IDENTIFYING GENETIC FACTORS PROMOTING FITNESS AND RIFAMPICIN TOLERANCE IN MYCOBACTERIUM TUBERCULOSIS BIOFILMS

by

Jacob P. Richards

BS in Biological Sciences, University of Pittsburgh, 2010

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Jacob P. Richards

It was defended on

September 23, 2013

and approved by

Jeremy Martinson, DPhil, Assistant Professor, Department of Infectious Diseases and Microbiology, Graduate School of Public Health, University of Pittsburgh

Ronald C. Montelaro, PhD, Co-Director, Center for Vaccine Research, Professor, Department of Microbiology & Molecular Genetics, University of Pittsburgh Medical Center

Gerard J. Nau, MD, PhD, Assistant Professor, Center for Vaccine Research, Department Microbiology & Molecular Genetics, University of Pittsburgh Medical Center

Thesis Director: Anil K. Ojha, PhD, Assistant Professor, Department of Infectious Diseases and Microbiology, Graduate School of Public Health, University of Pittsburgh

Anil K. Ojha, PhD

IDENTIFYING GENETIC FACTORS PROMOTING FITNESS AND RIFAMPICIN TOLERANCE IN MYCOBACTERIUM TUBERCULOSIS BIOFILMS

Jacob P. Richards, MS

University of Pittsburgh, 2013

ABSTRACT

OBJECTIVE: In 2011, the World Health Organization reported 8.7 million new cases of tuberculosis and 1.4 million tuberculosis-related deaths worldwide. The causative agent of this disease *Mycobacterium tuberculosis* is a notoriously persistent pathogen whose treatment requires a 6-9 month course of multiple antibiotics to clear the infection. Here, we used next generation DNA sequencing to identify several genetic factors involved in the tolerance of *M. tuberculosis* to the antibiotic rifampicin in the context of the biofilm, a common bacterial stress survival strategy.

METHODS: A transposon insertion mutant library was constructed in the Erdman strain *of M. tuberculosis*. The library was then grown planktonically or in a biofilm and exposed to 5µg/ml of rifampicin. After exposure samples were processed and plated for colony formation. DNA was extracted from the colonies and prepared for sequencing by PCR amplification of transposon junction sites. DNA of these sites was sequenced using Illumina Hi-Seq technology and analyzed to find differential representation of transposon mutants between the unexposed and exposed library samples.

RESULTS: Biofilms of *M. tuberculosis* are more tolerant than planktonic samples *in vitro* to rifampicin treatment. Seven genes at eight genomic positions were found to contain insertion

sites that were ubiquitous in the samples unexposed to rifampicin, but found to be greatly decreased in the exposed biofilm samples: Rv0385, Rv1508c (at two positions), Rv1819c, Rv2779c, Rv3164c, Rv3796 and Rv3868.

DISCUSSION: Further analysis of these mutants by using knockouts and studying the effects *in vitro* or in a mouse model will be necessary to confirm their role in rifampicin tolerance in an actual infection and determine any possible clinical benefits in exploiting these genetic factors. As latent infection remains a significant problem in global public health, exploitation of targets that contribute the persistence of infection could be a valuable tool in clearing infections.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Firstly, I would like to thank my thesis advisor and principal investigator, Dr. Anil K. Ojha for his guidance and support while mentoring me throughout the duration of this project. I would also like to thank my committee members Dr. Jeremy Martinson, Dr. Ronald C. Montelaro and Dr. Gerard J. Nau.

For providing a protocol for the library preparation for transposon site sequencing and for answering questions pertaining to junction site sequencing, I would like to thank Dr. Chris Sassetti of the University of Massachusetts Medical School. Also from the Sassetti lab, I would like to thank Dr. Richard Baker for providing a short computer program for processing the raw fastq sequences and for answering many questions about computer programming for DNA sequencing data analysis.

I would also like to thank Dr. Albert Tai from Tufts University Genomics for his help with troubleshooting our sequencing data and for answering questions regarding sequencing results and data analysis. For helping with the shearing of DNA at University of Pittsburgh's Genomics and Proteomics Core Laboratories, I thank Bryan Thompson.

I would also like to thank the members of Dr. Ojha's laboratory, both past and present, for all their support and help since I joined the lab in fall of 2011: Dr. Mohammad Islam for training me on the biofilm project and his work on the beginning phases of this project for

acquiring data on drug tolerance of biofilms in the portable membrane-based model; Dr. Yong Yang for training me on various lab equipment, his insights on my research and for making sure I was prepared for my comprehensive examination; fellow students in the Ojha lab Jessica Battaglia, Aram Lee, Alexandra Bhatti, Jennifer Stock, Helen McGuirk, Joseph Thomas and Joseph Napora; Kathleen Kulka for excellent technical support, making media, assistance with ordering reagents, being responsible for making sure everything and everyone in the lab is functional and also talking about Pittsburgh Penguins hockey whenever a distraction from science was needed.

I must also thank Dr. Yvette Conley and Sandra Deslouches of the School of Nursing where I work as a research technician part-time for patiently allowing me to work around my graduate school project. I must also thank my parents Barb and Tom Richards, my sister Jennifer, my brother Zack, my girlfriend Melanie Panutsos and the rest of my family and friends for their continued encouragement and support throughout my graduate career. Lastly, my sincerest gratitude to my fellow Fall 2011 IDM students who over the course of this project I have seen become not only scientific colleagues, but amazing friends.

1.0 INTRODUCTION

1.1 GLOBAL DISEASE BURDEN OF TUBERCULOSIS

Tuberculosis (TB) in humans is caused by infection with the bacterium *Mycobacterium tuberculosis* (MTB). According to the most recently available data by the World Health Organization, there were 8.7 million newly reported cases of TB and 1.4 million TB-related deaths in 2011 [1]. The global burden of disease remains at an alarmingly high level; TB is the leading cause of death by infectious disease worldwide, and it has been estimated that one third of the world's total population has been infected. Multidrug-resistant tuberculosis (MDR-TB), defined as being resistant to the frontline antibiotics isoniazid and rifampicin, represents roughly five percent of active TB cases. There were approximately 60,000 new cases of MDR-TB reported in 2011, a number which has increased steadily over the past eight years. Eighty-four countries have also now reported at least one case of extensively drug-resistant tuberculosis (XDR-TB), defined as being resistant to both isoniazid and rifampicin as well as fluoroquinolones and at least one injectable second line antibiotic [2]. Human Immunodeficiency Virus (HIV) and TB co-infection also remains a substantial public health concern, especially in Africa where 80% of the total HIV-tuberculosis co-infections were reported. Approximately 400,000 of the 1.4 million TB-related deaths reported in 2011 involved co-infections with HIV [1]. Geographically, TB persists as a major public health threat

throughout much of the developing world including South Asia, Sub-Saharan Africa and parts of the Middle East, South America and Russia [3].

1.2 MTB: A PERSISTENT PATHOGEN

Although it is estimated that one third of the global population has been infected by MTB, only 5-10% of tuberculosis-infected individuals develop clinical disease symptoms. How these bacteria remain latent in the host is an important question in the study of the disease. Once an individual is infected with the bacterium, macrophages in the lung engulf the bacteria by phagocytosis, but bacterial immune evasion factors prevent the phagosome from fusing with the lysosome that can lyse the bacteria [4]. The pathogen cannot be cleared by either the innate immune response or the T-cell adaptive immune response that is instigated by dendritic cells after infection [5]. Usually this adaptive response is triggered by 4-6 weeks into the infection. In many cases, this is still not enough to clear the bacteria, and instead the bacteria become sequestered in the lung inside aggregates of host cells called granulomas. Granuloma architectures are typically comprised of an acellular center surrounded by large multi-nucleated macrophages, called Langhans giant cells, which are further surrounded by an outer ring of T-leukocytes [6]. Long-term persistence of MTB in this hostile environment of immunocompetent granluomas is considered an underlying factor in the development of asymptomatic infection.

Clearance of MTB takes 6-9 months of directly observed therapy-short course (DOTS), which is comprised of antibiotic treatment with rifampicin, isoniazid and usually ethambutol under the direct observation of a community healthcare worker. This multi-drug intervention was developed nearly forty years ago, and its usefulness is increasingly being challenged by the

development of MDR-TB. This is especially true in resource-limited regions where MDR-TB is a common result of incomplete treatment [7, 8]. Latent infections of MTB and the aggressive and lengthy treatment remains a major obstacle in treating this wicked problem in global health.

1.3 MYCOBACTERIA READILY FORM MULTICELLULAR STRUCTURES

Mycobacteria, when grown *in vitro* will readily form a pellicle at the air-liquid interface of the growth media. In fact, historically this has posed challenges in culturing strains of mycobacterial species *in vitro.* Since the 1950's, the solution to this problem has been to use Tween-80, a surfactant and emulsifier, to obtain a homogenous suspension of bacteria in culture without affecting virulence [9]. However when grown in this condition, certain biological characteristics of the bacteria are altered. Being neither truly Gram-positive nor Gram-negative, mycobacteria are unique for their waxy, impermeable and lipid-rich cell wall. When grown in the presence of detergent, the cellular lipids on the outer wall are altered to be more permeable to small molecules, including antibiotics [10]. Furthermore, bacteria grown in a homogenous suspension experience homogenous access to nutrients. In the macroscopic structures that these bacteria would form normally, a gradient of nutrient and oxygen access would create a subpopulation of cells at the core of the structure experiencing greater stress than cells dispersing off of the structures' outer edges. It has been shown that *M. tuberculosis* expresses a unique genetic profile under starvation conditions, and it is logical that the differing stress levels would produce phenotypically unique subpopulations that have adapted to tolerate the harsher conditions at the core of such structures [11]. The phenotypic diversity and biological adaptations that result from

the accumulation of bacteria in multicellular structures remains largely overlooked in the study of persistence of pathogenic mycobacteria.

1.4 BIOFILMS AS A DISTINCT BACTERIAL LIFESTYLE

Bill Costerton coined the term “biofilm” in 1978 to describe an aggregate of bacterial cells attached to a substratum [12]. Later, his group described microcolonies encapsulated in an extracellular matrix (ECM) of polysaccharides of *Pseudomonas aeruginosa* in lung tissue of cystic fibrosis patients [13]. Also, bacterial structures of *Staphylococcus aureus* attached to medical implants were discovered microscopically [14]. These studies provide the foundation of a branch of bacteriology that focuses on cells grown as a citizen of a sessile community rather than separated individuals. Biofilms consist of many cells aggregated into a highly structured community enclosed in ECM that contains fluid channels and even multiple bacterial species [15, 16].

Biofilm growth and development follows several specific stages. First, bacterial cells attach to a substratum. This is followed by non-motile, sessile growth and maturation of the biofilm through formation of architecture and ECM leading to stress gradients and phenotypic diversity. Finally, cells use quorum sensing and begin to disperse from the biofilm, to perhaps attach elsewhere and begin the process anew. Biofilm formation occurs through distinct genetic reprogramming of bacteria that occurs immediately after the attachment of a planktonic cell onto a substratum [17, 18]. In *Escherichia coli* for example, motility genes are suppressed by transcriptional regulation and up-regulating genes that produce components of the ECM [19]. Quorum sensing genes, a form of bacterial intercellular communication, were highly activated

while virulence genes were suppressed in biofilms of *Vibrio cholera* [20]. In *Bacillus subtilis*, a Gram-positive bacterium, another inverse correlation in gene regulation is observed during biofilm development. Regulators of ECM *SinR* and *SinI* negatively and positively regulate these genes respectively, while activation of sporulation genes negatively regulate biofilm formation by preventing sessile growth [21]. These data from gene expression and transcriptomic studies illustrates how bacteria forming biofilms and their distinct growth programs could give rise to heterogeneity in biofilms. These microenvironments present in separate portions of the architecture of the biofilm create phenotypic diversity across these subpopulations that is more tolerant to stress or antibiotic treatment.

1.5 BIOFILMS AS A METHOD OF ANTIBIOTIC TOLERANCE

As mentioned above, *P. aeruginosa* and *S. aureus*, two pathogenic bacteria, readily form biofilms in clinical settings. Pathogenic *E. coli* also form biofilms during infections of the urinary tract [22]. Across species, biofilms of pathogenic bacteria have proven to be both tolerant of antibiotic treatment as well as evasion of clearance by the host immune system [23, 24]. This makes effective treatment of bacterial infections caused by biofilm-forming bacteria a difficult prospect. *Staphylococci* species are 20-50 times more susceptible to antibiotic treatment in planktonic culture compared to biofilms [25]. This factor increases to planktonic cells being 100-1000 times more susceptible than biofilm samples in pathogenic *E. coli* and *P. auruginosa* [26]. This leads to a more aggressive antibiotic treatment of biofilm infections or coating medical implants with antimicrobial substances, although there remains an absence of treatment for dispersing bacteria from a biofilm prior to antibiotic treatment [27]. A recent study identified the

D-amino acids as a potent dispersal agent of pathogenic biofilms, although this is yet to be tested clinically or with other antibiotics [28].

1.6 POSSIBILITY OF CLINICAL MTB BIOFILMS

As mentioned above, mycobacteria will naturally form multicellular communities in culture. Biofilms of mycobacteria have been observed in both the environment and with pathogenic species in clinical settings [29, 30]. Environmental species of mycobacteria have been observed in showerheads as well as water supply systems [31, 32]. This is especially true in the case of the opportunistic pathogen species *M. avium.* In the model non-pathogenic, fast-growing and non-pathogenic model species *M. smegmatis*, it has been shown that during *in vitro* biofilm development glycopeptidolipids (GPL) are required for attachment and sliding motility [33]. Two other surface components necessary for biofilm development in *M. smegmatis* are free mycolic acids and mycolyl-diacylglycerol [34, 35]. Mycolyl-diacylglycerol is suspected to be key in the early attachment phase of biofilm development. This suggests cellular lipids play an important role in biofilm development in mycobacteria. A nucleoid-associated protein and suspected transcription regulator, *lsr2*, contributes to mycolic acid synthesis and biofilm development [36]. A regulator of genes involved in ECM synthesis would be in line with the distinct genetic expression profiles seen during biofilm development in other species of bacteria. Similar results of control of biofilm development at the transcriptional level is further demonstrated by the induction of 82 genes during biofilm maturation in *M. smegmatis* [37].

Growth of MTB *in vitro* without Tween-80 also leads to development of mature biofilm structures rich in free mycolic acid content [34]. Three genes, *pks16, pks1* and *helY* have been

linked to biofilm development in MTB by the failure of their mutants to form mature biofilms, yet grow normally in planktonic culture [34, 38]. Biofilms of MTB and *M. smegmatis* grown *in vitro* also harbor populations of drug-tolerant cells compared to biofilm-deficient mutants or planktonic cultures. These findings support the distinct and multi-phasic growth pattern of biofilms in other species of bacteria also occur in mycobacteria. Also, *in vitro*, MTB has proven to be highly tolerant to antibiotic treatment much like other pathogenic bacteria that form biofilms clinically. These results set the stage for the exploration of whether or not clinical biofilms of MTB could harbor a reservoir of drug-tolerant persisters in chronic and latent infections.

1.7 USING TRANSPOSON JUNCTION SEQUENCING AS A METHOD OF DISCOVERING GENES CONTRIBUTING TO RIFAMPICIN TOLERANCE

MTB has typically not been considered a very genetically tractable organism, however since the late 1990's transposon insertion mutagenesis has proven to be a useful tool in elucidating a comprehensive view of genetic requirements for growth [39]. This approach was largely developed by Eric Rubin and Chris Sassetti. A highly active mariner transposon-based bacteriophage, ΦMycoMarT7, inserts itself into genomic DNA randomly at any "TA" site in the sequence. There are 74,417 of these sites in MTB (Erdman). This is used to create a large pool of mutants in colonies that can be harvested and stored in freezer stocks as a mutant library. This mutant library is then grown under a given growth condition. Bacteria from each condition are then plated on media and colonies are harvested for DNA extraction. Junction sites of transposon insertion with genomic DNA are amplified by PCR using a primer with transposon homology

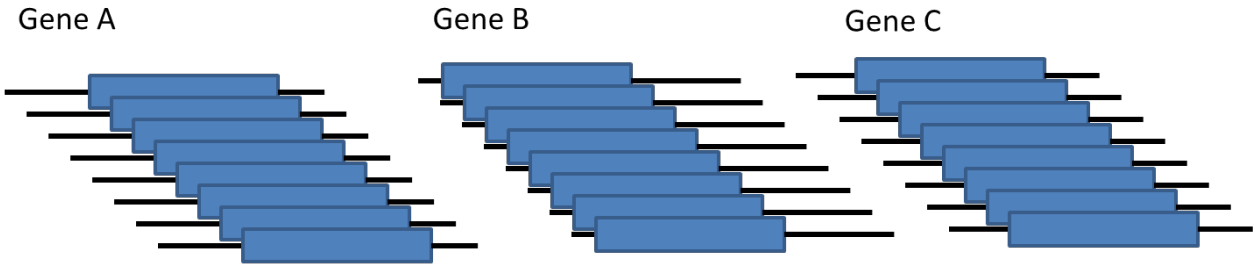
(see Methods below). Originally, this approach used microarray hybridization and was known as TraSH for transposon site hybridization [40]. More recently, next generation DNA sequencing of the amplified transposon junction sites in conjunction with mapping to a reference genome has become more commonly utilized (called Tn-Seq) [41]. This method has been used to identify genes in MTB essential for *in vitro* growth, growth in mice and macrophages and cholesterol catabolism. Here, we utilize this approach in the context of *in vitro* MTB biofilms when exposed to the frontline antibiotic rifampicin.

2.0 SPECIFIC AIMS

In the context of the biofilm, bacteria at different levels of the structure express different levels of stress. For example, those at the surface with freer access to oxygen and nutrients will experience considerably less stress than the bacteria attached to the substratum at the bottom of the biofilm. The response to this stress, at the genetic level, could be the source of the phenotypic persistence to antibiotics observed during biofilm growth. By making a transposon mutant library of MTB, growing it in biofilms (as well as comparing it to planktonically grown samples), exposing the bacteria to rifampicin and sequencing colony DNA at transposon junction sites, our aim is to elucidate genetic factors that contribute to rifampicin tolerance in the context of a multicellular community of MTB. Our approach involves negative selection of non-essential genes by looking for insertion sites that are not represented or underrepresented in the mutant library for the rifampicin-treated samples as compared to the untreated samples (Figure 1).

We predict that biofilms of the MTB mutant library challenged with brief exposure to rifampicin will identify genes which are most important for rifampicin-tolerance through their loss of representation in the sequencing data after exposure to the antibiotic. A decrease in the number of reads in the 24-hour, rifampicin-exposed samples at genomic sites where read numbers were high in the unexposed samples will be genes targeted in future studies for contributing to fitness and rifampicin tolerance in biofilms.

Transposon insertions into three different genes grown in biofilm condition—unexposed to rifampicin, followed by extraction of DNA, library preparation and amplified by PCR.



If same sample is exposed to RIF, amplification of “Gene C” disappears—this suggests “Gene C” is necessary for tolerance to RIF in a biofilm.

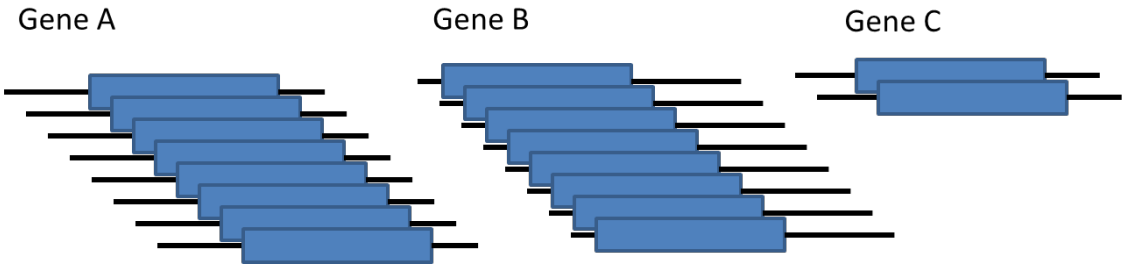


Figure 1. Demonstration of the determination of genes related to rifampicin tolerance by negative selection.

3.0 METHODS

3.1 A SIMPLIFIED BIOFILM MODEL

To begin to study the behavior of MTB within the context of a natural biofilm, a model was developed to grow the bacteria after attachment to a substratum and yet still be easily manipulated and handled for purposes of experimentation. This was done by first growing a culture of MTB (Erdman strain) in Middlebrook 7H9 media, Oleic Acid, Albumin, Dextrose Complex (OADC) and tween80 to an Optical Density (OD) at 600nm of 0.8-1.0 and then pipetting 10µl from the bacterial culture onto a 13mm in diameter polycarbonate membrane placed on top of a stack of cardstock strips inside of a Petri dish. After inoculation, membranes were allowed to dry for 45-60 minutes. 25ml of 7H9 and OADC were added to the Petri dish and absorbed by the cardstock up to the membranes through capillary action. The biofilm cultures were then incubated at 37ºC while the bacteria grew into mature biofilms for 18 days. Growth media was replenished as needed. Biofilms were processed by placing them into 15ml conical tubes containing 5ml of 1xPBS followed by mixing by vortex, sonication, a second vortexing and plating (Figure 2). All sonication (with the exception of the sonicator used for DNA shearing) was performed for ten minutes at 25ºC on the Branson 1510 model.

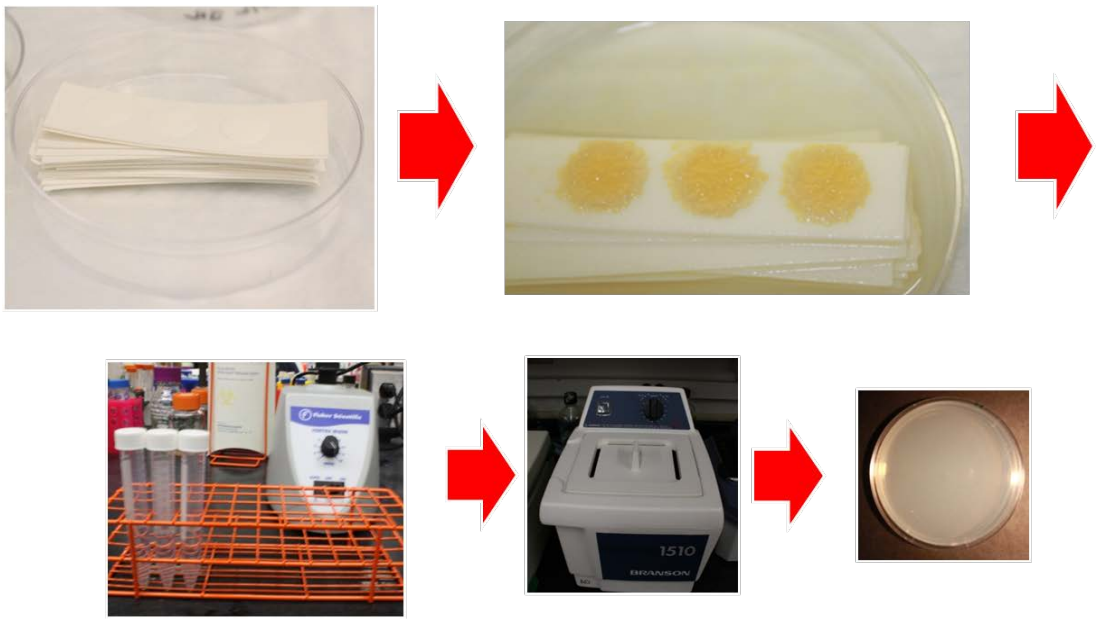


Figure 2. Flow chart of membrane-attached biofilm model.

3.2 TESTING DRUG TOLERANCE IN BIOFILMS

To test drug tolerance of the MTB (Erdman) in biofilms, mature biofilms were grown as described above and then exposed to 5µg/ml of rifampicin for a period of 6 days. Biofilms were exposed to rifampicin by moving the top cardstock with the membranes to a new Petri dish containing a fresh stack of cardstock and 5µg/ml rifampicin in 25ml of 1xPBS. Unexposed biofilms were transferred to cardstock in a Petri dish containing 25ml of only 1xPBS. Following exposure, membranes were removed from the cardstock and placed into a 15ml conical tube containing 5ml of 1xPBS. The tubes were vortexed until visible bacteria were dislodged from the membrane, and then tubes were sonicated for 10 minutes. After sonication, the membranes were vortexed for another two minutes. The bacteria from the membranes were then centrifuged for 10 minutes at 3488xg and 22ºC to wash any antibiotic off of the cells. Supernatant was

discarded, 5ml of fresh 1xPBS was added to the tubes, and the bacteria were re-suspended. At this point, the OD_{600} of the samples was measured. Serial dilutions were made and 10µl were pipetted at several dilutions onto 7H11 agar plates containing OADC plates for colony counts. For comparative analysis, planktonic samples of MTB were grown to an equal OD_{600} of the processed biofilm samples (described above) in 7H9 and OADC and subsequently underwent identical treatment as the biofilm samples for 5µg/ml rifampicin in 1xPBS exposure, vortexing, sonication, additional vortexing, centrifugation, washing and re-suspension in 1xPBS. The planktonic samples were diluted identically to the biofilm samples, and 10µl were pipetted onto 7H11 and OADC plates for colony counts.

3.3 TESTING RIFAMPICIN TOLERANCE OF BIOFILMS OVER 24 HOURS

To acquire a timeframe of rifampicin exposure of mutants that would only see the weakest mutants die off from antibiotic killing, the above protocol was followed for biofilm samples of MTB (Erdman) and MTB $mc^2$7000 (an attenuated strain), but with 5µg/ml of rifampicin exposure only lasting 24 hours before biofilm processing and subsequent plating for counting colonies and calculating percent survival.

3.4 CONSTRUCTION OF THE TRANSPOSON MUTANT LIBRARY

A planktonic culture of MTB (Erdman) was grown in 7H9, OADC and tween80 to an OD_{600} of 1.0. This culture was then sub-cultured into ten bottles containing 1ml from the original culture

grown in 25ml of 7H9, OADC and tween80 to OD_{600} 0.8-1.0. The MTB stocks were then split into two sets of four 50ml conical tubes containing approximately 30ml each. The 50ml conical tubes were centrifuged at 2348xg for 5 minutes at 22ºC. Supernatant was removed and the pellet was washed and re-suspended in 20ml 1xPBS. The bacteria were centrifuged again under the same conditions and re-suspended in 5ml of warm 7H9 and OADC. The cultures were then each infected with 1ml of the temperature sensitive ΦMycoMarT7 mycobacteriophage with a titer of 10^{11} and incubated for four hours at 37ºC. The phage contains a *himar1* mariner-based transposon that can insert into any "TA" site into the MTB genome. After incubation, each of the phage-transduced cultures were plated with one ml from each culture (5 plates from each tube, totaling 40 plates) on 15cm plates containing Middlebrook 7H11 agar containing OADC, tween80 with 20µg/ml of Kanamycin and incubated at 37ºC for 21 days. Approximately 12,000 colonies of the transduced MTB were collected by pipetting a 2 to 5 milliliters of 7H9 and tween80 onto the 15cm 7H11, OADC, tween80 and Kanamycin plates and scraping the colonies off into the media [42]. From this, glycerol stocks of the mutant library were collected to be grown in culture.

3.5 GROWTH CONDITIONS AND RIFAMPICIN EXPOSURE OF THE MUTANT LIBRARY

The MTB transposon mutant library was grown in 7H9, OADC and Kanamycin to an OD_{600} of 0.8-1.0. From there, the culture was either split into a new planktonic culture or 10µl were inoculated onto polycarbonate membrane for biofilm growth as stated above. Planktonic samples were grown to an OD_{600} of 0.8-1.0 and exposed to 5µg/ml rifampicin for 24 hours or left

unexposed to antibiotic, processed as stated above, plated on 15cm 7H11, OADC tween80 and Kanamycin and incubated for 21 days at 37ºC. Biofilm samples were grown to maturity at 18 days, and then samples were exposed to 5µg/ml rifampicin for 24 hours or left unexposed to antibiotic and processed as stated above. After processing, cells from the biofilm samples were plated on 15cm 7H11, OADC, tween80 and Kanamycin and incubated at 37ºC for 21 days.

3.6 EXTRACTION OF MUTANT GENOMIC DNA

Colonies were harvested from the 15cm plates by pipetting 2 to 5 milliliters of 7H9 and tween80 onto the plates, then scraping off colonies and collecting them into the 7H9 and tween80 media [42]. The colonies and media were centrifuged at 2348xg for 10 minutes at 22ºC. The supernatant was discarded and the pellet was re-suspended in 5ml 10mM Tris-HCl, 1 mM EDTA at a pH of 9. The re-suspended cells were mixed with an equal volume of chloroform and methanol in 2:1 ratio and rocked for 5 minutes on a shaker. The suspension was centrifuged at 3488xg for 10 minutes at 22ºC. Both the aqueous and the organic phases were removed from the 50mL conical tube. The solid bacterial mass was dried by leaving the tube open in the biosafety cabinet for 3 hours. Ten milliliters of TE containing 0.1M Tris-HCl at a pH of 9 was added to the pellet. The bacteria were then re-suspended by vortexing. A one hundredth volume of 10mg/ml lysozyme was added, and the cells were incubated overnight at 37ºC. One milliliter of 10% Sodium dodecyl sulfate (SDS) solution was added to the incubated cells. Proteinase K was added to the cells to a final concentration of 100g/mL and mixed via vortexing. The samples were incubated at 50ºC for 3 hours. The viscous solution was transferred into a clean tube containing an equal volume of phenol and chloroform in an equal 1:1 ratio. After mixing, the cells were left

to stand for 30 minutes. Then, the cells were rocked on a shaker for 30 minutes at, followed by centrifugation at 12000xg for 15 minutes at 22ºC. The upper aqueous phase was pipetted off to a new tube with an equal volume of chloroform and the centrifugation was repeated. Again, the upper aqueous phase was removed and pipetted into a new tube with an equal volume of isopropanol and a one tenth volume of 3M sodium acetate at a pH of 5.2. The DNA was spooled out and washed with 70% ethanol, and dissolved in 0.5 to 1 mL TE.

3.7 PREPARATION OF DNA LIBRARY FOR HI-SEQ 2500

DNA from untreated planktonic cultures, untreated biofilm cultures and biofilm cultures exposed to rifampicin for 24 hours were prepared for DNA sequencing. The genomic library preparation was adapted from a protocol used for transposon site sequencing by the Sassetti Laboratory at the University of Massachusetts [40]. The entire process of preparing the DNA library from shearing to amplification is depicted at the sequence level (Figure 3). To prepare the extracted genomic DNA from the transposon mutant library for sequencing at the transposon junction site, the DNA first had to be sheared into segments ranging 400-600 base pairs in length. This was done on a Covaris S2 Focused-ultrasonicator at the University of Pittsburgh Genomics and Proteomics Core Laboratories. To obtain the target base pair peak of 500bp, 130µl of each genomic DNA sample was placed in the sonicator set to an intensity of 3, 5% duty cycle, 200 cycles per burst for 90 seconds at 7ºC with a water level of 12. The DNA was then run on a 1% agarose gel for 60 minutes at 100 volts, then the 400-600bp range of the smear was excised, gel extracted and column purified using Qiagen Gel Extraction kit to the manufacturer’s instructions.

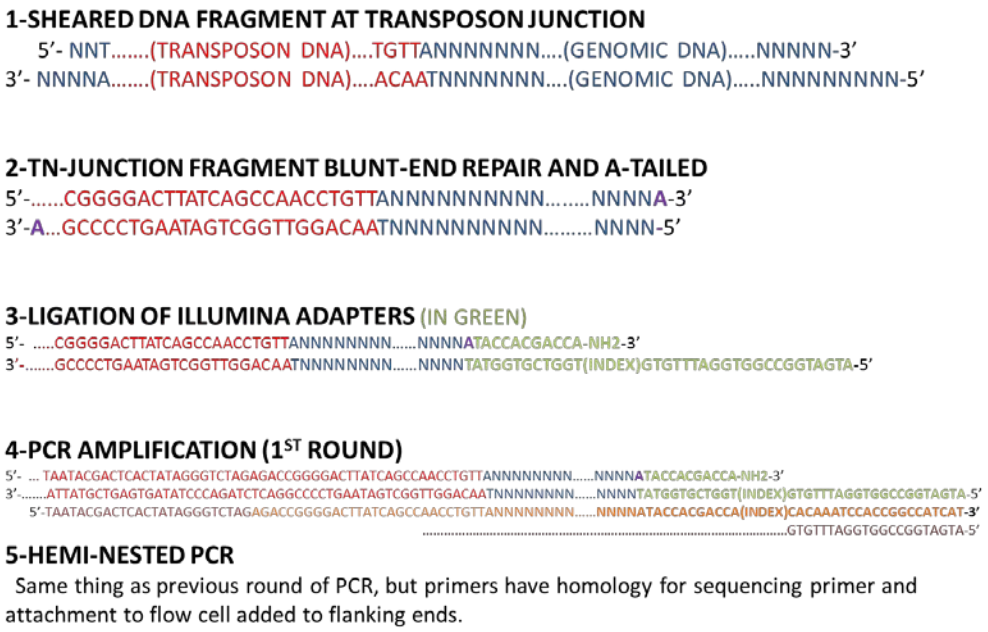


Figure 3. Preparation of DNA library for sequencing.

The sheared DNA was blunt ended using Epicentre's End-It DNA end repair kit according to the manufacturer's instructions. An adenosine nucleotide was added to the 3' end of each strand of blunt ended DNA by adding 5µl of Invitrogen 10x PCR buffer, 10µl of 100mM dATP and 3µl of Invitrogen Taq polymerase to each DNA sample and then incubating at 72ºC for 45 minutes. The DNA was then purified by using the Qiagen enzyme reaction clean up kit following the manufacturer's instructions and eluted into 50µl of distilled water after column purification.

To allow adapters to hybridize and prepare them for ligation to the genomic DNA fragments, 48µl from a 100µM stock of each adapter and 4µl of 50mM $MgCl_2$ were mixed into a microcentrifuge tube and heated in a 95ºC water bath for ten minutes and were then cooled to room temperature. The adapter sequences were 5'-TACCACGACCA-NH_2-3' and 5'-ATGATGGCCGGTGGATTTGTGTGGTCGTGGTAT-3'. For each ligation reaction, 1µl DNA ligase, 1.5µl DNA ligase buffer and 1µl of the hybridized adapter mix were added to 11.5µl of

the A-tailed DNA samples and incubated at room temperature for an hour. Then, the reaction was spiked with 1µl DNA ligase, 1µl DNA ligase buffer and 8µl water and incubated at room temperature for two more hours. The Qiagen enzyme reaction clean up kit was used as per the manufacturer's instructions with 3 additional washes of the DNA with PE buffer before elution into 100µl of water.

Next, genomic DNA fragments were amplified to enrich transposon junction sites by PCR amplification using a primer with transposon homology (5'-TAATACGACTCACTATAGGGTCTAGAG-3') and a primer with homology to the adapter sequence (5'-ATGATGGGCCGGTGGATTTGTG-3'). The reaction mix for the PCR was 100ng of DNA, 5µl of 10x PCR buffer, 1.5µl 50mM $MgCl_2$, 2.5µl of 5mM dNTPs, 2.5µl DMSO, 0.5µl Taq polymerase and 1.5µl of each primer from a 10µM stock. Water was added to each reaction to bring it up to a volume of 42µl. PCR parameters were 95ºC for 10 minutes; 20 cycles of 95ºC for 30 seconds, 58ºC for 30 seconds and 72ºC for 45 seconds; and 72ºC for 5 minutes. The amplified DNA was visualized on a 2% agarose gel run at 60 volts for 140 minutes. Smears of amplified DNA in the 400-600bp range were excised, placed in the negative 20ºC freezer for 30 minutes and then gel extracted with the Qiagen Gel Extraction kit and eluted into 50µl of water. A quality control PCR was also run on first PCR product using the same parameters with a primer with transposon homology downstream of the original primer (5'-GACTTATCAGCCAACCTGTTA-3') and the same adapter primer as the first round.

The last process in library preparation was further amplification of transposon junction site DNA with staggered primers with homology to the Illumina Tru-Seq adapter as well as transposon homology (Table 1) and primers with Illumina Tru-Seq adapter homology, a six-base long Illumina Index bar code for multiplexing and homology to the original adapters (Table 2).

Table 1. Staggered PCR primers with Transposon and Illumina Tru-Seq adapter homology used in hemi-nested PCR for sequencing preparation.

	Primer Sequences
1	5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTCGGGGACTTATCAGCCAACC 3'
2	5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTTCGGGGACTTATCAGCCAACC 3'
3	5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTGATACGGGGACTTATCAGCCAACC 3'
4	5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTATCTACGGGGACTTATCAGCCAACC 3'

Table 2. Primers with adapter homology and Illumina barcode sequence for multiplexing

	Sample	Primer Sequences	Barcode Sequence
1	Untreated planktonic	5'CAAGCAGAAGACGGCATACGAGATATCACGGTG ACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTCA ATGATGGCCGGTGGATTTGTG 3'	CGTGAT
2	Untreated biofilm 1	5'CAAGCAGAAGACGGCATACGAGATCGATGTGTG ACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTCA ATGATGGCCGGTGGATTTGTG 3'	ACATCG
3	Untreated biofilm 2	5'CAAGCAGAAGACGGCATACGAGATTTAGGCGTG ACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTCA ATGATGGCCGGTGGATTTGTG 3'	GCCTAA
4	24 hour RIF exposure 1	5'CAAGCAGAAGACGGCATACGAGATTAGCTTGTG ACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTCA ATGATGGCCGGTGGATTTGTG 3'	AAGCTA
5	24 hour RIF exposure 1	5'CAAGCAGAAGACGGCATACGAGATGGCTACGTG ACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTCA ATGATGGCCGGTGGATTTGTG 3'	GTAGCC
6	24 hour RIF exposure 1	5'CAAGCAGAAGACGGCATACGAGATCTTGTAGTG ACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTCA ATGATGGCCGGTGGATTTGTG 3'	TACAAG

The staggered primers were mixed together in equal concentration prior to being added to the PCR mix. The reaction mix for hemi-nested PCR included 6µl of amplified DNA product from the first round of PCR, 2µl of 10x PCR buffer, 0.5µl of 50mM $MgCl_2$, 0.5µl of dNTPs, 0.2µl Taq

polymerase, 0.2µl of the staggered transposon primer mix, 0.2µl of the adapter barcoded primer and 10.4µl of water for a final reaction volume of 20µl. PCR parameters were 95ºC for 5 minutes; 10 cycles of 95ºC for 30 seconds, 58ºC for 30 seconds and 72ºC for 45 seconds; and 72ºC for 5 minutes. Amplified DNA was run on a 1% agarose gel at 90 volts for 60 minutes. The 400-600bp smear was excised and extracted using the Qiagen Gel Extraction kit as per the manufacturer's instructions.

3.8 ILLUMINA HI-SEQ 2500 DNA SEQUENCING

After PCR amplification of transposon junction sites, 20ng of each sample of prepared DNA at a volume of 20nM was sent to Tufts University Core Facilities for Illumina Hi-Seq 2500 DNA sequencing [43]. The universal Illumina sequencing primer (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3') was used for single-end sequencing of the samples from the transposon end.

The single stranded DNA fragments bind randomly to the surface of the flow cell channel, then added nucleotides facilitate bridge amplification of the input prepared DNA into double-stranded amplified fragments on the solid flow cell surface. The DNA is denatured again leaving a template for complete amplification of up to 40 million clusters of DNA per flow cell. Then, the actual sequencing process begins with four labeled reversible terminators, the sequencing primer, the index primer and DNA polymerase. The sequencing primer binds to the indexed adapter and fluorescently-labeled bases are added individually and the result is recorded by a camera. The index primer has adapter homology up to the six-base bar code sequence,

which is then read by the machine to differentiate between the individuals samples present in the multiplex. Ideally, this process can yield up to 200 million reads.

3.9 ANALYSIS OF SEQUENCING DATA

A combination of small perl-based computer programs, the open-source, browser-based Galaxy project bioinformatics tool and the Broad Institute's Integrated Genome Viewer were used to analyze the results of DNA sequencing [44-46]. A small computer script provided by Dr. Richard Baker was used to check that the raw fastq files were checked to contain the "TGTTA" transposon junction site and then cleave off the beginning bases which contained transposon homology leaving the "TA" sequence where the transposon originally inserted itself into MTB' genome during mutant library construction [41]. These processed fastq sequences were groomed using the Galaxy browser and quality statistics for the reads were calculated [47]. Also using Galaxy browser system, the fastq reads were mapped to the MTB (Erdman) genome using Bowtie [48] tool. Galaxy was then also used to convert mapped sam files to binary "bam" files for graphical viewing [49]. A second computer script was used to count the number of reads for each sample at each possible "TA" site across the entire genome. Mapped reads were viewed graphically along the genome in the Integrated Genome Viewer. Normalizations and differentiation of sequences were calculated using Microsoft Excel.

3.10 QUANTITATIVE PCR AND ANALYSIS

To attempt to confirm the presence of rifampicin-sensitive mutants observed from the sequencing data in the genomic DNA, quantitative PCR was performed on the DNA samples from the unexposed and rifampicin-exposed biofilms and planktonic cultures. A single transposon-based primer (5'-CGACTCACTATAGGGTCTAGAGAC-3') was used as a forward primer in all reactions. The reverse primers for each individual gene, as well as two endogenous control sequences that had mapped insertions across all samples, had sequence homology with a downstream sequence of the gene after transposon insertion (Table 3). DNA input was normalized across samples by viewing genomic aliquots on a 1% agarose gel coupled with spectrophotometer readings. Each reaction contained 1µl DNA, 6.25µl SYBR Green Master Mix, 4µl water and 0.625µl of each primer for a final reaction volume of 12.5µl.

Table 3. Reverse primers used in quantitative PCR for each gene

Position of insertion	Erdman gene #	H37rv ortholog gene #	Primer Sequences
462570	E_0425	Rv0385	5' GAGCTTAGAATCGAGGTCCG 3'
1691647	E_1681	Rv1508c	5' CTTGCCCGATTGATAGTTCTTG 3'
2055036	E_2009	Rv1819c	5' GGCGGTGAAGATATCGACG 3'
3073068	E_3045	Rv2779c	5' GTCGGCATTGAGGTTTTCG 3'
3521539	E_3465	Rv3164c	5' GGAGAGGTAGCGCAGTTC 3'
4231539	E_4162	Rv3796	5' ACAGAGCGCCAAAGATGAG 3'
4324565	E_4241	Rv3868	5' GGCTGTACGGCGATATCAC 3'
3072909			5' GTGATCAGCCCGTATACCAG 3'
3787076			5' GGCAAACCCGAATCAATGG 3'

4.0 RESULTS

4.1 THE PORTABLE BIOFILM MODEL AND DRUG-TOLERANT BACTERIA

By growing MTB on a membrane, we were able to both manipulate the growth conditions of the bacteria, but also ensure that all bacteria were processed similarly, and all bacteria harvested were part of a biofilm population. In the past, our lab has assayed biofilms by using pellicles grown at the air-liquid interface [50], but this method often includes cells that are submerged in the liquid or settled at the bottom of the dish. The membrane-attached method allows for growth and manipulation of a biofilm attached to a solid substrate and processing of all the biofilms' cells without accidental inclusion of suspended cells.

Percent survival of cells grown in the membrane-attached biofilm method revealed biofilm cells to contain more persisters than their planktonic counterparts after exposure to the antibiotic rifampicin (Figure 4). MTB (Erdman) biofilms had a mean percent survival of 0.295% percent in the 18-day biofilms compared to 0.006% percent in the planktonic samples. MTB is 48 times more tolerant in biofilms than in planktonic culture. The student's t-test calculation resulted in a test statistic of 3.31 for a p-value of <0.05, but slightly greater than 0.01 (a test statistic of 3.36 would result in a p-value of 0.01). Mean percent survival was calculated for planktonic (without tween-80) and 18-day biofilms of MTB after 6-day exposure to 5µg/ml rifampicin. Bacteria concentration was normalized by measuring OD_{600} prior to plating. Each

bar is the result of 3 independent experiments performed in triplicate with standard error bars showing range between samples. These experiments were performed in collaboration with a former post-doctoral fellow in the Ojha laboratory, Dr. Mohammad Islam.

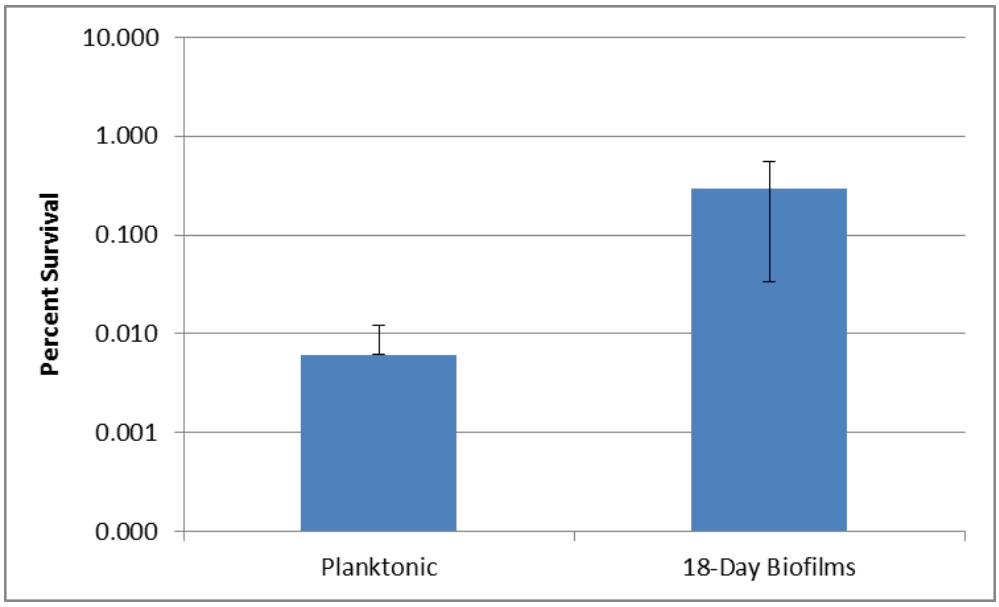


Figure 4. MTB mean percent survival to rifampicin in biofilm and planktonic culture.

To obtain a clear picture of how long we could expose biofilms of MTB to 5µg/ml of rifampicin and only lose the most rifampicin-sensitive mutants from the population, 18-day biofilms of MTB (Erdman) and the attenuated $mc^2$7000 strain were exposed to 5µg/ml of rifampicin for a 24 hour period. This time period was revealed to kill off only a small percentage of the biofilm cells with a mean 88.57 percent survival in the $mc^2$7000 strain and 98.48 percent survival in the Erdman strain (Figure 5). The mean percent survival was calculated from three replicate experiments with standard error bars. This high percentage survival ensures that any loss of mutants through killing in the 24-hour exposure window represents the most rifampicin-sensitive mutants in the library and not a loss from normal rifampicin treatment.

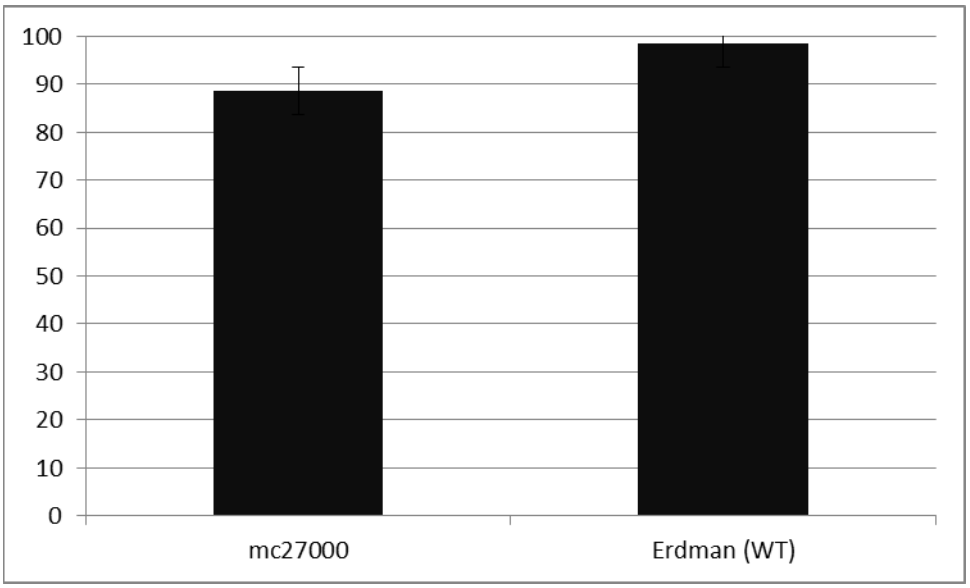


Figure 5. Mean percent survival after 18-day biofilms were exposed to 5µg/ml rifampicin for a 24 hour period.

4.2 CONSTRUCTION OF THE MUTANT LIBRARY AND AMPLIFICATION OF TRANSPOSON JUNCTION SITES

After infection with the ΦMycoMarT7 phasmid, up to 40 7H11, OADC, Kanamycin and twee80 plates were incubated for 21 days resulting in a mutant library consisting of approximately 12,000 colonies of clones carrying the transposon inserted into their DNA. This represents a limited, but still significant number of clones in the mutant library. The MTB (Erdman) genome has potentially 74,417 possible "TA" insertion sites, and an ideal library would consist of clones at least 2 to 3 times that number (the 200,000-300,000 range). The 12,000 clone library was used for the rifampicin-exposed biofilms and planktonic samples.

The yield from genomic DNA extraction measured with a spectrophotometer varied from sample to sample as can be seen in the table below (Table 4). The measure of absorbance of

A260:A280 showed most samples still contained some protein contamination (A 100% nucleic acid preparation will have an A260:A280 of 1.8). Also, four of the samples had an A260:A230 under 2.0, which is most likely the result of residual phenol in the ethanol precipitation. These contaminations were reduced upon subsequent column purification following the shearing of DNA for library preparation for sequencing. Because protein contamination contributes little to the absorbance, it is still unlikely that this affected the readings of the DNA concentrations by a spectrophotometer.

Table 4. Genomic DNA concentrations and absorbance by nanodrop after extraction.

Sample	DNA concentration (ng/µl)	A260:A280	A260:A230
Planktonic	80	1.524	1.882
0hr Biofilm 1	1000	1.802	2.116
0hr Biofilm 2	205	1.708	1.783
24hr Biofilm 1	205	1.745	2.000
24hr Biofilm 2	85	1.360	1.308
24hr Biofilm 3	335	1.558	1.861

After DNA extraction, genomic DNA was sheared into fragments in the 400-600bp range. While the Covaris sonicator yielded a wide range of sheared DNA, it did successfully shear away all of the whole genomic DNA (as can be seen on the 1% agarose gel image) and only the DNA in the desired range was excised and gel extracted. (Figure 6).

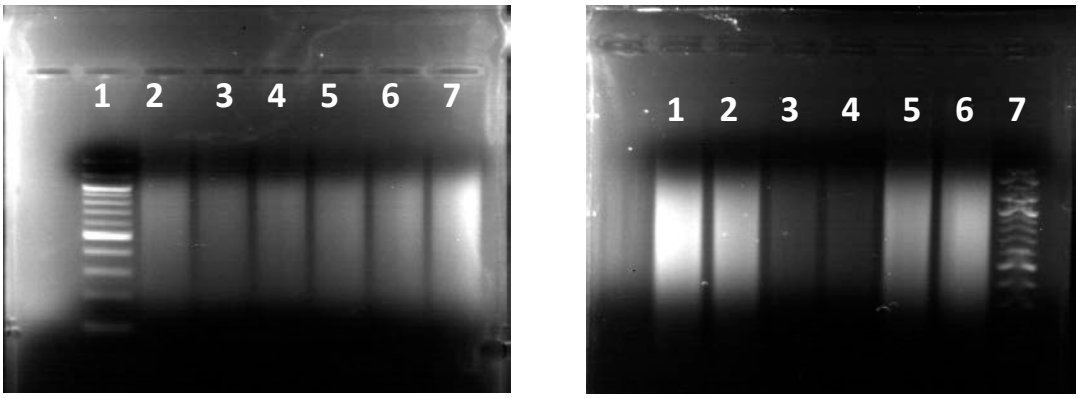


Figure 6. Genomic DNA after shearing on a 1% agarose gel.

In the gel image, lane 1 is a 100bp ladder, lanes 2 and 3 are from the unexposed planktonic sample, lanes (4 and 5) and (6 and 7) are from two separate unexposed biofilm samples. In the gel on the right, lanes (1 and2), (3 and 4) and (5 and 6) are each sheared DNA from 3 separate biofilm samples that were exposed to rifampicin for 24 hours and lane 7 is a 100bp ladder. After gel extraction, the sheared mutant library DNA from all samples was, blunt ended, A-tailed and Illumina adapters were ligated to the sheared fragments before PCR amplification of the transposon junction site in the DNA (See Methods). The first round of PCR used short primers to amplify transposon junction sites (Figure 7).

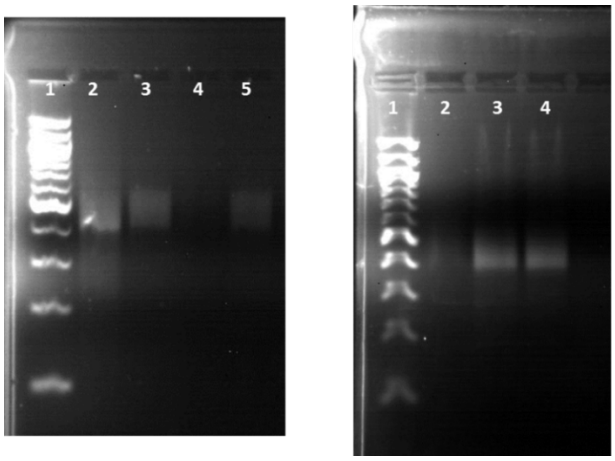


Figure 7. First round of PCR with short primers.

In the gel image on the left, lane 1 is a 100bp ladder, lane 2 is the unexposed planktonic sample, lane 3 is the first unexposed biofilm sample, lane 4 is a failed and unused reaction and lane 5 is the second unexposed biofilm sample. In the gel image on the right, lane one is a 100bp ladder, lane 2 is the first 24-hour rifampicin-exposed sample, lane 3 is the second, and lane 4 is the third. DNA from this PCR reaction was excised, gel extracted, column purified and then the 2nd round of PCR was performed on the amplified product to yield amplified transposon junction sites flanked by sequences with homology to Illumina sequencing and index primers (Figure 8). This was the DNA sent to Tufts University Core Facility for Illumina Hi-Seq 2500. In the gel image on the left lane 1 is a 100bp ladder, lane 2 is a reaction with water instead of DNA, lane 3 is the unexposed planktonic sample, lane 4 is the first unexposed biofilm sample and lane 5 is the second unexposed biofilm sample. In the gel image on the right, lane 1 is a 100bp ladder, lane 2 is a reaction with water rather the DNA, lanes 3, 4 and 5 are the first, second and third 24-hour rifampicin-exposed biofilm samples, respectively.

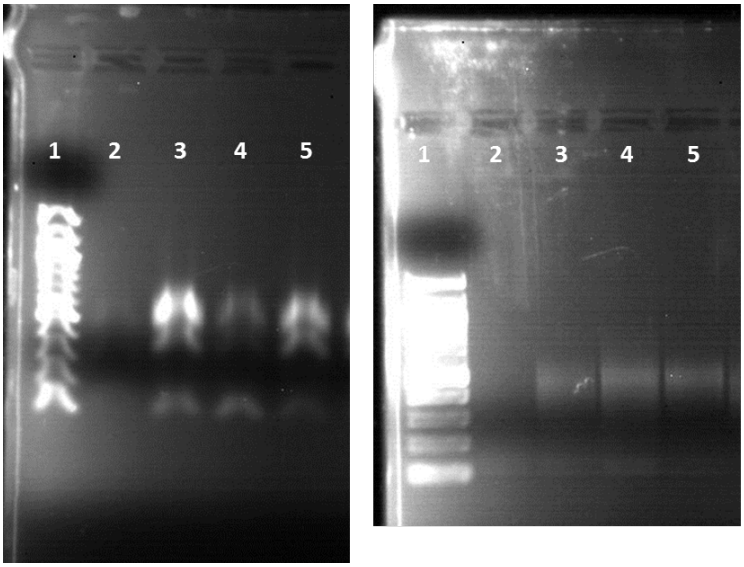


Figure 8. Hemi-nested PCR with primers that have Illumina sequencing primer and indexing primer homology.

After preparation for sequencing, DNA concentration of the prepared fragments and absorbance were measured by nanospectrophotometer (Table 5).

Table 5. DNA concentration and absorbance measured by nanodrop after library preparation for sequencing.

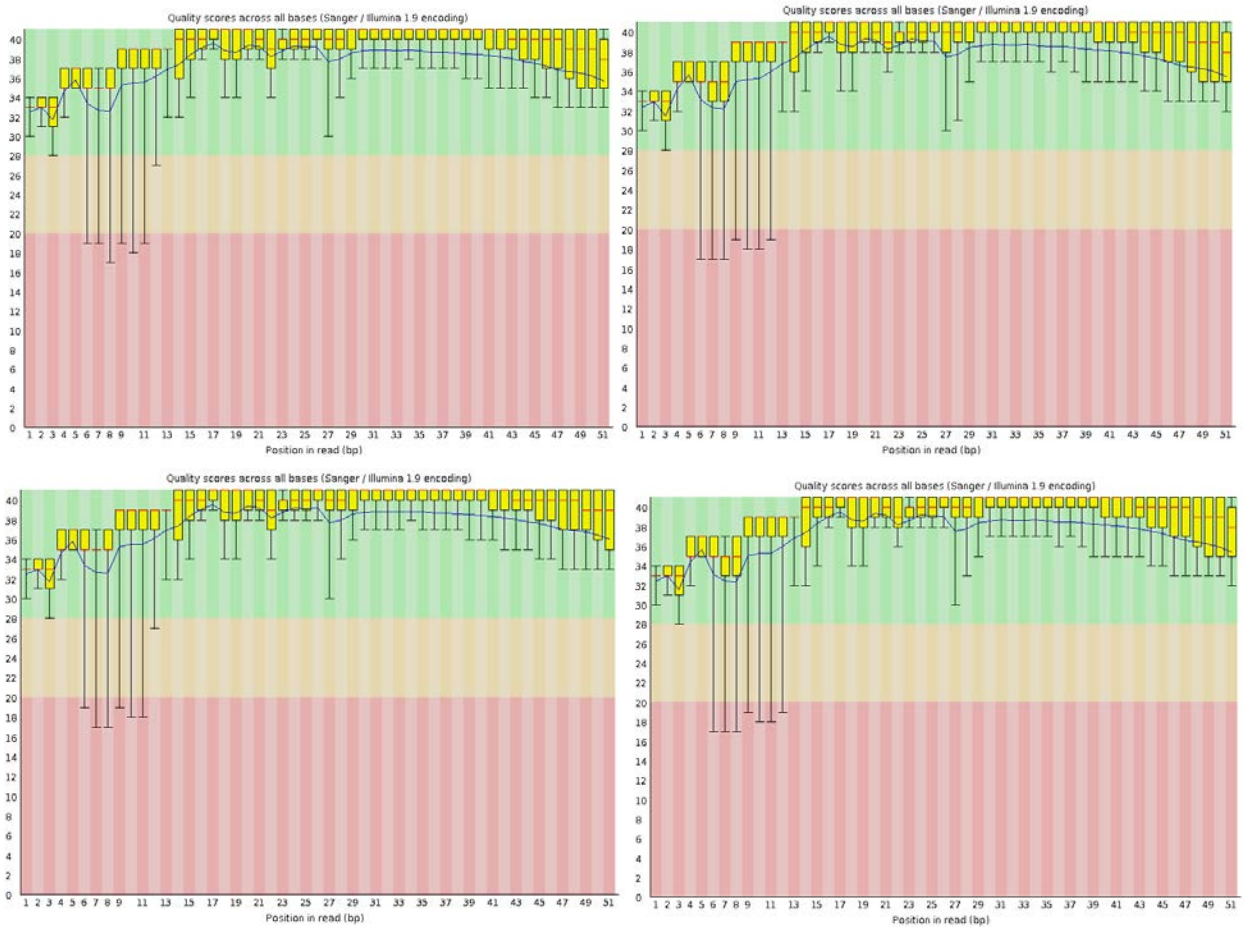
Sample	DNA concentration (ng/µl)	A260:A280	A260:A230
Planktonic	17.5	1.915	2.228
0hr Biofilm 1	20.0	1.833	2.000
0hr Biofilm 2	12.5	1.776	1.915
24hr Biofilm 1	7.5	1.811	2.400
24hr Biofilm 2	17.5	1.824	2.116
24hr Biofilm 3	17.5	1.870	2.003

4.3 DNA SEQUENCING REVEALS POTENTIAL MUTANTS SENSITIVE TO RIFAMPICIN TREATMENT

The sequencing data returned from Tufts showed the number of reads ranging from 629,825 reads for the second unexposed biofilm sample to 5,331,968 for third 24-hour rifampicin-exposed biofilm sample. While this is not an overly large number of reads, the most rifampicin-sensitive mutants may still be detected by testing only the mutants that show the largest differential in presence between the rifampicin-exposed and unexposed samples. The returned sequences did show high mean quality scores according to the fastqc report provided by Tufts for each sample and the mean quality score by base in each sequence (Table 6 and Figure 9).

Table 6. Basic statistics from DNA sequencing run.

Sample	# of Reads	% of Raw Clusters per Lane	% Perfect Index Reads	% One Mismatch Reads (Index)	% of >= Q30 Bases	Mean Quality Score
Planktonic	850,447	4.15	100.0	0.0	93.65	37.29
0hr Biofilm-1	2,366,430	11.54	100.0	0.0	92.94	37.09
0hr Biofilm-2	629,825	3.07	100.0	0.0	93.74	37.27
24hr Biofilm-1	2,859,380	13.95	100.0	0.0	93.06	37.11
24hr Biofilm-2	2,397,451	11.69	100.0	0.0	93.55	37.22
24hr Biofilm-3	4,172,700	26.00	100.0	0.0	93.37	37.17



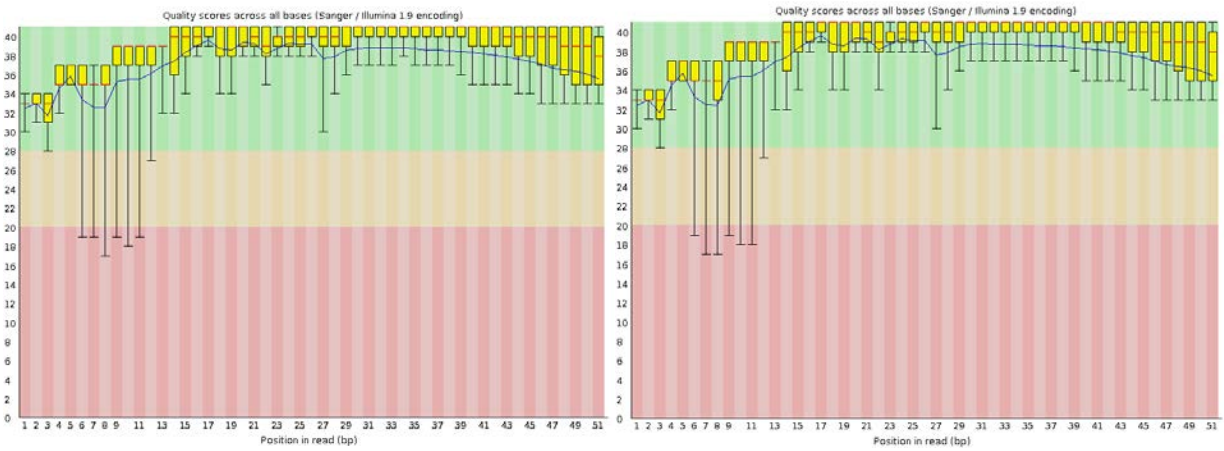


Figure 9. Quality scores across all bases for each sample. Top left: planktonic, top right: 0hr Biofilm-1, Middle left: 0hr Biofilm-2, Middle right: 24hr Biofilm-1, Bottom left: 24hr Biofilm-2, Bottom right: 24hr Biofilm-3.

After processing the raw fastq files and mapping with Bowtie to the MTB (Erdman) genome, roughly two-thirds of the returned reads aligned to a TA site. The exact values for each sample are set out in the table below (Table 7). The number of specific "TA" sites that had insertions mapped to them ranged from 12,265 for the planktonic sample and 22,720 for the third 24-hour rifampicin-exposed sample.

Table 7. Results for the number of reads mapped to TA sites for each sample.

Sample Name	Total Reads	Reads processed	Reads discarded	Insertions mapped to top strand	Insertions mapped to bottom strand	Unaligned reads	Mismatched at TA reads	Total TA sites hit
Planktonic	850447	667528	182919	286169	278750	102224	385	12265
0hr exp. 1	2366430	1810105	556325	800679	788436	219993	997	20976
0hr exp. 2	629825	497303	132522	226720	237319	33013	251	12526
24hr exp. 1	2859380	2237301	622079	1023866	1005918	206552	965	20287
24hr exp. 2	2397451	1885675	511776	876007	857451	151646	571	19295
24hr exp. 3	5331968	4181889	1150079	1935076	1897399	347300	2114	22720

The first unexposed biofilm sample (0hr exp. 1), had the most aligned reads among the unexposed (therefore being the most representative), and was hence used as the reference sample. Reads for each insertion in the other samples were normalized to the number of reads from this reference sample. Eight insertion sites that were located inside of seven different open reading frames (E_1681 had two insertion points which showed great differentiation between the exposed and unexposed samples) were selected for having insertion mutants highly present in the unexposed samples but much less so in rifampicin-exposed samples. Other TA sites either had too little insertions (<100) to determine a meaningful difference or insertion counts were even across exposed and unexposed samples or were too varied among the exposed and unexposed classes.

The eight genomic positions and their respective Erdman gene numbers are: 462570 at E0425, 1691647 and 1691886 at E1681, 2055036 at E2009, 3073036 at E3045, 3521539 at E3465, 4231539 at E4162 and 4324565 at E4241. These sites had insertions on both strands of the genome, but were especially diminished in samples exposed to 24 hours of rifampicin treatment. Alignments were visualized in Broad's Integrated Genome Viewer (Figure 10 and 11 are two examples captured by screenshots). Figure 10 shows a zoomed out screenshot of alignments mapped to E1681-E1689 (each gray marker is an individual alignment). Figure 11 shows this area zoomed in to E1681 with each gray bar showing alignments to specific points in the genome. Mismatched nucleotides are shown in color inside the gray area of an alignment. Note the multiple positions and orientations of alignments at this site. These images were taken from viewing the bam file showing the mapping of the 0hr Biofilm-1 sample (the unexposed reference). The number of insertions after normalization across all samples is displayed in the bar graph in Figure 12.

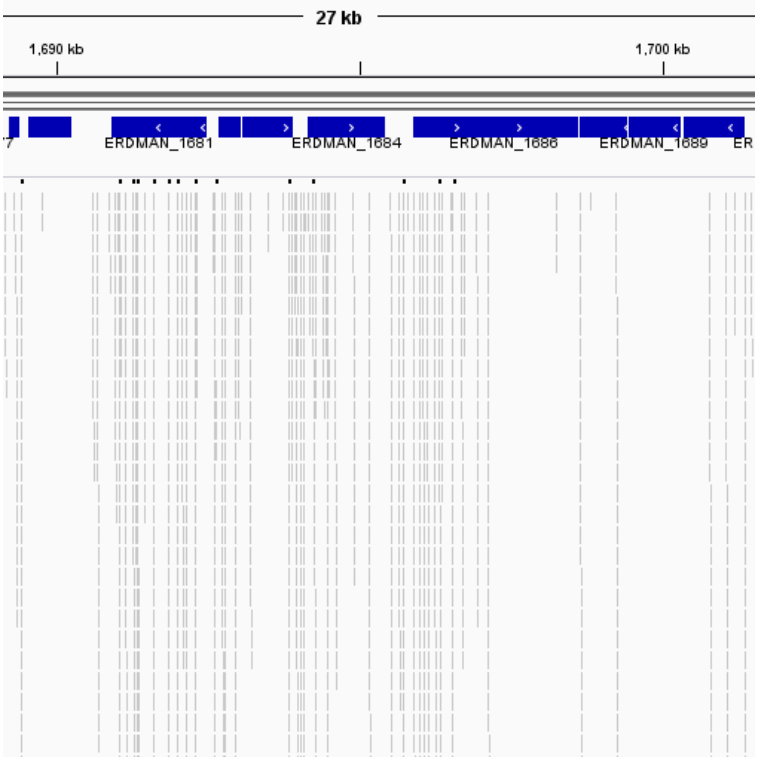


Figure 10. Alignments of insertions sites in the 1,690 to 1,700kb area of the MTB (Erdman) genome in IGV.

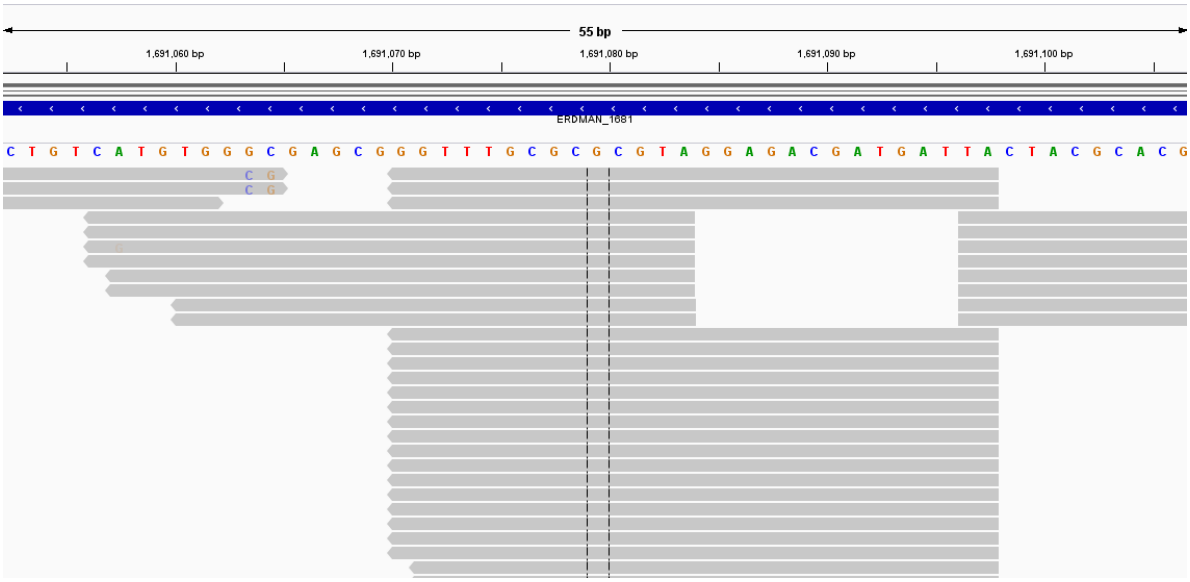


Figure 11. Alignments of insertions sites zoomed into 55bp window showing alignments in the unexposed biofilm sample to E1681.

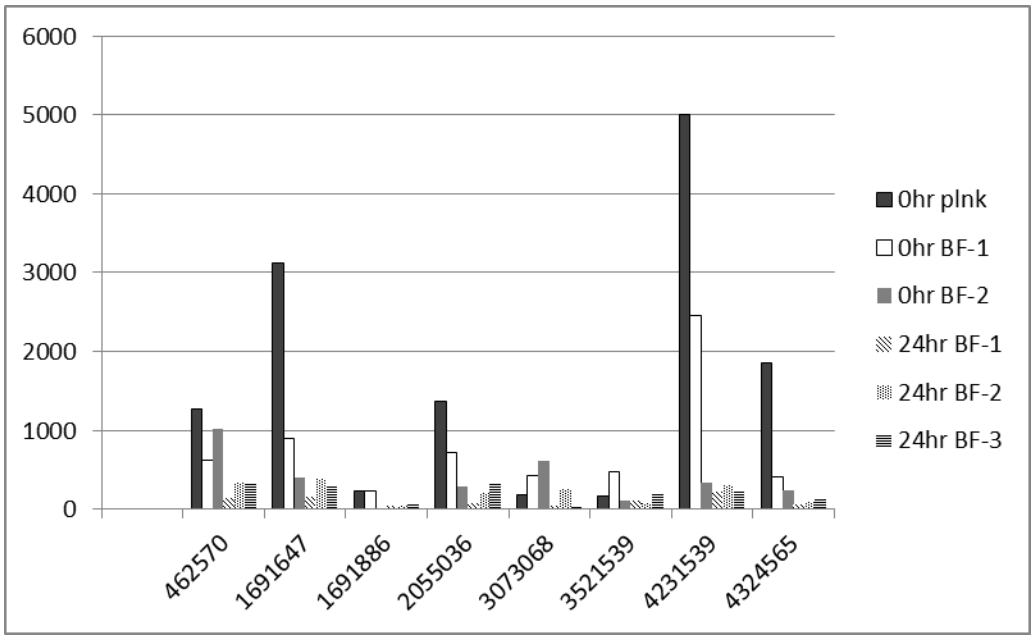


Figure 12. Total insertions mapped to eight different genomic positions in unexposed and exposed DNA samples.

A ratio was calculated of the difference between insertions present in the reference sample to the mean number of insertions across the three rifampicin-exposed samples (Figure 13). The error bars represent the standard deviation between the three exposed samples. Genomic positions have been labeled by their corresponding locus from the H37Rv strain of MTB. The gene names and functions for each gene (or proposed functions for poorly characterized genes) being investigated is also summarized (Table 8). Rv3796 had the highest differential between unexposed and the exposed samples at 9.94. The ratios in descending order are Rv3796: 9.94, Rv2779c: 6.90, Rv1508c (position 2): 5.15, Rv1819c: 4.79, Rv3868: 4.73, Rv3164c: 4.24, Rv1508c (position 1): 3.67 and Rv0385: 2.66. These seven genes comprise the group with the greatest difference in insertions between unexposed and exposed samples.

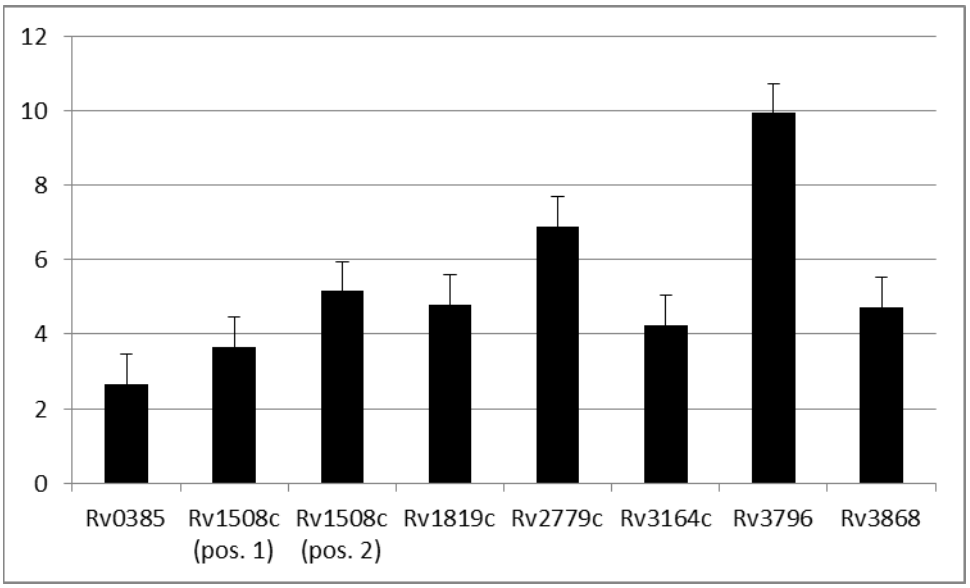


Figure 13. Ratio of number of insertions in the unexposed reference sample to mean of the rifampicin-exposed samples.

Table 8. Proposed genetic factors that contribute to rifampicin tolerance and their functions.

Genome position	Erdman gene #	Similar H37rv#	Gene Name/Proposed Function
462570	E0425	Rv0385	probable monooxygenase
1691647	E1681	Rv1508c	probable membrane protein of glycosyltransferase family
1691886	E1681	Rv1508c	probable membrane protein of glycosyltransferase family
2055036	E2009	Rv1819c	drug-transport transmembrane ATP-binding protein ABC transporter
3073068	E3045	Rv2779c	LRP/AsnC family transcriptional regulator
3521539	E3465	Rv3164c	moxR3 - methanol dehydrogenase transcriptional regulatory protein
4231539	E4162	Rv3796	putative arylsulfatase
4324565	E4241	Rv3868	AAA ATPase (ESX conserved component EccA1. ESX-1 type VII secretion system protein)

4.4 CONFIRMATION OF MUTANTS USING QUANTITATIVE PCR

Despite several attempts to confirm a differential representation in mutants by quantitative PCR, we were unable to accurately measure the presence of insertions across samples involved in the experiment. Even in the case of the insertion sites chosen for endogenous controls which represented similar insertion numbers across samples in both unexposed and exposed conditions, amplification in duplicate PCR runs did not return consistent CT values for even identical samples. It is possible that this is due to some inherent difficulty in measuring the level of transposon junction site DNA at a specific location among a pool of mutants.

5.0 DISCUSSION

The results of the DNA sequencing identified seven genes that showed a high number of insertions at positions in their genes in the unexposed samples, but that were diminished in the rifampicin exposed samples. Transposon insertion mutagenesis has been a useful and reliable tool in demonstrating the essentiality of genes under certain growth conditions for the past twelve years. With the limited pool of mutants, only 22,720 of a possible 74,417 transposon mutants were examined, we were able to identify eight specific genomic positions that could contribute to rifampicin tolerance. To develop a more comprehensive view as to which genes contribute to rifampicin tolerance, the experiment will be repeated with more replicates in a denser library of mutants that can be guaranteed to represent all possible "TA" sites outside of genes essential for *in vitro* growth. It is possible that this could be achieved using the H37Rv MTB strain, which is more often used for genetic manipulation than the Erdman strain. The more "TA" sites that have been disrupted by transposon insertion, the more representative the pool of mutants with dysfunctional genes will be. However, this project demonstrates the feasibility of the transposon mutant-based approach.

Because of the relatively low concentration of and brief exposure time to antibiotic, it is plausible that these mutants are among the hyper sensitive clones to rifampicin clones. This is further supported by the fact that the biofilms of MTB (Erdman) wildtype when exposed to the same antibiotic treatment saw 98 percent survival of the bacteria. Overall, we believe these seven

mutants warrant further investigation and characterization in their roles in rifampicin tolerance and biofilm fitness.

The gene identified to have the greatest differentiation between the unexposed reference sample and the biofilms exposed to rifampicin was E4162, corresponding to Rv3796. Rv3796 is a poorly characterized, non-essential hypothetical protein and putative arylsulfatase [51]. Arylsulfatases are a class of enzymes that catalyze the hydrolysis of a phenol sulfate. The amino acid sequence of Rv3796 shows similar sequence with the metallo-beta-lactamase protein domain and twin arginine translocation (Tat) pathway signal sequence [52-54]. The metallo-beta-lactamase protein domain is often a component of proteins involved in antibiotic resistance, while the Tat sequence is thought to play a role in transporting folded proteins across the lipid bilayer, and it has been suggested that phospholipase virulence factors in MTB are exported in a Tat-dependent manner [55-57]. Four genes in MTB encoding Phospholipases C were up-regulated in the first 24 hours of macrophage infection and triple and quadruple mutants of these genes attenuated the bacteria in a mouse model [58]. This makes Rv3796 an intriguing gene to play a role in rifampicin tolerance in MTB as its amino acid sequence is linked to both the pathogen's ability to breakdown antibiotic and to remain virulent.

The genetic factor with the second-highest differential ratio was E3045, or its corresponding H37rv gene loci, Rv2779c. Rv2779c is a non-essential, DNA-binding protein and is most likely a transcriptional regulatory protein of the Lrp/AsnC family (leucine responsive regulatory protein/regulator of asparagine synthase C gene product) [51, 53]. Lrp/AsnC family regulators have been highly induced during nutrient starvation, a state relative to the biofilm model, and have been linked to persistence as well as the regulation of pili synthesis, amino acid metabolism and DNA repair and recombination [11, 59].

E1681, or Rv1508c, had the third highest differential ratio, and the data showed high differential in the number of insertions in the unexposed to exposed samples at two genomic positions. The Rv1508c gene is a non-essential gene that codes for a hypothetical protein that has been proposed to be a membrane-bound glycosyltransferase [41, 53, 60]. A portion of the peptide this gene codes for has been previously proposed among a group of potential targets for diagnosis because of their MTB specificity and absence in *Mycobacterium bovis* BCG [61, 62]. The gene has also been noted to be possibly up-regulated in macrophage infection [63]. The cluster of biosynthetic genes in this region of the genome, including Rv1508c and known glycosyltransferases have been linked to synthesis of mycolic acids, which play an important role in the virulence of MTB [64].

E2009, or Rv1819c, another gene identified by its rifampicin sensitive mutant, codes for bacA, a gene for a probable ATP-binding cassette transporter protein and has speculated to play a role in export of antibiotic export across the cellular membrane as an efflux pump making its possible involvement in rifampicin tolerance intuitive [65, 66]. The gene's role was deduced by its sequence containing known ABC transport protein domains and further illustrated to be active during antibiotic challenge. Mice infected with this mutant showed tolerance to the cancer medication bleomycin and survived longer after infection [67]. ABC transporter's role in antibiotic resistance has been well-documented in many bacteria including MTB [68, 69]. Rv1819c has been shown to be up-regulated during isoniazid exposure, but not rifampicin, and it along with seven other efflux pump genes have shown increased activity in MDR strains of the bacteria [70, 71]. It has also been implicated in the transport of lipids across the cellular membrane as it contains a lipid attachment site [72]. The gene is non-essential for growth [51].

E4241, or Rv3868, is a gene encoding for an EccA1, an AAA ATPase (ATPases associated with diverse cellular activities) with conserved domains of a type VII secretion system [73]. These enzymes hydrolyze ATP to move macromolecular substrates across the cell wall. The type VII secretion system is unique to mycobacteria and is necessary for moving proteins across the generally impermeable mycobacterial cell wall. EccA1, while non-essential for *in vitro* growth, is required for growth in the mouse model or in primary macrophages making it difficult for further characterization [74-76]. MTB has several variants of type VII secretion systems and ATPase components, and EccA1 is one of two involved with transporting virulence factors outside of the cell [77, 78]. The EccA1 gene is required for the secretion of T-cell and B-cell antigens ESAT-6 and CFP-10 [79]. This system has also been referred to as the ESX-1 virulence factor secretion pathway. A deletion mutant of EccA1 in *Mycobacterium marinum*, an opportunistic pathogen, was shown to have decreased synthesis of mycolic acids, a core component of the mycobacterial cell wall [80].

Rv3164c, corresponding to E3465 was the sixth mutant identified in this study. This gene encodes the probable methanol dehydrogenase transcriptional regulator protein MoxR3 which is thought to regulate genes involved in methanol oxidation reactions. MoxR3 is non-essential for *in vitro* growth and like Rv3868 shows homology in sequence with proteins associated with AAA ATPases [51, 53, 65]. However, any link of this gene product to antibiotic tolerance remains unclear, although its role in the MTB regulatory network has been proposed [81].

E0425, or Rv0385. Rv0385 encodes a probable monooxygenase enzyme protein based on its similarities in sequence to other monoxygenase genes at the C-terminus [53]. The gene is non-essential for *in vitro* growth [51]. Monooxygenase enzymes function by adding hydroxyl groups to their substrates in many metabolic pathways. Rv0385 also has approximately 25

percent sequence homology with flavohemoglobin proteins [82] and could act as a dioxygenase with nitric oxide as many bacterial flavohemoglobin proteins do [83]. Bacteria overexpressing this gene had an advantage in countering oxidative stress, and it has been suggested Rv0385 allows MTB to balance oxidative stress levels and reduce damage to the cellular membrane [84]. This could provide a unique function to this protein in the biofilm context where many cells at the lower level of the structure are exposed to oxidative stress. It is possible that this stress management helps the bacteria persist in the presence of antibiotics like rifampicin.

This study, still an initial assessment, provides a valuable proof of principal for the approach of using a transposon mutant library screen in MTB biofilms for examining rifampicin tolerance and fitness from a genomic perspective, and it has identified several genes for investigation that when disrupted by transposon insertion caused rifampicin sensitivity. It has generated several hypotheses as to the various mechanisms to this sensitivity for each gene. A full characterization of each mutant and the mechanism by which it contributes to rifampicin tolerance would provide powerful insight into the persistence of MTB as a pathogen. If possible, these gene products could be further targeted or exploited to compromise the persistence ability of MTB, thereby facilitating a shorter antibiotic intervention for tuberculosis.

6.0 FUTURE DIRECTIONS

Moving forward from these findings, the next step is a characterization of the mutants identified and confirmed to be highly represented in the unexposed samples, but less so in the samples exposed to rifampicin. First, deletion or loss of function mutants for the genes observed in the sequencing data must be obtained followed by testing rifampicin sensitivity both planktonically and in the membrane-based portable biofilm model here. From there, these mutants will be studied *in vitro* in macrophages and *in vivo* in the mouse model for response to rifampicin treatment after infection with each rifampicin-sensitive mutant. The mutants could be tagged with fluorescent protein and mouse lungs could be processed and viewed under the microscope to characterize the locale of persistent infection and check for the presence of acellular biofilms in the mouse lung tissue. Beyond that, the next steps would be complementing the mutants to restore regular rifampicin tolerance and characterizing the mechanism by which the genes' products confer rifampicin tolerance or sensitivity when knocked out.

To further generate possible genes that contribute to rifampicin tolerance, we have already begun to replicate the same experiment in a much denser transposon mutant library in MTB $mc^2$7000 strain. This library was previously made by Anil Ojha and contains approximately 80,000 transposon mutants. The DNA from the cultures exposed to rifampicin will be prepared using the same protocol here sent off for sequencing and undergo the same protocol for data analysis, allowing us to generate hypotheses and quickly analyze results based

off of the sequencing data. A denser mutant library is also being prepared in the virulent MTB H37rv strain for repeat experiments. This will provide the laboratory with a set of genes to explore the role and mode of rifampicin tolerance in not only MTB pathogenesis, but also the extent to which this tolerance is related to biofilm behavior and development in tuberculosis infection. It would also benefit to see if these genes each contribute to antibiotic tolerance for multiple drugs or if it is rifampicin-specific. Further down the pipeline, the products of these genes could be exploited in the treatment of persistent tuberculosis infection to potentially decrease the time required for pharmaceutical intervention, DOTS and eventually an overall decrease in the burden of disease.

BIBLIOGRAPHY

1. *Global Tuberculosis Report 2012*, W.H. Organization, Editor 2012, World Health Organization: http://www.who.int/tb/publications/global_report/gtbr12_main.pdf.
2. Abubakar, I., et al., *Drug-resistant tuberculosis: time for visionary political leadership.* The Lancet Infectious Diseases, 2013. **13**(6): p. 529-539.
3. Dye, C., et al., *Trends in tuberculosis incidence and their determinants in 134 countries.* Bulletin of the World Health Organization, 2009. **87**(9): p. 683-91.
4. Goren, M.B., et al., *Prevention of phagosome-lysosome fusion in cultured macrophages by sulfatides of Mycobacterium tuberculosis.* Proceedings of the National Academy of Sciences, 1976. **73**(7): p. 2510-2514.
5. Henderson, R.A., S.C. Watkins, and J.L. Flynn, *Activation of human dendritic cells following infection with Mycobacterium tuberculosis.* The Journal of Immunology, 1997. **159**(2): p. 635-43.
6. Lay, G., et al., *Langhans giant cells from M. tuberculosis-induced human granulomas cannot mediate mycobacterial uptake.* The Journal of Pathology, 2007. **211**(1): p. 76-85.
7. Mitchison, D.A., *How drug resistance emerges as a result of poor compliance during short course chemotherapy for tuberculosis Counterpoint.* The International Journal of Tuberculosis and Lung Disease, 1998. **2**(1): p. 10-15.
8. Ormerod, L.P., *Multidrug-resistant tuberculosis (MDR-TB): epidemiology, prevention and treatment.* British Medical Bulletin, 2005. **73-74**(1): p. 17-24.
9. Dubos, R.J. and G. Middlebrook, *The effect of wetting agents on the growth of tubercle bacilli.* The Journal of experimental medicine, 1948. **88**(1): p. 81-88.
10. Van Boxtel, R., R. Lambrecht, and M. Collins, *Effects of colonial morphology and tween 80 on antimicrobial susceptibility of Mycobacterium paratuberculosis.* Antimicrobial agents and chemotherapy, 1990. **34**(12): p. 2300-2303.
11. Betts, J.C., et al., *Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling.* Molecular Microbiology, 2002. **43**(3): p. 717-731.
12. Costerton, J.W., et al., *Microbial biofilms.* Annual Reviews in Microbiology, 1995. **49**(1): p. 711-745.
13. Lam, J., et al., *Production of mucoid microcolonies by Pseudomonas aeruginosa within infected lungs in cystic fibrosis.* Infection and immunity, 1980. **28**(2): p. 546-556.
14. Costerton, J., et al., *The role of bacterial surface structures in pathogenesis.* Critical reviews in microbiology, 1981. **8**(4): p. 303-338.
15. Stoodley, P., et al., *Biofilms as complex differentiated communities.* Annual Reviews in Microbiology, 2002. **56**(1): p. 187-209.
16. Allison, D.G., *The biofilm matrix.* Biofouling, 2003. **19**(2): p. 139-150.

17. Blankenship, J.R. and A.P. Mitchell, *How to build a biofilm: a fungal perspective.* Current opinion in microbiology, 2006. **9**(6): p. 588-594.
18. O'Toole, G.A., et al., *[6] Genetic approaches to study of biofilms.* Methods in enzymology, 1999. **310**: p. 91-109.
19. Ogasawara, H., et al., *Regulation of the Escherichia coli csgD promoter: interplay between five transcription factors.* Microbiology, 2010. **156**(8): p. 2470-2483.
20. Hammer, B.K. and B.L. Bassler, *Quorum sensing controls biofilm formation in Vibrio cholerae.* Molecular microbiology, 2003. **50**(1): p. 101-104.
21. Chai, Y., et al., *Bistability and biofilm formation in Bacillus subtilis.* Molecular microbiology, 2008. **67**(2): p. 254-263.
22. Anderson, G.G., et al., *Intracellular bacterial communities of uropathogenic< i> Escherichia coli</i> in urinary tract pathogenesis.* TRENDS in Microbiology, 2004. **12**(9): p. 424-430.
23. Høiby, N., et al., *Antibiotic resistance of bacterial biofilms.* International journal of antimicrobial agents, 2010. **35**(4): p. 322-332.
24. Jensen, P.Ø., et al., *The immune system vs. Pseudomonas aeruginosa biofilms.* FEMS Immunology & Medical Microbiology, 2010. **59**(3): p. 292-305.
25. Saginur, R., et al., *Multiple combination bactericidal testing of staphylococcal biofilms from implant-associated infections.* Antimicrobial agents and chemotherapy, 2006. **50**(1): p. 55-61.
26. Ceri, H., et al., *The Calgary Biofilm Device: new technology for rapid determination of antibiotic susceptibilities of bacterial biofilms.* Journal of clinical microbiology, 1999. **37**(6): p. 1771-1776.
27. Banin, E., K.M. Brady, and E.P. Greenberg, *Chelator-induced dispersal and killing of Pseudomonas aeruginosa cells in a biofilm.* Applied and environmental microbiology, 2006. **72**(3): p. 2064-2069.
28. Kolodkin-Gal, I., et al., *D-amino acids trigger biofilm disassembly.* Science, 2010. **328**(5978): p. 627-629.
29. Marsollier, L., et al., *Colonization of the salivary glands of Naucoris cimicoides by Mycobacterium ulcerans requires host plasmatocytes and a macrolide toxin, mycolactone.* Cellular microbiology, 2005. **7**(7): p. 935-943.
30. Carter, G., et al., *Characterization of biofilm formation by clinical isolates of Mycobacterium avium.* Journal of medical microbiology, 2003. **52**(9): p. 747-752.
31. Feazel, L.M., et al., *Opportunistic pathogens enriched in showerhead biofilms.* Proceedings of the National Academy of Sciences, 2009. **106**(38): p. 16393-16399.
32. Berry, D., C. Xi, and L. Raskin, *Microbial ecology of drinking water distribution systems.* Current opinion in biotechnology, 2006. **17**(3): p. 297-302.
33. Recht, J. and R. Kolter, *Glycopeptidolipid acetylation affects sliding motility and biofilm formation in Mycobacterium smegmatis.* Journal of bacteriology, 2001. **183**(19): p. 5718-5724.
34. Ojha, A.K., et al., *Growth of Mycobacterium tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria.* Molecular microbiology, 2008. **69**(1): p. 164-174.
35. Ojha, A.K., et al., *Enzymatic hydrolysis of trehalose dimycolate releases free mycolic acids during mycobacterial growth in biofilms.* Journal of Biological Chemistry, 2010. **285**(23): p. 17380-17389.

36. Ojha, A., et al., *GroEL1: a dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria.* Cell, 2005. **123**(5): p. 861-873.
37. Ojha, A. and G.F. Hatfull, *The role of iron in Mycobacterium smegmatis biofilm formation: the exochelin siderophore is essential in limiting iron conditions for biofilm formation but not for planktonic growth.* Molecular microbiology, 2007. **66**(2): p. 468-483.
38. Pang, J.M., et al., *The polyketide Pks1 contributes to biofilm formation in Mycobacterium tuberculosis.* Journal of bacteriology, 2012. **194**(3): p. 715-721.
39. Akerley, B.J., et al., *Systematic identification of essential genes by in vitro mariner mutagenesis.* Proceedings of the National Academy of Sciences, 1998. **95**(15): p. 8927-8932.
40. Sassetti, C.M., D.H. Boyd, and E.J. Rubin, *Comprehensive identification of conditionally essential genes in mycobacteria.* Proceedings of the National Academy of Sciences, 2001. **98**(22): p. 12712-12717.
41. Griffin JE, Gawronski JD, DeJesus MA, Ioerger TR, Akerley BJ, et al., *High-Resolution Phenotypic Profiling Defines Genes Essential for Mycobacterial Growth and Cholesterol Catabolism.* PLoS Pathogens, 2011. **7**(9): p. e1002251.
42. Murry, J., et al., *Transposon Site Hybridization in Mycobacterium tuberculosis,* in *Microbial Gene Essentiality: Protocols and Bioinformatics,* A. Osterman and S. Gerdes, Editors. 2008, Humana Press. p. 45-59.
43. Bentley, D.R., et al., *Accurate whole human genome sequencing using reversible terminator chemistry.* Nature, 2008. **456**(7218): p. 53-59.
44. Goecks, J., et al., *Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.* Genome Biology, 2010. **11**(8): p. R86.
45. Blankenberg, D., et al., *Galaxy: A Web-Based Genome Analysis Tool for Experimentalists,* in *Current Protocols in Molecular Biology*. 2001, John Wiley & Sons, Inc.
46. Giardine, B., et al., *Galaxy: A platform for interactive large-scale genome analysis.* Genome Research, 2005. **15**(10): p. 1451-1455.
47. Blankenberg, D., et al., *Manipulation of FASTQ data with Galaxy.* Bioinformatics, 2010. **26**(14): p. 1783-1785.
48. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biology, 2009. **10**(3): p. R25.
49. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-2079.
50. Kulka, K., G. Hatfull, and A.K. Ojha, *Growth of Mycobacterium tuberculosis Biofilms.* 2012(60): p. e3820.
51. Sassetti, C.M., D.H. Boyd, and E.J. Rubin, *Genes required for mycobacterial growth defined by high density mutagenesis.* Molecular Microbiology, 2003. **48**(1): p. 77-84.
52. Malen, H., et al., *Definition of novel cell envelope associated proteins in Triton X-114 extracts of Mycobacterium tuberculosis H37Rv.* BMC Microbiology, 2010. **10**(1): p. 132.
53. de Souza, G.A., et al., *Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway.* Journal of Proteomics, 2011. **75**(2): p. 502-510.
54. Song, H., et al., *Identification of outer membrane proteins of Mycobacterium tuberculosis.* Tuberculosis, 2008. **88**(6): p. 526-544.

55. Shaw, R.W. and S.K. Kim, *Inhibition of metallo-β-lactamase*, 2010, Google Patents.
56. Saint-Joanis, B., et al., *Inactivation of Rv2525c, a Substrate of the Twin Arginine Translocation (Tat) System of Mycobacterium tuberculosis, Increases β-Lactam Susceptibility and Virulence.* Journal of Bacteriology, 2006. **188**(18): p. 6669-6679.
57. De Buck, E., E. Lammertyn, and J. Anné, *The importance of the twin-arginine translocation pathway for bacterial virulence.* Trends in Microbiology, 2008. **16**(9): p. 442-453.
58. Raynaud, C., et al., *Phospholipases C are involved in the virulence of Mycobacterium tuberculosis.* Molecular Microbiology, 2002. **45**(1): p. 203-217.
59. Deng, W., H. Wang, and J. Xie, *Regulatory and pathogenesis roles of Mycobacterium Lrp/AsnC family transcriptional factors.* Journal of Cellular Biochemistry, 2011. **112**(10): p. 2655-2662.
60. Liu, J. and A. Mushegian, *Three monophyletic superfamilies account for the majority of the known glycosyltransferases.* Protein Science, 2003. **12**(7): p. 1418-1431.
61. Mustafa, A.S., *Proteins and Peptides Encoded by M. tuberculosis-Specific Genomic Regions for Immunological Diagnosis of Tuberculosis.* Mycobacterial Diseases, 2012.
62. Al-Khodari, N.Y., R. Al-Attiyah, and A.S. Mustafa, *Identification, Diagnostic Potential, and Natural Expression of Immunodominant Seroreactive Peptides Encoded by Five Mycobacterium tuberculosis-Specific Genomic Regions.* Clinical and Vaccine Immunology, 2011. **18**(3): p. 477-482.
63. Hobson, R.J., et al., *Use of an arrayed promoter-probe library for the identification of macrophage-regulated genes in Mycobacterium tuberculosis.* Microbiology, 2002. **148**(5): p. 1571-1579.
64. Berg, S., et al., *The glycosyltransferases of Mycobacterium tuberculosis—roles in the synthesis of arabinogalactan, lipoarabinomannan, and other glycoconjugates.* Glycobiology, 2007. **17**(6): p. 35R-56R.
65. Mawuenyega, K.G., et al., *Mycobacterium tuberculosis Functional Network Analysis by Global Subcellular Protein Profiling.* Molecular Biology of the Cell, 2005. **16**(1): p. 396-404.
66. Domenech, P., et al., *BacA, an ABC Transporter Involved in Maintenance of Chronic Murine Infections with Mycobacterium tuberculosis.* Journal of Bacteriology, 2009. **191**(2): p. 477-485.
67. da Silva, P.E.A., et al., *Efflux as a mechanism for drug resistance in Mycobacterium tuberculosis.* FEMS Immunology & Medical Microbiology, 2011. **63**(1): p. 1-9.
68. Higgins, C.F., *Multiple molecular mechanisms for multidrug resistance transporters.* Nature, 2007. **446**(7137): p. 749-757.
69. Pasca, M.R., et al., *Rv2686c-Rv2687c-Rv2688c, an ABC Fluoroquinolone Efflux Pump in Mycobacterium tuberculosis.* Antimicrobial Agents and Chemotherapy, 2004. **48**(8): p. 3175-3178.
70. Gupta, AK, VM Katoch, DS Chauhan, R Sharma, M Singh, K Venkatesan, and VD Sharma., *Microarray Analysis of Efflux Pump Genes in Multidrug-Resistant Mycobacterium tuberculosis During Stress Induced by Common Anti-Tuberculous Drugs* Microbial Drug Resistance, 2010. **16**(1): p. 21-28.
71. Xin Jiang, WZ, Y Zhang, F Gao, C Lu, X Zhang, and H Wang, *Assessment of Efflux Pump Gene Expression in a Clinical Isolate Mycobacterium tuberculosis by Real-Time Reverse Transcription PCR.* Microbial Drug Resistance, 2008. **14**(1): p. 7-11.

72. Braibant, M., P. Gilot, and J. Content, *The ATP binding cassette (ABC) transport systems of Mycobacterium tuberculosis.* FEMS Microbiology Reviews, 2000. **24**(4): p. 449-467.
73. Gey van Pittius, N., et al., *The ESAT-6 gene cluster of Mycobacterium tuberculosis and other high G+C Gram-positive bacteria.* Genome Biology, 2001. **2**(10): p. research0044.1 - research0044.18.
74. Sassetti, C.M. and E.J. Rubin, *Genetic requirements for mycobacterial survival during infection.* Proceedings of the National Academy of Sciences, 2003. **100**(22): p. 12989-12994.
75. Rengarajan, J., B.R. Bloom, and E.J. Rubin, *Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(23): p. 8327-8332.
76. Kruh NA, J Troudt, A Izzo, J Prenni, and KM Dobos, *Portrait of a Pathogen: The Mycobacterium tuberculosis Proteome In Vivo.* PLoS One, 2010. **5**(11): p. e13938.
77. DiGiuseppe Champion, P.A., et al., *ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria.* Molecular Microbiology, 2009. **73**(5): p. 950-962.
78. Simeone, R., D. Bottai, and R. Brosch, *ESX/type VII secretion systems and their role in host–pathogen interaction.* Current Opinion in Microbiology, 2009. **12**(1): p. 4-10.
79. Gao, L.Y., et al., *A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion.* Molecular Microbiology, 2004. **53**(6): p. 1677-1693.
80. Joshi, Shilpa A., et al., *EccA1, a Component of the Mycobacterium marinum ESX-1 Protein Virulence Factor Secretion Pathway, Regulates Mycolic Acid Lipid Synthesis.* Chemistry & Biology, 2012. **19**(3): p. 372-380.
81. Sanz J, J Navarro, A Arbués, C Martín, PC Marijuán and Y Moreno, *The Transcriptional Regulatory Network of Mycobacterium tuberculosis.* PLoS One, 2011. **6**(7): p. e22178.
82. Gupta, S., et al., *Novel flavohemoglobins of mycobacteria.* IUBMB Life, 2011. **63**(5): p. 337-345.
83. Namboori, S., et al., *Enhanced functional and structural domain assignments using remote similarity detection procedures for proteins encoded in the genome ofMycobacterium tuberculosis H37Rv.* Journal of Biosciences, 2004. **29**(3): p. 245-259.
84. Gupta, S., et al., *An Unconventional Hexacoordinated Flavohemoglobin from Mycobacterium tuberculosis.* Journal of Biological Chemistry, 2012. **287**(20): p. 16435-16446.