Genomic Analysis of Pathogenicity Determinants in *Mycobacterium kansasii* Type I

Thesis by

Qingtian Guan

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

King Abdullah University of Science and Technology, Thuwal

Kingdom of Saudi Arabia

© May 2016

Qingtian Guan

All rights reserved
The thesis of Qingtian Guan is approved by the examination committee.

Committee Chairperson: Dr. Arnab Pain
Committee Members: Dr. Abdallah M. Abdallah, Dr. Liming Xiong, Dr. Peiying Hong
ABSTRACT

Genomic Analysis of Pathogenicity Determinants in *Mycobacterium kansasii* Type I

Qingtian Guan

Mycobacteria, a genus within *Actinobacteria* Phylum, are well known for two pathogens that cause human diseases: leprosy and tuberculosis. Other than the obligate human mycobacteria, there is a group of bacteria that are present in the environment and occasionally cause diseases in immunocompromised persons: the non-tuberculosis mycobacteria (NTM). *Mycobacterium kansasii*, which was first discovered in the Kansas state, is the main etiologic agent responsible for lung infections caused by NTM and raises attention because of its co-infection with human immunodeficiency virus (HIV).

Five subspecies of *M. kansasii* (Type I-V) were described and only *M. kansasii* Type I is pathogenic to humans. *M. kansasii* is a Gram-positive bacteria that has a unique cell wall and secretion system, which is essential for its pathogenicity. We undertook a comparative genomics and transcriptomic approach to identify components of *M. kansasii* Type I pathogenicity. Our previous study showed that *espA* (ESX-1 essential protein) operon, a major component of the secretion system, is exclusively present in *M. kansasii* Type I. The purpose of this study was to test the functional role of the *espA* operon in pathogenicity and identify other components that may also be involved in pathogenicity.
This study provides a new molecular diagnostic method for *M. kansasii* Type I infection using PCR (Polymerase Chain Reaction) technique to target the *espA* operon. With detailed manual curation of the comparative genomics datasets, we found several genes exclusively present in *M. kansasii* Type I including *ppsA/ppsC* and *whiB6*, that we believe are involved, or have an effect on ESX-mediated secretion system. We have also highlighted, in our study, the differences in genetic components coding for the cell membrane composition between the five subspecies of *M. kansasii*. These results shed light on genetic components that are responsible for pathogenicity determinants in Type I *M. kansasii* and may help to design better control measures and rapid diagnostic tools for monitoring these group of pathogens.
ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Prof. Arnab Pain, my supervisor, who gave me the opportunity to work in his lab and provided me with suggestions and valuable comments. I wish to thank also Dr. Abdallah M. Abdallah, who is an expert in mycobacteria field and generated the DNA and RNAseq libraries in this study. He gave me much help and instruction and taught me everything I know about mycobacteria; the thesis could not be completed without him. I would also like to express my thanks to Dr. Fathia Ben Rached, who has spent a lot of time reading through the thesis draft, provided me with inspiring advices, she also helped me during experimental work and encouraged me.

I want to thank Prof. Liming Xiong and Prof. Peiying Hong for accepting to be my committee members. Thanks to Dr. Roy Ummels for sending the bacteria from the Netherlands and Shoaib Amini for providing the assembled and annotated genome. I am also deeply indebted to all the other tutors and teachers in Pathogen Genomics Lab group.

Finally, I would like to thank my beloved family and girlfriend. Thank you for your all loving consideration and confidence in me through these last two years, even though 8,000 miles and 5 hours separate us. Thanks to my friends in KAUST, I enjoyed the time we spent here together and thanks to the simple, easy and comfortable life provided by KAUST.
# TABLE OF CONTENTS

Genomic Analysis of Pathogenicity Determinants in *Mycobacterium kansasii* Type I................................................................. 1
EXAMINATION COMMITTEE APPROVALS FORM .......................................................... 2
ABSTRACT ......................................................................................................................... 3
ACKNOWLEDGMENTS .................................................................................................... 5
TABLE OF CONTENTS .................................................................................................... 6
LIST OF ABBREVIATIONS ............................................................................................. 8
LIST OF FIGURES ............................................................................................................ 9
LIST OF TABLES ............................................................................................................. 10

## CHAPTER 1 INTRODUCTION ............................................................................................ 11

1.1 The *Mycobacterium* genus .................................................................................... 11
1.2 *M. tuberculosis* ...................................................................................................... 13
1.3 Non-tuberculous *Mycobacteria* ........................................................................... 17
1.4 The secretion system types .................................................................................... 19
1.5 Aim of this study .................................................................................................... 24
   1.5.1 Evaluation of the espA operon as a diagnostic marker for *M. kansasii* Type I .......................................................... 24
   1.5.2 Functional role of the espA operon in pathogenesis of *M. kansasii* Type I .................. 24
   1.5.3 Comparative genomic analysis of *M. kansasii* Type I to V .......................... 25
   1.5.4 Comparative transcriptomic analysis of *M. kansasii* Type I to V ................ 25

## CHAPTER 2 EVALUATION OF THE ESPA OPERON AS A DIAGNOSTIC MARKER AND PATHOGENICITY DETERMINANT FOR *M. KANSASII* TYPE I ... 27

2.1 Materials and methods ......................................................................................... 27
   2.1.1 PCR of espA operon ...................................................................................... 27
   2.1.2 Growth curve of *M. kansasii* and effects of different suspension methods 28
   2.1.3 Bacterial strains and culture conditions ...................................................... 28
   2.1.4 THP-1 cells and culture conditions ............................................................. 29
   2.1.5 Infection of THP-1 macrophages ................................................................ 29
2.2 Results and discussion ......................................................................................... 30
   2.2.1 espA as a diagnostic marker for *M. kansasii* Type I ................................. 30
   2.2.2 Growth curve of *M. kansasii* Type I .......................................................... 32
   2.2.3 Comparison of different suspension methods ............................................. 32
   2.2.4 espA operon and pathogenicity of *M. kansasii* Type I ............................. 33

## CHAPTER 3 UNIQUE GENES OF *M. KANSASII* TYPE I .............................................. 36

3.1 Material and methods ............................................................................................ 36
   3.1.1 DNA isolation, DNA reads assembly and genome annotation .................... 36
3.1.2 Comparative genomics of the five types of M. kansasii............................. 37
3.1.3 Phylogeny of the five subspecies............................................................... 37
3.1.4 Unique genes in M. kansasii Type I............................................................ 38
3.2 Results and discussion .................................................................................. 39
  3.2.1 Orthologous clusters of the five M. kansasii types ..................................... 39
  3.2.2 Phylogeny of the five M. kansasii subspecies ........................................... 40
  3.2.3 Unique genes in M. kansasii Type I............................................................ 40
3.3 Future Work ................................................................................................. 46

CHAPTER 4 COMPARATIVE TRANSCRIPTOMIC ANALYSIS OF M. KANSASII
TYPE I TO V ........................................................................................................ 47
4.1 Material and methods .................................................................................. 47
  4.1.1 RNA extraction and library preparation ............................................... 47
  4.1.2 RNAseq reads trimming ................................................................. 48
  4.1.3 RNAseq reads mapping ................................................................. 48
  4.1.4 Determination of the Differentially Expressed (DE) genes ................ 49
  4.1.5 Functional analysis of the DE/UDE genes............................................ 49
4.2 Results and discussion ................................................................................ 50
  4.2.1 Quality control of RNAseq data .......................................................... 50
  4.2.2 Read mapping results ........................................................................ 51
  4.2.3 Identification of the DE genes ............................................................. 53
  4.2.4 Functional analysis of the DE/UDE genes ............................................ 54
4.3 Future Work ................................................................................................. 63

CONCLUSION ..................................................................................................... 65
APPENDICES ....................................................................................................... 67
BIBLIOGRAPHY ................................................................................................. 71
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCG</td>
<td>Bacillus Calmette-Guérin Vaccine</td>
</tr>
<tr>
<td>CFP-10</td>
<td>10 kDa culture filtrate antigen</td>
</tr>
<tr>
<td>CFU</td>
<td>Colony forming unit</td>
</tr>
<tr>
<td>ESAT-6</td>
<td>The 6 kDa early secretory antigenic target</td>
</tr>
<tr>
<td>FCS</td>
<td>Fetal calf serum</td>
</tr>
<tr>
<td>HIV</td>
<td>Human immunodeficiency virus</td>
</tr>
<tr>
<td>HSP65</td>
<td>Heat Shocking Protein 65</td>
</tr>
<tr>
<td>IM</td>
<td>Inner membrane</td>
</tr>
<tr>
<td>MAC</td>
<td>Mycobacterium avium complex</td>
</tr>
<tr>
<td>MAI</td>
<td>Mycobacterium avium-intracellulare infection</td>
</tr>
<tr>
<td>MDR</td>
<td>Multi-drug resistant</td>
</tr>
<tr>
<td>MK</td>
<td>Mycobacterium kansasii</td>
</tr>
<tr>
<td>MM</td>
<td>Mycomembrane</td>
</tr>
<tr>
<td>MOI</td>
<td>Multiplicity of infection</td>
</tr>
<tr>
<td>Mtb</td>
<td>Mycobacterium tuberculosis</td>
</tr>
<tr>
<td>MTBC</td>
<td>Mycobacterium tuberculosis complex</td>
</tr>
<tr>
<td>NGS</td>
<td>Next generation sequencing technologies</td>
</tr>
<tr>
<td>NTM</td>
<td>Nontuberculosis Mycobacteria</td>
</tr>
<tr>
<td>OM</td>
<td>Outer membrane (OM)</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate-buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PMA</td>
<td>Phorbol myristate acetate</td>
</tr>
<tr>
<td>RD1</td>
<td>Region of Difference 1</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction fragment length polymorphism</td>
</tr>
<tr>
<td>RPMI</td>
<td>Roswell Park Memorial Institute medium</td>
</tr>
<tr>
<td>RGM</td>
<td>Rapidly Growing Mycobacteria</td>
</tr>
<tr>
<td>SGM</td>
<td>Slowly Growing Mycobacteria</td>
</tr>
<tr>
<td>T7S</td>
<td>Type VII secretion system or ESX secretion system</td>
</tr>
<tr>
<td>Tat</td>
<td>Two Arginine Translocation System</td>
</tr>
<tr>
<td>TB</td>
<td>Tuberculosis</td>
</tr>
<tr>
<td>TDR</td>
<td>Totally Drug Resistance</td>
</tr>
<tr>
<td>XXDR</td>
<td>Extremely drug-resistant</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Fig. 1 Phylogenetic tree of 96 Mycobacteria based on 16S rRNA gene .................................. 15
Fig. 2 Estimated TB incidence rates map in 2014. ................................................................. 16
Fig. 3 Transmission and infection of the tubercle bacilli......................................................... 18
Fig. 4 Different types of gram-negative secretion system. ..................................................... 20
Fig. 5 Current view of *M. tuberculosis* cell envelope............................................................ 22
Fig. 6 Hypothetical ESX-1 secretion system. ........................................................................ 23
Fig. 7 Presence of the *espA* operon as a diagnostic marker for *M. kansasii* Type I .......... 31
Fig. 9 *M. kansasii* suspensions visualized by light microscopy (200X) after Ziehl–Neelsen staining........................................................... 33
Fig. 10 The florescence before and after infection ................................................................... 34
Fig. 11 Functional complementation experiment reveals that *espA* operon plays a significant role in *M. kansasii* Type I pathogenicity ............................................. 35
Fig. 12 The workflow of OrthoMCL and downstream analysis ........................................... 39
Fig. 13 Venn Diagram of the Orthologs and Paralogs shared by the five *M. kansasii* subspecies. .................................................................................................................. 41
Fig. 14 Phylogenetic tree of the five *M. kansasii* subspecies determined by concatemer sequences of 2188 single copy genes covering 72,7331 amino acids............................... 43
Fig. 15 Unique genes in *M. kansasii* Type I ......................................................................... 45
Fig. 16 The workflow of the RNAseq data analysis.............................................................. 50
Fig. 17 Quality control of the RNAseq .................................................................................. 52
Fig. 18 Up-regulated or down-regulated DE genes ............................................................... 56
Fig. 19 Up-regulated and down-regulated overlapping DE genes ...................................... 57
Fig. 20 Heatmaps of the DE genes ....................................................................................... 58
Fig. 21 Up-regulated DE genes GOstats analysis result ........................................................ 60
Fig. 22 Down-regulated DE genes GOstats analysis result .................................................. 62
Table. 1 Reads Mapping Percentage of Three Replicates of *M. kansasii* Type I to V ..... 58
Table. 2 Log2 Fold Change of the UDE Genes and Their Functions............................... 58
CHAPTER 1 INTRODUCTION

1.1 The *Mycobacterium* genus

Mycobacteria are nonmotile bacteria, and they were given this name because of their mold-like pellicle formation when they grow in liquid media. *Mycobacterium* is a genus within *Actinobacteria* phylum\(^1\) that consists of pathogens and saprophytes\(^2\). Other than advanced classification methods, such as 16S rRNA gene sequencing\(^3\), most mycobacteria are classified via physiological and morphological characteristics, which are related to their thick, waxy and lipid-rich cell wall. They can be also recognized as gram-positive bacteria and can be discriminated by Ziehl–Neelsen staining\(^4\).

All mycobacteria are strictly aerobic, and the pathogens within the genus are mainly involved in lung infections. Mycobacteria are well known for two famous diseases: tuberculosis (TB), which is caused by *M. tuberculosis*, and Hansen’s disease caused by *M. leprae*. From a clinical perspective, mycobacteria can be divided into three different groups\(^5\). The first group is composed of obligate human pathogens, including *M. tuberculosis* complex (MTBC), *M. lepraemurinum* and *M. leprae*. These bacteria generally cannot be found in the environment. Bacteria in the second group, such as *M. avium*, are mostly opportunistic pathogens that can infect humans or animals with immune dysfunctions. The third group of bacteria includes non-pathogenic subtypes. Those in the second group and third group are known also as non-tuberculosis mycobacteria (NTM). World Health Organization (WHO) claims that more than two billion people, one-third of the global population, are infected with tuberculosis\(^6\).
fact, TB was one of the most fatal diseases in Europe before 1900\textsuperscript{7}. With the increase in the HIV-infected human population, more and more attention has been given to the opportunistic non-tuberculosis mycobacteria throughout the last decades.

Based on their growth rates, mycobacteria can also be classified into two groups: rapidly growing bacteria (RGM) and slowly growing bacteria (SGM) (Fig.1). The rapidly growing bacteria are the descendants of slowly growing bacteria from an evolutionary perspective\textsuperscript{8}. Evolution of the genus remains controversial because of interspecies similarity based on the different phylogenetic tree building methods\textsuperscript{9}. However some species have been confirmed as closely related, for example, \textit{M. avium} complex (MAC): \textit{M. avium}, \textit{M. intracellulare}\textsuperscript{9}; \textit{M. tuberculosis} complex (MTBC): \textit{M. bovis}, \textit{M. canettii}, \textit{M. tuberculosis}, \textit{M. microti}, \textit{M. caprae}, \textit{M. pinnipedii}, and \textit{M. africanum}\textsuperscript{10}.

Because bacteria have a relatively short doubling time and a simplified genetic communication, they have the ability to evolve quickly and create a large and diverse sample pool for natural selection. \textit{Mycobacterium} species have acquired a various amount of subspecies all related to a single ancestor as a result of such evolution. These bacteria have similar genomes and even share the same hosts\textsuperscript{11}. For example, MAC can cause \textit{Mycobacterium avium-intracellulare} infections (MAI) and zoonotic infections in a variety of tissues and organs, including the lungs, bone marrow, and lymph nodes. Clinical symptoms include a single nodule, nodular bronchiectasis, nodular infiltration and diffusion in immunocompromised patients. It is reported that the lots of MAC strains showed multidrug resistance, preventing the treatment of MAC infections\textsuperscript{12}. Pathogens within the \textit{Mycobacterium} phylum have distinct host
preference\textsuperscript{13}. For example, the host of \textit{M. bovis} is cattle and the host of \textit{M. tuberculosis} is human, while \textit{Mycobacterium pinnipedii} has been isolated from marine mammals. But, other than bacteria that have a strict host preference, some species have also been isolated from non-primary hosts\textsuperscript{14}. The adaptation and co-evolution ability partially explains the spreading of mycobacteria pathogens and why mycobacteria have been “successful” pathogens for a long history. Important host components, such as cytokines (e.g. IL-1/10) or macrophages, forge or regress the inflammatory events. To survive in different host cells, the mycobacteria pathogens need to adapt to the intracellular environment of macrophages and benefit from it. On the other hand, the host cells comprise the replicative niche for most of the life cycle\textsuperscript{15}.

1.2 \textit{M. tuberculosis}

\textit{M. tuberculosis} is a pathogenic bacteria and is the primary cause of most tuberculosis (TB). \textit{M. tuberculosis} has been identified from a pre-Columbian Peruvian mummy, which presented the evidence that \textit{M. tuberculosis} has existed for several thousands of years\textsuperscript{16}. The ancestor of \textit{M. tuberculosis} is still unclear, and no reports have been found in the environment. \textit{M. tuberculosis} is believed to be emerged in Africa, most probably from the Horn of Africa and spread with human migration\textsuperscript{17}. TB is a serious global infectious disease (Fig.2), and, according to Centers for Disease Control and Prevention, TB is often found in people suffering from AIDS\textsuperscript{18}.

“\textit{Mycobacterium tuberculosis} complex" refers to genetically closely related pathogens that are \textit{M. tuberculosis} (obligate human pathogen), \textit{M. caprae} (isolated from goats), \textit{M. bovis} (mostly isolated from cattles), \textit{M. microti} (isolated from vales) and \textit{M.
africanum (isolated from human-beings in Africa). M. tuberculosis owes its pathogenicity to the ability to survive in macrophages and avoid the immune system response. Detailed infection pathology of tuberculosis is illustrated in Fig.3. TB disease usually starts with the inhalation of M. tuberculosis that is later ingested by macrophages. Then, the tubercle bacilli multiplies in macrophages, attracting more macrophages and lymphocytes to migrate to this region and forms an early tubercle. The intracellular life cycle of the bacteria makes detection and elimination by the immune system difficult. After a few weeks, the macrophages die and form a liquefied center in the middle of tubercle, where tubercle bacilli will be released. Dormancy of the disease may occur after this stage. After two to four weeks, most of the bacteria will be eliminated by the immune system. In cases in which the immune system is undermined, for instance, by HIV infection, the tubercle becomes active. Surrounding the granule, a fibroblast layer is formed, which is composed of macrophages and lymphocytes. The inner liquefied center will be enlarged as a result of the multiplication of the bacteria. Finally, the tubercle bursts, thus releasing bacteria. M. tuberculosis is an airborne pathogen and may spread and cause more tuberculosis infection by inhalation.

The detection of tuberculosis starts from the active pulmonary tuberculosis clinical symptoms and abnormal radiographs. The detection should be further confirmed by isolation of M. tuberculosis from clinical materials. Now, more and more molecular methods are available to detect the tuberculosis M. tuberculosis genome, such as the Xpert MTB/RIF, which can also be used for drug susceptibility detection.
Fig. 1 Phylogenetic tree of 96 mycobacteria based on 16S rRNA gene

The phylogenetic tree was generated using the NJ (neighbor-joining method) and using Kimura’s two-parameter model. The tree was rooted using *Nocardia abscessus* DSM 44432T.4.
The Bacillus Calmette–Guérin (BCG) vaccine is currently used against tuberculosis (TB). The vaccine was developed from attenuated *M. bovis* and is one of the most important measures used for prevention. However, the vaccine is not always efficient and drugs must be used to fight *Mycobacterium* infections. Isoniazid, rifampin, ethambutol, streptomycin and pyrazinamide are some of the effective antibiotics for treatment of tuberculosis. Together, they form the “first line” for TB treatment. Other medicines that have side effects or little efficacy and that are not suitable for long-term treatment, such as thioacetazone, kanamycin, and capreomycin, are called “second-line” drugs. Clofazimine, linezolid, amoxicillin and clavulanate, which have no defined roles in tuberculosis treatment, are named the “third line” for tuberculosis treatment. However, drug resistance is still a serious issue. Totally drug resistant MTB is used to refer to the *M. tuberculosis* that are resistant to most of the
first line drugs and at least one of the second line drugs. The term "Totally Drug Resistance MTB (XXDR)" was first used in 2007 when two Italian women were infected with \textit{M. tuberculosis} resistant to both of the first and second line anti-TB drugs\textsuperscript{24}. In the last few years, XXDR cases have also been found in Iran\textsuperscript{25}, the United States\textsuperscript{26}, India\textsuperscript{27} and South Africa\textsuperscript{28}, and no other proper treatments, except third line anti-TB drugs, which may have harmful side effects, are available.

1.3 Non-tuberculous Mycobacteria

Non-tuberculosis mycobacteria (NTM) refer to a group of bacteria outside of \textit{Mycobacteriaceae} complex and \textit{M. leprae}. Reports have shown that more than 50 species of mycobacteria are associated with other human diseases, and up to 25\% to 50 \% of AIDS cases are complicated with NTM infections\textsuperscript{29}.

The niches of NTM are found in various aquatic and terrestrial environments. They infect people by ingestion, inhalation or inoculation from the environment instead of human-to-human contact. NTM are not obligate human pathogens and can survive in a water distribution system\textsuperscript{30}. Mycobacteria are not yet considered as a parameter in Drinking Water Quality Management System, leading mycobacteria to be a potential biohazard for the immunocompromised population\textsuperscript{31}.

Infections by NTM come mainly from the environmental mycobacteria, and no evidences support that NTM could be transmitted from person to person\textsuperscript{32}. NTM particularly cause lung, regional lymph nodes, and skin infections. Although systemic disseminated infections can rarely happen in non-AIDS patients, in Europe, there are
up to 25% to 50% of AIDS patients infected with NTM with often systemic disseminated infections. Five groups of NTM are often linked to the infections of patients with immunodeficiency, including MAC, *M. kansasii*, *M. scrofulaceum*, *M. xenopi* and *M. haemophilum*. Primarily, NTM lung infections are caused by *M. kansasii*.

Fig. 3 Transmission and infection of the tubercle bacilli. Details of the infection are described in the figure showing that after macrophages infections, the pH deviation will inhibit the fusion of early phagosome and lysosome to form phagolysosome.

Nature Reviews | Microbiology
*M. kansasii* (MK) belongs to photochromogenic and slowly growing mycobacteria. The growth temperature ranges from 32°C to 42°C, and the severity of *M. kansasii* infection is proportional to that of HIV infection. In addition to systemic infection, *M. kansasii* more likely causes lung, cervical lymph node, and skin infections. Minocycline, pyrazinamide, and cycloserine are often used as anti-MK drugs. It is also noted that MAC together with *M. kansasii* usually cause systemically disseminated infections in AIDS patients. Drug-resistant *M. kansassii* is rarely reported, because it is always susceptible to rifampin in vitro. Clarithromycin remains active against *M. kansasii* with 100% of isolates displaying susceptibility in vitro.

In 1996, M. Picardeau pointed that there were five different subspecies of *M. kansasii*. He used PCR to amplify Heat Shock Protein 65 (HSP65) and then digested the fragments by *Hae*III and *Bst*II restriction enzymes. Five patterns were shown, revealing five different subspecies. But, only *M. kansasii* Type I has pathogenicity for human. The pathogenicity of *M. kansasii* is poorly understood, and the phylogeny of the five subspecies has not been discussed before.

1.4 The secretion system types

Bacterial proteins secretion is one of the important factors that determine the pathogenicity of bacteria. After the synthesis and secretion by the bacteria, these proteins are either transported to the surface of bacteria, released into the extracellular space or invade host cells. The transportation of typical secretary proteins with signal sequences are mediated by the Sec secretion pathway, while some other proteins are typically secreted by the Sec2 arginine translocation pathway or the Two Arginine...
Translocation (Tat) system\textsuperscript{42}. A molecular study has revealed multiple mechanisms of Gram-negative bacteria secreted proteins, which include six secretion systems, namely the Type I, Type II, Type III, Type IV, Type V and Type VI secretion systems (Fig.4)\textsuperscript{43}.

![Fig.4 Different types of gram-negative secretion system\textsuperscript{44}.](image)

Gram-negative bacteria have two membranes: the inner membrane (IM) and the outer membrane (OM). To pump out particular substances from cytoplasm to extracellular space, the secretion system of bacteria needs different types of proteins to fulfill this task, which can be resolved by either a one-step system or two-steps system. In the one step system, there are three essential components: the inner membrane ATP-binding cassette and the fusion proteins within the cytoplasm and the outer membrane pore. In the two step translocation system, namely Type II and V, proteins located in the inner membrane can be either SecA or Tat translocons. The proteins are folded in the periplasm and then transported out by the outer membrane secretion, which will only allow proteins that are correctly folded to go across the outer membrane\textsuperscript{43}.

Gram-positive bacteria and Gram-negative bacteria have different cell structures. Gram-negative bacteria are coated with lipopolysaccharides and a plasma membrane,
while Gram-positive bacteria contain only one layer constructed by a plasma membrane and peptidoglycans (Fig.5). Indeed, the two types of bacterial protein secretion mechanisms are entirely different\textsuperscript{44}. The cell wall of \textit{Mycobacterium} is the most significant feature of the \textit{Mycobacterium} genus. The covalent linkages between mycolipids provide a thick and hydrophobic layer, which results in drug and stress resistance characters of mycobacteria, and the mycolic acids linked to the cell wall matrix all form a highly hydrophobic region. The peptidoglycan and mycolipid layers are held together by polysaccharides. These overall structures make the \textit{Mycobacterium} cell wall complex and unique, leading to a need of a unique secretion system to release components. Therefore, the biosynthesis of the unique mycobacteria cell wall components can be considered as a potential target for the treatments of mycobacterial infections. The differences between the biochemical composition of Gram-positive and Gram-negative cell envelopes indicate that another type of secretion system in the Mycobacterial membranes must exist.

The ESX-1 secretion system has been previously described using a DNA chip, which has predicted its existence\textsuperscript{45}. The first experimental evidence of this secretion system was determined when Pym and colleagues discovered that ESAT-6 was unable to be secreted outside of the cell envelope until the RD1 (Region of Difference 1) was inserted into BCG vaccine strains\textsuperscript{46}. Any gene knockdowns in the RD1 region (especially Rv3870, Rv3871, and Rv3877) influence the secretion of ESAT-6 and CFP-10. This system is known as the ESAT-6 secretion system, ESX-1 secretion system, or Type VII secretion system (T7S)\textsuperscript{47}. 
Fig. 5 Current view of *M. tuberculosis* cell envelope.
The complex cell wall has an inner membrane and a cell wall complex. The cell wall complex contains three different layers and three different covalent linkages: the linkage between peptidoglycan and arabinoglycan, the linkage between arabinoglycan and mycomembrane layer, and the covalent linkage between mycomembrane. Also, there are many different types of free lipids in the mycomembrane, that are linked to the capsule layer, which is composed by polysaccharide.

There is no sound base model illustrating how the ESX-1 secretion system works. The stabilities of ESAT-6 and CFP-10 depend on each other, and these two proteins form a 1:1 dimer structure. Yeast two-hybrid experiments reveal that Rv3870 may be combined with Rv3871, and Rv3871 is combined with CFP-10. Therefore, Rv3871 firstly binds to ESAT-6 / CFP-10 complex and then binds to the carbon terminal of CFP-10. Next, Rv3871 interacts with Rv3870 and forms an active ATPase in the membrane (Fig. 6). Chaperone Rv3868 may further promote the secretion of ESAT-6 and CFP-10 proteins. The Rv3871 / Rv3870 complex resembles Type I and Type IV secretion systems by forming a six-transmembrane structure and having a central hollow structure, which helps the secretion of the substrate.
The ESX-1 secretion mechanism became more complicated when the second gene cluster (Rv3614c ~ Rv3616c) was found. A previous study, in our lab, showed that \( espA \) operon is exclusively present in \( M. kansasii \) Type I, which may be involved in its pathogenicity. Rv3616c (\( espA \)), CFP-10, and ESAT-6 are co-secreted through the ESX-1 secretion system. Rv3614~Rv3616c are homologous to Rv3864 ~ Rv3867 in the RD1 region, and \( espA \) may form an operon together with Rv3864~Rv3867. Rv3614c interacts with Rv3882c protein and promotes the secretion of ESX secretion proteins. Rv3815c is a secreted protein and, in the absence of \( espA \) or Rv3815c, ESAT-6 / CFP-10 complex can be produced but cannot be secreted, which means that all substrates of ESX-1 (\( espA \), Rv3815c, ESAT-6, and CFP -10) are interdependent in the secretion process.

Due to the distinct structure of the cell wall of \( M. tuberculosis \), mycobacteria have been speculated to have unique secretion system. Surprisingly, \( M. tuberculosis \)
contains five such secretion systems. ESAT-6 family genes in *M. tuberculosis* genome include 11-12 gene clusters, in which four clusters have a higher homology with ESX-1: ESX-2 (Rv8884c ~ Rv8895c), ESX-3 (Rv0282 ~ Rv0292), ESX-4 (Rv3444c ~ Rv3450c) and ESX-5 (Rv782 ~ Rv798). Comparative genomics phylogeny revealed ESX gene duplication during evolution generating thus five different ESX systems in mycobacteria, and the order is ESX-4, ESX-1, ESX-3, ESX-2 and ESX-5\(^49\). ESX-1 is related to pathogenicity, and evidences show that ESX-3\(^50\) and ESX-5\(^51\) are also associated with protein secretion. No associations have been reported for the other two ESX systems. The ESX-1 system is also absent in *M. avium* and *M. ulcerans*, which indicates other systems can also help to build up pathogenicity, which is rare\(^52\). However, the ESX-1 secretion system is widespread among *Mycobacterium* pathogens and only ESX-4 and ESX-3 are omnipresent in all mycobacteria\(^53\). The ESX-1 secretion system is also found in *M. marinum* and non-pathogenic mycobacteria, *M. smegmatis*. Studies of the secretion system of *M. kansasii*, who have five ESX systems, will help to understand ESX system.

1.5 Aim of this study

1.5.1 Evaluation of the *espA* operon as a diagnostic marker for *M. kansasii* Type I

As we discussed above, *espA* operon is present exclusively in *M. kansasii* Type I, and, in this study, we have used *espA* as a diagnostic marker for *M. kansasii* Type I, which is the only pathogenic one to humans amongst the five sub-types.

1.5.2 Functional role of the *espA* operon in pathogenesis of *M. kansasii* Type I
We have then explored the function of espA by using functional complementation experiments in wild-type \textit{M. kansasii} Type V. We transfected a \textit{M. kansasii} pMT3-espA-GFP plasmid, which we believe will complement the ESX-1 secretion system, and compared the intracellular survival of the engineered bacteria with WT \textit{M. kansasii} Type I and WT \textit{M. kansasii} Type V, which will help us to understand the function of espA operon in pathogenicity.

1.5.3 Comparative genomic analysis of \textit{M. kansasii} Type I to V

The genome size of the five \textit{M. kansasii} subspecies are above 6.00 Mb (6.50 Mb for \textit{M. kansasii} Type I, 6.40 Mb for \textit{M. kansasii} Type II, 6.08 Mb for \textit{M. kansasii} Type III, 6.45 Mb for \textit{M. kansasii} Type IV and 7.11 Mb for \textit{M. kansasii} Type V) with around 6,000 genes for each of the subspecies (6165 genes for \textit{M. kansasii} Type I, 6280 genes for \textit{M. kansasii} Type II, 5959 genes for \textit{M. kansasii} Type III, 6223 genes for \textit{M. kansasii} Type IV and 6767 genes for \textit{M. kansasii} Type V)\textsuperscript{54}. In this part of the study, we uncovered the phylogenetic relationships between the five subspecies and identified genes that are uniquely present across all species from a genome-wide orthology analysis followed by manual curation. We particularly focused our analysis on genes uniquely present in \textit{M. kansasii} Type I, which may provide some genomic clues to answer why \textit{M. kansasii} Type I is the only pathogenic subspecies for humans.

1.5.4 Comparative transcriptomic analysis of \textit{M. kansasii} Type I to V
We used transcriptomic profiling by RNA-seq to analyze the gene expression catalogue of all *M. kansasii* subtypes, and, after comparative analysis of the transcripts at genome-wide level, we wanted to specifically analyze the subset of genes that are either up- or down-regulated in *M. kansasii* Type I compared with other four subspecies with an aim to identify genetic components that may be differentially expressed in *M. kansasii* Type I.
CHAPTER 2 EVALUATION OF THE ESPA OPERON AS A DIAGNOSTIC MARKER AND PATHOGENICITY DETERMINANT FOR M. KANSASII TYPE I

2.1 Materials and methods

2.1.1 PCR of espA operon

The clinical bacteria samples were gathered from Radboud University Medical Center in the Netherlands from patients and the genomic DNA were isolated in the same way as described in 3.1.1 in our lab and genotyped based on their 16S-23S space. The primers were designed using Primer Blast based on the unique regions of the espA operon, and the primers were examined with BLAST and showed a strong specificity only to M. kansasii Type I, which will make sure the fragment will be only amplified in M. kansasii but not with other types. The amplification by PCR of the espA gene was conducted with 1 µL of PfuTurbo Cx hotstart DNA polymerase (2.5 U), 1X Pfu reaction buffer with MgSO4, 20 ng of DNA template (extracted using protocol described next chapter), dNTPs, espA specific primers (Forward: 5’-GTTCGTCTCGATTTTCGCAGC-3’ & Reverse: 5’-GAATCACGCGCCTGGATGAC-3’) and distilled water to a final volume of 100µL. 

i pr M primers (Forward: 5’-ACCATCGTGGGCTACTTCAC-3’ & Reverse: 5’-CTCACCACCTACACCCTGGTTCC -3’) were used as a control which should give a 194bp fragment. The i pr M gene, encoding for a glutamate-ammonia ligase adenylytransferase, is present in all of the five subspecies, which allows it to be used as a control to show the presence of the genomic DNA. The cycling conditions for the espA operon consisted of preheating at 95°C for 2 minutes, followed by 30 cycles of
denaturation at 95°C for 30s, annealing at 62°C for 30s and extension at 72°C for 30s. The final extension was performed at 72°C for 10 minutes after 30 cycles. For *iprM* amplification, the steps were the same as in the amplification process used for *espA* operon, except the annealing temperature for *iprM* primers was 67°C.

2.1.2 Growth curve of *M. kansasii* and effects of different suspension methods

*M. kansasii* belongs to SGM and plotting the growth curve will help to understand the growth status. *M. kansasii* was grown in Middlebrook 7H9 broth with 0.5% glycerol, 0.05% polysorbate 80 and ADC (0.85% Sodium Chloride, 5% Bovine Albumin, 2% Dextrose, 0.003% Catalaseper) at 37°C. The OD$_{600nm}$ was measured at different time points. Two different methods were used to disrupt the mycobacteria clumps that are generated because of the waxy cell wall. Before homogenization, the bacteria were cultivated until the OD$_{600nm}$ reached 0.8 to 1.0. In the first method, the bacterial culture was vortexed for five minutes. The second method was to expel the bacteria culture several times through a 25.5 gauge needle. Then, Ziehl–Neelsen staining was performed after each treatment and was visualized under an optical microscope (Axiovert 40 CFL, Carl Zeiss).

2.1.3 Bacterial strains and culture conditions

The bacterial strains, *M. kansasii* Type I, *M. kansasii* Type V, *M. kansasii* Type V were transfected with the plasmid pMT3-espA-GFP$^{58}$, which were obtained from Dr.
Roy Ummels (Department of Medical Microbiology and Infection Control, VU University Medical Center, Amsterdam, Netherlands). The \textit{M. kansasii} Type V had been transformed with pMT3-eapA-GFP to complement the ESX system by expressing \textit{in vivo} GFP-tagged EspA. Mycobacterial cultures were grown (OD$_{600\text{nm}}$ = 0.8–1.0) in Middlebrook 7H9 broth with 0.5% glycerol, 0.05% polysorbate 80 and ADC (0.85% Sodium Chloride, 5% Bovine Albumin, 2% Dextrose, 0.003% Catalaseper) at 37$^\circ$C.

2.1.4 THP-1 cells and culture conditions

The human monocytic cell lines THP-1 that derived from an acute monocytic leukemia patient were grown in Roswell Park Memorial Institute medium (RPMI)-1640 complemented with 10% Fetal Bovine Serum, 100 $\mu$g/mL penicillin and 100 $\mu$g/mL streptomycin at 37$^\circ$C, 5% CO$_2$.

2.1.5 Infection of THP-1 macrophages

Cultures of the different types of \textit{M. kansasii} (Type I, Type V, Type V+ pMT3-eapA-GFP) with an OD$_{600\text{nm}}$ = 0.8-1.0 were prepared. Simultaneously, THP-1 cells were counted using a hemocytometer and diluted to seed half a million cells per well in 24 well plates in the presence of 25 ng/mL phorbol myristate acetate (PMA), in order to allow the cells to differentiate into macrophages and adhere overnight at 37$^\circ$C. After their differentiation, the complete media was replaced with RPMI+10% FCS without antibiotics and incubated for three hours. The macrophages were incubated in
triplicate with the three types of *M. kansasii* (Type I, Type V, Type V+ pMT3-eapA-GFP), with a multiplicity of infection (MOI) = 5, or incubated with only media. After two hours of infection, the macrophages were washed with PBS and incubated with fresh media without antibiotics. The cells were then treated with 1 mL of 1% TritonX-100 for ten minutes at 37°C and collected at different time points after infection (0, 24, 48 and 72 hours). Then, the obtained lysates were diluted to 1:10, 1:100 and 1:1,000, and 100 µL of each was subsequently spread on Middlebrook 7H10 agar plates. The survival rate of the bacteria was evaluated as the percentage of CFU at different time points taking the number of CFU with time point 0 as the reference.

2.2 Results and discussion

2.2.1 espA as a diagnostic marker for *M. kansasii* Type I

Our previous study\(^{54}\) showed that the *espA* operon exists only in *M. kansasii* Type I, which prompted us to evaluate it as a molecular diagnostic marker. After DNA extraction from clinical samples and cultures of the five *M. kansasii* subtypes, PCR was performed. As depicted in Fig.7, the *iprM* gene, which is present in all of the five subtypes of *M. kansasii*, was amplified in all of the samples by PCR, while only *M. kansasii* Type I showed the presence of the *espA* operon by PCR using two primers: (Forward: 5’-GTTCGTCTCGATTTCGCAGC-3’ & Reverse: 5’-GAATCACGCGCCTTGATGAC-3’), giving a size product equal to 199bp. We can conclude that the designed primers for the amplification of the *espA* operon are specific and could be used as a new target for a quick molecular diagnostic of *M.*
kansasii Type I. However, the primers will also be verified in other mycobacteria that are close related to M. kansasii such as M. tuberculosis who has the espA operon as well.

Fig. 7 Presence of the espA operon as a diagnostic marker for M. kansasii Type I.
The result were separated in two gels due to the close molecular weight of the PCR products of iprM fragment (194bp) and espA operon fragment (199bp). The gel showed on the top is the PCR result using espA operon primers while the lower one is PCR result using iprM primers. Lane M: DNA marker, Lane 1: negative control with water as the template. Lane 2: PCR result with genomic DNA from the culture of M. kansasii Type I as template. Lane 3-5: DNA template from clinical samples infected by M. kansasii Type I. Lane 6: PCR result with genomic DNA from the culture of M. kansasii Type II as the template. Lane 7, 8: DNA template from clinical samples infected by M. kansasii Type II. Lane 9: PCR result with genomic DNA from the culture of M. kansasii Type III as the template. Lane 10, 11: DNA template from clinical samples infected by M. kansasii Type III. Lane 12: PCR result with genomic DNA from the culture of M. kansasii Type IV. Lane 13, 14: DNA template from clinical samples infected by M. kansasii Type IV. Lane 15: PCR result with genomic DNA from the culture of M. kansasii Type V as the template. Lane 16-18: DNA template from clinical samples infected by M. kansasii Type V. The lanes with "+" on the top are the positive PCR controls which used genomic DNA from M. kansasii cultures as template while the "-" lanes are the negative controls using water as template.
2.2.2 Growth curve of *M. kansasii* Type I

![Growth curve of M. kansasii Type I in Middlebrook 7H9 media.](image)

*Bacteria were grown to late log phase before being used for the experiment. To quantify bacterial numbers, 500μL of bacterial suspension in 7H9 was measured by spectrophotometry at 600nm. Values were blanked using bacteria-free 7H9.*

*M. kansasii* belongs to SGM, which typically takes a long time to grow. It is important to understand that the doubling time of *M. kansasii* takes 19.6 to 20.9 hours\(^5\). Thus, in the first few days, the culture has no obvious changes. After around eight days, the *M. kansasii* goes to exponential phase (Fig.8). Bacteria are suitable for further experiments when the OD\(_{600}\) value is equal to 0.8-1.0 ranges.

2.2.3 Comparison of different suspension methods

*M. kansasii* forms clumps during culture because of their waxy cell wall. They create a durable hydrophobic layer that could influence the infection process. Before treatment, the bacteria were stuck together, and some bacteria formed large clumps, which make it difficult for macrophages to engulf them. This phenomenon can, thus, influence the CFU count accuracy during macrophage infection experiments. So,
before applying the culture for further experimentation, the bacteria cultures were treated by vortexing or passaging through a needle. Thus, the clumps have been dispersed and show much more homogeneity compared with the original culture, as shown in Fig.9.

Fig.9 M. kansasii suspensions visualized by light microscopy (200X) after Ziehl–Neelsen staining. (A) M. kansasii suspension before treatment. (B) M. kansasii bacteria suspension after 5 minutes of vortex. (C) M. kansasii bacteria suspension after 20 times passages through a 25.5 gauge needle.

2.2.4 espA operon and pathogenicity of M. kansasii Type I

The pMT3-espA-GFP is expressed before and after infection (Fig.10). As shown in Fig.11 the colony forming units (CFU) per milliliter did not present a significant difference in the first two days, which indicates that the espA operon would not influence the invasion ability of M. kansasii. After two days, we can observe a significant survival rate decrease for wild type M. kansasii Type V. The amount of growth of each strain does not differ significantly, due to their relative long doubling time, as we can see from the figure. Indeed, the M. kansasii Type V complemented with the espA operon did not fully recover as well as wild-type M. kansasii Type I, which suggests that espA operon itself cannot explain the entire pathogenicity of M. kansasii Type I. Thus, other factors may also be involved in the M. kansasii Type I invasion or intracellular survival process. However, complementation with the espA
operon facilitates for the better survival of *M. kansasii* Type V against digestion from macrophages. It is probable that this increased survival is caused by the complementation by the ESX secretion system, which might be important for surviving within macrophages, based on our results. A further *espA*-operon knock down experiment is also needed to support our result.

Fig. 10 The fluorescence before and after infection. *M. kansasii* was observed under fluorescence microscopy (800X, Zeiss). Fluorescence observed before (A) and after (B) infection, the bacteria suspension from the Type V with pMT3-*espA*-GFP plasmid.
Fig. 11 Functional complementation experiment reveals that espA operon plays a significant role in M. kansasii Type I pathogenicity. The percentage survival curve of three strains at various time points post infection. The value of time point 0 has been set as the initial value. *: significant difference $p<0.05$, using Student’s t –test with three replicates. ***: highly significant difference $p<0.01$ using Student’s t –test with three replicates.
CHAPTER 3 UNIQUE GENES OF *M. KANSASI*I TYPE I

3.1 Material and methods

3.1.1 DNA isolation, DNA reads assembly and genome annotation

The DNA library was generated by Yara Alzahid\textsuperscript{54}, and the process will be briefly discussed here. The *M. kansasii* Type I-V DNA molecules were isolated using phenol-chloroform protocol\textsuperscript{57}. Briefly, after centrifugation of the bacteria cultures at 3,000 rpm for 10 minutes and decantation of the supernatant, the bacterial pellet was stored at -80°C overnight. After thawing, the bacteria were homogenized in 10mL of TE buffer. We, then, added 10mL of 2:1 chloroform-methanol mixture, and centrifugation at 2500g for 20 minutes was performed to discard the organic-aqueous layer. After an air-drying step, we added 5mL of TE buffer and 1M Tris-HCl to adjust the pH at 7.2. Lysozyme was incorporated to a final concentration of 100μg/μL and incubated at 37°C overnight. After digestion of the cell wall, we added 1/10 of 10% SDS buffer and 1/100 of Proteinase K and 10μL RNase (Invitrogen), and the mixtures were incubated for 3 hours. An equal volume of 25:24:1 phenol-chloroform-isoamyl alcohol was added, and the mixture was placed on the platform rocker (Qiagen) for 30 minutes. After centrifugation, the upper layer was transferred, and an equal volume of 24:1 chloroform-isoamyl alcohol was added to remove the phenol. Then, the aqueous layer was discarded by centrifugation, and 3M NaAc (pH 5.2) and 1 volume of isopropanol were incorporated. The mixture was placed at 4°C overnight, and cold 70% ethanol was added before another centrifugation step. The obtained DNA pellet was then air-dried and rehydrated in TE buffer.
Then, the obtained genomic DNA molecules from *M. kansasii* Type I-V were sheared into 500bp fragments using Covaris system. Paired-end, Matepair and PCR-Free libraries were obtained by following manufacturers’ instructions (Illumina kits) and were sequenced with HiSeq 2000 platform (Illumina). IDBA\(^60\) was performed for De-novo DNA assembling and followed by scaffolding with SSPACE\(^61\). Gap-filling was performed using IMAGE and Gap Filler\(^62\). The assembled genomes were valued by REAPR\(^63\) and annotated by PROKKA\(^64\) and RAST\(^65\).

3.1.2 Comparative genomics of the five types of *M. kansasii*

To determine the paralogous groups of the five *M. kansasii* species, the OrthoMCL \(^66\) program was used. Predicted protein sequences of the five *M. kansasii* subspecies were analyzed using OrthoMCL with a 50% identity cut-off and inflation parameter of 1.5. Fig.12 shows the workflow of OrthoMCL and downstream analysis. The OrthoMCL groups’ results were carefully examined by the BAM file, generated using the PCR-free library reads that mapped to *M. kansasii* Type I assembled genome using SMALT unique mapping with k-mer word length 11 and stepsize 2 (-k 11 and -s 2).

3.1.3 Phylogeny of the five subspecies

To produce phylogeny of the five *M. kansasii* subspecies, the *M. tuberculosis* H37Rv genome\(^67\) and annotation GFF file (RefSeq: NC_000962.3) were used for clustering.
The clustering of the protein sequences was done using OrthoMCL\textsuperscript{66}, in the same manner as shown in Fig.12, and the orthologs that contain six genes, one from \textit{M. tuberculosis} and one from each of the \textit{M. kansasii} subtypes, also called one to one orthologs, were concatenated together as concatemers to reveal the phylogeny of the five subspecies. These groups contain 2,188 genes, and they were aligned with Muscle\textsuperscript{68}. The phylogenetic tree was, then, generated using the Maximum Likelihood Method with RAxML\textsuperscript{69}. Bootstrapping was performed using RAxML’s bootstrapping algorithm (50,000 iterations).

3.1.4 Unique genes in \textit{M. kansasii} Type I

Genes unique to the \textit{M. kansasii} Type I genome were found by comparing its DNA to that of the other four types. Further analysis was performed through GOstats\textsuperscript{70}, a functional annotation tool for analyzing protein function in which they might have significant enrichment. Briefly, all of the manually curated unique CDSs were searched online via the NCBI Non-redundant database, using BlastP. Then, the matched sequences were scanned through InterproScan\textsuperscript{71}. Subsequently, the BLAST hits of each sequence were mapped to the GO database\textsuperscript{72} using a 0.00001 E-value cutoff. The GO terms generated after this step will be used to generate a GO term matrix for GOstat analysis.
Fig. 12 The workflow of OrthoMCL and downstream analysis. OrthoMCL program analyzes the result from the BLAST and groups these genes based on the similarities: both the similarities of the genes within the same species, which are called paralogs, and the genes in different species, called orthologs. To verify the uniqueness of the genes that are present only in *M. kansasii* Type I, the PCR-free Illumina reads are further mapped against *M. kansasii* Type I. The BAM files generated were viewed in Artemis and the uniqueness was examined.

3.2 Results and discussion

3.2.1 Orthologous clusters of the five *M. kansasii* types
The orthologous clusters can be seen in Fig.13. Both of the unique orthologs that exist only in one subspecies and orthologs shared by different subspecies are shown in the figure. The total numbers of clusters that are present in all types (3,816) have been further exploited in Chapter 4, and the number of CDSs that are uniquely present within types I-V is, respectively, 341, 427, 350, 533 and 498. All of the results within each group have been examined. Unique genes present within the unique cluster of *M. kansasii* Type I that are shown in Fig. 13 have further been investigated.

3.2.2 Phylogeny of the five *M. kansasii* subspecies

The phylogenetic tree shown in Fig.14 reveals that *M. kansasii* Type V has the closest linkage with *M. kansasii* Type I and that they are sharing the same putative ancestor, while Type II, III, and IV are grouped together. This tree reveals the potential evolution pathway of the five *M. kansasii* subspecies and how close they are to each other.

3.2.3 Unique genes in *M. kansasii* Type I

From comparing the five *M. kansasii* types, using OrthoMCL, unique orthologous groups were identified. In type I, there are eight unique orthologous groups and 311 genes that are not present in the orthologous groups, which are also called singleton genes. Similarly, type II has 18 orthologous groups, with 50 genes amongst those groups, and 377 singleton genes. Type III has 16 groups, with 40 genes within these groups, and 310 singletons. Type IV has 13 unique orthologous groups, containing 45
genes, and 588 singleton genes. Type V has 40 groups, with 92 genes, and 406 singleton genes. All of the unique genes were manually checked for their uniqueness with the BAM file that mapped against the *M. kansasii* Type I genome, therefore supporting the fact that they are unique to each type. We prepared a Venn diagram showing the numbers of shared and unique orthologous clusters while comparing the five types against each other.

![Venn Diagram of the Orthologs and Paralogs shared by the five *M. kansasii* subspecies.](https://www.dropbox.com/s/i4w7o6s4fsan8tr/OrthoMCL_gene.xlsx?dl=0)

Gene ontology analysis reveals that among these genes (368 unique genes in 341 clusters), a large number are hypothetical proteins or without a significant hit (125 sequences). Among the sequences that had been annotated, 35 of them are involved in
cellular part. Interestingly, among these 35 genes, 31 of them are related to the cell membrane or cell wall which have also been examined via Phobius\textsuperscript{74} while only four of them have relation to the cell part. This result agrees with our hypothesis, which is that the main difference between \textit{M. kansasii} Type I and the other four subspecies lies in the ESX secretion system. The other genes are involved in several biological processes and molecular functions, such as the gene expression regulation process, nucleic acid binding, endonuclease, and nucleotide biosynthesis functions. Since these genes are enriched, the unique genes may be highly regulated when compared with other genes in the genome. Also, several epigenome-related genes, such as those responsible for DNA methylation and demethylation, and DNA alkylatation genes are enriched compared with the whole genome. These biological processes are enriched in unique Type I genes, which implies extensive regulation of the genes which may be involved in the pathogenicity of \textit{M. kansasii} Type I (Fig.15).

Furthermore, among the unique genes, a large number of them come from members of the PE/PPE family. PE/PPE family proteins are a group of proteins, which have ProGlu or ProProGlu motifs in their N-terminal. This family is present only in mycobacteria and are widely expanded in SGM\textsuperscript{75}. They may promote the antigenic diversity of mycobacteria\textsuperscript{76}. PE and PPE proteins usually form a dimer, and their functions are still poorly understood\textsuperscript{77}. The presence of the PE/PPE proteins is not crucial for pathogenicity, as we have also found the PE/PPE family proteins in the other types. The uniqueness of these genes is also supported in the appendices Fig. S1.
Fig. 14 Phylogenetic tree of the five *M. kansasii* subspecies determined by concatemer sequences of 2188 single copy genes covering 72,7331 amino acids. *M. tuberculosis* H37Rv strain has been used as an outgroup. The scale bar represents 1% sequence divergence. The numbers on each branch represent the bootstrapping values. MK1: *M. kansasii* Type I, MK2: *M. kansasii* Type II, MK3: *M. kansasii* Type III, MK4: *M. kansasii* Type IV, MK5: *M. kansasii* Type V, Mtb: *M. tuberculosis*. The concatenated sequences could be got from the following website: https://www.dropbox.com/s/58vjfotni00casy/Concatemers.fasta?dl=0.

WhiB protein, which regulates ESX-1 secretion system upstream of the ESX-1 locus, is considered as a component of the ESX secretion system and has been found uniquely present in *M. kansasii* Type I. WhiB family proteins are known as both transcription regulators and redox state controllers of targeted proteins. It has been shown that WhiB protein acts differently according to the metal cation environment. Indeed, after mycobacteria have been engulfed by the macrophages, they will be present in the phagovacuole, which is rich in metal cations. The *whib6* gene is found in the *M. kansasii* Type I unique genes pool, which provides the evidence that the Whib6 protein may have an important role in pathogenicity of *M. kansasii* Type I.

Other than *whiB* family members, other interesting genes, including the *ppsA, ppsC* genes, were also found in the unique gene pool. *ppsA/C* genes are involved in the
phenolpthiocerol and phthiocerol dimycocerosate biosynthesis, which are linked to malonyl CoA and found to participate in the cell membrane lipids biosynthesis process. It is reported that the \textit{pps} gene family plays an important role in cell envelope architecture and permeability.

The \textit{irtA} gene, which has proved to be essential for \textit{M. tuberculosis}, is also a unique gene in \textit{M. kansasii} Type I. The protein was first discovered in membrane fractions in 2003 and the mutant showed reduced uptake of Fe-carboxymycobactin, which indicates that \textit{irtA} gene is responsible for iron homeostasis and energy coupling for transporting the substrates across the membrane.

Meanwhile, other genes, including ESX-1 secretion system components \textit{EspB}/\textit{EspD}/\textit{EspF}/\textit{EspH}, the \textit{mtbH} gene (involved in biogenesis of hydroxyphuyloxazoline-containing siderophore mycofactin), and lipase \textit{lipU}, are also interesting potential virulence components of \textit{M. kansasii} Type I.
Fig. 15 Unique genes in *M. kansasii* Type I GOstats result of the enriched GO terms in cellular component (green), molecular functions (red) and biological process (blue) of *M. kansasii* Type I compared with the whole genome. The color refers to the Log2OddsRatio, which is an indicator of numbers of genes enriched in particular GO terms. The size of the circle represents the P-value.
3.3 Future Work

First, the unique genes that we have explored using OrthoMCL and manually examined need to be confirmed using PCR. Also, other than the \textit{espA} operon diagnostic marker that targets Type I, other genes that are unique to each type can also be used as typing markers, which can be used to reveal the presence of different subspecies. Also, the drug-resistance gene candidates can be explored, which may be useful to illustrate the drug-resistance of \textit{M. kansasii}. However, due to lack of available time for this thesis, it is difficult to examine more these genes. Indeed, the function of the genes that are unique to \textit{M. kansasii} Type I should be examined using gene knock down or complementary molecular approaches. Additionally, the essential genes of \textit{M. kansasii} should be interesting to investigate, which can be explored using Transposon-directed insertion site sequencing (TraDIS)\textsuperscript{88}.
CHAPTER 4 COMPARATIVE TRANSCRIPTOMIC ANALYSIS OF *M. KANSASII* TYPE I TO V

4.1 Material and methods

4.1.1 RNA extraction and library preparation

The RNAseq library was generated by Dr. Abdallah M. Abdallah and Yara Alzihid. In brief, the five *M. kansasii* types were grown on Middlebrook 7H10 agar supplemented with 0.05% Tween 80, Glycerol and 10% OADC. Single colonies were then taken from the agar culture and transferred to Middlebrook 7H9 broth, which is supplemented with 0.05% Tween 80, glycerol and ADC enrichment. After the bacteria had reached exponential phase (OD$_{600nm}=0.8-1.0$), the bacteria were then separated into three biological replicates. Then, the RNA was extracted using the Trizol protocol. Briefly, the bacteria cultures were centrifuged at 3500rpm for 15 minutes, suspended in 1mL Trizol and incubated for five minutes. Then, 500µL of zirconia beads were added and treated with beating at maximum speed for 30 seconds, six times. Then, the mixture was centrifuged, and the upper layer was incubated with 200µL of chloroform. After centrifuging at 4°C at maximum speed for 20 minutes, an equal volume of isopropanol was added to the aqueous layer. The mixture was centrifuged at 4°C at full speed and the supernatant was discarded. 1.5mL of 70% cold ethanol was added and centrifuged for 10 minutes. The ethanol was discarded and the RNA was air-dried. The RNA was suspended in the proper amount of RNase-free water and incubated at 60°C until all of the RNA became totally dissolved. The RNA was then stored in -80°C. DNA was removed using Turbo DNase, and rRNA
was removed using the Introgen Ribominus Kit. Then, the strand-specific libraries were achieved using TruSeq kit following the manufactory manual (Illumina), and the RNAseq was performed on Hiseq 2000 platform (Illumina).

4.1.2 RNAseq reads trimming

The RNAseq reads obtained from Hiseq 2000 were first trimmed using the Trimomatic program\textsuperscript{89} (LEADING: 3 TRAILING: 3 ILLUMINAACLIP: adapters list: 2:30:10 SLIDINGWINDOW: 4:15 MINLEN: 36). After this process, the reads for which leading or tailing quality was below 3 or contains N bases had been removed. The program trimmed the reads in a 4-base window and cut the fragments once the average quality in the window was below 15. The Illumina adapters were also removed in this process, and the reads with a length below 36 were withdrawn. After this step, the reads became "clean reads" and were ready for mapping to the genome sequence.

4.1.3 RNAseq reads mapping

The “clean reads” were further mapped to the annotated genome \emph{M. kansasii} Type I generated above after trimming. To minimize the errors of cross mapping, we took only the one-to-one orthologous genes generated above of the five species, which are 3416 genes out of 3816 core genes. To compare the difference between \emph{M. kansasii} Type I and the four other types, the reads were mapped against the \emph{M. kansasii} Type I annotated genome using Tophat\textsuperscript{90}. Tophat2 uses bowtie2\textsuperscript{91} as the core mapping tools.
Because prokaryotes do not have alternative splicing sites, the parameter “--no-novel-juncs” was used and the library parameter “--library-type=fr-firststrand” was used for the strand-specific library.

4.1.4 Determination of the Differentially Expressed (DE) genes

To identify the genes that are differentially expressed, three methods based on different algorithms were used. The first method, Cuffdiff, is based on Fragments Per Kilobase of transcript per million mapped reads (FPKM); it was used with q value <0.01 and Log2 Fold Change >2. The two other methods, DEseq2 and BaySeq, are based on the reads counting of each gene. We selected the genes with padj<0.01 or False Discovery Rate (FDR) <0.01 and Fold Change >2. The selected genes were then separated into two parts: the up-regulated genes and down-regulated genes based on their expression levels. The list of the overlapping DE genes of these three methods have been produced and further analyzed in the following steps.

4.1.5 Functional analysis of the DE/UDE genes

The functions of the overlapping DE genes were analyzed via the Blast2GO program and GOstats in the same process as described in 3.1.4.
Fig. 16 The workflow of the RNAseq data analysis.

Trimomatic was used to generate clean reads (Parameters: LEADING:3 TRAILING:3 ILLUMINAALP:adapters.list:2:30:10 SLIDINGWINDOW: 4:15 MINLEN:36). RUST and PROKKA were used to annotate the genome (Parameters: e value cutoff:1e-06 Kingdom: bacteria Rnammer was used to removal rRNA). Reads were mapping against the genome using TopHat2 (--no-novel-juncs --library-type=fr-firststrand) and the read count was performed with HTseq(Parameters: Mode:union Strand: reverse Type:CDS) and analyzed with DEseq2 and Bayseq (BaySeq: FDR< 0.01, log2 FC>2; DESeq: padj <0.01, log2 FC>2). The RNAseq data was also analyzed with CuffDiff2 (q value<0.01 log2 FC>2).

4.2 Results and discussion

4.2.1 Quality control of RNAseq data

The quality of the RNAseq data is crucial for further study. To examine the quality of the RNAseq reads, RNAseq reads clustering plots and volcano matrix plots were generated. Fig.17 shows that the three biological replicates of each subtype are clustering together, which suggests the three replicates within each type correlate with each other. The similarity of each replicate in the different subspecies is necessary for
the analysis below. The volcano matrix shows that genes are differentially expressed, as we can see from the red dots in Fig.17 (B).

4.2.2 Read mapping results

The read mapping results can be seen in Table.1. The entire read-mapping rate in our case is satisfactory. Indeed, the majority of the genes are shared by the five subspecies, and the number of genes that we selected to reveal the RNAseq profile is 3416, which is more than half of the annotated genes in all of the subspecies. Also, the statistic Student's test was performed and no significant differences have been found among the three replicates of each type, which means that the result may be used for further analysis.
Fig. 17 Quality control of the RNAseq.
(A) Clustering of the 15 replicates. (B) Volcano Matrix of the reads expression level comparison. The red dots are the significant genes (p<0.01) differently expressed based on Student’s test.
4.2.3 Identification of the DE genes

The set of DE genes determined by BaySeq, DEseq, and Cuffdiff2 are different because of the differences in the algorithm. The number of up-regulated genes compared with other four types in *M. kansasii* Type I determined by DEseq2 is 497, while BaySeq found 1474 genes and TopHat2 found 743 genes. The number of down-regulated genes in *M. kansasii* Type I determined by Cuffdiff2 is 942, while DESeq2 determined 412 genes and BaySeq determined 1551 genes. To explore the genes that are most differentially expressed, we have selected the genes that were overlapping with the three methods, respectively. 402 down-regulated genes and 497 up-regulated genes were shared (Fig.18). In the analysis process that we discussed above, we compared *M. kansasii* Type II-Type V expression data against *M. kansasii* Type I expression data (FPKM data or read counts data). However, one gene could be up-regulated in one subtype but down-regulated in another. So, the number of genes, respectively, 402 down-regulated genes and 497 up-regulated genes, represent the genes that are at least differently expressed in one subtype. Meanwhile, exploring the genes that are uniquely up-regulated or down-regulated (UDE genes) in *M. kansasii* may also give us some clues about the genes involved in *M. kansasii* pathogenicity. Further analysis of the genes pool generated in the last step, was performed. A close look of the DE state of each gene was illustrated in Fig.19. Most of the genes are not uniquely up or down-regulated in *M. kansasii* Type I. There are five genes only up-regulated in *M. kansasii* Type I, while ten genes are exclusively down-regulated. The expression level of all of one-to-one orthologs genes can also be seen from the heatmaps (Fig.20).
4.2.4 Functional analysis of the DE/UDE genes

The UDE genes were first investigated via BLAST. From the five genes that are only up-regulated in *M. kansasii* Type I, three of them do not have high scoring segment pairs (HSPs), while the remaining two genes’ functions are known. It is very probable that these unknown genes are related to the ESX secretion system or the organism’s general pathogenicity, making it interesting to investigate the functions of the genes via molecular approaches. There are three genes that are well annotated in down-regulated UDE genes pool, which is a small group of genes. A detail of the five UDE genes is shown in Table 2. Functional analysis of the DE genes reveals that they are quite different from each other. For example, holiday junction resolvase complex is enriched in the when *M. kansasii* Type I comparing with *M. kansasii* Type II and *M. kansasii* Type III and groups, while not in the other two types. A large amount of differences in cellular components has been found in *M. kansasii* Type IV compared with *M. kansasii* Type I, including protein complexes, integral components of the membrane, ABC transporters and Hbp35-ATPase complexes.

Interestingly, a large amount of redox enzymes genes have been found to be differently expressed. For example, 2-alkenal reductase is enriched in *M. kansasii* Type I when compared with *M. kansasii* Type II, redox enzymes acting on NAD(P)H is enriched when compared with *M. kansasii* Type III, oxidoreductase acting on CH-CH group is enriched when compared with *M. kansasii* Type IV and the redox enzymes are enriched when compared with *M. kansasii* Type V. From the biological processes that are highly presented, *M. kansasii* Type I contains several genes that respond to hypoxia. As we know, intracellular survival is important for *M. kansasii*
Type I. The superoxide generated by NAPDH oxidase and xanthine oxidase systems is an importance defense system for mammal cells\textsuperscript{95}, the superoxide radical can interacts with its targets such as lipids on the membrane of the bacteria, proteins, the microbial DNA, for example, iron-sulfur cluster proteins. After entering the phagocytes, \textit{M. kansasii} have to deal with the anti-microbial oxygen and nitrogen radicals produced by phagocytes\textsuperscript{96}. Dealing with the oxidative environment after infection is an important task for mycobacteria\textsuperscript{96}. For example, \textit{SOD, KatG, Tpx, AhpC/E} are important antioxidant genes in \textit{M. tuberculosis} and important for intracellular survival\textsuperscript{97}. Additionally, the iron homeostasis is an important issue to successfully survive for mycobacteria\textsuperscript{98}. These high-regulated genes may give us some clues of the intracellular survival process of \textit{M. kansasii} Type I. As for the UDE genes, waxy ester synthease-like Acyl-acyltransferase is an enzyme that catalyzes the final steps in triacylglycerol (TAG) and wax ester (WE) biosynthesis. It has been reported that a large number of WS/DGAT-homologous genes are in pathogenic mycobacteria and they are probable important for the pathogenesis and latency of these bacteria\textsuperscript{99}. Interestingly, glutaredoxin, TSA family proteins, LLM class of F420-dependent reductase and 2-nitropropane deoxygenate are some other redox enzymes, which strengthen our argument that the redox enzymes play an important role in \textit{M. kansasii} pathogenicity.
Fig. 18 Up-regulated or down-regulated DE genes.
(A) UP-regulated DE genes Venn diagrams. Numbers within each circle represent the number of genes that are DE genes determined by different programs. (B) Up-regulated DE genes Venn Diagrams.
Fig. 19 Up-regulated and down-regulated overlapping DE genes.
(A) Up-regulated DE genes Venn Diagrams. The gene numbers in each circle represent the DE state in each subspecies compared to *M. kansasii* Type I expression level. (B) Down-regulated DE genes Venn Diagrams. The UDE genes are the overlapping genes between all the four *M. kansasii* subtypes.
Fig. 20 Heatmaps of the DE genes. Heatmap of the DE genes within 3460 one to one orthologs. The value was calculated using normalized count read values in each subspecies against *M. kansasii* Type I.

### Table 1 Reads Mapping Percentage of Three Replicates of M. kansasii Type I to V

<table>
<thead>
<tr>
<th>Feature</th>
<th>M. kansasii Type I</th>
<th>M. kansasii Type II</th>
<th>M. kansasii Type III</th>
<th>M. kansasii Type IV</th>
<th>M. kansasii Type V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate I</td>
<td>94.70%</td>
<td>97.20%</td>
<td>96.60%</td>
<td>91.80%</td>
<td>91.10%</td>
</tr>
<tr>
<td>Replicate II</td>
<td>94.60%</td>
<td>96.90%</td>
<td>96.50%</td>
<td>93.90%</td>
<td>91.00%</td>
</tr>
<tr>
<td>Replicate III</td>
<td>93.40%</td>
<td>97.20%</td>
<td>96.90%</td>
<td>92.10%</td>
<td>90.90%</td>
</tr>
</tbody>
</table>

### Table 2 Log2 Fold Change of the UDE Genes and Their Functions

<table>
<thead>
<tr>
<th>Gene_ID</th>
<th>MK2</th>
<th>MK3</th>
<th>MK4</th>
<th>MK5</th>
<th>DE State</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MK1_3550</td>
<td>2.14</td>
<td>3.27</td>
<td>5.58</td>
<td>3.27</td>
<td>Down</td>
<td>Waxy ester synthase-like Acyl-acyltransferase domain</td>
</tr>
<tr>
<td>MK1_4007</td>
<td>2.63</td>
<td>3.13</td>
<td>3.94</td>
<td>3.36</td>
<td>Down</td>
<td>Glutaredoxin</td>
</tr>
<tr>
<td>MK1_4008</td>
<td>2.07</td>
<td>2.45</td>
<td>2.97</td>
<td>2.51</td>
<td>Down</td>
<td>TSA family</td>
</tr>
<tr>
<td>MK1_4508</td>
<td>-2.02</td>
<td>-3.00</td>
<td>-2.73</td>
<td>-2.15</td>
<td>Up</td>
<td>LLM class F420-dependent oxidoreductase</td>
</tr>
<tr>
<td>MK1_5003</td>
<td>-2.26</td>
<td>-3.29</td>
<td>-3.41</td>
<td>-2.10</td>
<td>Up</td>
<td>2-nitropropane dioxygenase</td>
</tr>
</tbody>
</table>
Fig. 21 Up-regulated DE genes GOstats analysis result.
(A) Cellular component and molecular functions GO terms enriched in the up-regulated DE genes. (B) Biological process enriched in the up-regulated DE genes. The color refers to Log2OddsRation which is an indicator of the number of the genes while the size of the circle represent P-values of each GO term. MK2, *M. kansasii* Type II, MK3, *M. kansasii* Type III, MK4, *M. kansasii* Type IV, MK5, *M. kansasii* Type V.
Fig. 22 Down-regulated DE genes GOstats analysis result. 
(A) Cellular component and molecular functions GO terms enriched in the down-regulated DE genes. 
(B) Biological process enriched in the down-regulated DE genes. The color refers to Log2OddsRatio which is an indicator of the number of the genes while the size of the circle represents P-values of each GO term. MK2, *M. kansasi* Type II, MK3, *M. kansasi* Type III, MK4, *M. kansasi* Type IV, MK5, *M. kansasi* Type V.
4.3 Future Work

With advancements in DNA sequencing technology, a more precise genome can now be obtained from PacBio sequencing, whose reading fragments can be as long as 20kb and the depth can reach 50X. The annotated genes may also differ from the annotation files we used in this study, which are generated based on the assembled genomes from the Illumina platform. New genes may be discovered and more DE candidates may also come up, which will provide us with a more accurate view of the RNAseq profile. Meanwhile, the RNAseq reads in this study were not fully explored. Around half of the genes in five different species that are not present in the one-to-one orthologs were not explored. These data may also be interesting if we have multiple conditions of each type. For example, we can infect THP-1 macrophage cells with \textit{M. kansasii} Type I and isolate the RNA from it and compare the RNAseq profile only in \textit{M. kansasii} Type I, in which all of the genes will be investigated. This will also give us some interesting results about the adaptive strategy used by \textit{M. kansasii} Type I to survive in macrophages.

Meanwhile, as we have discussed before, the diverse RNAseq profiles may support the argument to treat the five strains as species, not only subspecies. Other than the RNAseq profile data, a more accurate sequencing of the genome data by PacBio machines may help to support this conclusion. A large number of undefined genes as we can see from our results reveal that the pathogenicity and gene functions of \textit{M. kansasii} are still a mystery. Although the functions of these genes are relatively understudied, we can obtain information by using molecular approaches, such as generating mutations to uncover their cellular roles. The redox enzymes in \textit{M.
*kansasii* Type I should be further explored as important components that may help the intracellular survival process.
In this study, we have designed specific primers for targeting the \textit{espA} operon exclusively present in \textit{M. kansasii} Type I, which could be used as a diagnostic molecular marker for \textit{M. kansasii} Type I infections in a clinical environment. This will represent a new, quick and cheap solution to rapidly determine if a patient is infected by this human pathogen.

Also, we have confirmed that the \textit{espA} operon plays a crucial role in the pathogenicity of \textit{M. kansasii} Type I. Using functional complementation by the \textit{espA} operon in wild-type \textit{M. kansasii} Type V, we were able to observe almost the same phenotype of growth with that of \textit{M. kansasii} Type I. However, the \textit{espA} operon alone was not enough for recovering the same pathogenicity as \textit{M. kansasii} Type I, suggesting that other components are also involved. These missing components, whose functions can help to explain the pathogenicity of \textit{M. kansasii} Type I, may be identified from our comparative genomic studies and RNAseq analysis.

The data, obtained from comparative genomics approaches and RNAseq analysis of the five \textit{M. kansasii} subspecies, gave us few explanations of why \textit{M. kansasii} Type I is considered as a human pathogen while the other four types are environmental. Unique genes that are exclusively present in Type I, including \textit{whiB6}, \textit{ppsA/C}, \textit{irtA} and \textit{PE/PPE} family genes and several components involved in or related to secretion systems, that we believe to be a major component in \textit{M. kansasii} pathogenicity were discovered. Indeed, we showed that the main difference between the five subspecies
is found in the cell membrane composition, in which the ESX secretion system is located. The RNAseq profile reveals a large difference that has been found in the oxidative-reductive process of *M. kansasii* Type I from others types, which may help intracellular survival. However, these important DE genes such as *irtA, mthH* need to be further verified using qPCR and their functions in Pathogenicity are also needed to be confirmed using knock-off approaches.

In the future, these results need to be exploited further and take advantage of the differential membrane composition to identify new therapeutic targets and be able to find curative treatments against pathogenic mycobacteria species.
Fig. S1 Artemis BAMview of Illumina reads from PCR-free libraries’ reads of five subspecies of *M. kansasii* types I-V mapped against type I.

The genes *whiB6* (A)(B), *PPE62*(C), *ppsA_2*(D), *mtbH* (E), *irtA* (F) *ppsC_2*(G) that are absent in other four types except *M. kansasii* Type I, are highlighted in the figure.
BIBLIOGRAPHY


41. Taillard, C. *et al.* Clinical Implications of Mycobacterium kansasii Species Heterogeneity : Swiss National Survey Clinical Implications of Mycobacterium


