

UNIVERSITI TEKNOLOGI MARA

**GENOMIC EPIDEMIOLOGY OF
SPORADIC HUMAN *Brucella*
melitensis CASES IN MALAYSIA
THROUGH WHOLE GENOME
ANALYSIS**

**JAMA'AYAH BINTI MOHAMED ZAHIDI @
AHMAD BAKRI**

Thesis submitted in fulfillment
of the requirements for the degree of
Master of Science
(Molecular Biology)

Faculty of Applied Science

February 2023

ABSTRACT

Brucella melitensis is the most pathogenic and genetically diverged species within the genus *Brucella*. In Malaysia, human infections by *B. melitensis* are reported yearly and the cases are primarily sporadic. Whole genome sequencing (WGS) analysis of *B. melitensis* strains can provide high-quality information for genotyping, monitoring and trace-back purposes in epidemiological studies. This research aim is to determine the genetic relatedness of local *B. melitensis* strains compared to strains from other geographical locations and to study the genetic variation of these local *B. melitensis* strains isolated from sporadic human infections. In this study, whole genomic analysis of 25 sporadic human *B. melitensis* cases in Malaysia and 75 genome sequences of *B. melitensis* retrieved from the NCBI database. From a total of 100 genome sequences, thirty-four Malaysian strains (25 sporadic isolates in this study, six human isolates from other studies and three animal isolates) can be grouped into genotype II and segregated into a distinct lineage clade of South-East Asian (SEA). Phylogenetics analysis showed that 24 of these 25 strains clustered within a single clade supported by 95% bootstrap. One of the isolate was clustered closely with the Indian isolates strain Bm IND1. The strain also had different SNPs number as compared to 24 sporadic *B. melitensis* strain. MLST typing indicated that 24 out of 25 isolates (96%) are grouped as type ST39, except for strain BMM36/11 which is ST8. WGS analysis indicated that the genome of all 25 strains appears to be highly conserved. However, SNP analysis indicates that the Malaysian strains are diverged from the reference *B. melitensis* 16M strain by about 2234 SNPs. WGS analysis revealed 44 virulence factors genes of which eleven are from the type IV secretion system (*virB1-virB11*). AMR genes *adeF* and *mprF* were detected in all 25 strains. In conclusion, the WGS analysis provides useful data with high-quality information and allows us to help identify the source of origin in *B. melitensis* strains, which beneficial to clinicians, public health, researcher and scientist to overcome the trace-back epidemiology investigation.

ACKNOWLEDGEMENT

Firstly, I wish to thank God for giving me the opportunity to embark on my Msc and for completing this long and challenging journey successfully. My gratitude and thanks go to my supervisor Professor Dr. Mohd Faiz Foong Abdullah.

My appreciation goes to the Head Unit of Bacteriology Institute for Medical Research (IMR), NIH Dr Rohaidah Hashim as my co supervisor and Head Centre Infectious Disease, IMR, NIH Dr Norazah Ahmad who provided the facilities and support during my study. Special thanks to my colleagues and friends for helping me with this project.

Finally, this thesis is dedicated to the loving memory of my very dear late father, mother and supportive husband for the vision and determination to educate me and to achieve my journey. This piece of victory is dedicated to both of you. Alhamdulillah.

TABLE OF CONTENTS

	Page
CONFIRMATION BY PANEL OF EXAMINERS	ii
AUTHOR'S DECLARATION	iii
ABSTRACT	iv
ACKNOWLEDGEMENT	v
TABLE OF CONTENTS	vi
LIST OF TABLES	ix
LIST OF SYMBOLS	xii
LIST OF ABBREVIATIONS	xiii
LIST OF NOMENCLATURE	xiv
CHAPTER ONE INTRODUCTION	1
1.1 Research Background Malaysia Scenario	1
1.2 Problem Statement	2
1.3 Objectives	3
1.4 Significance of Study	3
1.5 Scope and limitation	4
CHAPTER TWO LITERATURE REVIEW	5
2.1 <i>Brucella</i> species	5
2.1.1 Morphological and Biochemical Characteristic	8
2.2 Pathogenic <i>Brucella</i> species	12
2.2.1 <i>Brucella melitensis</i>	12
2.2.2 <i>Brucella abortus</i>	12
2.2.3 <i>Brucella suis</i>	13
2.3 Antibiotic Sensitivity	13
2.4 Global Distribution of Human Brucellosis	14
2.5 Transmission of Brucellosis	16
2.6 Molecular characterisation of <i>Brucella</i> spp.	17
2.7 Genome Organization of <i>B. melitensis</i>	19

2.8	Molecular Typing of <i>B. melitensis</i>	21
2.9	Comparative Genomics of <i>Brucella</i> spp.	24
2.9.1	Virulence Factor Genes	24
2.9.2	Antimicrobial Resistance Genes	25
2.9.3	SNP Analysis and Whole Genome Multilocus Sequence Typing	25
2.9.4	Core Genome Multilocus Sequence Typing	26
CHAPTER THREE RESEARCH METHODOLOGY		28
3.1	Bacterial Samples and Genomic Sequence	28
3.2	Bacterial Strains	28
3.3	Whole Genome Sequencing of Bacterial Isolates	29
3.3.1	Genomic DNA Sample Preparation	29
3.3.2	Genomic Library Preparation	29
3.3.3	Whole Genome Sequencing	31
3.4	Post-sequencing Data Processing	34
3.5	Exploratory Data Analysis	34
3.5.1	Genome Assembly and Annotation	34
3.5.2	Speciality Gene Analysis	35
3.5.3	Multilocus Sequence Typing (MLST)	36
3.6	Comparative Genome Analysis Among Sporadic <i>B. melitensis</i>	36
3.6.1	Data Retrieval and Global Clustering	36
3.6.2	Genome Phylogeny	36
3.6.3	Variation Analysis	37
CHAPTER FOUR RESULTS AND DISCUSSION		38
4.1	Phenotypic Characterisation of <i>Brucella melitensis</i> Strains	38
4.2	Genomic Sequencing Data	43
4.3	<i>De novo</i> Assembly and Genome Features	44
4.3.1	Gene Composition	48
4.4	Phylogenetic Analysis	50
4.5	In silico MLST	57
4.5.1	Trace-back studies for strain BMM 36/11-ST8	57
4.5.2	Defining a New MLST type for <i>B. melitensis</i> – ST160	59