

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

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.

ACKNOWLEDGEMENT

Alhamdulillah, first and foremost all praises and thank to Allah azza wajalla, the Almighty for gracing me with strength to complete this thesis with a great help from my Supervisor, Co-supervisor and my families.

My utmost gratitude and appreciation go to my supervisor Professor Dr. Mohd Zaki Salleh and Professor Dr. Teh Lay Kek for their encouragement, support, personal guidance, patient through the period of this research.

I am also grateful to my family and friends who had supported me physically, morally, love, financial and sharing knowledge during this period.

My special thank goes to Ministry of Higher Education Malaysia (MOHE) for the financial support for this research through they grant; Exploring the Metabolomics Regulatory Networks of Virulence and Resistance Behaviour of Pathogenic *Klebsiella pneumoniae* and *Proteus mirabilis* using Whole Genome Sequencing, Proteomics and Metabolomics Approaches, 600-RMI/ERGS 5/3 [50/2012].

TABLE OF CONTENTS

	Page
AUTHOR'S DECLARATION	ii
ABSTRACT	iii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	v
LIST OF TABLES	x
LIST OF FIGURES	xii
LIST OF ABBREVIATION	xiv
LIST OF FORMULA	xvii
CHAPTER ONE: INTRODUCTION	
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
CHAPTER TWO: LITERATURE REVIEW	
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

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND OF STUDY

Proteus sp. is a Gram-negative bacterium ubiquitous in the environment. This species can be isolated from soil, sewage, or mammal gastrointestinal tracts. Currently, there are five *Proteus* species identified, namely *Proteus mirabilis*, *Proteus hauseri*, *Proteus myxofaciens*, *Proteus pennari*, and *Proteus vulgaris*. *P. mirabilis* is a common flora in the human gastrointestinal tract which is a common cause of upper urinary tract infection and re-current infection, compared to *Escherichia coli* among urinary catheter patients. This species, with other urease producer pathogens causes formation of urinary stones in the host's urinary tract system (Chenet *al.*, 2012). *P. mirabilis* together with the stones will then cause an episode of acute pyelonephritis and cystitis in the host. Aside from urinary tract infection, this species also causes bacteriosperma, meningitis and other diseases (Zainab, Aziz, Haque, Samad and Cheung, 2003). All the genes contribute to its pathogenicity and virulence are believed to lie in the genome.

In this study, the genomic architecture of a local clinical isolate of *P. mirabilis* strain PR03 was explored to understand the mechanisms responsible for pathogenesis and virulence using whole-genome sequencing and bioinformatics approaches.

1.2 PROBLEM STATEMENT

It's up to date; there is only one complete whole genome sequence of this species in the NCBI database, which is *P. mirabilis* HI4320. However, the isolate was taken from a catheterised European patient, which may differ from local septicaemia isolates due to differences in host parameters, environment, and lifestyle. Hence, it is crucial to obtain the complete whole genome sequence of a local isolate in order to select specific antibiotic treatment, avoiding unnecessary resistance and develop a rapid detection kit by targeting its unique genes.