



UNIVERSITI PUTRA MALAYSIA

***A DRAFT GENOME SEQUENCES AND MOLECULAR CHARACTERISATION
OF GLYCOPROTEIN B AND LOWER MATRIX PHOSPHOPROTEIN OF RAT
CYTOMEGALO VIRUS ALL-03 STRAIN***

QUAH YI WAN

FPV 2013 5



**A DRAFT GENOME SEQUENCES AND MOLECULAR
CHARACTERISATION OF GLYCOPROTEIN B AND LOWER MATRIX
PHOSPHOPROTEIN OF RAT CYTOMEGALOVIRUS ALL-03 STRAIN**

By

QUAH YI WAN

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

July 2013

COPYRIGHT

All material contained within the thesis, including without limitation text, logos, icons, photographs and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright © Universiti Putra Malaysia



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

**A DRAFT GENOME SEQUENCES AND MOLECULAR
CHARACTERISATION OF GLYCOPROTEIN B AND LOWER MATRIX
PHOSPHOPROTEIN OF RAT CYTOMEGALOVIRUS ALL-03 STRAIN**

By

QUAH YI WAN

July 2013

Chairman: Zeenathul Nazariah Allaudin, PhD

Faculty: Veterinary Medicine

Congenital cytomegalovirus (CMV) infection causes neurological damages to newborn infants and persists over the years in infants. Rat CMV (RCMV) ALL-03 strain (RCMV ALL-03) was the first RCMV isolated from the placenta and uterus of rat. This RCMV ALL-03 was hypothesised to mimic congenital human CMV (HCMV) infection. Overall, the study signifies the genome characterisation of a novel local transplacental strain RCMV ALL-03, as well as the glycoprotein B (gB) and lower matrix phosphoprotein (pp65) RCMV ALL-03 gene characterisation. The study was commenced by propagating RCMV ALL-03 extensively to a large amount alongside with the reference RCMV English strain (RCMV-E) in rat embryonic fibroblast (REF) cells. Infected cells exhibiting advance cytopathic effect were harvested and concentrated by 8% (w/v) PEG6000. Concentrated viruses were then purified through sucrose gradient. Based on the similarity between HCMV and

RCMV ALL-03, monoclonal antibody against gB of HCMV was used to immunoprecipitate RCMV ALL-03 gB through Western blot and immunoperoxidase assay. Likewise, monoclonal antibody pp65 of HCMV was able to immunoprecipitate pp65 of RCMV ALL-03 by immunoperoxidase assay, indicating there are shared homology between HCMV and RCMV ALL-03. The Illumina Genome Analyzer IIx generated a total of 3.8 million paired end sequence reads from multiplexing of both genomic viral DNA libraries. The raw sequencing results were assembled via *de novo* approach into contigs, producing a total of 6 large CMV-related contigs for RCMV ALL-03 and 11 large CMV-related contigs for RCMV-E. These contigs were aligned and arranged based on reference RCMV Maastricht strain (RCMV-M), followed by scaffolding into draft genomes. The estimated size of RCMV ALL-03 was 198,895 bp while RCMV strain English was estimated to be 175,071 bp. Phylogenetic tree of RCMV ALL-03 draft genome with other reference CMVs revealed that RCMV ALL-03 belongs to the *Muromegalovirus* genus and that it is closest to RCMV-E, with an estimated difference of 8.2 substitution per 1 kbp from RCMV-E. Based on gene prediction software GeneMarkS, there were 136 genes predicted for RCMV ALL-03 and 112 genes predicted for RCMV-E. All predicted genes were subjected to BlastX analysis, followed by mapping and annotation of the genes using Blast2Go software. For the RCMV ALL-03, there were 5 genes without any blast hits and a total of 46 genes were annotated. The gB and pp65 genes were among the annotated genes. Phylogenetic analysis of RCMV ALL-03 gB and pp65 showed that RCMV ALL-03 branched individually from other reference CMVs despite having almost similar branch distance as the reference CMVs. By designing and examining the RCMV gB and pp65 specific primers for polymerase chain reaction (PCR) via *in silico* and lab verification, RCMV ALL-03

gB and pp65 gene were amplified. This study revealed the genome sequence of RCMV ALL-03, which is essential in the understanding of shared homology between RCMV ALL-03 and HCMV. The relatedness of RCMV ALL-03 to RCMV-E showed the conserve regions of rodent CMV but their divergence revealed potential RCMV ALL-03 unique transplacental characteristics. Future exploitation study on the RCMV ALL-03 could help provide a better understanding of congenital HCMV.



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

**DRAF URUTAN GENOM DAN CIRI-CIRI MOLEKUL GLIKOPROTEIN B
DAN FOSFOPROTEIN MATRIX RENDAH STRAIN ALL-03
SITOMEGALOVIRUS TIKUS**

Oleh

QUAH YI WAN

Julai 2013

Pengerusi: Zeenathul Nazariah Allaudin, PhD

Fakulti: Perubatan Veterinar

Jangkitan kongenital sitomegalovirus (CMV) boleh mengakibatkan kerosakan sistem saraf pada bayi baru lahir dan ia akan berkekalan pada bayi tersebut. Strain CMV tikus (RCMV) ALL-03 adalah RCMV pertama yang dipencilkan daripada uri dan rahim tikus. Jangkitan RCMV ALL-03 ke atas tikus menyerupai jangkitan kongenital CMV manusia (HCMV). Secara keseluruhannya, kajian ini adalah berkaitan pencirian genom virus transplasenta tempatan, strain RCMV ALL-03, serta pencirian gen glikoprotein B (gB) dan gen fosfoprotein matriks rendah (pp65). Kajian ini dimulakan dengan pembiakkan RCMV ALL-03 pada jumlah yang besar secara

intensif bersama-sama dengan RCMV rujukan daripada strain English (RCMV-E) di dalam sel fibroblas embrio tikus (REF). Sel terjangkit yang mempamerkan kesan sitopatik telah dituai dan dipekatkan dengan menggunakan 8% (w/v) PEG6000. Virus yang pekat kemudiannya dituliskan melalui proses kecerunan sukrosa. Berdasarkan persamaan di antara HCMV dan RCMV ALL-03, antibodi monoklonal terhadap gB HCMV telah digunakan untuk memendakkan tindak balas imun RCMV ALL-03 gB melalui Western blot dan ujian immunoperoksidase. Antibodi monoklonal pp65 daripada HCMV mampu memendakkan tindak balas imun pp65 daripada RCMV ALL-03 melalui ujian immunoperoksidase. Ini menunjukkan perkongsian homologi di antara HCMV dan RCMV ALL-03. Penjujuk Illumina Genome Analyzer Iix menjana sebanyak 3.8 juta jujukan akhir berpasangan daripada koleksi multipleks kedua-dua genom DNA virus. Hasil jujukan dibentuk melalui kaedah *de novo* kepada kontig yang menghasilkan 6 kontig berkaitan CMV untuk RCMV ALL-03 dan 11 kontig berkaitan CMV untuk RCMV-E. Kesemua kontig disusun dan dipadankan berdasarkan RCMV rujukan strain Maastricht (RCMV-M), dan kemudian diperancahkan menjadi draf genom. Saiz anggaran RCMV ALL-03 adalah 198,895 bp manakala saiz RCMV-E dianggarkan berjumlah 175,071 bp. Pokok filogenetik di antara draf genom RCMV ALL-03 dengan CMV rujukan lain menunjukkan RCMV ALL-03 tergolong di bawah genus *Muromegalovirus* dan ia mempunyai persamaan terdekat dengan RCMV-E (anggaran perbezaan sebanyak 8.2 penggantian pada setiap 1 kbp). Berdasarkan gen ramalan pada perisian GeneMarkS, terdapat 136 gen yang dijangkakan untuk RCMV ALL-03 dan 112 gen untuk RCMV-E. Semua gen ramalan dianalisis dengan BlastX, diikuti dengan pemetaan dan anotasi gen menggunakan perisian Blast2Go. Terdapat 5 gen RCMV ALL-03 yang tidak dapat dikenal pasti dan sebanyak 46 gen RCMV ALL-03 yang dianotasi.

Gen-gen gB dan pp65 adalah di antara gen yang dianotasi. Analisis filogenetik RCMV ALL-03 gB dan pp65 menunjukkan bahawa RCMV ALL-03 mempunyai jarak cabang yang hampir sama dengan rujukan CMV lain tetapi bercabang sendirian dan jauh daripada rujukan CMV lain. Pencetus khas bagi RCMV ALL-03 gB and pp65 telah direka dan diuji melalui tindak balas berantai polymerase (PCR) *in silico* dan gen RCMV ALL-03 gB dan pp65 telah berjaya diamplifikasi. Urutan genom RCMV ALL-03 yang dihasilkan adalah penting bagi memahami persamaan di antara RCMV ALL-03 dengan HCMV. Perkaitan antara RCMV ALL-03 dengan RCMV-E menunjukkan kawasan pemuliharaan pada CMV tikus tetapi perbezaannya adalah ciri transplasenta yang unik bagi RCMV ALL-03. RCMV ALL-03 boleh dieksploitasikan untuk kajian masa depan bagi pemahaman yang lebih mendalam mengenai tabiat kongenital HCMV.

ACKNOWLEDGEMENTS

This thesis would have remained a dream without the guidance and help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

First and foremost, my utmost gratitude to my supervisor, Assoc. Prof. Dr. Zeenathul Nazariah Allaudin, for her patience guidance, encouragement and valuable advice she has provided throughout my study. Her perseverance and patience attitude spurs me onwards. I would also like to thank my co-supervisor, Prof. Mohd Azmi Mohd Lila, for his support and valuable insights to RCMV ALL-03. Special gratitude goes to Prof. Abdul Rahman Omar, who allowed me to use the Dell workstation in the Bioinformatics Laboratory of Institute of Bioscience, UPM.

I would like to acknowledge my senior and partner, Ms.Siti Nazrina Camalxaman, for her kindness, friendship and continuous support. In my daily work, her unwavering inspirations accompany and cheer me up throughout this exhausting study. In addition, I am indebted to Mr. Wan Fahmi, who helped tutor me in the use of various bioinformatics softwares and the analysis of bioinformatics data.

Completing this work would have been all the more difficult were it not for the support and friendship provided by the other members from the Virology Laboratory of Faculty of Veterinary Medicine and Institute of Bioscience: Mr. Mohd Kamarudin, Ms. Noraini, Ms. Norzatul Iffah, Mr. Mohd Nazari, Mr. Homayoun Hani, Dr. Amer Alazawy, Mr. Tam Yew Joon, Dr. Tan Seok Shin, Ms. Lo Sewn Cen, Dr. Nik Mohd Afizan, Ms. Ruzila, Ms. Hidayah, Ms. Movarid, Ms. Parisa, Ms.

Caryn, Ms. Faezah and Ms. Fatemeh. In addition, I wish to extend my appreciation to other unnamed individuals who had contributed directly and indirectly throughout my study.

Last but not least, special thanks go to my family and the one above all of us, the omnipresent God, for answering my prayers for giving me the strength to plod on despite my constitution wanting to give up and throw in the towel, thank you so much Dear Lord.



I certify that a Thesis Examination Committee has met on 5th July 2013 to conduct the final examination of Quah Yi Wan on her thesis entitled "**A draft genome sequences and molecular characterisation of glycoprotein B and lower matrix phosphoprotein of Rat Cytomegalovirus ALL-03 strain**" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science.

Members of the Thesis Examination Committee were as follows:

HASSAN HJ. MOHD. DAUD, PhD

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

ABDUL RAHMAN OMAR, PhD

Professor
Institute of Bioscience
Universiti Putra Malaysia
(Internal Examiner)

RAHIM ABDUL MUTALIB, PhD

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

MOHD. ZAKI SALLEH, PhD

Professor
Faculty of Pharmacy
Universiti Teknologi Mara
Malaysia
(External Examiner)

NORITAH OMAR, PhD

Assoc. Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 16 August 2013

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

Zeenathul Nazariah Allaudin, PhD

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

Mohd Azmi bin Mohd Lila, PhD, MBA, LLM

Professor
Faculty of Veterinary Medicine
Universiti Putra Malaysia
(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 at Universiti Putra Malaysia or at any other institution.

QUAH YI WAN

Date: 5 July 2013

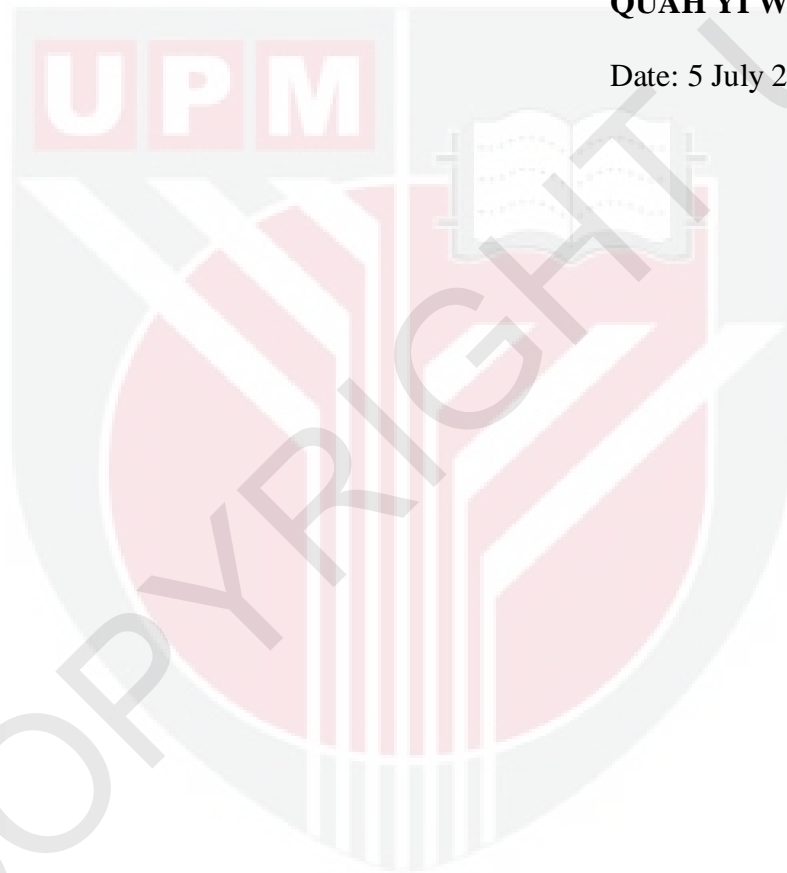


TABLE OF CONTENTS

	Page
ABSTRACT	ii
ABSTRAK	v
ACKNOWLEDGEMENTS	viii
APPROVAL	x
DECLARATION	xii
LIST OF TABLES	xvi
LIST OF FIGURES	xvii
LIST OF ABBREVIATIONS	xix
CHAPTER	
1 INTRODUCTION	1
2 LITERATURE REVIEW	4
2.1 Herpesvirus	4
2.1.1 Herpesvirus genome	6
2.2 Cytomegalovirus	10
2.2.1 History of cytomegalovirus	10
2.2.2 Structure of cytomegalovirus	12
2.2.3 Epidemiology of CMV infection	15
2.2.4 CMV infection and disease	16
2.2.5 Pathogenesis and persistency of CMV	18
2.3 Animal models of CMV	19
2.3.1 Rodent models of CMV	21
2.4 CMV genome	23
2.4.1 Importance of CMV genome	25
2.4.2 Kinetic classes of CMV genes	26
2.4.3 Immunogenic proteins of CMV	27
2.4.3.1 Glycoprotein B	27
2.4.3.2 Lower matrix phosphoprotein	28
2.5 Genome sequencing	29
2.5.1 Trends in genome sequencing	29
2.5.2 NGS platform	31
2.5.3 Bioinformatics analysis	33
3 MATERIALS AND METHODS	35
3.1 Experimental design	35
3.2 Cell culture	39
3.2.1 Cell propagation	39
3.2.2 Subculture of cells	39
3.2.3 Cryopreservation and resuscitation of cells	40
3.3 Virus culture	41
3.3.1 Virus propagation	41
3.3.2 Titration of virus stock	42
3.3.3 Virus concentration	42
3.3.3.1 Ultracentrifugation	42
3.3.3.2 Tangential flow filtration	43

3.3.3.3	Polyethylene glycol (PEG) precipitation	44
3.3.4	Virus purification	44
3.4	Extraction of virus protein	45
3.5	Protein quantification	45
3.6	SDS PAGE	46
3.7	Western blotting	47
3.8	Indirect immunoperoxidase staining	48
3.8.1	Detection of glycoprotein B (gB)	48
3.8.2	Detection of lower matrix phosphoprotein (pp65)	49
3.9	Virus genome DNA extraction	50
3.10	Quantitative analysis of genomic DNA	51
3.11	Agarose gel electrophoresis	51
3.12	Genomic DNA sequencing using Illumina next generation sequencing (NGS)	52
3.13	Genome informatics analysis	52
3.13.1	Sequencing data from Illumina Genome Analyzer IIX	53
3.13.2	Genomic <i>de novo</i> assembly	53
3.13.3	Genomic reference assembly	54
3.13.4	Rearrangement of contigs	54
3.13.5	Scaffolding	55
3.13.6	Phylogenomic analysis	55
3.13.7	Gene prediction	55
3.13.8	Blast2GO	56
3.13.9	Artemis	56
3.14	<i>In silico</i> characterisation of glycoprotein B (gB) and lower matrix phosphoprotein genes (pp65)	57
3.14.1	Phylogenetic analysis	57
3.14.2	Primer design	57
3.14.3	Primer validation	58
3.14.3.1	Motif search	58
3.14.3.2	<i>In silico</i> PCR via MFEprimer	59
3.14.3.3	<i>In silico</i> PCR via Simulated PCR (SPCR)	59
3.15	Polymerase Chain Reaction	60

4

RESULTS

4.1	Cytopathogenicity of RCMV ALL-03 and RCMV-E	61
4.2	Virus concentration and purification	63
4.3	Protein profile of RCMV ALL-03 and RCMV-E	65
4.4	Glycoprotein B detection	67
4.5	Lower matrix phosphoprotein detection	68
4.6	Genome concentration and purity	70
4.7	Sequencing data from Illumina Genome Analyzer	71
4.8	Bioinformatics analysis	71
4.8.1	<i>De novo</i> assembly of raw data into contigs	71
4.8.2	Genomic reference assembly	72

4.8.3	Rearrangement of contigs	73
4.8.4	Scaffolding, gene prediction and annotation	75
4.8.5	Phylogenomic analysis	79
4.9	<i>In silico</i> characterisation of glycoprotein B	94
4.9.1	Multiple sequence alignment and phylogenetic tree	97
4.9.2	Designed primers and <i>in silico</i> PCR	98
4.10	<i>In silico</i> characterisation of lower matrix phosphoprotein genes	100
4.10.1	Multiple sequence alignment and phylogenetic tree	103
4.10.2	Designed primers and <i>in silico</i> PCR	104
4.11	Laboratory verification of primers	106
5	DISCUSSIONS	108
6	CONCLUSION AND RECOMMENDATIONS	123
6.1	Conclusion	123
6.2	Recommendations for future research	124
	REFERENCES	126
	APPENDICES	145
A	Buffers and Solutions	145
B	Cell count by haemocytometer	147
C	Virus titer determination by TCID ₅₀	148
D	Protein Standard Curve	149
E	SDS PAGE and Western blot	150
F	Polymerase chain reaction (PCR)	154
G	Bioinformatics analysis	155
	BIODATA OF STUDENT	208
	PUBLICATIONS	209