CHAPTER IV

RESULTS

4.1 Size distribution and mapping of small RNA species from gill and hepatopancreas

In order to identify the miRNAs in *M.rosenbergii*, two separate libraries of small RNAs (36 nucleotides in length) from the gill (G) and hepatopancreas (H) tissues of *M.rosenbergii* were sequenced using paired-ends sequencing (Illumina Inc). The total number of raw sequencing reads obtained for the G and H small RNA libraries were 16,300,958 reads (90.13% high quality reads; 9.87% low quality reads) and 16,744,853 reads (87.81% high quality reads; 12.19% low quality reads) respectively (Table 4.1). The raw reads obtained were subjected to filtering of low quality reads, trimming of adaptor sequence at the 3' end, cleaning up 5' end adaptor contaminants and filtering of reads smaller than 18 nucleotides to obtain high quality clean reads (18 to 30 nucleotides in length) of 11,718,644 reads for the G and 12,874,553 reads for the H small RNA libraries.

The size distribution of the total sequence reads varied for each library. The peaks in the size distribution at 22 nucleotides that were shown for both small RNA of G and H (Figure 4.1) indicate an enrichment of mature miRNAs in the constructed small RNA libraries. The length of 22 nucleotides was also consistent with the common size of miRNAs from RNase III endonuclease (Dicer) digestion (Cai *et al*., 2010). From the size distribution of total reads for the H small RNA library, a distinct bimodal distribution,

Table 4.1 Trim clean reads (18 to 30 nucleotides in length) from sequenced gill and hepatopancreas small RNA libraries.

Categories	Gill		Hepatopancreas	
	Number of reads	% Number of reads	Number of reads	% Number of reads
Total raw reads	**16,300,958**	**100**	**16,744,853**	**100**
High quality reads	14,692,791	90.13	14,704,371	87.81
Low quality reads	1,608,167	9.87	2,040,482	12.91
High quality reads	**14,692,791**	**100**	**14,704,371**	**100**
Adaptor 3'_null	17,083	0.12	27,997	0.19
Insert_null	71,359	0.49	115,880	0.79
Adaptor 5'_contaminants	400,046	2.72	422,051	2.87
Size of reads less than 18 nt	2,485,468	16.92	1,263,531	8.59
Poly-A-tail	191	0	359	0
Clean reads (18-30nt)	11,718,644	79.76	12,874,553	87.56

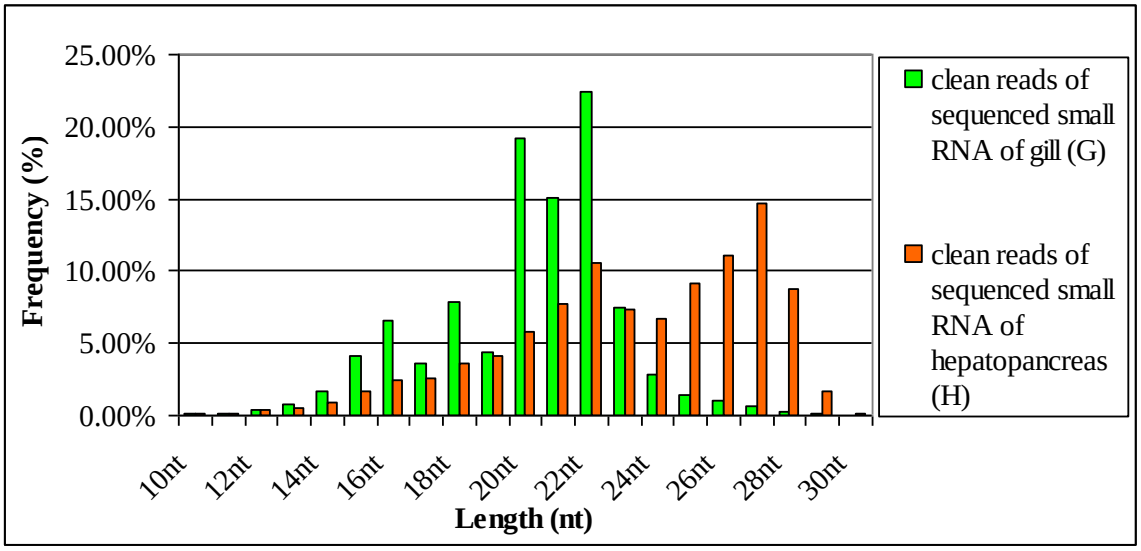


Figure 4.1 Size distribution of sequenced small RNA of gill and hepatopancreas.

with a peak around 21 to 23 nucleotides likely to represent miRNAs and another distinct peak around 26 to 29 nucleotides likely to represent longer piRNA-like small RNA, as shown in silkworm, *Bombyx mori* (Jagadeeswaran *et al*., 2010). In contrast, the G small RNA library had a single peak at around 21 to 23 nucleotides probably corresponding to miRNA. The enrichment of peak around 26 to 29 nucleotides was not expected to be found in the G small RNA library because gill was not germline cell so is not expected to contain piRNA.

The clean reads (18 to 30 nucleotides in length) from both libraries (G and H) were analyzed using the *de novo* transcriptome of *M.rosenbergii* as the reference (Mohd-Shamsudin *et al*., 2012). The draft transcriptome of *M.rosenbergii* was assembled *de novo* from 75bp paired end Ilumina reads of gill, hepatopancreas and muscle tissues, derived from RNA-Seq using SOAP2 (Li *et al*., 2009) with k-mer size of 25. This assembly has generated 102,230 unigenes (Mohd-Shamsudin *et al*., 2012). Only a small percentage of the clean 18 to 30 nucleotide sized reads from each library could be mapped to the *de novo* transcriptome of *M.rosenbergii* without any mismatches; 201,344 reads (1.72 %) for G and 1,468,901 reads (11.41%) for the H small RNA library (Table 4.2). Subsequently, the clean reads that mapped perfectly to the *de novo* transcriptome of *M.rosenbergii* (Mohd-Shamsudin *et al*., 2012) were further annotated and classified into different categories of non-coding RNA as listed in Table 4.2. The remaining sequences that could not map to the *de novo* transcriptome of *M.rosenbergii* (Mohd-Shamsudin *et al*., 2012) were discarded. The sequences matched to known miRNAs of other animal species (13,786 reads for the G and 162,827 reads for H small RNA library) were used to identify conserved miRNA orthologs in *M.rosenbergii* whereas the remaining

unannotated reads; 176,440 reads for the G and 1,290,624 reads for the H small RNA library were used to predict novel miRNA in *M.rosenbergii*.

Table 4.2 Distribution of small RNA reads in the sequenced gill and hepatopancreas small RNA libraries.

Categories	Gill		Hepatopancreas	
	Number of distinct sequences	Number of reads	Number of distinct sequences	Number of reads
Total small RNA (18-30nt)	491,419	11,718,644	2,280,263	12,874,553
Mapping to *de novo* transcriptome	81,564	201,344	251,108	1,468,901
known miRNAs (miRBase 15.0)	1,674	13,786	2,865	162,827
rRNA	1,115	10,371	1,027	14,070
snRNA	123	345	152	765
snoRNA	81	257	93	513
tRNA	69	145	46	102
Unannotated	78,502	176,440	246,925	1,290,624

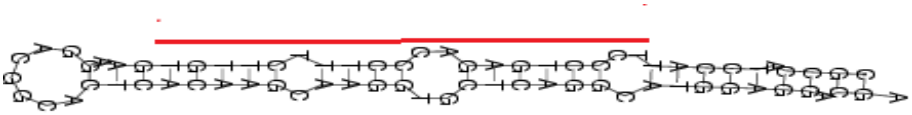
4.2 Orthologous miRNA and conserved miRNA-variants

Sequence comparisons with reads from G and H small RNA libraries respectively to the miRBase 15.0 (Griffiths-Jones, 2004; Griffiths-Jones *et al.,* 2006; Griffiths-Jones *et al.,* 2008, Kozomara and Griffiths-Jones, 2011) allowing at most 2 mismatches outside of the seed region revealed that 327 miRNA candidate sequences from *M.rosenbergii* were orthologous to those of mature miRNA identified in other animal species. Uridine residues are predominant in the first nucleotide position of the

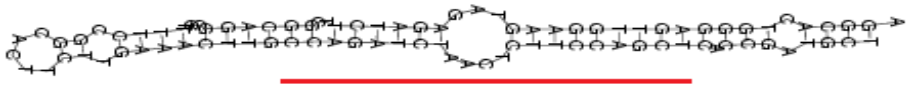
mature miRNA and might play an important role in recognition of the miRNAs by the effector complexes i.e. miRNA-containing ribonucleoprotein complex (miRNP) or miRNA-containing RNA induced silencing complex (miRISC) (Mi *et al*., 2008). As reported for other animals such as pea aphid, *Acyrthosiphon pisum* (Legeai *et al*., 2010) and human parasite, *Schistosoma mansoni* (de Sauza Gomes *et al.,* 2011) most of the predicted *M.rosenbergii* miRNA mature sequences (58.75% from the H library and 77.68% from the G library) start with a uridine residue at their 5'end.

One of the important features that differentiate miRNA from other endogenous small RNA is the ability of the precursor miRNA (pre-miRNA) sequence to adopt a stem-loop hairpin structure with mature miRNA present in one of its arms (Chen *et al*., 2009 Wei *et al*., 2011). MIREAP analysis of the *de novo* transcriptome of *M.rosenbergii* (Mohd-Shamsudin *et al*., 2012) for hairpin structures, identified three conserved orthologs predicted to form a hairpin from the sequenced small RNA libraries derived from messenger like non coding transcript (Figure 4.2). The conserved miRNA-variants were assigned to family miR-125, taxon: Crustacean (identity score of 20/22 as referred to MIMAT0000397, dme-miR-125, taxon: Hexapoda), miR-750, taxon: Crustacean (identity score of 21/23 as referred to MIMAT0012704, isc-miR-750, taxon: Chelicerata) and miR-993, taxon: Crustacean (identity score of 22/23 as referred to MIMAT0012676, dpu-miR-993, taxon; Crustacean) based on sequence similarity. In the current study, miR-125 and miR-993 family members were found in both G and H small RNA populations whereas the miR-750 candidate was found only in the H small RNA library. All three conserved miRNAs (miR-125, miR-993 and miR-750) were found in the transcriptome of *M.rosenbergii* (taxon: Crustacean).

miR-125



miR-750



miR-993

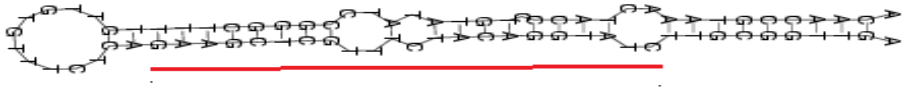


Figure 4.2 Predicted hairpin structures of three conserved orthologs. The RNA secondary structures of pre-miRNA hairpins were drawn using RNAfold (Hofacker, 2009). Red line represents the mature miRNA position.

4.3 Novel miRNA candidates

MIREAP analysis of the *de novo* transcriptome of *M.rosenbergii* (Mohd-Shamsudin *et al*., 2012) for predicted hairpin structures with the unannotated reads from the G and H small RNA libraries identified a total of 59 novel miRNA candidates with hairpin formation but lacking significant identity with any known mature miRNA sequence. After the removal of 32 pseudo precursor miRNA sequences on the basis of their overlap with protein coding region and homology to known non-coding RNA (based on sequence similarity and secondary structure), the remaining 27 candidates were considered as real novel miRNAs (Table 4.3). These 27 novel miRNA candidates were derived from messenger like non coding RNA transcript (mlncRNA) loci. The fold-back sizes of the predicted novel miRNAs ranged from 60 to 120 nucleotides (Appendix A).

Table 4.3 The mature and precursor miRNA sequences of 27 predicted novel miRNA candidates

Predicted novel miRNA candidates	Mature miRNA sequence (5' to 3')	Precursor miRNA sequences (5' to 3')
G-m0002/H-m0009	ACAAGCAAGGTGCTCAGGCAT	CTCGCCATCCATTCCCTGAGACCCTTTCTTGTGAAGG ACGGCACTCACAAGCAAGGTGCTCAGGCATGGAGGA CGAT
G-m0005	AGTACTAAACACGGCGAAACAG	CCGAGGGGCTAGTACTAAACACGGCGAAACAGTGAT TCATCCAATTTGTTTCGCCGCATTTAGTACTAGCCACC CGGG
G-m0006	CTGTTTCGCCGTGTTTAGTACTA	CCCCGGGTGGCTAGTACTAAATGCGGCGAAACAAAT TGGATGAATCACTGTTTCGCCGTGTTTAGTACTAGCC CCTCGGG
G-m0007	TACACTGTAGGCATTACTTAA	TGCGTGTGTCACCGTAGGTGGTAGTGCCGTCAGTGCA TCTCACGCACAGTACACTGTAGGCATTACTTAAGATT CTTTGC
G-m0008/ H-m0016	AGGGCTGCAATTTGGTATGTTT	CTGAGGCTACAGGGCTGCAATTTGGTATGTTTTGGTG ATTGGAGGGTGGGTGACCAACACACTAATTTGCAGC CCTCTAGCCTCAGT
G-m0009	TGGGGTCTGCGGTGGGAGGTTG	TGCTGCAGTGTGGGGTCTGCGGTGGGAGGTTGAAAC CAACATTCTTTGGAAACTTAAATTTCAAGTCGGTGAC CCCTGCGTGCTTGT

G-m0011/ H-m0027	AGAGGGCTGCAATTTGGTATGT	TGAGGAGGCTAGAGGGCTGCAATTTGGTATGTTTGAT GATTGGAGAATGGATGATCAACCAATTTGCAGCCGC AGTAGTTTTTAAG
G-m0013	GAGGGCTGAAAATTGGTATGTT	ACTGAGGCTGGAGGGCTGAAAATTGGTATGTTAGTC ATCCACCCTCCAATCATCCAACATACCAAATTGCAAC CCTCTATCCTCGGTA
G-m0014/ H-m0031	AGTTGGAAGTAGAGATCTTCGGCA	AGGCACTGGGAGTTGGAAGTAGAGATCTTCGGCAGG CCTTTTCCGGCACTTCTTGAAACTTGCCAGATCTAACT CTTCCAGCTCACCGATGCTC
G-m0015	GAACGAGACGGCGAGAGAGTGA	GGAACGAGACGGCGAGAGAGTGAAAGAGGAACGTTT CTCTCTCGTTTCCTCTCGTTTCA
H-m0005	TCGTTAACCGGATTTTCAGTG	GCTAGCGACGTCGTTAACCGGATTTTCAGTGCCAGTA AGCAGTCGGCGCCGATATCCGGTTAACGGTGCTGAT
H-m0015	TTGGTTTGATCGGGTTAGGCAG	GCGGGCTGGTTTGGTTTGATCGGGTTAGGCAGGCCTT GGAGGAGGCGGATGGTCGTCTGGCTTCGGTGGTGAC TTCTAATAGGCAGGA
H-m0019	TGTCTTTAAGATCCGAGGGATA	GCCGAAGCCTTGTCTTTAAGATCCGAGGGATAGAGG AGACCCTGTCCCTGAGCTTGGTGTTTGCTCGGCT
H-m0020	TGGAGATTGAGACAGGGTGATGT	TCTCCACGTCTGGAGATTGAGACAGGGTGATGTCCAA TTATGGGCCCAACTACGGCGCTGGCCTATACCTCCAC CAATTCACAGGCCGTGATCCAACTG

H-m0021	CTTTCGAGCAGGACGGATTTCGC	GCAGTGGCAGACCAGACTCGGTCCTACAATGAAGTG CACACTATAAGCAATTTTGAAGATATCAGCCCTTTCG AGCAGGACGGATTTCGCCGCCGCTGGG
H-m0022	TACGAGCGGAAGAGCAGCAGTTTC	CACGGCGATGCGTCTCGTGGTTTCTCGTGCATAGTTC GTCCGGAGGGAACGAACACCTGTGGCAGTACGAGCG GAAGAGCAGCAGTTTCGTGTCTCGGT
H-m0023	TGTAAAGAATAACGAACCCAATA	ATTGTTCCCATGTAAAGAATAACGAACCCAATAATCT TCTTGTAAAGGGTAGCGAACCTTGCTGCAAGTGGGTT TGCTAATCTTGACACGATCCTTTC
H-m0025	CACGCGACTAGACTACAGTTG	AACCGTTTATCACGCGACTAGACTACAGTTGGTGCTC GAGGTTTCTATCGATTGTTGTCACTCGTTGGTCAAAC TATA
H-m0030	AGAACCTGTAGCTTTAGTTTCT	AAAAAAATAGAGAACCTGTAGCTTTAGTTTCTGCGGA AGTACCGCGTAAGGATGAGTCAAAATGACTAGGGTT TCAAAGGTACATATCGAGAGG
H-m0032	AGAGGGCTGCAAATTGATATGT	CATTTTGTCCGCACTTTTTCTGTCCTCCCTCAGATGTT AAAAACCACTGAGGCCAGAGGGCTGCAAATTGATAT GTTGATCTTCCA
H-m0033	TTGAACTGTATTTGGTGGAAA	TTTTGAAAGCCTTTCATCGCCATTCACTCAATGATTA AGGAAGATTCATTTTGTCCAATGCCGAATTGAACTGT ATTTGGTGGAAAGCTTTCCTGT
H-m0034	TGTCTCTAAATCTGGTCTTCTTT	GGAGGAATACTAAGGTTAAGTCCTGTTTTAGGCGCCC TTCAGCTGTTGCAAAGGAACACACGTGTCTCTAAATC TGGTCTTCTTTGACTGTCCTT

H-m0038	TGAGCGTGATGTCGCATTTTCGTGA	TGCCACAGAATGAGCGTGATGTCGCATTTTCGTGACT CCAGCCTTGAATTCGACGTTGCAGCCTTCGCAATGAG TGCTCGCTTCAAACTTACGGGTGGATTCAG
H-m0039	ATCCGGATTAAGCATTGTTGTCA	TATTAGACTCTTAAACAACACTTCTTAATCCTACCTTC TGAAAGATAAAGAAATCCGGATTAAGCATTGTTGTC AAAAGTTTGAT
H-m0041	TAGGATTGCTGTGGGGTACTCAAG	TTCTGCAGGGCTTGCATCTTCATTACTGTGGTCCTACA TGCTGGGGCTAGGATTGCTGTGGGGTACTCAAGCTGT GGCCTC
H-m0042	TTGACTCGAGCACCGGCGGAAACA	CGACACAGTGTTGACTCGAGCACCGGCGGAAACAGA GTGAGGTTGTCTGCTAGTGGTTCCTGGTACAGTATTC CCGTTAGTGGGC
H-m0044	TAAGAACATGCCGGTTTATGG	ACTTCTTTCCAAAATGCAGTATGTTCACTGAGTCACA ATAGACACAGTCATGAAGAAAGCATCATTAAGAACA TGCCGGTTTATGGTAAAATAAGG

4.4 Multiple distinct miRNA-like RNA arise from a single miRNA precursor

Unigene15704_All as shown in Figure 4.3 was a subset of 102,323 unigenes representing the draft transcriptome datasets of *M.rosenbergii* (unpublished data, Mohd Shamsudin). From transcriptome alignments of Unigene15704_All, it was observed that one of the novel miRNA candidates, G-m0002/H-m0009 derives from the same precursor as miR-125 (Figure 4.3 and Figure 4.4). This shows that a single precursor may produce two distinct miRNAs (miR-125 and G-m0002/H-m0009) as has been reported for other miRNA such as sma-mir-8 in human parasite, *Schistosoma mansoni* (de Souza Gomes *et al.*, 2011). On the other hand, since the numbers of reads detected in small RNA library sequencing for the mature miRNA sequence of miR-125 (Figure 4.3A and 4.3B) are more than that of G-m0002/H-m0009 (Figure 4.3C and 4.3D) this might indicate that the mature sequence of G-m0002/H-m0009 is the miR-125 star sequence. In addition, there was complementarity between the mature sequence of miR-125 and G-m0002/H-m0009 which would be expected if G-m0002/H-m0009 is the miR-125 star sequence (Figure 4.4). Nevertheless, based on current data, it is not possible to confirm which of these two possibilities is correct.

A

```
mir-125_G Unigene15704_All:212:285:- 74(nt) -34.80(kcal/mol)
CGCCATCCATTCCCTGAGACCCTTTCTTGTGAAGGACGGCACTCACAAGCAAGGTGCTCAGGCATGGAGGACGA mir-125_G 7523
((((.(((((..((((((..((((.(((((((.(.......)))))))).))))..))))))).)))))).)).
**********TCCCTGAGACCCTTTCTTGT********************************************* mir-125_G-5p 7523
-------CATTCCCTGAGACCCTTTCTTGTGA---------------------------------------------- t0198429 1
----------TCCCTGAGACCCTTTCTT------------------------------------------------- t0001302 379
----------TCCCTGAGACCCTTTCTTG------------------------------------------------ t0000515 921
----------TCCCTGAGACCCTTTCTTGT----------------------------------------------- t0000201 2678
----------TCCCTGAGACCCTTTCTTGTG---------------------------------------------- t0000549 862
----------TCCCTGAGACCCTTTCTTGTGA--------------------------------------------- t0000211 2558
----------TCCCTGAGACCCTTTCTTGTGAA-------------------------------------------- t0004353 99
-----------CCCTGAGACCCTTTCTTGTG---------------------------------------------- t0134522 1
------------CCTGAGACCCTTTCTTGTGA---------------------------------------------- t0012636 24
```

B

```
mir-125_H Unigene15704_All:212:285:- 74(nt) -34.80(kcal/mol)
CGCCATCCATTCCCTGAGACCCTTTCTTGTGAAGGACGGCACTCACAAGCAAGGTGCTCAGGCATGGAGGACGA mir-125_H 969
((((.(((((..((((((..((((.(((((((.(.......)))))))).))))..))))))).)))))).)).
**********TCCCTGAGACCCTTTCTTGTGA********************************************* mir-125_H-5p 969
-------CATTCCCTGAGACCCTTTCTTGTGA---------------------------------------------- t1692616 1
----------TCCCTGAGACCCTTTCTT------------------------------------------------- t0021470 50
----------TCCCTGAGACCCTTTCTTG------------------------------------------------ t0011511 100
----------TCCCTGAGACCCTTTCTTGT----------------------------------------------- t0004925 244
----------TCCCTGAGACCCTTTCTTGTG---------------------------------------------- t0008425 139
----------TCCCTGAGACCCTTTCTTGTGA--------------------------------------------- t0002970 394
----------TCCCTGAGACCCTTTCTTGTGAA-------------------------------------------- t0040258 25
----------TCCCTGAGACCCTTTCTTGTGAAG------------------------------------------- t0778993 1
-----------CCCTGAGACCCTTTCTTGTGA---------------------------------------------- t1083431 1
------------CCTGAGACCCTTTCTTGTGA---------------------------------------------- t0065653 14
```

C

```
G-m0002 Unigene15704_All:211:287:- 77(nt) -36.30(kcal/mol)
CTCGCCATCCATTCCCTGAGACCCTTTCTTGTGAAGGACGGCACTCACAAGCAAGGTGCTCAGGCATGGAGGACGAT G-m0002 586
.(((((.(((((..((((((..((((.(((((((.(.......)))))))).))))..))))))).)))))).))).
*********************************************ACAAGCAAGGTGCTCAGGCAT********** G-m0002-3p 586
----------------------------------------------ACAAGCAAGGTGCTCAGG------------- t0023813 10
----------------------------------------------ACAAGCAAGGTGCTCAGGC------------ t0187895 1
----------------------------------------------ACAAGCAAGGTGCTCAGGCA----------- t0008829 39
----------------------------------------------ACAAGCAAGGTGCTCAGGCAT---------- t0000923 530
-----------------------------------------------CAAGCAAGGTGCTCAGGCAT---------- t0119954 2
-----------------------------------------------AAGCAAGGTGCTCAGGCA----------- t0242048 1
-----------------------------------------------AAGCAAGGTGCTCAGGCAT---------- t0121007 2
------------------------------------------------AGCAAGGTGCTCAGGCAT---------- t0348035 1
```

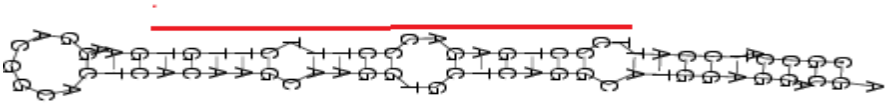
D

```
H-m0009 Unigene15704_All:211:287:- 77(nt) -36.30(kcal/mol)
CTCGCCATCCATTCCCTGAGACCCTTTCTTGTGAAGGACGGCACTCACAAGCAAGGTGCTCAGGCATGGAGGACGAT H-m0009 170
.(((((.(((((..((((((..((((.(((((((.(.......)))))))).))))..))))))).)))))).))).
*********************************************ACAAGCAAGGTGCTCAGGCAT********** H-m0009-3p 170
---------------------------------------------CACAAGCAAGGTGCTCAGGCAT---------- t1025292 1
----------------------------------------------ACAAGCAAGGTGCTCAGG------------- t0306429 3
----------------------------------------------ACAAGCAAGGTGCTCAGGCA----------- t0047826 20
----------------------------------------------ACAAGCAAGGTGCTCAGGCAT---------- t0008231 143
-----------------------------------------------CAAGCAAGGTGCTCAGGCAT---------- t0776918 1
-----------------------------------------------AAGCAAGGTGCTCAGGCAT---------- t0394999 2
```

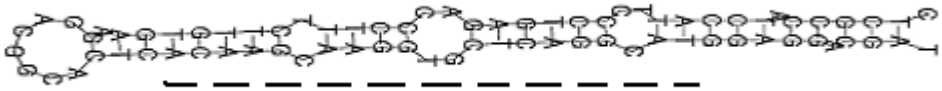
Figure 4.3 Transcriptome alignments of miR-125 and G-m0002/H-m0009. The MIREAP was used to align short reads from small RNA library to the transcripts from transcriptome library. The miRNA precursor predicted secondary structures are shown in the form of dot-bracket notation. Short RNA reads are followed by id numbers (starting with the letter‘t’) and number of reads detected in small RNA library sequencing.

(A). Shorts reads of miR-125 from the small RNA library of gill aligned to single transcript, Unigene15704_All from the transcriptome library.
(B). Shorts reads of miR-125 the small RNA library of hepatopancreas were aligned to single transcript, Unigene15704_All from the transcriptome library.
(C). Shorts reads of G-m0002/H-m0009 from the small RNA library of gill aligned to single transcript, Unigene15704_All from the transcriptome library.
(D). Shorts reads of G-m0002/H-m0009 from the small RNA library of hepatopancreas aligned to single transcript, Unigene15704_All from the transcriptome library.

miR-125



G-m0002/H-m0009



Complementarity between the mature sequence of miR-125 and G-m0002/H-m0009

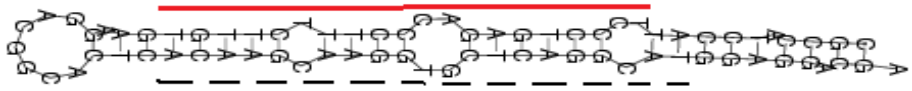


Figure 4.4 Predicted hairpin structures of miR-125 and G-m0002/H-m0009. The RNA secondary structures of pre-miRNA hairpins were drawn using RNAfold (Hofacker, 2009). The red lines and broken lines represent the positions of the mature miRNA of miR-125 and G-m0002/H-m0009 respectively.

4.5 Expression profile

4.5.1 Deep sequencing and semi-quantitative RT-PCR

The representation of each unique sequence in deep parallel sequencing can be used to reflect its abundancy (Chen *et al*., 2009) and was therefore used to establish differential miRNA expression between gill and hepatopancreas (from the G and H libraries respectively) as shown in Figure 4.5. Having computationally identified three conserved and 27 novel miRNA candidates from small RNA libraries with reference to the draft transcriptome of *M.rosenbergi* (unpublished data, Mohd Shamsudin), three miRNA candidates (miR-125, miR-750 and G-m0002/H-m0009) were selected for experimental validation with semi-quantitative PCR and conventional sequencing. The main purpose was to verify the presence of the precursor miRNAs since the draft transcriptome of *M.rosenbergii* (Mohd-Shamsudin *et al*., 2012) may have assembly errors since it was assembled *de novo* without a reference genomic sequence (Wang *et al*., 2009). Forward and reverse primers were designed from the unigene from the draft transcriptome of *M.rosenbergii* (Mohd-Shamsudin *et al*., 2012) that flanked the miRNA candidate region. Semi-quantitative RT-PCR was performed in gill and hepatopancreas tissues. Amplicons (162bp) from primers for pre-miR-125/pre-miR-G-m0002/pre-miR-H-m0009 were observed in gill (Lane 1 and 2 in Figure 4.6A) and hepatopancreas (Lane 3 and 4 in Figure 4.6A) whilst amplicons (297bp) from primers for pre-miR-750 was only observed in hepatopancreas (Lane 3 and 4 in Figure 4.6B). Additionally, the expression of miR-125 was 32 fold higher in G compared to H whilst miR-750 was expressed in H only as reflected by the deep sequencing frequencies. This shows that the expression patterns of both transcripts determined by PCR, correlated to the deep sequencing frequencies for each miRNA (Figure 4.6). The amplicons of 162bp and

297bp were also observed in Lane 5 and 6 in Figure 4.6 when genomic DNA was used as the template with the primer for pre-miR-125/pre-miR-G-m0002/pre-miR-H-m0009 and pre-miR-750 repectively. This result implies that there was no intron resided between the regions that primers targeted. Sequencing confirmed the identities of amplicons from semi-quantitative RT-PCR (Figure 4.7).

4.5.2 MiRNA dot-blot

In this study, a membrane-based miRNA profiling where a positive charge nylon membrane spotted with oligodeoxynucleotides (oligos) antisense to mature miRNAs, mature-star miRNA and other small RNA sequences was used to profile miRNA expression in tissues i.e. hepatopancreas (Figure 4.8A), gill (Figure 4.8B), muscle, pleopod and eye. Small molecular weight RNAs from tissues were radiolabeled and used as probes to hybridize with the oligos on the membrane with dot-blot. Hybridization signals were only observed in miRNA dot-blot arrays hybridized with RNA from hepatopancreas and gill tissues. No hybridization signal was observed on the non probe control (negative control) or for the positive control which was a serial dilution of beta-actin (22 nucleotides) on an additional strip of nylon membrane (Figure 4.8). Therefore, a more accurate and reliable tool such as stem-loop quantitative real time PCR (qRT-PCR) was used to determine the expression of candidate miRNA genes of *M.rosenbergii*.

A

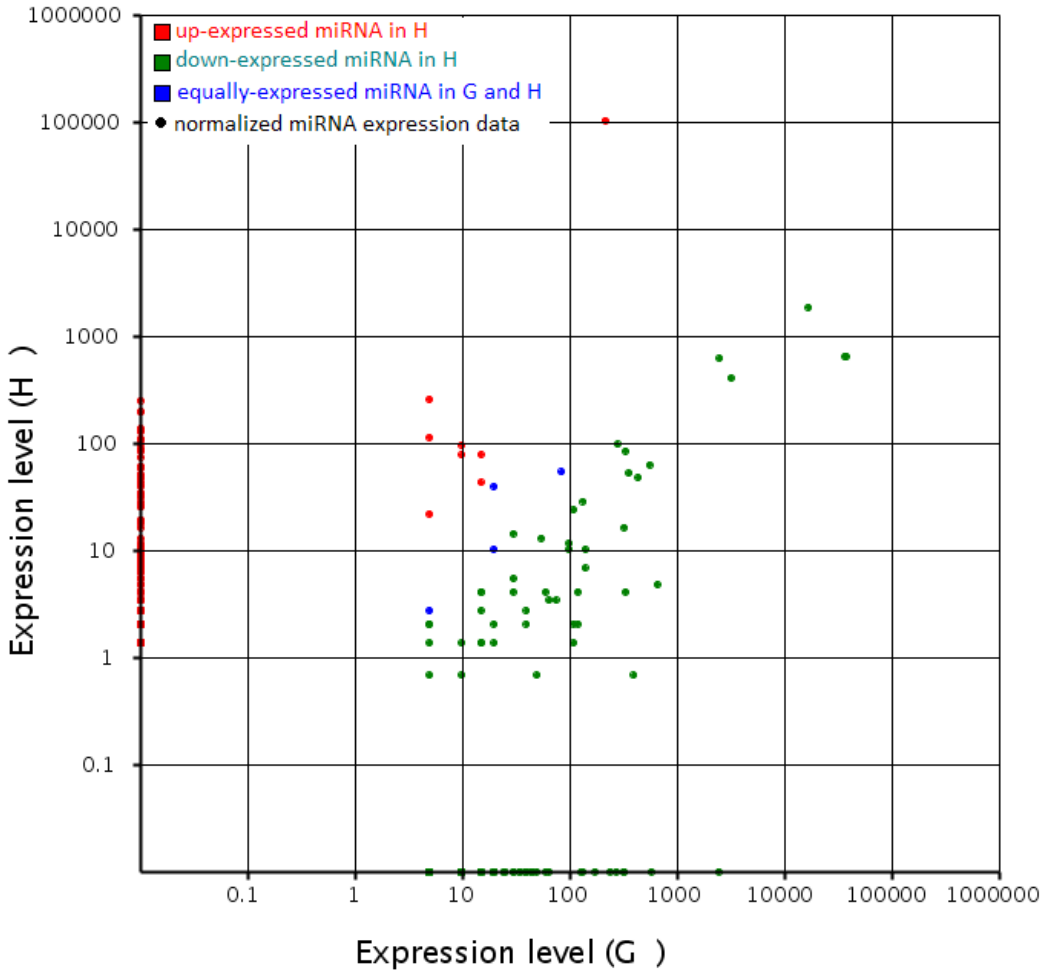


Figure 4.5, continued

B

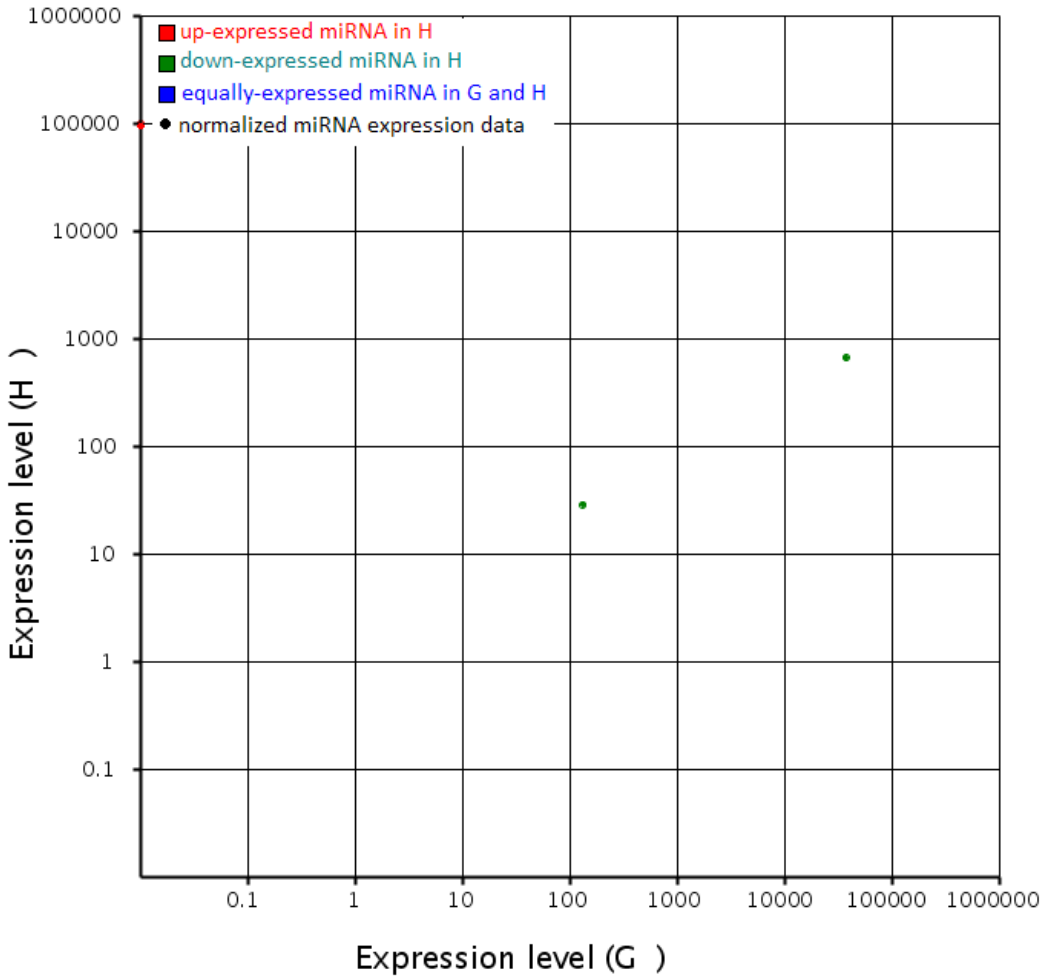


Figure 4.5 Relative miRNA expression. Each data point represents a unique miRNA and is based on number of occurrences of a particular mature miRNA in gill (G) and hepatopancreas (H) tissues. Normalized expression data for each unique miRNA in G and H was used to plot the graph of differential miRNAs expression level in G and H of *M.rosenbergii*. (A). Differential expression profile for orthologous miRNA candidates in G and H of *M.rosenbergii*, (B). Novel miRNA candidates in G and H of *M.rosenbergii***.**

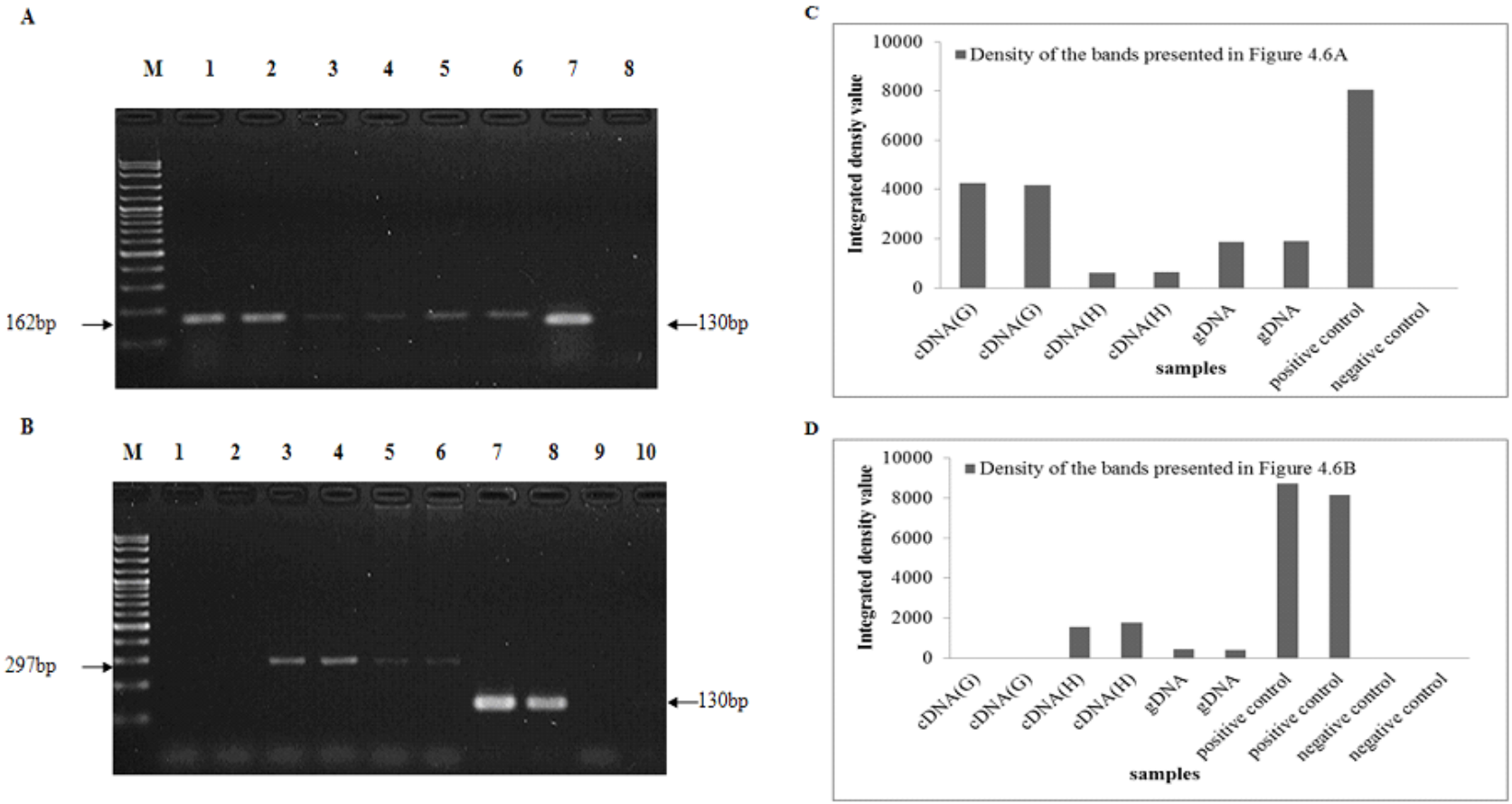


Figure 4.6 Semi-quantitative RT-PCR determination of the expression levels.

(A). Amplicons from primers for pre-miR-125 / pre-miR-G-m0002/ pre-miR-H-m0009 (162bp) in G and H. M: 100bp DNA marker, Lane 1 and 2: cDNA (G), Lane 3 and 4: cDNA (H), Lane 5 and 6: genomic DNA, Lane 7: positive control (130bp), Lane 8: negative control.

(B). Amplicons from primers for pre-miR-750 (297bp) in G and H. M: 100bp DNA marker, Lane 1 and 2: cDNA (G), Lane 3 and 4: cDNA (H), Lane 5 and 6: genomic DNA, Lane 7 and 8: positive control (130bp), Lane 9 and 10: negative control.

(C). The density of the bands presented in Figure 4.6A were analyzed using Image J.

(D). The density of the bands presented in Figure 4.6B were analyzed using Image J.

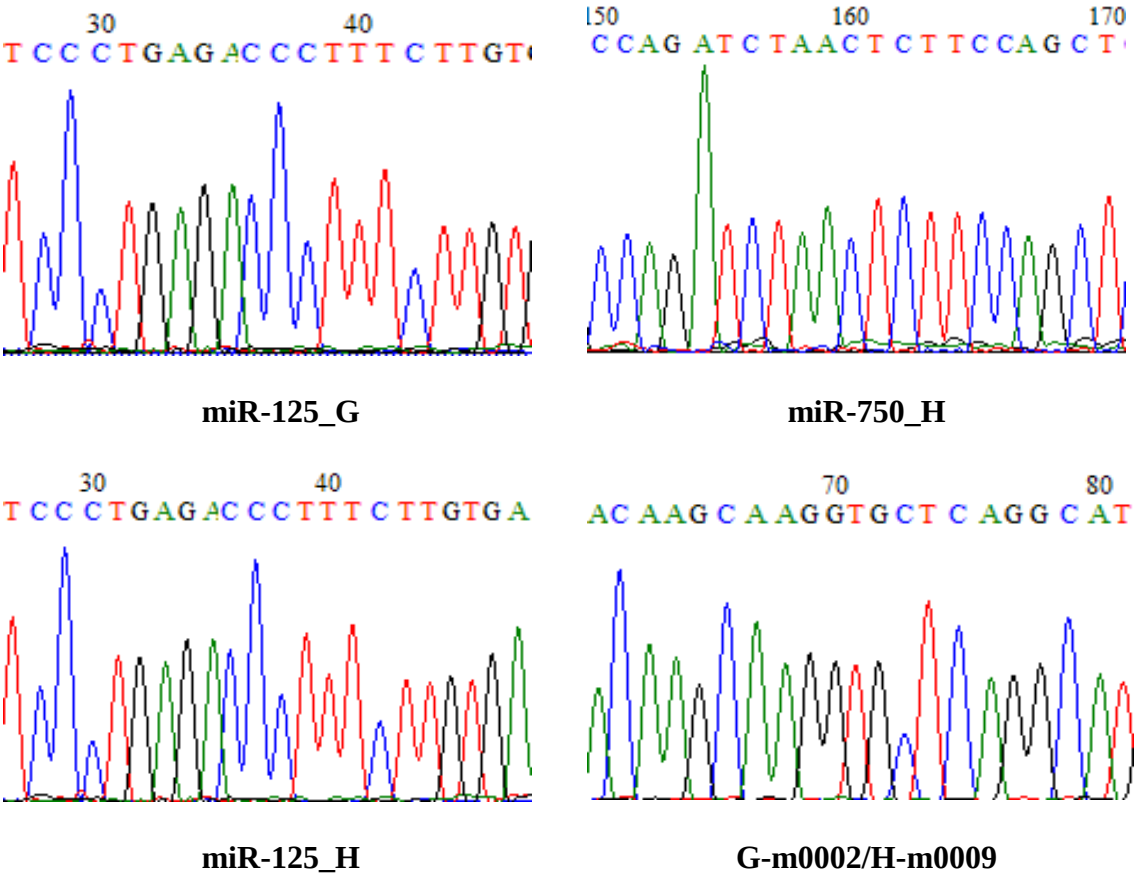
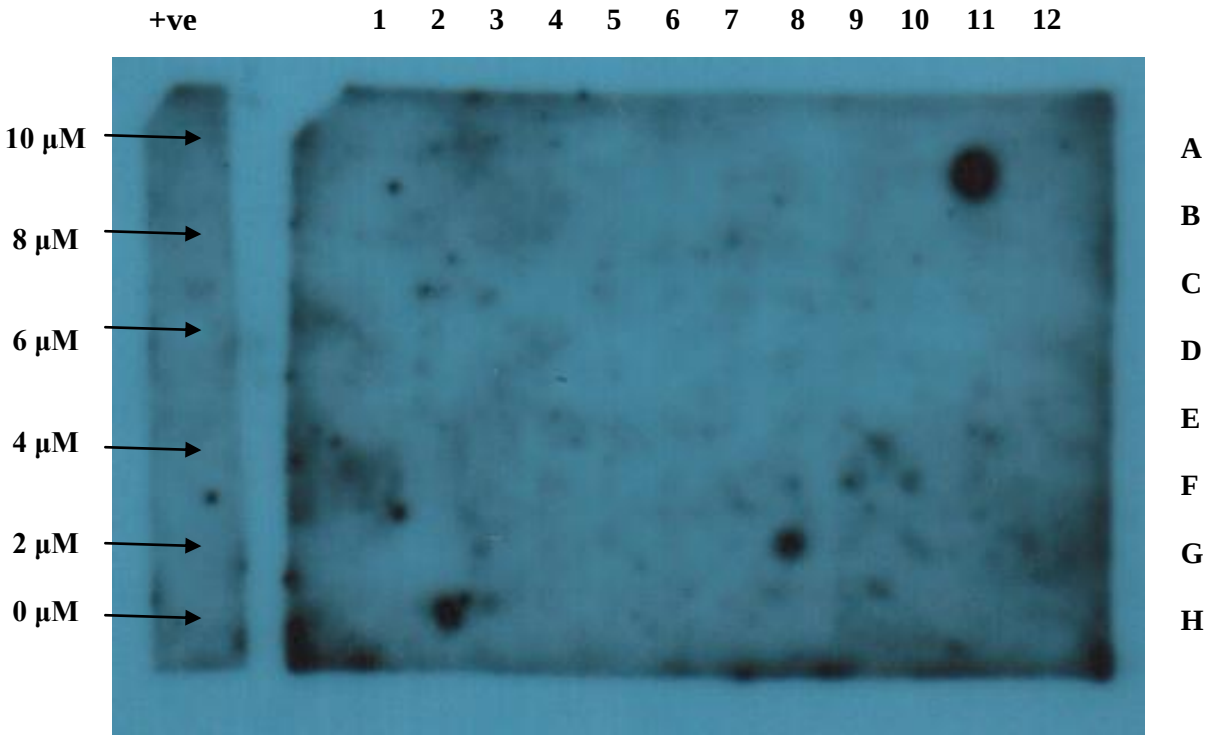


Figure 4.7 Sequences of the semi quantitative RT-PCR products for four miRNA precursors. Only the mature miRNA sequence is shown in each case and represents the dominant mature miRNA sequence from the illumina deep-sequencing data.

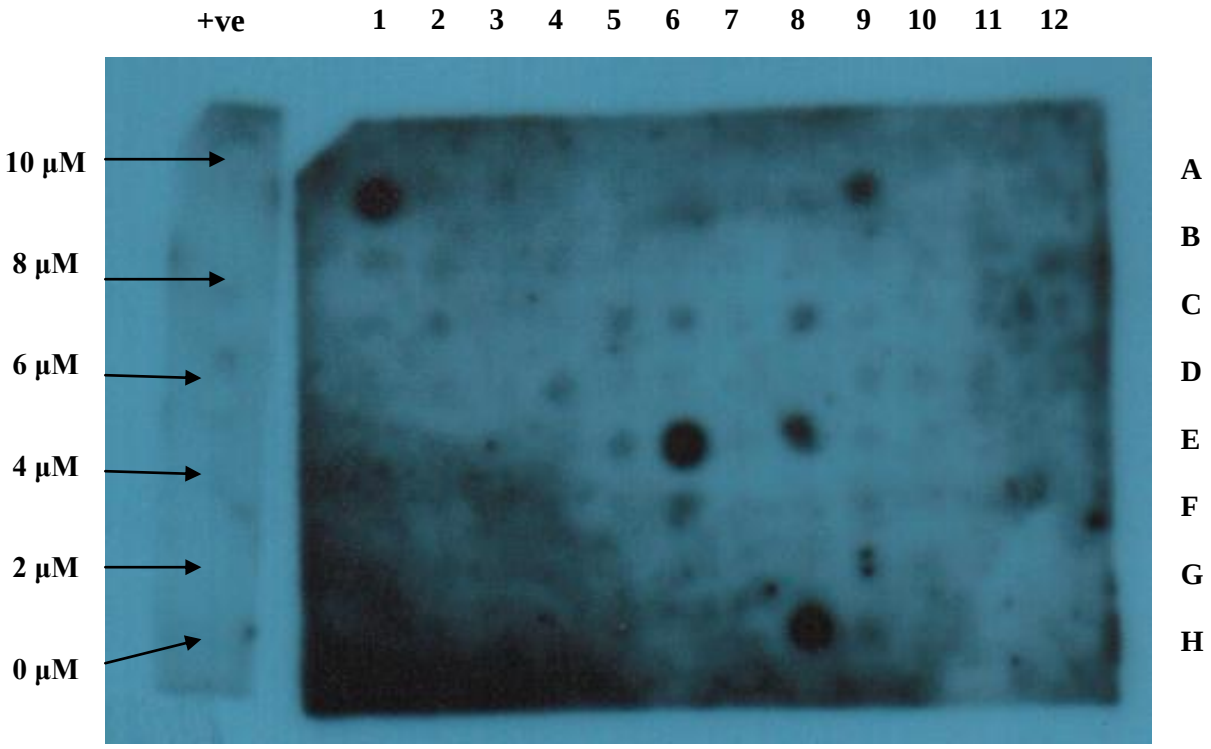
A



	1	2	3	4	5	6	7	8	9	10	11	12
A	miR-125_G	miR-125_H	miR-750_H	G_m0003	G_m0005	G_m0006	miR-125_G	miR-125_H	miR-750_H	G_m0003	***G_m0005***	G_m0006
B	UCR	UCR	G_m0009	G_m0011	G_m0007	G_m0019	UCR	UCR	G_m0009	G_m0011	G_m0007	G_m0019
C	G_m0008	G_m0004	G_m0004*	G_m0002	UCR	UCR	G_m0008	G_m0004	G_m0004*	G_m0002	UCR	UCR
D	UCR	UMG	H_m0012	UMG	H_m0019	UMG	UCR	UMG	H_m0012	UMG	H_m0019	UMG
E	UMG	H_m0020	H_m0020*	H_m0021	H_m0022	UMG	UMG	H_m0020	H_m0020*	H_m0021	H_m0022	UMG
F	UCR	UCR	H_m0025	H_m0034	H_m0030	H_m0041	UCR	UCR	H_m0025	H_m0034	H_m0030	H_m0041
G	H_m0044	UCR	UMG	UMG	H_m0033	UCR	H_m0044	***UCR***	UMG	UMG	H_m0033	UCR
H	UCR	UMG	UMG	UMG	UMG	-ve	UCR	UMG	UMG	UMG	UMG	-ve

Figure 4.8, continued

B



	1	2	3	4	5	6	7	8	9	10	11	12
A	***miR-125_G***	miR-125_H	miR-750_H	G_m0003	G_m0005	G_m0006	miR-125_G	miR-125_H	***miR-750_H***	G_m0003	G_m0005	G_m0006
B	UCR	UCR	G_m0009	G_m0011	G_m0007	G_m0019	UCR	UCR	G_m0009	G_m0011	G_m0007	G_m0019
C	G_m0008	G_m0004	G_m0004*	G_m0002	UCR	UCR	G_m0008	G_m0004	G_m0004*	G_m0002	UCR	UCR
D	UCR	UMG	H_m0012	UMG	H_m0019	UMG	UCR	UMG	H_m0012	UMG	H_m0019	UMG
E	UMG	H_m0020	H_m0020*	H_m0021	H_m0022	***UMG***	UMG	***H_m0020***	H_m0020*	H_m0021	H_m0022	UMG
F	UCR	UCR	H_m0025	H_m0034	H_m0030	H_m0041	UCR	UCR	H_m0025	H_m0034	H_m0030	H_m0041
G	H_m0044	UCR	UMG	UMG	H_m0033	UCR	H_m0044	UCR	UMG	UMG	H_m0033	UCR
H	UCR	UMG	UMG	UMG	UMG	-ve	UCR	***UMG***	UMG	UMG	UMG	-ve

Figure 4.8 miRNA dot-blot expression array (A) hybridized with RNA from hepatopancreas and (B) gill tissues. Arrangement of the antisense oligodeoxynucleotide on the nylon membrane was shown. Bold and italic represent the position of the hybridization signal on the membrane. UCR represent unannotated clean read that unmapped to draft trancriptome of *M.rosenbergii*. UMG represent unannotated clean

read that mapped to draft transcriptome of *M.rosenbergii*. Both UCR and UMR are unannotated small RNA. –ve indicate negative control (non probe control). +ve indicate positive control which represent a serial concentration, 10 μM, 8 μM, 6 μM ,4 μM ,2 μM, and 0 μM of beta-actin (22 nucleotides) on an additional strip of nylon membrane.

4.5.3 Real time expression of seven differentially expressed miRNA

Seven differentially expressed miRNAs were randomly selected for validation with stem-loop quantitative real time PCR (qRT-PCR) for pearson correlation analysis (Section 4.5.4). The miR-125, miR-750, G-m0002/H-m0009, G-m0005, G-m0008/H-m0016, G-m0011/H-m0027 and G-m0015 have diverse expression across the hepatopancreas, gill, muscle, pleopod and eye tissues as assayed by qRT-PCR (Figure 4.9). The reference gene, small nuclear RNA (snRNA) of *M.rosenbergii* was used to normalize the relative abundance of the miRNA. The quantification analysis was carried out relative to miRNA expressed in gill tissue as the calibrator. The G-m0011/H-m0027, miR-125 and G-m0002/H-m0009 were abundantly expressed in muscle tissue with differences of 4-, 9- and 10- fold respectively, relative to gill tissue whilst miR-750 was most abundantly expressed in hepatopancreas tissue with differences of 58,210- fold relative to gill tissue. Both G-m0005 and G-m0008/H-m0016 with differences of 101- and 196- fold respective to gill tissue were abundantly expressed in eye tissues whereas G-m0015 was absent in eye tissues.

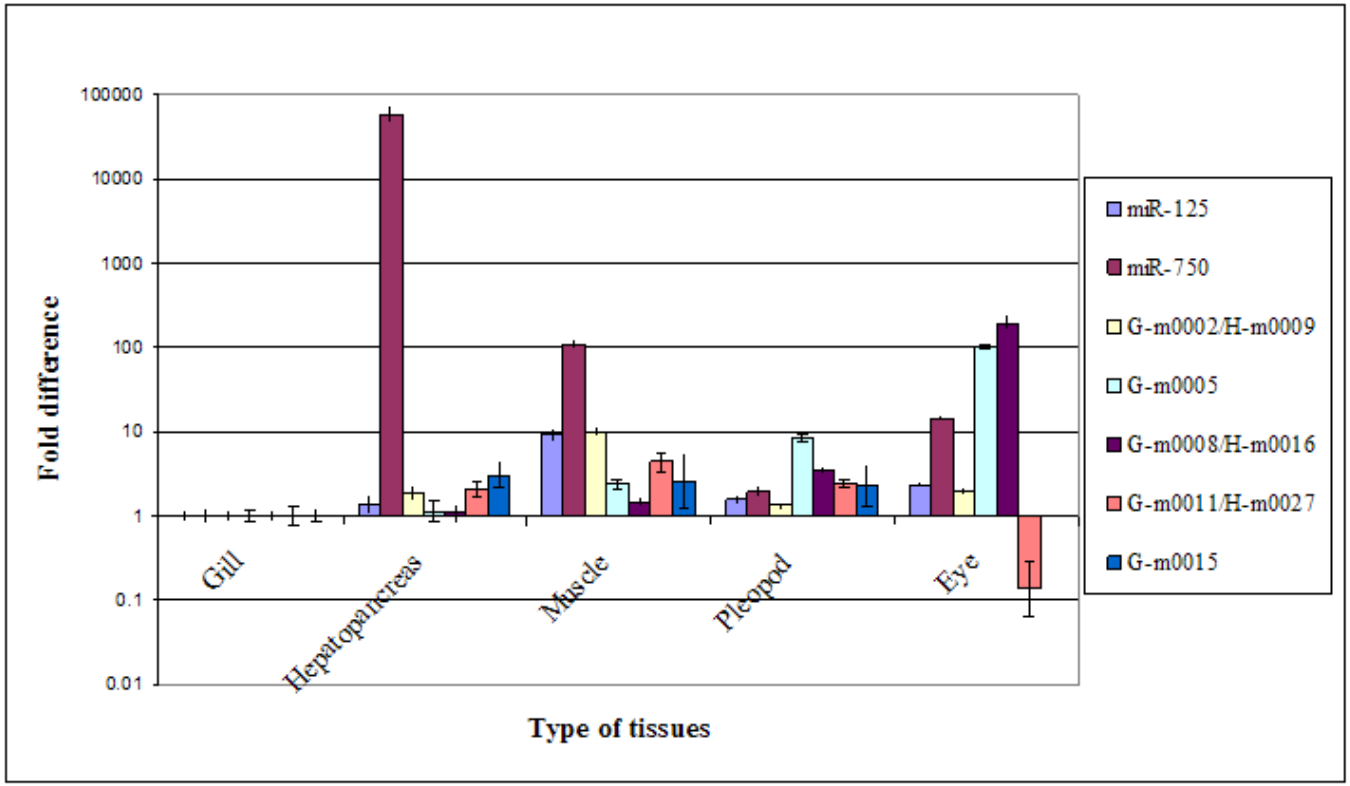


Figure 4.9 Quantitative Real-time PCR (qRT-PCR) determination of mature miRNA expression levels. The expression of mature miRNAs (miR-125, miR-750, G-m0002/H-m0009, G-m0005, G-m0008/H-m0016, G-m0011/H-m0027 and G-m0015) was measured at fixed threshold level of 0.2 in tissues i.e. hepatopancreas, gill, muscle, pleopod and eye. Reference gene (snRNA) was used to normalize the relative abundance of miRNA. Quantification analysis was relative to that miRNA expressed in gill tissue (calibrator). Fold difference was generated using the equation 2(-Delta Delta Ct). Bars indicate the fold difference of mature miRNA in each type of tissues relative to gill tissue. The error bars display the calculated maximum and minimum expression level that represent standard error of the mean expression level.

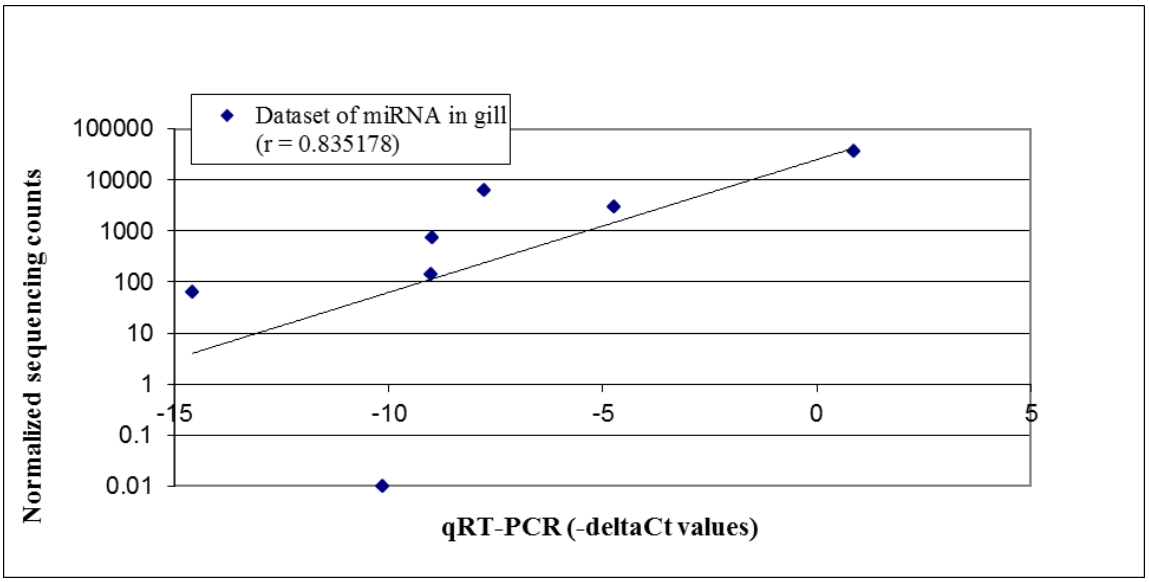
4.5.4 Correlation of miRNA sequencing count and qRT-PCR

To validate miRNA deep sequencing data, the correlation between normalized sequencing counts (expression data) (Section 3.9.1) and expression determined by qRT-PCR was examined. The Pearson's correlation coefficient between sequencing (logarithmized normalized sequencing counts) and qRT-PCR data (–DeltaCt values) was calculated for seven miRNAs (miR-125, miR-750, G-m0002/H-m0009, G-m0005, G-m0008/H-m0016, G-m0011/H-m0027 and G-m0015) in gill and hepatopancreas tissues. The relative expression levels of miRNA candidates in gill and hepatopancreas tissues evaluated using qRT-PCR were was also consistent with the deep sequencing data (Figure 4.10). Strong correlation was observed between the miRNA deep sequencing data and qRT-PCR as evaluated with Pearson's correlation coefficient (r = 0.835178 for miRNA in gill, r = 0.724131 for miRNA in hepatopancreas) which indicates good agreement between both methods.

4.6 Inference of novel miRNA function from pathway analysis

In this study, computational prediction of miRNA targets was performed using a support vector machine (SVM) developed at the Beijing Genomic Institute, BGI (unpublished) and RNAhybrid 2.2 (Kruger *et al*., 2006). However, thousands of potential target genes were identified for 27 novel miRNA candidates as shown in Figure 4.11. Twelve out of 27 novel miRNA candidates were differentially expressed between gill and hepatopancreas tissues [(fold change (log_2) >1 or fold change (log_2) >-1 and p-value <0.05)]. In order to further narrow down the potential mRNAs target genes of miRNAs, pathway enrichment and inverse expression correlation analysis was performed.

A



B

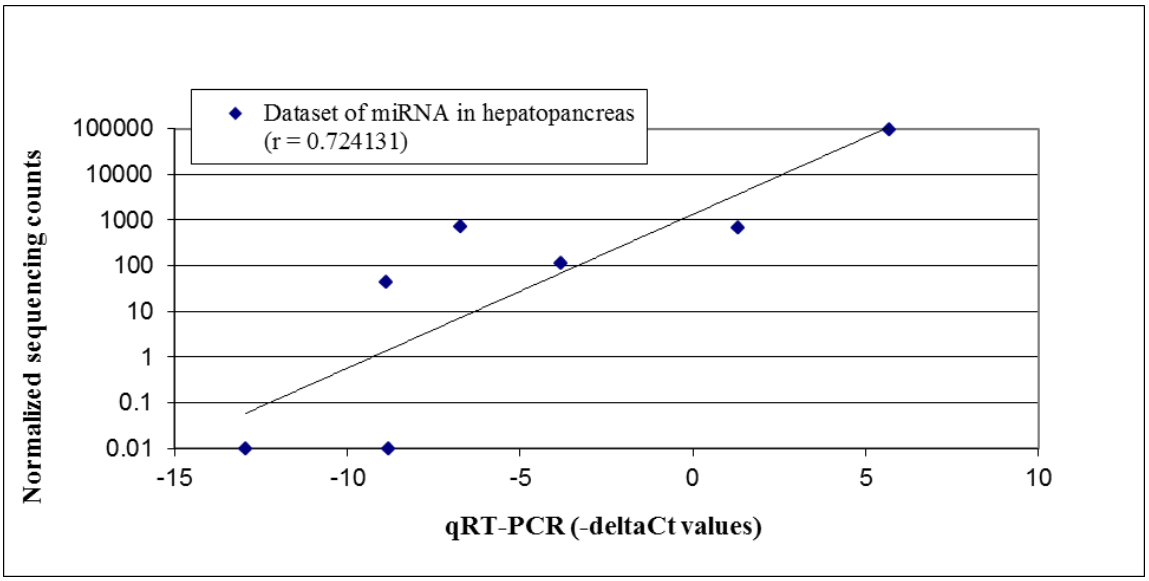


Figure 4.10 Confirmation of accuracy of illumina deep sequencing with qRT-PCR. Seven miRNAs i.e. miR-125, miR-750, G-m0002/H-m0009, G-m0005, G-m0008/H-m0016, G-m0011/H-m0027 and G-m0015 were selected and their expression level was measured by illumina deep sequencing and qRT-PCR. A Pearson correlation scatter plot of (logarithmized) normalized sequencing counts (deep sequencing data) against – DeltaCt values (qRT-PCR) was drawn. The Ct (cycle threshold) is defined as the PCR cycle at which the fluorescent signal cross the threshold. Hence, Ct values are inversely related to the logarithmized amount of target nucleic acid in the sample. Conversion of the Ct values by multiplication with (-1) results in negative Ct values which is proportional to the logarithmized amount of target nucleic acid.

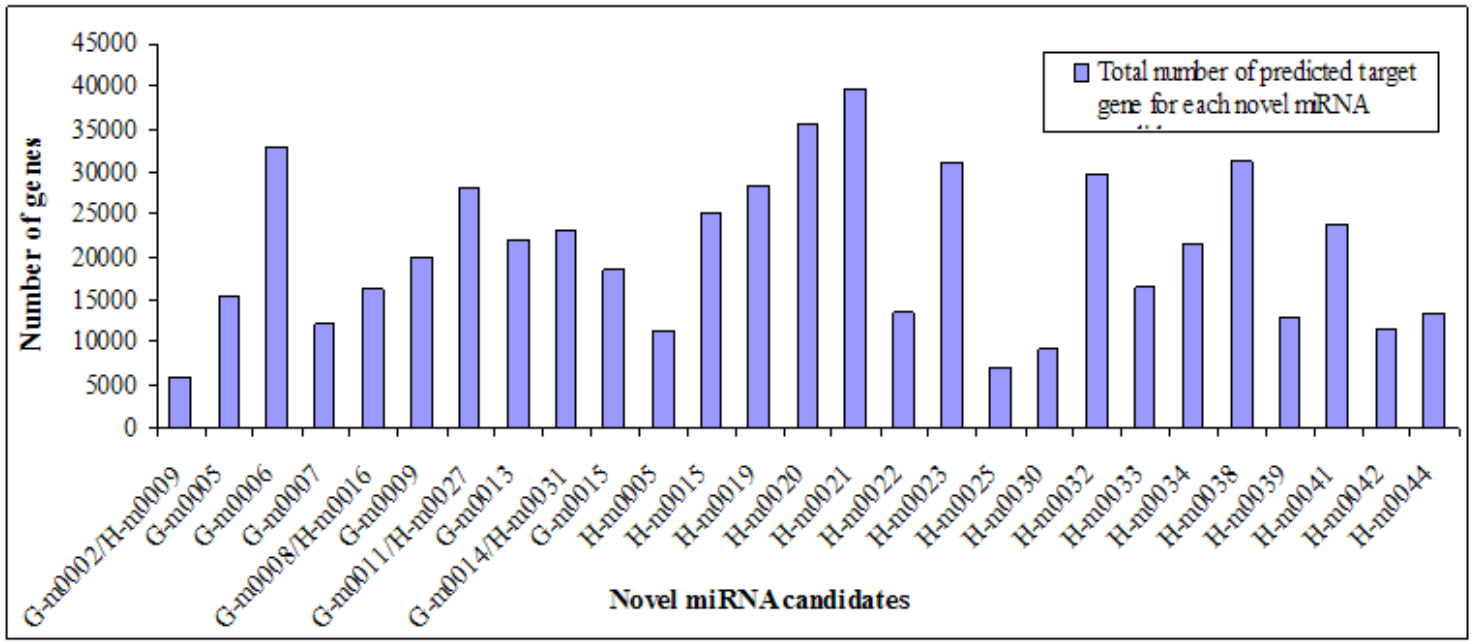


Figure 4.11 Total number of predicted target genes for each novel miRNA candidates from G and H small RNA libraries. miRNA targets prediction was performed using support vector machine (SVM) developed at the Beijing Genomic Institute, BGI (unpublished) and RNAhybrid 2.2 (Kruger *et al*., 2006).

To further analyze the functional correlation of each differentially expressed novel miRNA [(fold change (log_2) >1 or fold change (log_2) >-1 and p-value <0.05)] between gill and hepatopancreas and their target genes (mRNAs) (false discovery rate, FDR< 0.001), the related canonical pathways were determined using Fisher's exact test with Bonferroni correction (p-value <0.05). The parameters considered in this test are: (1) Total number of miRNA target genes; (2) Total number of differentially expressed miRNA target genes; (3) Number of miRNA target genes associated with a specific pathway; (4) Number of differentially expressed miRNA target genes associated with a specific pathway.

Pathway enrichment with Bonferroni correction (p-value <0.05) was successfully applied to four out of twelve novel miRNA candidates that showed significant differential expression [(fold change (log_2) >1 or fold change (log_2) >-1 and p-value <0.05)] between gill and hepatopancreas of *M.rosenbergii*. The differentially regulated target genes of four co-expressed miRNAs (G-m0005, G-m0008/H-m0016, G-m0011/H-m0027, and G-m0015) were associated with energy metabolism and were significantly enriched in the oxidative phosphorylation pathway (Table 4.4 and Figure 4.12). The predicted target genes that function in the mitochondria include V-type H+-transporting ATPase subunit H and three subunits of the cytochrome c oxidase (COX) complexes i.e. cytochrome c oxidase subunit 1, cytochrome c oxidase subunit XV assembly protein and cytochrome c oxidase subunit XVII assembly protein. In order to identify potential mRNAs which might be targeted by specific miRNAs, inverse correlation analysis was performed. In general, miRNAs negatively regulate the expression of their targets via translational repression or mRNA destabilization at the posttranscriptional level (Olena *et al*., 2010; Fabian *et al*., 2010). Selbach *et al.,* (2008) and Baek *et al.,* (2008) have

found that most targets substantially repressed translationally by miRNA also show mRNA destabilization. Therefore, if a given miRNA is low in expression, the expression of its target gene is likely to be high (and vice-versa) resulting in an inverse expression relationship between each miRNA and its target mRNA (Pandey *et al*., 2008; Peng *et al*., 2009). As shown in Figure 4.13, V-type H+-transporting ATPase subunit H was most likely to be a true target for G-m0005, G-m0008/H-m0016 and G-m0015; cytochrome c oxidase subunit 1 (COX1) for G-m0005 and G-m0015; cytochrome c oxidase subunit XV assembly protein (COX15) for G-m0005, G-m0011/H-m0027 and G-m0015; cytochrome c oxidase subunit XVII assembly protein (COX17) for G-m0005. Based on the mRNA profiling results (Figure 4.13) the expression of target genes (i.e. V-type H+-transporting ATPase subunit H and three subunits of the cytochrome *c* oxidase complexes i.e. cytochrome *c* oxidase subunit 1, cytochrome *c* oxidase subunit XV assembly protein and cytochrome *c* oxidase subunit XVII assembly protein) were high in hepatopancreas compared to gill tissues. This may correlate to the relative abundance of mitochondria in the hepatopancreas as the site of multiple oxidative reactions (Arun *et al*., 1998; Felgenhauer, 1992).

Table 4.4 KEGG pathways enriched with targets of each miRNAs in gill and hepatopancreas tissues of *M.rosenbergii*.

Novel miRNA candidates	KEGG Pathway term	Total no. of miRNA target genes	Total no. of differentially expressed miRNA target genes	No. of miRNA target genes associated with specific pathway	No. of differentially expressed miRNA target genes associated with specific pathway	Bonferroni corrected p-value
G-m0005	Oxidative phosphorylation	15339	8513	44	4	2.60E-05
G-m0008/ H-m0016	Oxidative phosphorylation	16176	9325	32	2	5.20E-05
G-m0011/ H-m0027	Oxidative phosphorylation	28049	15553	56	4	0.00E+00
G-m0015	Oxidative phosphorylation	18560	10215	51	5	8.00E-06

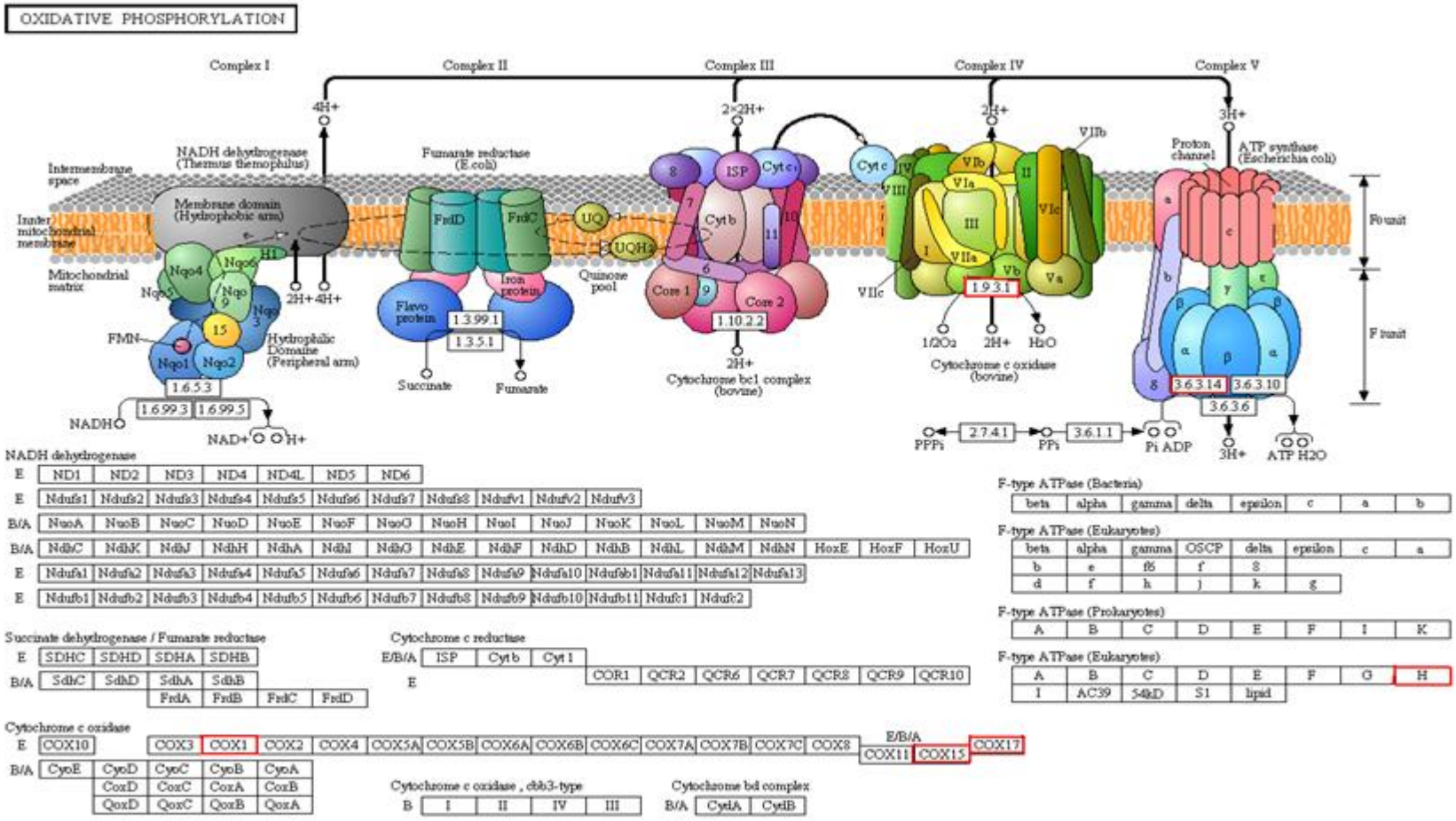


Figure 4.12 Oxidative phosphorylation pathway. Red box represent the predicted target genes of novel miRNA candidates. Adopted from KEGG Pathway database retrived on 14 February 2011 from http://www.genome.jp/kegg/pathway.html.

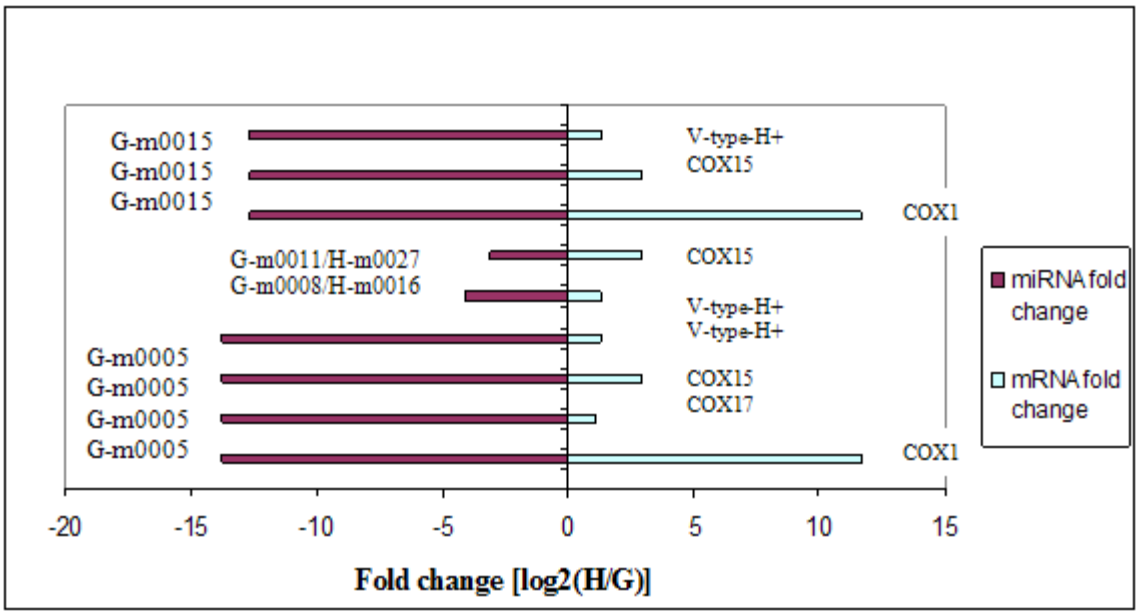


Figure 4.13 Predicted target gene based on inverse expression relationship of miRNA and mRNA profiling. miRNA target genes (mRNA) of G-m0005, G-m0008/H-m0016, G-m0011/H-m0027, and G-m0015 were significantly enriched in the oxidative phosphorylation pathway. The pathway enrichment was determined using Fisher's exact test with Bonferroni correction (p-value <0.05). The fold change of miRNA and their miRNA target genes (mRNA) between hepatopancreas (H) and gill (G) tissues were used to plot the graph of inverse expression relationship. If a given miRNA is low in expression, the expression of its target gene is likely to be high (and vice-versa). V-type H+ represent V-type H+-transporting ATPase subunit H; COX1 represent cytochrome c oxidase subunit 1; COX15 represent cytochrome c oxidase subunit XV assembly protein; COX17 represent cytochrome c oxidase subunit XVII assembly protein.

4.7 Inference of novel miRNA function with preliminary IHHNV challenge test

The temporal expression of miRNA candidates i.e. miR-125, miR-750, G-m0002/H-m0009, G-m0005, G-m0008/H-m0016, G-m0011/H-m0027 and G-m0015 relative to an internal control gene (snRNA) at various time points after infection (infectious hypodermal and haematopoietic necrosis virus, IHHNV challenge test) was studied. The expression of the miRNA candidates was reduced across the various time points of infection. The expression of G-m0005, G-m0008/H-m0016 and G-m0011/H-m0027 was significantly reduced at 9 and 24 hours post infection (Figure 4.14).

The predicted target genes for G-m0005, G-m0008/H-m0016 and G-m0011/H-m0027 include V-type H+-transporting ATPase subunit H and three subunits of the cytochrome c oxidase (COX) complexes i.e. cytochrome c oxidase subunit 1, cytochrome c oxidase subunit XV assembly protein and cytochrome c oxidase subunit XVII assembly protein (Section 4.6). Interestingly, the predicted genes involved in energy metabolism targeted by G-m0005, G-m0008/H-m0016 and G-m0011/H-m0027, which expression was reduced due to infection in this preliminary IHHNV infected *M.rosenbergii* study (sub-adult, 10 g, injection challenge) indicates that miRNAs might play a role in immune response to virus challenge.

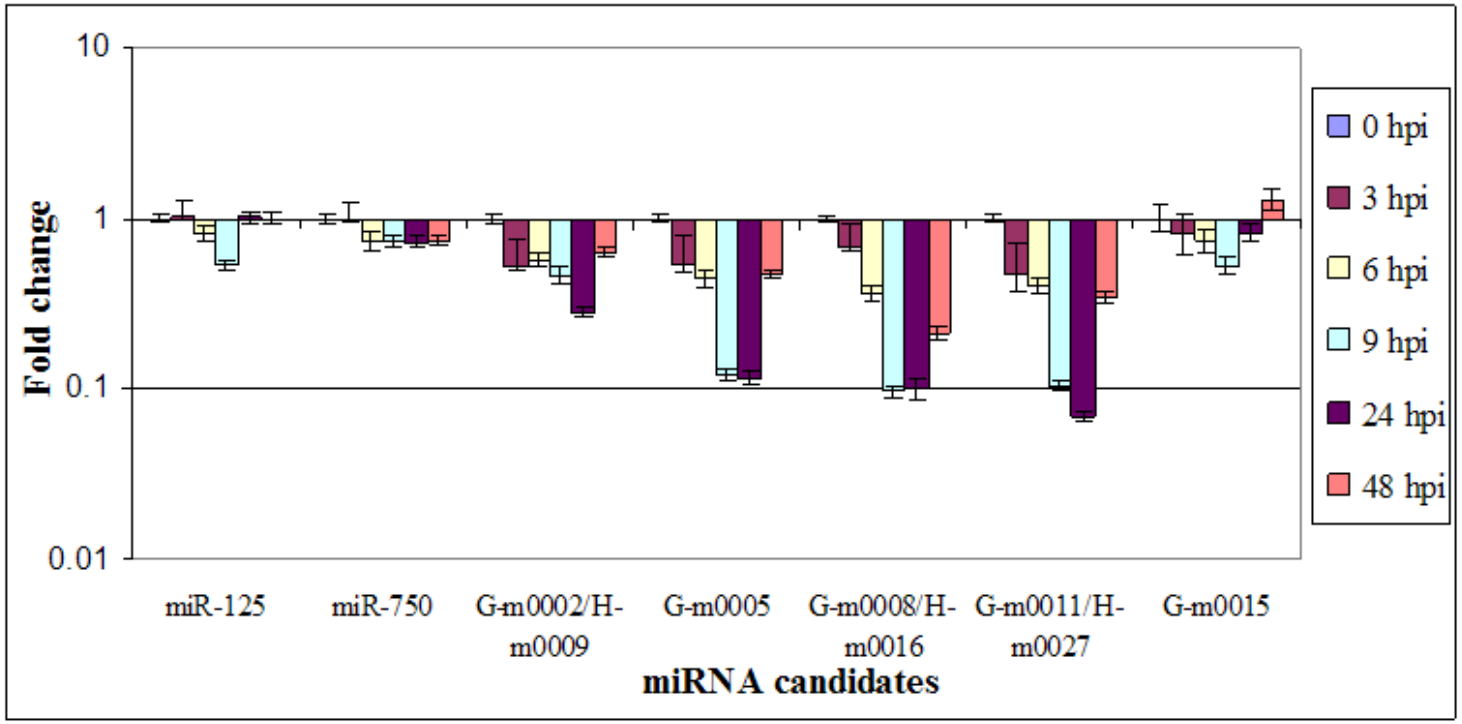


Figure 4.14 Quantitative Real-time PCR (qRT-PCR) determination of mature miRNA expression levels at various time-point of infection. The expression of mature miRNAs (miR-125, miR-750, G-m0002/H-m0009, G-m0005, G-m0008/H-m0016, G-m0011/H-m0027 and G-m0015) was measured at a fixed threshold level of 0.2 in hepatopancreas tissues. Reference gene (snRNA) was used to normalize the relative abundance of miRNA. Fold change was generated using the equation 2^(-Delta Delta Ct) = [Ct(miRNA) – Ct(reference gene)] $_{time\ x}$ – [Ct(miRNA) – Ct(reference gene)] $_{time\ 0}$. Bars indicate the fold change of mature miRNA. The error bars display the calculated maximum and minimum expression level that represent standard error of the mean expression level.

CHAPTER V

DISCUSSION

5.1 Overview

Crustacean miRNAs are far less understood compared to their mammal and plant counterparts. In fact, out of the total number of miRNA in miRBase Release 17.0 (Griffiths-Jones, 2004; Griffiths-Jones *et al.*, 2006; Griffiths-Jones *et al.*, 2008; Kozomara and Griffiths-Jones, 2011), only 43 crustacean miRNAs (*Daphnia pulex*) have been reported, accounting for less than 0.3% of total reported miRNAs. In the present study, the Illumina deep sequencing of small RNAs technology was used to produce two separate small RNA libraries from gill (G) and hepatopancreas (H) tissues of *M.rosenbergii* (taxon: crustacean). In the identification and characterization of miRNA in *M.rosenbergii*, the *de novo* transcriptome of *M.rosenbergii* (Mohd-Shamsudin *et al*., 2012) was used as the reference for the analysis of the G and H small RNA libraries. The *de novo* transcriptome was assembled from raw reads of gill, hepatopancreas and muscle tissues derived from RNA-Seq. The rationale for combining the transcriptomes was to obtain as much information as possible since the assembled genome sequence of *M.rosenbergii* was not available at the time of this study.

5.2 miRNA encoding messenger like non-coding loci

In the present study, the precursors of the three conserved and 27 novel miRNA candidates were derived from messenger like non-coding transcript (mlncRNA). Similar to messenger RNA, mlncRNA are the product of RNA polymerase II, and are often

polyandenylated, capped, spliced and lack long ORFs (Rodriguez *et al.*, 2004; Kim *et al*., 2009; Olena *et al*., 2010). The advantages of identifying miRNAs from *de novo* transcriptome are: (1) the possibility of finding novel miRNAs in the mlnc-RNA genes, and (2) adding to the draft transcriptome of *M.rosenbergii*.

5.3 Classification of pseudo miRNA and real miRNA

A draft transcriptome of *M.rosenbergii* (Mohd-Shamsudin *et al*., 2012) was assembled *de novo* without a reference genomic sequence to generate 102,323 unigenes. In addition, the draft transcriptome of *M.rosenbergii* (Mohd-Shamsudin *et al*., 2012) was not fully annotated. Further, the MIREAP software predicts stem-loop structure regardless of the flanking region (e.g. protein coding region, non-coding region) of the small RNA on the draft transcriptome of *M.rosenbergii* (Mohd-Shamsudin *et al*., 2012). Therefore, to eliminate false positive pseudo-pre-miRNA sequences, the following stepwise filtering were used: (1) the pre-miRNA sequences were analyzed against the Non-redundant NCBI database (Nr) (URL: http://www.ncbi.nlm.nih.gov/genbank/) and Swiss-Prot (The UniProt Consortium, 2010) to exclude pre-miRNA sequences that have similarities to any known, putative, and hypothetical proteins; (2) getORF (EMBOSS program) was used to exclude possible open reading frame (ORF) (Rice *et al*., 2000); (3) ESTScan was used to detect potential coding regions on the pre-miRNA sequences based on the codon bias of coding regions derived from the *de novo* transcriptome of *M.rosenbergii* (Mohd-Shamsudin *et al*., 2012) (Iseli *et al.,* 1999); (4) INFERNAL program (Nawrocki *et. al*., 2009) was used against the Rfam 10.0 database(Gardner *et al.,* 2009; Griffiths-Jones *et al.,* 2005; Griffiths-Jones *et al.,* 2003) to remove pre-miRNA sequences which were homologs of known non-coding RNA based on sequence

similarity and secondary structure. This is because non-coding RNAs often do not show significant sequence similarity, but rather conserve a base-paired secondary structure (Nawrocki *et. al*., 2009). In addition, repeat elements that are derived from a non-coding RNA gene could also fold into a dysfunctional hairpin (Chen *et al*., 2009); (5) Additional tRNA molecules were screened using tRNAscan-SE as tRNA sequences could also fold into secondary hairpin structures (Lowe *et al.,* 1997). Those pre-miRNA sequences that pass through the filtering criteria were considered to be real precursor miRNA candidates.

5.4 Conserved and novel miRNA candidates in *M.rosenbergii*

Three conserved miRNA identified in this study were miR-125, miR-750 and miR-993. These conserved miRNA in *M.rosenbergii* were orthologous to those of mature miRNA identified in other animal species and were able to form appropriate hairpin structures. miR-125 has been predicted to be involved in cell differentiation. In drosophila, miR-125 is specifically expressed during pupal and adult stages in response to induction by the steroid hormone ecdysone. In the drosophila life-cycle, the ecdysone pulse triggers the major post-embryonic developmental transition, including moulting and metamorphosis (Bashirullah *et al*., 2003; Riddiford *et al*., 1993). At present, no functional target has been predicted for either miR-993 or miR-750 in any animal species. Wheeler *et al.* (2009) used Roche 454 sequencing of miRNA libraries across 13 metazoan taxa for phylogenetic studies in which, the miR-993 characterized family was added into the taxon, Ecdysozoa. miR-750 was identified as a lophotroczoan-specific miRNA in phylogenetic analysis by Sempere *et al.* (2007) and later was found to be a protostome-specific miRNA by Wheeler *et al.* (2009).

Despite the lack of genomic sequence for *M.rosenbergii* the availability of the *de novo* transcriptome (Mohd-Shamsudin *et al*., 2012) has aided in the identification of 27 novel miRNA candidates of *M.rosenbergii*. Previously, the miRNA identification from deep-sequencing data for non-annotated genomes has resulted only in annotation of conserved miRNA families (e.g. cotton, *Gossypium hirsutum*; Ruan *et al*., 2009) whilst novel miRNA identification has been reported where there exists a reference genome (e.g. sea urchin, *Strongylocentrotus nudus*; Wei *et al*., 2011). In the current study, an assembled transcriptome was used as the best available substitute for a reference genome and was able to aid in the prediction of novel miRNA for *M.rosenbergii*. This might be due to the non-coding RNA sequences that were present in the transcriptome data indicate that they were transcribed, hence reduces the possibility of over prediction (Wen *et al*., 2007).

5.5 *In silico* identification of miRNA target genes

Computational predictions of miRNA targets were not consistent across different algorithms and usually yield hundreds of miRNA targets (Gusev, 2008; Zhu *et al*., 2011). Therefore, an integrated miRNA and mRNA gene expression analysis was used (Guimbellot *et al*., 2009; Zhu *et al*., 2011). The expression of miRNAs could be associated with the inverse expression of a subset of predicted target mRNAs in an enriched pathway in *M.rosenbergii*. The differentially expressed miRNA-mRNA target pairs between gill and hepatopancreas tissues were found to be enriched in the oxidative phosphorylation pathway. The predicted target genes for G-m0005, G-m0008/H-m0016, G-m0011/H-m0027, and G-m0015 include V-type H+-transporting ATPase subunit H and three subunits of the cytochrome c oxidase (COX) complexes i.e. cytochrome c

oxidase subunit 1, cytochrome c oxidase subunit XV assembly protein and cytochrome c oxidase subunit XVII assembly protein. V-type H+-transporting ATPase subunit H is involved in transporting protons across the membrane for acidifying the intracellular compartments in eukaryotic cells (The UniProt Consortium, 2010). Cytochrome *c* oxidase complexes are involved in the reduction of oxygen in the electron transport chain (The UniProt Consortium, 2010). Some of the ortholog target genes involved in this electron transport chain have also been reported to be involved in neurodegenerative diseases (cell senescence) including Parkinson's (Ekstrand *et al*., 2007; Tufekci *et al*., 2011) and Alzheimer's disease (Fukui *et al*., 2007; Strazielle *et al*., 2009) in a mouse model. However, it is unknown if similar or other unknown pathways involving oxidative phosphorylation are present in *M.rosenbergii*, yet it must be considered that other pathways in the organism may be targets of the novel miRNA from *M.rosenbergii.*

5.5.1 miRNA-target mRNA inverse expression correlation analysis

Although the approach of identifying an inverse correlation between miRNA and mRNA in *M.rosenbergii* model considerably reduces the number of potential miRNA target gene candidates, this approach excludes translational repression by miRNA, which has been considered one of the primary mechanisms of miRNA action (Olena *et al*., 2010). Therefore, additional miRNA target genes in *M.rosenbergii* model need to be considered at the protein level in the future. Zhu *et al*., (2011) demonstrated that by using an integration of the miRNA and mRNA expression profiling, the expression of miRNAs could be associated with the inverse expression of subset of predicted target mRNAs in mouse mammary tumor models. The approach of inverse correlation analysis between miRNA and mRNA has improved the identification of miRNA targets from

potential targets identified using TargetScan for all of the mammary tumors and normal tissues. For instance, the expression of only 9 out of 101 predicted targets via TargetScan were found to show an inverse correlation with expression of miR-412. Zhu *et al*., (2011) also mentioned that the current approach of inverse correlation analysis between miRNA and mRNA only able to predict which miRNA targets were regulated by transcript destabilization or degradation. Alternatively, in order to extend the approach to address translational repression, protein expression data is required. Likewise with transcriptional destabilization, if miRNA regulates target expression through inhibition of translation, then the miRNA expression profiles should explain any observed down-regulation of protein abundance compared to what would be expected based on the mRNA expression profiles.

5.6 Preliminary controlled experiment of *M.rosenbergii* challenged with IHHNV

The expression of miRNAs candidates (G-m0005, G-m0008/H-m0016, G-m0011/H-m0027, and G-m0015) were reduced across various hours post infection in a preliminary controlled experiment of *M.rosenbergii* (sub-adult, 10 g, injection challenge) challenged with infectious hypodermal and haematopoietic necrosis virus (IHHNV). The expression of G-m0005, G-m0008/H-m0016 and G-m0011/H-m0027 were found to be significantly reduced at 9 and 24 hours post infection. This indicates that the miRNAs might play a role in immune response to virus challenge. The predicted miRNA target gene, V-type H+-transporting ATPase subunit H is involved in transporting protons across the membrane for acidifying the endosomal compartments in eukaryotic cells i.e. catalyzing ATP hydrolysis and pumping activity (Beyenbach and Wieczorek, 2005). The acidification of the endosome plays a vital role for the entry of

pathogen to the cells. It was reported that the marked increase of V-type H+-transporting ATPase subunit H was observed in the lymphoid organ of yellow head virus infected *P.monodon* (tiger prawn) at 24 hours post infection (juvenile shrimp, 16-22g, injection challenge) (Bourchookarn *et al.,* 2008). The upregulation of the V-type H+-transporting ATPase subunit H could possibly reflect an attempt to balance the energy production in the infected cells (Bourchookarn *et al.,* 2008). Other target genes predicted for G-m0005, G-m0008/H-m0016, G-m0011/H-m0027 and G-m0015 include three subunits of the cytochrome c oxidase (COX) complexes i.e. cytochrome c oxidase subunit 1, cytochrome c oxidase subunit XV assembly protein and cytochrome c oxidase subunit XVII assembly protein. It was reported that an increased abundance of cytochrome c oxidase was observed in white spot syndrome virus infected *Penaeus monodon* (tiger prawn) at 66 hours post infection (postlarvae, 0.02 g, immersion challenge) (Leu *et al.*, 2007) as well as in the hepatopancreas of white spot syndrome virus infected *Fenneropenaeus indicus* (banana prawn) at 36 hours post infection (juvenile shrimp, 3.0-4.0 g, immersion challenge) (James *et al*., 2010).

5.7 MicroRNA dot-blot

MicroRNA dot-blot could only be used to detect the presence or absence of miRNA. The profiling of miRNA expression using a dot-blot system in the present study was not successful. Hybridization signal was not observed on the non probe control indicating no cross-contamination when preparing the miRNA oligo dot-blot nylon membrane. However, no hybridization signal observed on the positive control indicates that the hybridization step was not working or that the endogenous gene used was not a suitable positive control for this experiment. Due to the low sensitivity of the dot-blot

system, false negative (low abundance miRNA) and false positive (replicates not working) were encountered. Therefore, a few precautions should be taken into consideration for future profiling of miRNA using a dot-blot system: (1) More appropriate endogenous control such as small nuclear RNA (snRNA) or transfer RNA (tRNA) should be used; (2) The endogenous control should be dotted on the same nylon membrane as other antisense oligos; (3) The eluate of enriched small RNA should be labeled immediately with 32P-αATP after isolation due to the short length of miRNAs which results in their instability (Wang and Cheng, 2008). (4) Prolong the incubation time to 24 hours for RNA labeling at high radioactive specificity (Wang and Cheng, 2008).

5.8 Future Directions

Future work aimed at functional elucidation of the miRNA candidates such as G-m0005, G-m0008/H-m0016 and G-m0011/H-m0027 is warranted. This could be studied by over-expressing these miRNAs in IHHNV infected and non-infected *M.rosenbergii* cell lines or other shrimp or even insect cell lines (where these are suitable hosts). The downstream effect of differential miRNA expression (e.g. for G-m0005, G-m0008/H-m0016 and or G-m0011/H-m0027) on their predicted miRNA target genes could be observed at the mRNA level by quantitative real time PCR and at the protein level by western blot. mRNA expression data can be useful to address transcriptional inhibition whereas protein expression data can be used to determine changes in levels of target proteins e.g. the cytochrome c pathway members which are predicted targets for novel miRNAs in the current study. This could be very interesting because the mRNA and protein expression data would be useful in revealing the molecular events involved in the

IHHNV-*M.rosenbergii* interaction mediated by the novel miRNA candidates in the present study. In addition, it is also important to determine whether over-expression of the miRNA could modulate the amount of mitochondrial mRNA and whether such changes to the amount of mitochondrial mRNA could affect the mitochondrial oxidative phosphorylation and thus the generation of adenosine triphosphate (ATP).

CHAPTER VI

CONCLUSIONS

The discovery of small regulatory RNA molecules, miRNAs, is certainly one of the most essential recent findings in biological research. The highly parallel sequencing technologies (Solexa/Illumina platform) are powerful tools for miRNA discovery with high accuracy and efficiency. A deep parallel sequencing and *in silico* data analysis approach was used to identify key metabolic regulatory small RNA genes from the gill and hepatopancreas of *M.rosenbergii* (taxon: crustacean), the Malaysian Giant Freshwater Prawn, an economically important food animal. Despite the limitation of the unavailability of complete genome information for *M.rosenbergii,* bioinformatic prediction of the conserved miRNA and their targets based on the *de novo* trancriptome has successfully identified three precursors (miR-125, miR-750, miR-993) which are orthologs of known miRNA, 27 novel miRNA candidate encoding messenger like non-coding RNA loci and multiple potential miRNA targets. In the present study, miRNA expression profiling was carried out using miRNA deep sequencing and qRT-PCR methods. There was a strong correlation between the miRNA deep sequencing data and qRT-PCR as evaluated with Pearson's correlation coefficient (r = 0.835178 for miRNA in gill, r = 0.724131 for miRNA in hepatopancreas) which indicates good agreement between both methods. Through a combinatorial approach of pathway analysis and inverse expression relationship of miRNA and target mRNA profiles, four co-expressed novel miRNA candidates (G-m0005, G-m0008/H-m0016, G-m0011/H-m0027, and G-m0015) from the gill and hepatopancreas of *M.rosenbergii* were found to be associated

with energy metabolism. The predicted target genes that function in the oxidative phosphorylation pathway include V-type H+-transporting ATPase subunit H and three subunits of the cytochrome c oxidase (COX) complexes i.e. cytochrome c oxidase subunit 1, cytochrome c oxidase subunit XV assembly protein and cytochrome c oxidase subunit XVII assembly protein. The expression of the three novel miRNA candidates (G-m0005, G-m0008/H-m0016 and G-m0011/H-m0027) were found to be significantly reduced at the 9 and 24 hours post infection in a controlled experiment of *M.rosenbergii* (sub-adult, 10 g, injection challenge) challenged with infectious hypodermal and haematopoietic necrosis virus (IHHNV). This indicates that the miRNAs might play a role in immune response to virus challenge. Overall, these findings provide an integrative strategy combining deep parallel sequencing and *in silico* data analysis to identify miRNAs based on the *de novo* trancriptome of *M.rosenbergii* as well as miRNA target prediction based on a combinatorial approach of pathway enrichment analysis and inverse expression relationship of miRNA and mRNA. This study can also serve as a reference point to further improve the understanding of the repertoire of crustacean miRNAs. However, future work aiming at functional elucidation of the miRNA is necessary.