence of the annotated SNP tables generated by the pipeline gives us the ability to look for any gene in the *Mtb* genome.

Aim 2 of this dissertation focused on the characterization and the determination of the role of protein phosphorylation in association with the secretion of *Mtb* Acr protein. Beyond the experiments completed in this study, there are several more studies and future directions that can be proposed on the subject. First, the detection of Acr expressed from plasmids transformed into *M. marinum* was not completed. The completion of this experiment would help further the understanding of the regulation, production and transport of Acr and the requirement of other factors and in particular those encoded by adjacent genes in the operon. Because *M. marinum* pathogenesis has similar virulence characteristics to *Mtb*, such as granuloma formation, and *M. smegmatis* does not, potentially at least some similar Acr regulation, production and transport mechanisms are present in *Mtb* and *M. marinum*. Second, further experiments need to be done to determine if the culture medium used in the Wayne cultures has any potential effect on the production of Acr. In this study, the cultures were grown in 7H9 medium, a mycobacterial medium that is rich in nutrients. However, the original Wayne cultures were grown in Dubos medium which is similar to 7H9 but is lacking some nutrients. This would rule out any possibility that the type of media had any effect on protein production.

Furthermore, a study conducted by Kruh-Garcia et al. showed that Acr can be detected from exosomes isolated from serum samples of tuberculosis patients [100]. This can potentially add to our Acr transport data as we did not find any typical transport pathway motifs used for traditional secretion of a protein. A future experiment can be performed by isolating exosomes from *Mtb* cultures containing the serine-to-alanine mutations and probing the contents for the

presence of Acr. This would potentially help clarify if the Acr that is being detected in the whole cell lysates and supernatants are from the exosomes and if the mutation of Acr has any effect on the transport of the protein via an exosome.

Lastly, the determination of the phosphorylation state of Acr needs to be confirmed by using a bottom-up mass spectrometry approach. The phosphorylation results presented in this study were conducted using a top-down approach by analyzing the entire protein. The bottom-up approach would cleave the protein and analyze the residues in smaller segments, potentially increasing the accuracy of the analysis. Another phosphorylation experiment that can potentially be beneficial would be to transform the Acr serine-to-alanine mutation plasmids into a *Mtb* strain that has a deletion in the *acr* gene. Performing growth curves and a western blot with these variants would determine any defects in growth or protein expression due to phosphorylation.

REFERENCES

1. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic co-expansion of Mycobacterium tuberculosis with modern humans. Nat Genet. 2014;45(10):1176–82.
2. WHO. World Tuberculosis Report 2019. 2019.
3. WHO. World Tuberculosis Report 2018. 2018.
4. Maher D, Mikulencak M. What is DOTS? A Guide to Understanding the WHO-recommended TB Control Strategy Known as DOTS. 1999.
5. Lockwood D. Leprosy elimination—a virtual phenomenon or a reality? BMJ [Internet]. 2002 [cited 2019 Sep 9];324:1516–8. Available from: https://www-ncbi-nlm-nih-gov.proxy-remote.galib.uga.edu/pmc/articles/PMC1123450/pdf/1516.pdf
6. Mhimbira F, Hella J, Maroa T, Kisandu S, Chiryamkubi M, Said K, et al. Home-Based and Facility-Based Directly Observed Therapy of Tuberculosis Treatment under Programmatic Conditions in Urban Tanzania. PLoS One [Internet]. 2016 [cited 2019 Sep 9];11(8). Available from: https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0161171&type=printable
7. WHO. The End TB Strategy, World Health Organisation, Geneva. 2015 [cited 2019 Aug 28]; Available from: https://www.who.int/tb/strategy/End_TB_Strategy.pdf?ua=1
8. Guwatudde D, Nakakeeto M, Jones-Lopez EC, Maganda A, Chiunda A, Mugerwa RD, et al. Tuberculosis in Household Contacts of Infectious Cases in Kampala, Uganda. Am J Epidemiol [Internet]. 2003 [cited 2019 Sep 4];158(9):887–98. Available from: https://www-ncbi-nlm-nih-gov.proxy-

remote.galib.uga.edu/pmc/articles/PMC2869090/pdf/nihms-200021.pdf

9. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Genomics [Internet]. 2015 [cited 2019 Aug 20];107(1):1–8. Available from: http://dx.doi.org/10.1016/j.ygeno.2015.11.003

10. Ahmadian A, Gharizadeh B, Gustafsson A, Sterky F, Nyren P, Uhlen M, et al. Single-Nucleotide Polymorphism Analysis by Pyrosequencing. Anal Biochechemistry [Internet]. 2000 [cited 2019 Aug 19];280:103–10. Available from: https://pdf.sciencedirectassets.com/272363/1-s2.0-S0003269700X00425/1-s2.0-S0003269700944932/main.pdf?X-Amz-Security-Token=AgoJb3JpZ2luX2VjEPb%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJIMEYCIQCuAFouCWDTpt6nLViuBhOpBZ8XYVQau8OruxotpZKMYQIhAKhcY60aVfWn

11. Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis [Internet]. 2013 Feb [cited 2017 Oct 27];13(2):137–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23158499

12. Ling Lin P, Flynn JL. Understanding Latent Tuberculosis: A Moving Target. J Immunol [Internet]. 2010 [cited 2019 Sep 5];185(1):15–22. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3311959/pdf/nihms351174.pdf

13. Getahun H, Matteelli A, Chaisson R, Raviglone M. Latent Mycobacterium tuberculosis Infection. NEJM. 2015;372:2127–35.

14. Houben R, Dodd P. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. PLOS Med. 2016;13(10).

15. Lugo-Villarino G, Hudrisier D, Benard A, Neyrolles O, Boros DL, Philips JA. Emerging trends in the formation and function of tuberculosis granulomas. Front Immunol [Internet]. 2013 [cited 2019 Aug 19];3(405):1–9. Available from: www.frontiersin.org

16. Marakalala MJ, Raju RM, Sharma K, Zhang YJ, Eugenin EA, Prideaux B, et al. Inflammatory signaling in human tuberculosis granulomas is spatially organized. Nat Med [Internet]. 2016;22(5):531–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27043495

17. Ramakrishnan L. Revisiting the role of the granuloma in tuberculosis. Nat Rev Immun [Internet]. 2012 [cited 2019 Sep 9];12:352–66. Available from: https://www.nature.com/articles/nri3211.pdf

18. Lö Nnroth K, Jaramillo E, Williams BG, Dye C, Raviglione M. Drivers of tuberculosis epidemics: The role of risk factors and social determinants. Soc Sci Med. 2009;68:2240–6.

19. Horwitz J. a-Crystallin can function as a molecular chaperone. Proc Natl Acad Sci USA. 1992;89:10449–53.

20. Haslbeck M, Franzmann T, Weinfurtner D, Buchner J. Some like it hot: the structure and function of small heat-shock proteins. Nat Struct Mol Biol. 2005;12(10):842–6.

21. Bakthisaran R, Tangirala R, Rao CM. Small heat shock proteins: Role in cellular functions and pathology. Biochim Biophys Acta - Proteins Proteomics. 2015 Apr;1854(4):291–319.

22. Yuan Y, Crane D, Simpson R, Zhu Y, Hickey M, Sherman D, et al. The 16-KDa-crystallin (ACR) protein of Mycobacterium tuberculosis is required for growth in macrophages. Proc Natl Acad Sci USA [Internet]. 1998;95(August):9578–83. Available from: http://dx.doi.org/10.1073/pnas.95.16.9578

23. Hu Y, Liu A, Menendez MC, Garcia MJ, Oravcova K, Gillespie SH, et al. HspX knock-out in Mycobacterium tuberculosis leads to shorter antibiotic treatment and lower relapse rate in a mouse model - A potential novel therapeutic target. Tuberculosis. 2015;95(1):31–6.

24. Cave A. THE EVIDENCE FOR THE INCIDENCE OF TUBERCULOSIS IN ANCIENT EGYPT. Br J Tuberc [Internet]. 1939 [cited 2019 Aug 13];33(3):142–52. Available from: https://pdf.sciencedirectassets.com/273491/1-s2.0-S0366085039X00131/1-s2.0-S0366085039800163/main.pdf?X-Amz-Security-Token=AgoJb3JpZ2luX2VjEGcaCXVzLWVhc3QtMSJHMEUCID4XDySDaMe%2BVFGlQMgXUZVsX4Qs9R7tCJBF1pdTxP0tAiEAyV2wZfk3iz%2BmL0%2Bj9e4aMnscF6JP4ljsM5C7rn

25. Daniel TM. The history of tuberculosis. Respir Med. 2006;100:1862–70.

26. Rodriguez-Campos S, Smith NH, Boniotti MB, Aranaz A. Overview and phylogeny of Mycobacterium tuberculosis complex organisms: Implications for diagnostics and legislation of bovine tuberculosis. Res Vet Sci [Internet]. 2014 Oct [cited 2019 Aug 13];97:S5–19. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0034528814000435

27. Müller B, Dürr S, Alonso S, Hattendorf J, Laisse CJ, Parsons SD, van Helden PD ZJ. Zoonotic Mycobacterium bovis– induced Tuberculosis in Humans. Emerg Infect Dis [Internet]. 2013 [cited 2019 Aug 13];19(6):899=908. Available from: http://dx.doi.org/10.3201/eid1906.120543

28. Djemal S, Siala M, Smaoui S, Marouane C, Bezos J, Messadi-Akrout F, et al. Genetic diversity assessment of Tunisian Mycobacterium bovis population isolated from cattle.

BMC Vet Res [Internet]. 2017 [cited 2019 Sep 30];13(393). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5732386/pdf/12917_2017_Article_1314.pdf

29. Johnson MM, Odell JA. Nontuberculous mycobacterial pulmonary infections. J Thorac Dis [Internet]. 2014 [cited 2019 Aug 16];6(3):210–20. Available from: www.jthoracdis.com

30. Luca S, Mihaescu T. History of BCG Vaccine. Maedica (Buchar) [Internet]. 2013 [cited 2019 Aug 14];8(1):53–8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3749764/pdf/maed-08-53.pdf

31. Murray JF, Schraufnagel DE, Hopewell PC. Treatment of Tuberculosis: A Historical Perspective. Ann Am Thorac Soc [Internet]. 2015 [cited 2019 Aug 14];12(12):1749–59. Available from: www.atsjournals.org

32. Sensi P. History of the Development of Rifampin. Rev Infect Dis [Internet]. 1983 [cited 2019 Aug 14];5(3):402–6. Available from: https://www-jstor-org.proxy-remote.galib.uga.edu/stable/pdf/4453138.pdf?refreqid=excelsior%3A3a6a3907b991fb2b517da9eee416df0b

33. Whalen CC. Failure of Directly Observed Treatment for Tuberculosis in Africa: A Call for New Approaches. Clin Infect Dis. 2006;42(7):1048–50.

34. Gupta A, Kaul A, Tsolaki AG, Kishore U, Bhakta S. Mycobacterium tuberculosis: Immune evasion, latency and reactivation. Immunobiology. 2011;217:363–74.

35. Keane J, Katarzyna Balcewicz-Sablinska M, Remold HG, Chupp GL, Meek BB, Fenton MJ, et al. Infection by Mycobacterium tuberculosis Promotes Human Alveolar Macrophage Apoptosis. Infect Immun. 1997;65(1):298–304.

36. Flynn JL, Chan J. Immune evasion by Mycobacterium tuberculosis: living with the enemy. Curr Opin Immunol. 2003;15:450–5.

37. Hmama Z, Gabathuler R, Jefferies W, De Jong G, Reiner NE, Hmama Z, et al. Attenuation of HLA-DR Expression by Mononuclear Phagocytes Infected with Mycobacterium tuberculosis Is Related to Intracellular Sequestration of Immature Class II Heterodimers. J Immunol [Internet]. 1998 [cited 2019 Aug 15];(161):4882–93. Available from: http://www.jimmunol.org/content/161/9/4882

38. Brennan PJ. Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. Tuberculosis [Internet]. 2003 [cited 2016 Nov 17];83:91–7. Available from: www.elsevierhealth.com/journals/tube

39. Todar K. Mycobacterium tuberculosis [Internet]. Todar's Online Textbook of Bacteriology. 2019 [cited 2019 Aug 15]. Available from: http://textbookofbacteriology.net/tuberculosis.html

40. Lambert P a. Cellular impermeability and uptake of biocides and antibiotics in gram-positive bacteria and mycobacteria. J Appl Microbiol. 2002;92:46S-54S.

41. Kleinnijenhuis J, Oosting M, Joosten LAB, Netea MG, Crevel R Van. Innate Immune Recognition of Mycobacterium tuberculosis. Clin Dev Immunol [Internet]. 2011 [cited 2019 Aug 21];2011. Available from: https://www-ncbi-nlm-nih-gov.proxy-remote.galib.uga.edu/pmc/articles/PMC3095423/pdf/CDI2011-405310.pdf

42. Chen T, He L, Deng W, Xie J. The Mycobacterium DosR regulon structure and diversity revealed by comparative genomic analysis. J Cell Biochem. 2013;114(1):1–6.

43. Leistikow RL, Morton RA, Bartek IL, Frimpong I, Wagner K, Voskuil MI. The Mycobacterium tuberculosis DosR regulon assists in metabolic homeostasis and enables

rapid recovery from nonrespiring dormancy. J Bacteriol. 2010;192(6):1662–70.

44. Kim M-J, Park K-J, Ko I-J, Kim YM, Oh J-I. Different Roles of DosS and DosT in the Hypoxic Adaptation of Mycobacteria. J Bacteriol. 2010;192(19):4868–75.

45. Wayne L, Hayes LG. An In Vitro Model for Sequential Study of Shiftdown of Mycobacterium tuberculosis through Two Stages of Nonreplicating Persistence Downloaded from http://iai.asm.org/ on May 9 , 2015 by UNIVERSITY OF PENNSYLVANIA LIBRARY. Infect Immun. 1996;64(6):2062–9.

46. Ojha AK, Baughn AD, Sambandan D, Hsu T, Trivelli X, Guerardel Y, et al. Growth of Mycobacterium tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. Mol Microbiol. 2008;69(1):164–74.

47. Kerns PW, Ackart DF, Basaraba RJ, Leid J, Shirtliff ME. Mycobacterium tuberculosis pellicles express unique proteins recognized by the host humoral response HHS Public Access pellicle-specific proteins represent targets for the development of future diagnostic tests and vaccines. Pathog Dis. 2014;70(3):347–58.

48. Sambandan D, Dao DN, Weinrick BC, Vilchèze C, Gurcha SS, Ojha A, et al. Keto-Mycolic Acid-Dependent Pellicle Formation Confers Tolerance to Drug-Sensitive Mycobacterium tuberculosis. MBio. 2013;4(3).

49. Alnimr AM. Dormancy models for Mycobacterium tuberculosis: A minireview. Braz j microbiol [Internet]. 2015;46(3):641–7. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1517-83822015000300641

50. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature. 1998;393(6685):537–44.

51. Philipp WJ, Poulet S, Eiglmeier K, Pascopellat L, Balasubramaniant V, Heym B, et al. An integrated map of the genome of the tubercle bacillus, Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium leprae (genome mapping/contig mapping/ordered libraries/bacterial genomics/tuberculosis). Proc Natl Acad Sci USA. 1996;93:3132–7.

52. Wu S, Howard ST, Lakey DL, Kipnis A, Samten B, Safi H, et al. The principal sigma factor sigA mediates enhanced growth of Mycobacterium tuberculosis in vivo. Mol Microbiol. 2004;51(6):1551–62.

53. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, et al. A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun [Internet]. 2014;5:4812. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4166679&tool=pmcentrez&rendertype=abstract

54. Ford CB, Shah RR, Kato Maeda M, Gagneux S, Murray MB, Cohen T, et al. Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug resistant tuberculosis. Nat Genet. 2013;45(7):784–90.

55. Hernández-Pando R, López B, Aguilar D, Orozco H, Burger M, Espitia C, et al. A marked difference in pathogenesis and immune response induced by different Mycobacterium tuberculosis genotypes. Clin Exp Immunol. 2003;133:30–7.

56. Kimchi-Sarfaty C, Oh JM, Kim I-W, Sauna ZE, Calcagno AM, Ambudkar S V, et al. A “silent” polymorphism in the MDR1 gene changes substrate specificity. Science [Internet]. 2007;315(5811):525–8. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/17185560

57. Hamosh A, King tTerri M, Rosenstein BJ, Corey M, Levison H, Durie P, et al. Cystic Fibrosis Patients Bearing Both the Common Missense Mutation Gly-vAsp at Codon 551 and the AF508 Mutation Are Clinically Indistinguishable from AF508 Homozygotes, Except for Decreased Risk of Meconium lieus. Am J Hum Genet. 1992;51:245–50.

58. Ramaswamy S V, Reich R, Dou S-J, Jasperse L, Pan X, Wanger A, et al. Single nucleotide polymorphisms in genes associated with isoniazid resistance in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother [Internet]. 2003;47(4):1241–50. Available from: papers2://publication/uuid/F59DC034-E0BA-4A8D-8362-FA595F71A069

59. Hutchison CA. DNA sequencing: bench to bedside and beyond. Nucleic Acids Res [Internet]. 2007 [cited 2019 Aug 20];35(18):6227–37. Available from: http://www.ncbi.nlm.nih.gov/Genbank/

60. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-Genome Random Sequencing and Assembly of Haemophilus Influenzae. Science (80-) [Internet]. 1995 [cited 2019 Aug 20];269(5223):496–8. Available from: https://www-jstor-org.proxy-remote.galib.uga.edu/stable/pdf/2887657.pdf?refreqid=excelsior%3A0942f3fad5c5b7eb1224c548d6ce414a

61. Craig Venter J, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The Sequence of the Human Genome. Science (80-) [Internet]. 2001 [cited 2019 Aug 20];291(5507):1304–51. Available from: https://www-jstor-org.proxy-remote.galib.uga.edu/stable/pdf/3083494.pdf?refreqid=excelsior%3A19c528382dc069b1b01bfd10337766b4

62. NCBI. GenBank [Internet]. 2019. [cited 2019 Aug 20]. Available from: https://www.ncbi.nlm.nih.gov/genbank/

63. Guthrie JL, Gardy JL. A brief primer on genomic epidemiology: lessons learned from Mycobacterium tuberculosis. Ann N Y Acad Sci. 2017;1388(1):59–77.

64. Valdez MM, Clark JI, Wu GJS, Muchowski PJ. Functional similarities between the small heat shock proteins Mycobacterium tuberculosis HSP 16.3 and human ??B-crystallin. Eur J Biochem. 2002;269(7):1806–13.

65. Yuan Y, Crane DD, Barry CE. Stationary Phase-Associated Protein Expression in Mycobacterium tuberculosis: Function of the Mycobacterial ⎵-Crystallin Homolog. J Bacteriol. 1996;178(15):4484–92.

66. Siddiqui KF, Amir M, Gurram RK, Khan N, Arora A, Rajagopal K, et al. Latency-associated protein Acr1 impairs dendritic cell maturation and functionality: A possible mechanism of immune evasion by mycobacterium tuberculosis. J Infect Dis. 2014;209(9):1436–45.

67. Aaronson DS, Horvath CM. A Road Map for Those Who Don ' t Know JAK-STAT. Science (80-). 2002;296(5573):1653–5.

68. Chandel NS, Trzyna WC, McClintock DS, Schumacker PT. Role of Oxidants in NF- B Activation and TNF- Gene Transcription Induced by Hypoxia and Endotoxin. J Immunol [Internet]. 2000;165(2):1013–21. Available from: http://www.jimmunol.org/content/165/2/1013.full

69. Summers KC, Shen F, Potchanant EAS, Phipps EA, Hickey RJ, Malkas LH. Phosphorylation: The Molecular Switch of Double-Strand Break Repair. Int J Proteomics. 2011;8.

70. Preneta R, Papavinasasundaram KG, Cozzone AJ, Duclos B. Autophosphorylation of the 16 kDa and 70 kDa antigens (Hsp 16.3 and Hsp 70) of Mycobacterium tuberculosis. Microbiology. 2004;150(7):2135–41.

71. Drews SJ, Hung F, Av-Gay Y. A protein kinase inhibitor as an antimycobacterial agent. FEMS Microbiol Lett. 2001;205:369–74.

72. Yates TA, Khan PY, Knight GM, Taylor JG, Mchugh TD, Lipman M, et al. The transmission of Mycobacterium tuberculosis in high burden settings. Lancet Infect Dis [Internet]. 2016 [cited 2019 Dec 9];16(2):227–38. Available from: www.thelancet.com/infectionVol

73. Meertens RM, Mj Van De Gaar V, Spronken M, De Vries NK. Prevention praised, cure preferred: results of between-subjects experimental studies comparing (monetary) appreciation for preventive and curative interventions. BMC Med Inform Decis Mak [Internet]. 2013 [cited 2019 Dec 16];13(136). Available from: http://www.biomedcentral.com/1472-6947/13/136

74. WHO. Uganda Tuberculosis Profile 2018. 2019.

75. WHO. An Investigation of Household Contacts of Open Cases of Pulmonary Tuberculosis amongst the Kikuyu in Kiambu, Kenya. Bull World Heal Organ. 1961;25(6):831–50.

76. Guwatudde D, Nakakeeto M, Jones-Lopez EC, Maganda A, Chiunda A, Mugerwa RD, et al. Tuberculosis in Household Contacts of Infectious Cases in Kampala, Uganda. Am J Epidemiol. 2003;158(9):887–98.

77. Buu TN, Van Soolingen D, Huyen MNT, Lan NNT, Quy HT, Tiemersma EW, et al. Tuberculosis Acquired Outside of Households, Rural Vietnam. Emerg Infect Dis [Internet]. 2010 [cited 2019 Dec 9];16(9):1466–8. Available from: http://phil.cdc.gov/phil.

78. Classen CN, Warren R, Richardson M, Hauman JH, Gie RP, Ellis JHP, et al. Impact of social interactions in the community on the transmission of tuberculosis in a high incidence area. Thorax. 1999;54:136–40.

79. Yaganehdoost A, Graviss EA, Ross MW, Adams GJ, Ramaswamy S, Wanger A, et al. Complex Transmission Dynamics of Clonally Related Virulent Mycobacterium tuberculosis Associated with Barhopping by Predominantly Human Immunodeficiency Virus-Positive Gay Men. J Infect Dis. 1999;180(4):1245–51.

80. Verver S, Warren R, Munch Z, Richardson M, Van der spuy G, Borgdorff M, et al. Proportion of tuberculosis transmission that takes place in households in a high-incidence area. Lancet [Internet]. 2004 [cited 2019 Dec 9];363(9404):212–4. Available from: www.thelancet.com

81. Cavalcante SC, Durovni B, Barnes GL, A Souza FB, Silva RF, Barroso PF, et al. Community-randomized trial of enhanced DOTS for tuberculosis control in Rio de Janeiro, Brazil. Int J Tuberc Lung Dis. 2010;14(2):203–9.

82. Sharma K, Verma R, Advani J, Chatterjee O, Solanki HS, Sharma A, et al. Whole Genome Sequencing of Mycobacterium tuberculosis Isolates From Extrapulmonary Sites. Omi A J Integr Biol [Internet]. 2017 [cited 2019 Dec 9];21(7):412–25. Available from: http://phaster.ca

83. Sekandi JN, Zalwango S, Martinez L, Handel A, Kakaire R, Nkwata AK, et al. Four Degrees of Separation: Social Contacts and Health Providers Influence the Steps to Final Diagnosis of Active Tuberculosis Patients in Urban Uganda. BMC Infect Dis. 2015;

84. Pouseele H, Supply P. Accurate Whole Genome Sequencing Based Epidemiological Surveillance of Mycobacterium tuberculosis. Methods Microbiol. 2015;42:359–94.

85. UBOS. National Population and Housing Census 2014. National Population and Housing Census. 2017.

86. Sekandi JN, List J, Luzze H, Yin XP, Dobbin K, Corso PS, et al. Yield of undetected tuberculosis and human immunodeficiency virus coinfection from active case finding in urban Uganda. Int J Tuberc Lung Dis. 2014;18(6):754.

87. Talarico S, Durmaz R, Yang Z. Insertion- and deletion-associated genetic diversity of Mycobacterium tuberculosis phospholipase C-encoding genes among 106 clinical isolates from Turkey. J Clin Microbiol. 2005;43(2):533–8.

88. Kong D, Kunimoto D. Secretion of Human Interleukin 2 by Recombinant Mycobacterium bovis BCG. Infect Immun. 1995;63(3):799–803.

89. Bardarov S, Bardarov S, Pavelka M, Sambandamurthy V, Larsen M, Tufariello J, et al. Specialized transduction: An efficient method for generating marked and unmarked targeted gene disruptions in Mycobacterium tuberculosis, M. bovis BCG and M. smegmatis. Microbiology. 2002;148:3007–17.

90. Rodrigue S, Brodeur J, Jacques PÉ, Gervais AL, Brzezinski R, Gaudreau L. Identification of mycobacterial ?? factor binding sites by chromatin immunoprecipitation assays. J Bacteriol. 2007;189(5):1505–13.

91. Hu Y, Hu Y, Movahedzadeh F, Movahedzadeh F, Stoker NG, Stoker NG, et al. Deletion of the Mycobacterium tuberculosis ?-Crystallin-Like hspX Gene Causes Increased Bacterial Growth In Vivo. Infect Immun. 2006;74(2):861–8.

92. Jain P, Hsu T, Arai M, Biermann K, Thaler D, Nguyen A, et al. Specialized Transduction Designed for Precise High-Throughput Unmarked Deletions in Mycobacterium tuberculosis. MBio. 2014;5(3).

93. Swaim LE, Connolly LE, Volkman HE, Humbert O, Born DE, Ramakrishnan L. Mycobacterium marinum Infection of Adult Zebrafish Causes Caseating Granulomatous Tuberculosis and Is Moderated by Adaptive Immunity †. Infect Immun [Internet]. 2006 [cited 2020 Mar 17];74(11):6108–17. Available from: http://iai.asm.org/

94. Natale P, Brüser T, Driessen AJM. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-Distinct translocases and mechanisms. Biochim Biophys Acta - Biomembr. 2008 Sep;1778(9):1735–56.

95. Goosens VJ, Monteferrante CG, Van Dijl JM. The Tat system of Gram-positive bacteria. Biochim Biophys Acta - Mol Cell Res. 2014 Aug 1;1843(8):1698–706.

96. Hardie KR, Wells TJ, Linke D, Meuskens I, Saragliadis A, Leo JC. Type V Secretion Systems: An Overview of Passenger Domain Functions. Front Microbiol [Internet]. 2019 [cited 2020 Mar 11];10(1163). Available from: www.frontiersin.org

97. Bagos PG, Nikolaou EP, Liakopoulos TD, Tsirigos KD. Combined prediction of Tat and Sec signal peptides with hidden Markov models. Bioinformatics [Internet]. 2010 [cited 2020 Mar 11];26(22):2811–7. Available from: http://www.compgen.org/tools/PRED-TAT/.

98. Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: A retrospective observational study. Lancet Infect Dis. 2013 Feb;13(2):137–46.

99. Mougous JD, Gifford CA, Ramsdell TL, Mekalanos JJ. Threonine phosphorylation post-translationally regulates protein secretion in Pseudomonas aeruginosa. Nat Cell Biol. 2007;9(7):797–803.

100. Kruh-Garcia NA, Wolfe LM, Chaisson LH, Worodria WO, Nahid P, Schorey JS, et al.

Detection of Mycobacterium tuberculosis peptides in the exosomes of patients with active and latent M. tuberculosis infection using MRM-MS. PLoS One. 2014;9(7):1–11.