

IDENTIFICATION AND CHARACTERIZATION OF MICRO-  
RNA IN *MACROBRACHIUM ROSENBERGII*

TAN TIAN TIAN

DISSERTATION SUBMITTED IN FULFILMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF MASTER OF  
SCIENCE

INSTITUTE OF BIOLOGICAL SCIENCES  
FACULTY OF SCIENCE  
UNIVERSITY OF MALAYA  
KUALA LUMPUR

2012

UNIVERSITI MALAYA

**ORIGINAL LITERARY WORK DECLARATION**

Name of Candidate: Tan Tian Tian

(I.C/Passport No: 850611-01-6994)

Registration/Matric No: SGR090134

Name of Degree: Master of Science

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):  
IDENTIFICATION AND CHARACTERIZATION OF MICRO-RNA IN  
*MACROBRACHIUM ROSENBERGII*

Field of Study: Genetics and Molecular Biology

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any Copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature

Date

Subscribed and solemnly declared before,

Witness's Signature

Date

Name:

Designation:

## ABSTRACT

MicroRNAs are ~20-22nt non protein-coding RNA regulatory genes that post-transcriptionally regulate many protein-coding genes to influence critical biological and metabolic processes. While the number of known microRNA is increasing, there is no published data for microRNA in freshwater prawns (*Macrobrachium rosenbergii*) a commercially cultured and economically important food species. Through deep parallel sequencing and an *in silico* data analysis approach, 327 miRNA families have been identified from small RNA libraries with reference to both the *de novo* transcriptome of *M. rosenbergii* obtained from RNA-Seq and to miRBase (Release 18.0, November 2011). Based on the identified mature miRNA and recovered precursor sequences that form appropriate hairpin structures, three conserved miRNA (miR125, miR750, miR993) and 27 novel miRNA candidates encoding messenger-like non-coding RNA have been identified. miR-125, miR-750, G-m0002/H-m0009, G-m0005, G-m0008/H-m0016, G-m0011/H-m0027 and G-m0015 were selected for experimental validation with stem-loop quantitative RT-PCR and were found coherent with the expression profile of deep sequencing data as evaluated with Pearson's correlation coefficient ( $r = 0.835178$  for miRNA in gill,  $r = 0.724131$  for miRNA in hepatopancreas). Using a combinatorial approach of pathway enrichment analysis and inverse expression relationship of miRNA and mRNA, four co-expressed novel miRNA candidates (G-m0005, G-m0008/H-m0016, G-m0011/H-m0027, and G-m0015) were found to be associated with energy metabolism. In addition, the expression of the three novel miRNA candidates (G-m0005, G-m0008/H-m0016, and G-m0011/H-m0027) were also found to be significantly reduced at the 9 and 24 hours post infection in a controlled experiment of *M. rosenbergii* challenged with infectious hypodermal and haematopoietic necrosis virus. These findings provide a reference point to further improving the understanding of the repertoire of crustacean miRNAs.

## ABSTRAK

MikroRNAs adalah ~ 20-22 nt bukan protein-pengekodaan RNA gen pengawalseliaan yang mengawal selia pengekodaan gen protein pada peringkat post-transkripsi yang banyak mempengaruhi proses biologi dan metabolisme yang kritikal. Walaupun, bilangan mikroRNA dikenali semakin meningkat, tidak ada data yang diterbitkan bagi mikroRNA dari udang air tawar (*Macrobrachium rosenbergii*) yang merupakan makanan komersial yang penting dari segi ekonomi. Dengan menggunakan pendekatan jujukan dalam selari dan analisis data secara siliko, 327 keluarga miRNA telah dikenal pasti dari perpustakaan kecil RNA dengan merujuk kepada transkripsi data, *M.rosenbergii* yang dihimpun secara *de novo* yang diperolehi dari RNA-seq dan miRBase (Keluaran 18,0, November 2011). Berdasarkan miRNA matang yang dikenal pasti dan pelopor yang membentuk struktur jepit rambut yang sesuai, tiga miRNA konserve (miR125, miR750, miR993) dan 27 calon miRNA novel pengekodaan mesej seperti bukan pengekodaan RNA telah dikenal pasti. miR-125, miR-750, G-m0002/H-m0009, G-m0005, G-m0008/H-m0016, G-m0011/H-m0027 dan G-m0015 telah dipilih untuk pengesahan uji kaji dengan ranting-gelung kuantitatif RT-PCR dan ditemui koheren dengan profil ungkapan data penjujukan yang dalam seperti yang dinilai dengan pekali korelasi Pearson ( $r = 0,835178$  untuk miRNA di insang,  $r = 0,724131$  untuk miRNA dalam hepatopancreas). Menggunakan pendekatan kombinatorik analisis pengayaan laluan dan hubungan ungkapan songsang miRNA dan mRNA, empat calon miRNA novel (G-m0005, G-m0008/H-m0016, G-m0011/H-m0027, dan G-m0015) ditemui berkaitan dengan metabolisme tenaga. Di samping itu, ungkapan untuk tiga calon miRNA novel (G-m0005, G-m0008/H-m0016, dan G-m0011/H-m0027) juga didapati semakin berkurangan pada 9 dan 24 jam selepas jangkitan dalam eksperimen terkawal *M.rosenbergii* yang dicabar dengan virus nekrosis berjangkit hypodermal dan haematopoietic. Penemuan ini menyediakan titik rujukan untuk meningkatkan lagi pemahaman mengenai himpunan miRNAs krustacean.

## ACKNOWLEDGEMENTS

I am sincerely grateful to the people who have assisted me directly and indirectly, to complete my dissertation successfully.

I would like to express my deep appreciation to my supervisors, Professor Dr Jennifer Ann Harikrishna and Dr Subha Bhassu. Their constant encouragement, guidance, willingness to listen, discuss my results with me and provided helpful suggestion when I had problems or unexpected result as well as to share their times will always be greatly appreciated.

Beijing Genomic Institute (BGI), Shenzhen, China and Malaysia Palm Oil Board (MPOB) are acknowledged for granting me permission to use their respective bioinformatics facilities and dot-blot facilities. I would like to thanks Dr Zhang Guojie (BGI) for conceptual advice, Mr Chen Maoshan (BGI) for assistance in bioinformatics, Dr Rajinder Singh (MPOB) for permission to use chemical, consumable and equipment in his laboratory and Ms Leslie Ooi Cheng Li (MPOB) for technical assistance and advice in carrying out the miRNA dot-blot experiment.

This study was supported by grants from the Ministry of Science, Technology and Innovation, Malaysia (MOSTI Grant No: 53-02-03-1030, granted to Dr Subha Bhassu) and from University of Malaya (IPPP Grant No: PS260/2010A, granted to Ms Tan Tian Tian).

Thanks to my colleagues-laboratory mates for their generous assistance and patience throughout my research project and many thanks to my friends and family who always led me out of the darkness whenever I stumble as well as for their invaluable support and their belief in me throughout my life.

## TABLE OF CONTENTS

	<b>Page</b>
<b>ORIGINAL LITERARY WORK DECLARATION</b>	i
<b>ABSTRACT</b>	ii
<b>ABSTRAK</b>	iii
<b>ACKNOWLEDGEMENTS</b>	iv
<b>TABLE OF CONTENTS</b>	v
<b>LIST OF FIGURES</b>	viii
<b>LIST OF TABLES</b>	Ix
<b>LIST OF SYMBOLS AND ABBREVIATIONS</b>	x
<b>LIST OF APPENDICES</b>	xii
<b>CHAPTER I INTRODUCTION</b>	1
1.1 Overview	1
1.2 Importance of the study	3
1.3 Objectives	3
<b>CHAPTER II LITERATURE REVIEW</b>	4
2.1 Description of <i>Macrobrachium rosenbergii</i>	4
2.1.1 Problems associated with production of <i>M.rosenbergii</i>	6
2.2 Small non coding RNA	7
2.3 MicroRNA biogenesis	9
2.4 Mechanism of microRNA function	11
2.5 MicroRNA expression and function in animals	12
2.5.1 MicroRNAs mediate development timing	13
2.5.2 MicroRNAs mediate carcinogenesis	13
2.5.3 Interaction of host and viral miRNA	14
2.5.4 MicroRNAs involved in innate immunity of crustacean	15
2.6 Approaches in identification of microRNA	16
2.7 MicroRNA target prediction	18
<b>CHAPTER III MATERIALS AND METHODS</b>	22
3.1 Sample collection and dissection	22
3.2 DNA extraction	22

3.3	Total RNA extraction	24
3.4	DNase treatment	25
3.5	Preparation of small RNA library	26
3.6	Illumina deep sequencing	26
3.7	<i>In silico</i> analysis and small RNA annotation	27
3.8	MicroRNA target gene prediction and enrichment analysis	30
3.9	Statistical tests for differential expression	31
	3.9.1 Normalization of the expression data	31
	3.9.2 Fold change	32
	3.9.3 Audic-Claverie	32
3.10	MicroRNA dot-blot	33
	3.10.1 Preparation of microRNA oligo dot-blot nylon membrane	33
	3.10.2 Total RNA extraction and enrichment of microRNA	36
	3.10.3 RNA labeling	37
	3.10.4 Post-tailing clean up	38
	3.10.5 Prehybridization and hybridization	38
3.11	Reverse transcription	39
3.12	Validation of miRNA expression by semi quantitative RT-PCR and sequencing	39
3.13	PCR product purification	40
3.14	Validation of miRNA expression by stem-loop quantitative real-time PCR	41
3.15	Pearson correlation coefficient	43
3.16	IHHNV challenge test on <i>M.rosenbergii</i> and qRT-PCR of the miRNA expression	43
<b>CHAPTER IV</b>	<b>RESULTS</b>	<b>46</b>
4.1	Size distribution and mapping of small RNA species from gill and hepatopancreas	46
4.2	Orthologous miRNA and conserved miRNA-variants	49
4.3	Novel miRNA candidates	51
4.4	Multiple distinct miRNA-like RNA arise from a single miRNA precursor	56
4.5	Expression profile	59
	4.5.1 Deep sequencing and semi-quantitative RT-PCR	59
	4.5.2 MicroRNA dot-blot	60
	4.5.3 Real time expression of seven differentially expressed miRNA	67
	4.5.4 Correlation of miRNA sequencing count and qRT-PCR	69
4.6	Inference of novel miRNA function from pathway analysis	69
4.7	Inference of novel miRNA function with preliminary IHHNV challenge test	77
<b>CHAPTER V</b>	<b>DISCUSSION</b>	<b>79</b>

5.1	Overview	79
5.2	miRNA encoding messenger like non-coding loci	79
5.3	Classification of pseudo miRNA and real miRNA	80
5.4	Conserved and novel miRNA candidates in <i>M.rosenbergii</i>	81
5.5	<i>In silico</i> identification of miRNA target genes	82
	5.5.1 miRNA-target mRNA inverse expression correlation analysis	83
5.6	Preliminary controlled experiment of <i>M.rosenbergii</i> challenged with IHHNV	84
5.7	MicroRNA dot-blot	85
5.8	Future Directions	86
<b>CHAPTER VI CONCLUSIONS</b>		88
<b>REFERENCES</b>		90
<b>APPENDIX A</b>		112
<b>APPENDIX B</b>		116



## LIST OF FIGURES

Figure No.		Page
2.1	External anatomy of giant freshwater prawn, <i>Macrobrachium rosenbergii</i>	5
2.2	The miRNA biogenesis pathway in animal	10
3.1	Tissue sample collection of <i>M.rosenbergii</i>	23
4.1	Size distribution of sequenced small RNA of gill and hepatopancreas	47
4.2	Predicted hairpin structures of three conserved orthologs	51
4.3	Transcriptome alignments of miR-125 and G-m0002/H-m0009	57
4.4	Predicted hairpin structures of miR-125 and G-m0002/H-m0009	58
4.5	Relative miRNA expression	61
4.6	Semi-quantitative RT-PCR determination of the expression levels	63
4.7	Sequences of the semi quantitative RT-PCR products for four miRNAs precursors	64
4.8	miRNA dot-blot expression array	65
4.9	Quantitative Real-time PCR (qRT-PCR) determination of mature miRNAs expression levels	68
4.10	Confirmation of accuracy of illumina deep sequencing with qRT-PCR	70
4.11	Total number of predicted target genes for each novel miRNA candidates from G and H small RNA libraries	71
4.12	Oxidative phosphorylation pathway	75
4.13	Predicted target gene based on inverse expression relationship of miRNA and mRNA profiling	76
4.14	Quantitative Real-time PCR (qRT-PCR) determination of mature miRNAs expression levels at various time-point of infection	78

## LIST OF TABLES

Table No.		Page
2.1	Computational target prediction algorithms	19
3.1	DNase I Digestion Mixture	26
3.2	Antisense oligonucleotides for miRNA dot-blot	35
3.3	RNA labeling	38
3.4	Reverse Transcription Mixture	39
3.5	Primers for putative <i>M.rosenbergii</i> miRNA precursors	40
3.6	Custom taqman miRNA assay	43
4.1	Trim clean reads (18 to 30 nucleotides in length) from sequenced gill and hepatopancreas small RNA libraries	47
4.2	Distribution of small RNA reads in the sequenced gill and hepatopancreas small RNA library	49
4.3	The mature and precursor miRNA sequences of 27 predicted novel miRNA candidates	52
4.4	KEGG pathways enriched with targets of each miRNAs in gill and hepatopancreas tissues of <i>M.rosenbergii</i>	74

## LIST OF ABBREVIATIONS AND SYMBOLS

### **B**

bp Base pair

### **C**

cDNA Complementary deoxyribonucleic acid

°C Degree Celsius

### **D**

DEPC Diethyl pyrocarbonate

DNA Deoxyribonucleic acid

DNase Deoxyribonuclease

dNTP Deoxynucleoside triphosphate

### **E**

e.g. Example

### **G**

G Gravity

### **H**

hpi Hours post-injection

### **I**

i.e. That is

### **M**

M Molar

mM Miliampere

mg Miligram

ml	Mililiter
mM	Milimolar
<b>N</b>	
ng	Nanogram
<b>P</b>	
PCR	Polymerase chain reaction
Pmol	Picomole
%	Percentage
<b>R</b>	
RNA	Ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
<b>T</b>	
<i>Taq</i>	<i>Thermus aquaticus</i>
<b>U</b>	
μl	Microliter
USA	United States of America
<b>V</b>	
V	Voltage

## LIST OF APPENDICES

Appendix		Page
A	Predicted hairpin structures of novel miRNA candidate of G and H	112
B	Research publication or conference proceeding	116