



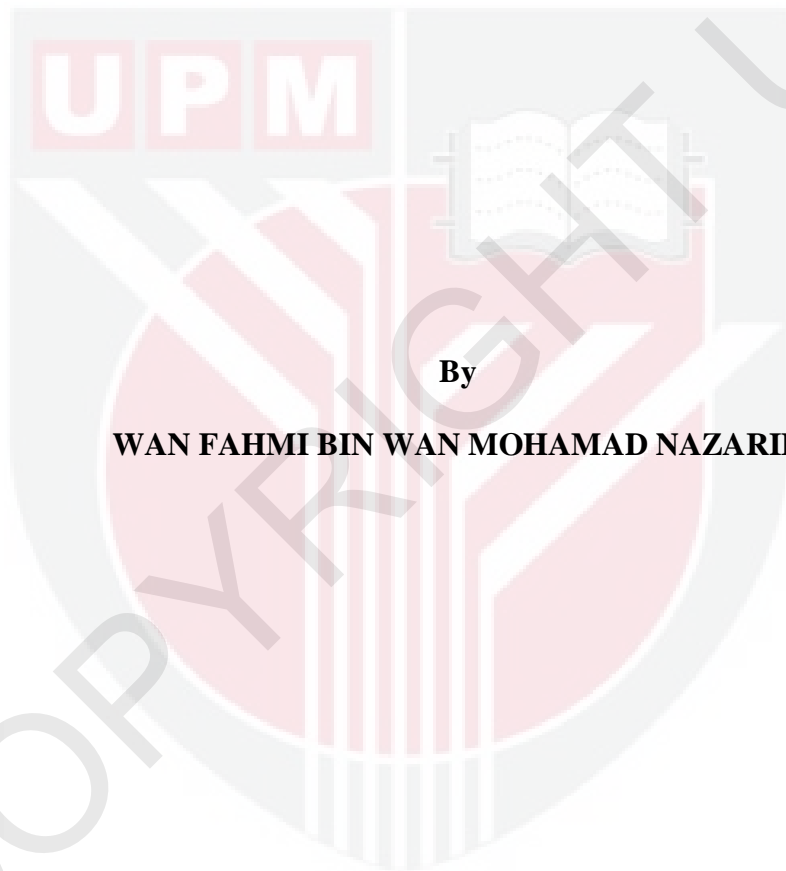
UNIVERSITI PUTRA MALAYSIA

**GENOME SEQUENCING AND BIOINFORMATIC ANALYSIS OF
PASTEURELLA MULTOCIDA SEROTYPE B:2 STRAIN *PMTB***

WAN FAHMI BIN WAN MOHAMAD NAZARIE

IB 2011 8

**GENOME SEQUENCING AND BIOINFORMATIC ANALYSIS OF
PASTEURELLA MULTOCIDA SEROTYPE B:2 STRAIN *PMTB***



By

WAN FAHMI BIN WAN MOHAMAD NAZARIE

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia,
in Fulfillment of the Requirements for the Degree of Master of Science**

August 2011

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirements for the degree of Master of Science

**GENOME SEQUENCING AND BIOINFORMATIC ANALYSIS OF
PASTEURELLA MULTOCIDA SEROTYPE B:2 STRAIN *PMTB***

By

WAN FAHMI BIN WAN MOHAMAD NAZARIE

August 2011

Chairperson : Abdul Rahman bin Omar, PhD

Institute : Institute of Bioscience

Pasteurella multocida is a Gram-negative bacterium, which is the causative agent of a wide range of diseases in animals. This organism usually resides in the mucous membrane of the intestinal, genital and respiratory tissues and is an opportunistic pathogen that causes fowl cholera, bovine haemorrhagic septicaemia and porcine atrophic rhinitis. So far, only complete genome of *P. multocida* serotype A:3 strain *Pm70* has been elucidated. This study was conducted to sequence the genome of *P. multocida* serotype B:2 strain *PMTB* and to compare with the complete genome of *Pm70*.

A total of 7.2 million sequence reads were generated from Illumina Genome Analyzer. *De novo* sequence assembly followed by comparison with *Pm70* reference sequence produced a partial near-complete genome of *PMTB* with missing nucleotide sequences located in 81 gaps. The partial genome of *P. multocida* strain *PMTB* is 97.78% identical to *Pm70*. The estimated size of the partial genome of *PMTB* is 2,208,894 bp while the *Pm70* genome is 2,257,487 bp. In addition, both

genomes contain similar % GC content 40 to 41%. Analysis using GeneMark software indicated the total genes of the partial genome of *PMTB* are 2078 while the reference genome *Pm70* has 2014 genes.

Gene comparison between *PMTB* and *Pm70* to construct *PMTB* sequences as a database blast against *Pm70* sequences showed there are 223 unique genes found in *PMTB* but absence in *Pm70*. The unique genes are probably specific to serotype B:2 only or the genes were not detected in sequence analysis since they are located in the missing sequences in the gaps. On the other hand, a total of 49 genes are not detected in partial *PMTB* genome but present in *Pm70*. Sequence analysis also showed the presence of genes with high similarity (99 to 100%) to the genes from previously characterized serotype B:2, genes that are also found in other *P. multocida* serotypes and genes found in other bacteria especially *Haemophilus influenza*, *Actinobacillus minor* and *Vibrio cholerae*. Based on the partial genome sequence analysis, there are probably several virulence genes and virulence-associated genes in the *P. multocida PMTB* genome which include adhesins protein [type 4 fimbria (*ptfA*)], serotype-specific capsular polysaccharide, lipopolysaccharide, iron acquisition related genes such as *Exbd* and *tonB*, gene associated hemoglobin binding protein (*HgbA*), gene encode for transferrin-binding protein (*tbpA*), and several uncharacterized secreted enzymes and proteins that play important role in the pathogenicity of the disease. Complete genome sequencing and genome-wide functional genomics studies on *P. multocida PMTB* genome will be able to provide valuable information on pathogenicity of haemorrhagic septicaemia in ruminants.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk Ijazah Master Sains

**PENJUJUKAN GENOM DAN ANALISIS BIOINFORMATIK
PASTEURELLA MULTOCIDA SEROTIP B:2 STRAIN *PMTB***

Oleh

WAN FAHMI BIN WAN MOHAMAD NAZARIE

Ogos 2011

Pengerusi : Abdul Rahman bin Omar, PhD

Institut : Institut Biosains

Pasteurella multocida adalah bakteria Gram-negatif, yang merupakan agen penyebab bagi pelbagai penyakit pada haiwan. Organisma ini biasanya berada di selaput lendir pada rangkaian usus, kemaluan dan tisu pernafasan dan merupakan patogen penyebab kolera unggas, hawar berdarah pada lembu dan rhinitis atrofi pada khinzir. Setakat ini, hanya genom *P. multocida* serotip A:3 strain *Pm70* telah dilengkapkan. Kajian ini dilakukan untuk menentukan jujukan genom *P. multocida* serotip B:2 *PMTB* dan membandingkannya dengan jujukan genom lengkap *Pm70*.

Sebanyak 7.2 juta jujukan bacaan dihasilkan dari *Illumina Genome Analyzer*. Himpunan jujukan secara *de novo* diikuti dengan perbandingan dengan jujukan rujukan, *Pm70* telah menghasilkan genom separa yang hampir lengkap *PMTB* dengan jujukan nukleotida yang hilang di 81 jurang. Genom separa *P. multocida* strain *PMTB* ialah 97.78% homologi dengan *Pm70*. Anggaran saiz genom separa *PMTB* adalah 2,208,894 pb manakala genom *Pm70* adalah 2,257,487 pb. Selain itu, kedua-dua genom mengandungi kandungan % GC yang hampir sama antara 40 hingga 41%. Analisis menggunakan perisian GeneMark menunjukkan, jumlah

keseluruhan gen bagi genom separa *PMTB* adalah 2078 gen sedangkan *Pm70* rujukan genom mempunyai 2014 gen.

Peramalan gen dengan membandingkan kedua-dua genom *PMTB* dan *Pm70* untuk membina jujukan *PMTB* sebagai pangkalan data terhadap jujukan *Pm70* menunjukkan terdapat 223 gen unik ditemui di *PMTB* tetapi tiada pada *Pm70*. Gen yang unik mungkin khusus hanya untuk serotip B:2 atau gen tersebut tidak dapat dikesan dalam analisis jujukan kerana jujukannya berada di bahagian jurang. Selain itu, terdapat sejumlah 49 gen yang tidak dikesan dalam genom separa *PMTB* tetapi hadir di dalam *Pm70*. Analisis jujukan juga menunjukkan adanya kesamaan gen dengan tinggi (99 hingga 100%) dengan gen dari serotip B:2 yang telah dikenal pasti sebelum ini, gen yang juga ditemui di *P. multocida* serotip lain dan gen yang ditemui pada bakteria lain, terutama *Haemophilus influenzae*, *Actinobacillus minor* dan *Vibrio cholerae*. Berdasarkan analisis jujukan genom separa, mungkin terdapat beberapa gen virulen dan gen berkaitan virulen seperti protein adhesin [jenis 4 fimbria (*PTFA*)], polisakarida kapsul serotip-khusus, lipopolisakarida, gen berkaitan pengambilalihan besi seperti *Exbd* dan *tonB*, gen berkaitan protein pengikat hemoglobin (*HgbA*), gen mengekod protein pengikat transferin (*tbpA*) dan beberapa protein dan enzim rembesan yang belum dikenal pasti yang berperanan penting dalam patogenesis penyakit. Kajian mengenai jujukan genom lengkap dan genomik berfungsi keseluruhan-genom pada genom *PMTB* akan dapat memberikan maklumat yang berharga terhadap patogenesis hawar berdarah di dalam ruminan.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to my supervisor Professor Dr. Abdul Rahman Omar for his enthusiasm towards my work and for giving fruitful suggestions and ideas. To my co-supervisors Professor Dr. Zamri Saad, Associate Professor Dr. Zunita Zakaria and Professor Dr. Wan Kiew Lian for their valuable advice, technical guidance, encouragement and thesis improvement.

I would like to dedicate to my family especially to my parents, Wan Mohamad Nazarie bin Wan Mahmood and Norsina binti Panot. I also would like deeply appreciation to my wife, Nur Annies binti Abd Hadi for her unconditional sacrifice and love. I wish to acknowledge the guidance and support from Mr. Mohd Noor Mat Isa and Mr. Ahmad Kamal Ghazali from Malaysia Genome Institute (MGI) for supplying ideas and guidance for this project. My gratitude also goes out to all my colleagues who are always appreciated.

I wish to extend my appreciation to everyone, although not individually named here, who had contributed directly or indirectly to my project and thesis. Without all of you, it would not be possible for me to complete my project and thesis. Thank you all for your support.

I certify that a Thesis Examination Committee has met on 25 August 2011 to conduct the final examination of Wan Fahmi bin Wan Mohamad Nazarie on his thesis entitled “Genome Sequencing and Bioinformatic Analysis of *Pasteurella multocida* Serotype B:2 Strain *PMTB*” in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The committee recommends that the student be awarded the Master of Science.

Members of the Examination Committee are as follows:

RASEDEE @ MAT BIN ABDULLAH, Ph.D

Professor
Faculty of Veterinary Medicine
Universiti Putra Malaysia
(Chairman)

SUHAIMI BIN NAPIS, Ph.D

Associate Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Internal Examiner)

SITI KHAIRANI BINTI BEJO, Ph.D

Associate Professor
Faculty of Veterinary Medicine
Universiti Putra Malaysia
(Internal Examiner)

MOHD NAZALAN MOHD NAJIMUDDIN, Ph.D

Professor
School of Biological Science
Universiti Sains Malaysia
(External Examiner)

SEOW HENG FONG, Ph.D

Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 22 November 2011

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirements for the Degree of Master of Science. The members of the Supervisory Committee were as follows:

ABDUL RAHMAN OMAR, PhD

Professor
Faculty of Veterinary Medicine
Universiti Putra Malaysia
(Chairman)

MOHD ZAMRI SAAD, PhD

Professor
Faculty of Veterinary Medicine
Universiti Putra Malaysia
(Member)

ZUNITA ZAKARIA, PhD

Associate Professor
Faculty of Veterinary Medicine
Universiti Putra Malaysia
(Member)

WAN KIEW LIAN, PhD

Professor
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
(External Member)

BUJANG BIN KIM HUAT, PhD

Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:
DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree in Universiti Putra Malaysia or other institution.



WAN FAHMI BIN WAN MOHAMAD NAZARIE

Date: 25 August 2011

TABLE OF CONTENTS

		Page
	ABSTRACT	ii
	ABSTRAK	iv
	ACKNOWLEDGEMENTS	vi
	APPROVAL	vii
	DECLARATION	ix
	LIST OF TABLES	xiii
	LIST OF FIGURES	xiv
	LIST OF ABBREVIATIONS	xvi
	CHAPTER	
1	INTRODUCTION	1
2	LITERATURE REVIEW	
	2.1 <i>Pasteurella multocida</i>	5
	2.2 <i>Pasteurella multocida</i> Classification	5
	2.3 Haemorrhagic Septicaemia	6
	2.3.1 Clinical Signs of Haemorrhagic Septicaemia	6
	2.3.2 Susceptible Animals	7
	2.3.3 Infection of <i>Pasteurella multocida</i>	7
	2.3.4 Demographic of Haemorrhagic Septicaemia Disease	8
	2.4 Pathogenicity of <i>Pasteurella multocida</i>	9
	2.4.1 Bacterial Survival	10
	2.5 <i>Pasteurella multocida</i> Virulence Factors and Its Associated Genes	11
	2.5.1 Virulence Genes	11
	2.5.2 Capsules	12
	2.5.1.2 Virulence Genes in Capsule Locus	13
	2.5.3 Lipopolysaccharide (LPS) Endotoxines	14
	2.5.4 Outer Membrane Proteins (OMP)	15
	2.5.4.1 Outer Membrane Protein (OMP) Related Genes	16
	2.5.5 Dermonecrotic Toxin (PMT)	17
	2.5.6 Iron Acquisition Factors	17
	2.5.6.1 Iron Acquisition Related Genes	19
	2.5.7 Adhesins	20
	2.5.7.1 Adhesion-related Genes	21
	2.6 Genome Sequencing of Organisms	22
	2.6.1 Genome Sequencing of <i>Pasteurella multocida</i>	23
	2.6.2 Genome Sequencing	24
	2.6.3 Next Generation Sequencing (NGS)	25
	2.6.4 Genome Sequencing Applications using Illumina Genome Analyzer	26

2.7	Bioinformatics Software Involved in Analysis Genome Sequencing	27
2.7.1	Velvet	27
2.7.2	GeneMark.HMM	28
2.7.3	BLAST	29
2.7.4	Gene Ontology Consortium	29
2.7.4.1	Blast2GO	30
2.7.5	MUMmer	31
2.7.5.1	Nucmer	32
2.7.5.2	MUMMerplot	33
2.8	CGViewer and Circular Genome	33
2.9	CLC bio	34
3	MATERIALS AND METHODS	
3.1	Bacterial Culture	35
3.2	Reconfirmation of <i>Pasteurella multocida</i>	35
3.2.1	Gram Staining	36
3.2.2	Genome DNA Extraction	36
3.2.3	Determination of DNA Concentration and Purity	37
3.2.4	Construction of Genomic Library	37
3.2.5	Genomic DNA Sequencing Using Next Generation Sequencing	39
3.3	Bioinformatics Analysis	40
3.3.1	Sequencing Data from Illumina Genome Analyzer	41
3.3.2	<i>De Novo</i> Assembly	41
3.3.3	<i>De Novo</i> Assembly Outcome	43
3.3.4	Artemis	43
3.3.5	Scaffolding	44
3.3.6	Alignment of Incomplete Genome	44
3.3.7	Gene Prediction	44
3.3.8	Annotation of the Genome	45
3.3.9	Comparing Genome Sequence	46
3.4	Gene Ontology (GO)	46
3.5	Detection of Genes	47
3.6	Physical Mapping	47
4	RESULTS AND DISCUSSIONS	
4.1	Introduction	48
4.2	Morphological and Biochemical Identification	48
4.3	Genome Concentration and Purity	49
4.4	Genomic Library	49
4.5	Bioinformatics Analysis	51
4.5.1	<i>De Novo</i> Assembly	52
4.5.2	MUMmer	52
4.5.3	Artemis Comparison Tools (ACT)	54
4.5.4	Gene Prediction Using GeneMark.hmm	55
4.5.5	Physical Map Using CGView Program	56
4.6	Comparison Between <i>Pasteurella multocida</i> Serotype B:2 Strain <i>PMTB</i> and <i>Pasteurella multocida</i> Serotype A:3 Strain <i>Pm70</i>	57

4.7	<i>Pasteurella multocida</i> PMTB Genome Size	59
4.7.1	Total Number of Genes in <i>Pasteurella multocida</i> PMTB	59
4.7.2	Unique Genes in <i>Pasteurella multocida</i> PMTB	60
4.7.2.1	Phage Sequence in Bacterial Genome	68
4.7.3	Genes Absence in <i>Pasteurella multocida</i> PMTB Compared to <i>Pasteurella multocida</i> Pm70	70
4.8	Virulence Genes and Its Associated Genes in <i>Pasteurella multocida</i> PMTB	72
4.8.1	Iron Acquisition Related Genes	73
4.8.2	Transferrin-Binding Protein A (<i>tbpA</i>)	74
4.8.3	Transferrin-Binding Protein B (<i>tbpB</i>)	75
4.8.4	Dermonecrotic Toxin (PMT) Genes	75
4.8.5	Capsule as Virulent Determinants	76
4.8.6	Lipopolysaccharide (LPS)	77
4.8.7	Adhesins	78
4.9	Annotation Analysis	79
4.9.1	Blast2GO analysis	79
4.9.2	Blast Analysis	80
4.9.2.1	Blast Statistic	81
4.9.3	Data Distribution	82
4.9.4	Annotation Statistics	84
6	CONCLUSION AND FUTURE RECOMMENDATION	86
	REFERENCES	89
	APPENDIXES	109
	BIODATA OF STUDENT	124