

**UNIVERSITI TEKNOLOGI MARA**

**DISSECTING THE GENOMICS  
STRUCTURE OF *Proteus mirabilis*  
strain PR03 USING WHOLE  
GENOME SEQUENCING  
APPROACH**

**MOHD IKHMAL HANIF BIN ABDUL  
KHALID**

Thesis submitted in fulfilment  
of the requirements for the degree of  
**Master of Science**

**Faculty of Pharmacy**

**November 2014**

## AUTHOR'S DECLARATION

I declared that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is a result of my own work, unless otherwise indicated or acknowledged as reference works. 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 : Mohd Ikhmal Hanif Bin Abdul Khalid

Student I.D No. : 2011482674

Programme : Master of Science

Faculty : Faculty of Pharmacy

Title : Dissecting The Genomics of *Proteus mirabilis*  
Strain PR03 Using Whole Genome Sequencing  
Approach

Signature of Student : 

Date : November 2014

## ABSTRACT

**Study background:** *Proteus mirabilis* is a common Gram-negative bacterium which causes upper urinary tract infection and re-current infection. With cutting-edge technology such as whole genome sequencing, the genome sequence could be fully explored to understand its pathogenic and virulence genes. This study aims to provide better understanding on its mechanisms to invade, infect, colonize host epithelial cells and evade host immune system. **Method:** DNA of local clinical isolate of *Proteus mirabilis* strain PR03 was extracted and subjected to whole genome sequencing using the Illumina second generation sequencer, Genome Analyzer II (Illumina, California, USA). The genomic data was trimmed, analyzed, assembled and annotated using bioinformatics pipeline to identify genes that contribute to the pathogenicity and virulence of the strain. The genome was compared with *P. mirabilis* strain HI4320 to identify genes of similarities and differences. **Results:** The genome size of *P. mirabilis* strain PR03 is 3.9 Mbp with a G+C content of 38.6%. This strain has 3 465 genes and 53 RNA. Flagella, fimbriae, capsule, cell membrane, cell wall, urease, invasion proteins and stress respond genes were identified that contribute to the pathogenic and virulence factors of this strain. Genomes comparison showed this species has 56.25% of essential genes, 39.25% of dispensable genes and 4.47% of strain specific genes. **Conclusion:** *P. mirabilis* strain PR03 was successfully sequenced, assembled and annotated. 23.39% of *P. mirabilis* strain PR03 total genes were identified to contribute it pathogenicity and virulence. The genome sequences were successfully deposited in NCBI genomic database.

# TABLE OF CONTENTS

	<b>Page</b>
<b>AUTHOR'S DECLARATION</b>	ii
<b>ABSTRACT</b>	iii
<b>ACKNOWLEDGEMENTS</b>	iv
<b>TABLE OF CONTENTS</b>	v
<b>LIST OF TABLES</b>	x
<b>LIST OF FIGURES</b>	xii
<b>LIST OF ABBREVIATION</b>	xiv
<b>LIST OF FORMULA</b>	xvii
<b>CHAPTER ONE: INTRODUCTION</b>	
1.1 Background of Study	1
1.2 Problem Statement	1
1.3 Significance of Study	2
1.4 Objectives	2
1.5 Scope and Limitation	2
<b>CHAPTER TWO: LITERATURE REVIEW</b>	
2.1 <i>Proteus</i> Mirabilis	3
2.1.1 Bacteria Characteristics	3
2.1.2 Pathogenesis of <i>P. mirabilis</i>	7
2.1.3 Infectious Diseases Associated with <i>P. mirabilis</i>	8
2.1.4 Difficulty in Treating <i>P. mirabilis</i> Infection	9
2.2 DNA Sequencing Technologies	10
2.2.1 First Generation DNA Sequencing	11
2.2.2 Second Generation DNA Sequencing	12
2.2.3 Whole Genome DNA Sequencing Impacts	14
2.3 Bioinformatics	16
2.3.1 Data Quality Control	16
2.3.2 Data Assembly	17

2.3.3	Genome Annotation	18
<b>CHAPTER THREE: METHODOLOGY</b>		
3.1	Study Design	22
3.2	Materials	23
3.2.1	Preparation of Media	23
3.2.2	Preparation of Buffers and Solutions	24
3.2.3	Bacteria Cultures, Growth Condition and Stock Preparation	25
3.3	Methods	
3.3.1	Sample Collection	25
3.3.2	Culturing of Bacteria	25
3.3.3	Bacterial Gram Staining	25
3.3.4	Bacterial Characterization Tests	26
3.3.5	Dienes Test	27
3.3.6	Growth Curve of <i>P. mirabilis</i>	28
3.3.7	Urinary Stone Production Test	28
3.4	Whole Genome Sequencing	29
3.4.1	Preparation of DNA Library	29
3.4.1.1	Preparation of Genomic DNA (gDNA) Library of <i>P. mirabilis</i> strain PR03	29
3.4.1.2	DNA Library Preparation	30
3.4.1.3	Quantitative PCR (qPCR)	30
3.4.2	Cluster Generation	31
3.4.2.1	Cluster Amplification	31
3.4.3	Sequencing by Synthesis	31
3.5	Data Analysis	31
3.5.1	Primary Analysis	32
3.5.2	Secondary Analysis	32
3.5.2.1	<i>de novo</i> Assembly	32
3.5.2.2	Basic Local Alignment Search Tool (BLAST)	33
3.5.2.3	Draft Genome	33
3.5.2.3	Formatdb File	34
3.5.2.4	Draft Genome Validation	35
3.5.2.5	Re-sequencing Assembly	36