

UNIVERSITI TEKNOLOGI MARA

**UNDERSTANDING THE GENOMIC
INFORMATION OF
MYCOBACTERIUM TUBERCULOSIS
CAUSING EXTRAPULMONARY
TUBERCULOSIS USING WHOLE GENOME
SEQUENCING APPROACH**

NUR AZWIN BTE ISMAIL

Thesis submitted in fulfillment
of the requirement for the degree of
Master of Science

Faculty of Pharmacy

November 2015

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of the Universiti Teknologi MARA. It is original and is the result of my own work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any other degree or qualification.

I hereby, acknowledge that I have been supplied with the Academic Rules and Regulations for Post Graduate, Universiti Teknologi MARA, regulating the conduct of my study and research.

Name of Student : Nur Azwin Bte Ismail
Student ID No : 2011834546
Programme : Master of Sciences
Faculty : Faculty of Pharmacy
Thesis Title : Understanding the Genomic Information of *Mycobacterium tuberculosis* Causing Extrapulmonary Tuberculosis Using Whole Genome Sequencing Approach

Signature of Student : 

Date : November 2015

ABSTRACT

Mycobacterium tuberculosis (MTB) causes serious disease of tuberculosis which threatens the world with 8 million individuals infected and 2 million deaths reported each year. MTB spreads by aerosol and infects the host through the lungs. This bacillus is phagocytised by macrophages in the lungs and replicate inside these cells. It invades through the macrophage by inhibiting the fusion between phagosome and bactericidal lysosome. The primary macrophage infection results in a pro-inflammatory response followed by recruitment of other cells that are essential for the innate and adaptive immune systems. It is known that MTB most commonly affects the lungs. However, it has emerged as an extrapulmonary pathogen that infects any organ system inside the human body. The bacterium is capable of infecting other organ system due to dissemination via lympho-haematogenous route in the early period of the pulmonary infection. A few tuberculosis cases were reported to occur at extrapulmonary sites especially among the immunocompromised. Despite the increase in extrapulmonary TB cases, the mechanisms by which the bacterium subvert host defense mechanism and invade deeper tissue of extrapulmonary site are still poorly understood. Compounding to this lack of knowledge, we therefore use whole genome sequencing platform to complete the genome sequence of extrapulmonary TB strain PR05. With the combination of genomics and bioinformatics analysis, study on the genes essential for virulence and pathogenesis of complete genome sequence of extrapulmonary TB strain PR05 was performed. Based on the annotation of the DNA sequence, proteins function, cellular localization and molecular features were predicted. All the identified genes that are essential for virulence and pathogenesis were grouped into eight categories. Few genes were identified to play role in invasion, and adhesion of extrapulmonary dissemination: *erp*, *fbp*, *mce* and *hbhA*. Two members (ESX-1 and ESX-5) of MTB unique secretion system, type seven secretion systems (T7SS) was found to be involved in the virulence of *M. tuberculosis* which affects cell to cell migration of this pathogenic mycobacterium. *En route* in looking for variations, a comparative analysis between extrapulmonary TB strain and five Pulmonary TB strain was performed to identify the similarities and differences between these strains. The comparative analysis and multiple genome alignment have identified few inversion rearrangement and eleven insertion region encoding 31 genes that play roles in pathogenicity and virulence of MTB. However, current project was not able to pinpoint the exact mechanism causing tropism. Further investigate on the functional importance of genetic variation identified in the extrapulmonary TB strain PR05 is required for better understanding in the mechanism of tropism for extrapulmonary tuberculosis.

ACKNOWLEDGEMENTS

Praise to Allah S.W.T, The most Righteous and Merciful, for giving me the strength and wisdom throughout my journeys to accomplish this study.

Special thanks to my beloved supervisor and co-supervisor Prof. Dr. Hj. Mohd Zaki Bin Salleh and Prof. Dr. Teh Lay Kek for their guidance, support, patience, courage and trust. I am grateful for the inspiration, wisdom and knowledge they gave me. My gratitude to Ministry of Education for funding the project to the completion [Grant no: 600-RMI/LRGS 5/3 (6/2011)]. My sincere gratitude and appreciation to iPROMISE for their cooperation.

My greatest appreciation to my family and my husband with their endless love, for making me stronger and giving me the courage to dream and to have aspirations. Not to forget, a deep sense of gratitude to all my friends for their assistance and courage, they are the real key supporters in my wonderful journey as a student.

CHAPTER ONE

INTRODUCTION

1.0 BACKGROUND

Mycobacterium tuberculosis is one of the major human pathogens that cause tuberculosis (TB), a multi-systemic disease with various manifestations. During the 18th and 19th centuries, TB reached epidemic in Europe and North America and was named as the "Captain Among these Men of Death" (Daniel, 2006). Apparently, the bacterium latently infected every part of the world and the largest cases was reported in Asia representing 60% of new TB cases worldwide as reported by WHO, 2013. This infectious disease spread through the air when a person breathes in tubercle bacilli from expelling droplets from an infected individual. The infection commonly occurs in the lungs nonetheless it also can affect other organs or systems outside the lungs: the central nervous system, lymphatic systems, genitourinary systems, or in bones and joints (Meena & Rajni 2010). Due to the increasing cases of HIV/AIDS and emergence of multi-drug resistant strain (MDR), extensively drug resistant strain (XDR) has promote the resurgence and reactivation of TB which also cause this disease to be incurable (Iseman, 1993; Spitznagel, 1999; Gandhi et al., 2006; Mandavilli, 2007). Determining the mechanism of virulence of *M. tuberculosis* is a continuous effort, however the mechanism involved is still poorly understood despite the knowledge obtained in the last 100 or more years. This is because *M. tuberculosis* has no classical virulence factor like other pathogenic bacteria such as *Shigella dysenteria*, *Vibrio cholera*, *Escherichia coli* 0157:H7, *Corynebacterium diphtheria* which produce toxins. However, these bacteria have progressed by creating few mechanisms to evade the hostile environment of the macrophage by impeding phagosome-lysosome fusion and escape acidic environment inside the phagolysosome (Meena & Rajni, 2010). The underlying mechanism and factors associated with virulence and disease progression have remained obscure until the completion of the first whole genome sequence of *M. tuberculosis* H37Rv by Stewart Cole and colleagues in 1998. This complete genome of *M. tuberculosis* strain H37Rv has