CHAPTER I

INTRODUCTION

1.1 Overview

MicroRNAs (miRNAs) are approximately 20 to 22 nucleotide noncoding RNA regulatory genes that posttranscriptionally regulate many protein–coding genes which indirectly have an important role in regulating biological and metabolic processes. These include development, signal transduction, cell differentiation and maintenance, stress responses, immune responses, carcinogenesis (cancer) and cell death (Watanabe and Kanai, 2011). Based on computational and experimental analysis, miRNAs are estimated to regulate about 50% to 60% of protein coding genes in mammalian system (Friedman *et al*., 2009; Krol *et al*., 2010).

Conventional experimental approaches to miRNA discovery have relied on cloning and Sanger sequencing protocols. A complementary approach, involving miRNA discovery by computational predictions, requires experimental follow-up using methods such as northern blotting and PCR which can be challenging due to the generally low expression level of miRNAs and when the mature miRNA region is unknown. Hence, the availability of highly parallel sequencing technologies have made it possible to detect small RNAs at unprecedented sensitivity, at higher magnitude and at lower cost than traditional approaches (Friedlander *et al.,* 2008; Wang *et al.* 2009; McCormick *et al*., 2011).

To date, more than 10,000 miRNAs found in animals and plants have been identified and deposited in the miRNA registry database (miRBase Release 18.0,

November 2011). Out of 4,600 genes predicted and identified from the kingdom Metazoa for invertebrates, there were no entries for miRNA genes from *Macrobrachium rosenbergii* (*M.rosenbergii*). Thus, there is a need to catalog the miRNA genes of *M.rosenbergii*, whereby the findings can contribute to improved models of metazoan gene regulatory circuitry in the future.

M.rosenbergii, also known as Malaysian giant river prawn, is a species of freshwater prawn, widely distributed in brackish water in Southeast Asia, northern Oceania and in the western Pacific islands (New, 2002). This giant freshwater prawn belongs to the family Palaemonidae is commercially important for its value as a high protein food source by the population of Asia and the Far East. It is also an economically important farmed crustacean (Sri Widada *et al.,* 2003). It was reported by Food and Agriculture Organization (FAO, 2002) of the United Nations that the production of farmed *M.rosenbergii* is expected to exceed 400, 000 tonnes by 2010.

To fully identify conserved and novel miRNAs in *M.rosenbergii*, it is necessary to sequence all expressed small RNAs. In this study, two small RNA libraries have been sequenced, i.e. from the gill and hepatopancreas tissues. Gills in crustaceans are a crucial organ that plays key roles in respiration, regulation of osmotic and ionic balance, detoxification, excretion and immune defense against microbial pathogens (Bhavan *et al.*, 2000; Martin *et al.,* 2000). Hepatopancreas is the most active organ in crustaceans and its multi-functional role is analogous to vertebrate's liver and insect's fat body. It is the major organ for metabolism of the crustaceans which involves digestive enzymes synthesis and secretion, digestion process, nutrient absorption, excretion, and reserve (lipid and glycogen) storage (Sousa *et al*., 2000; Franceschini-Vicentini *et al*., 2009). These metabolic activities are strongly influenced by the physiological demands,

including moulting, reproduction, digestion process, disease and by environmental changes such as hypoxia (Johnston *et al*., 1998).

1.2 Importance of the study

This study is essential and important in identifying the potential miRNA genes and their targets in *M.rosenbergii* (taxon: crustacean). This is because the crustacean miRNAs accounts for less than 0.3% of total reported miRNAs (miRBase Release 18.0, November 2011). The currently available draft transcriptome of *M.rosenbergii* which is represented by 102,230 unigenes (Mohd-Shamsudin *et al.,* 2012) provides good reference sequences for mining of miRNAs and target genes. The finding on miRNA-target interaction would be particularly useful for aquatic animals monitoring programs. In addition, the findings of miRNA in *M.rosenbergii* can contribute to improved models of metazoan gene regulatory circuitry in the future.

1.3 Objectives

1. To test an integrative strategy combining deep parallel sequencing and *in silico* data analysis to identify and characterize conserved and novel microRNAs based on *de novo* transcriptome of *Macrobrachium rosenbergii*.
2. To determine the expression and distribution of selected candidate microRNA genes of *Macrobrachium rosenbergii* with deep parallel sequencing, miRNA dot-blot and stem-loop quantitative real time PCR.
3. To infer the functional role of selected novel miRNAs candidates in *Macrobrachium rosenbergii*

CHAPTER II

LITERATURE REVIEW

2.1 Description of *Macrobrachium rosenbergii*

Macrobrahium rosenbergii (*M.rosenbergii*) (De Man, 1879) is one of the most commercially important decapod crustacean species cultured. It is also the largest genus of the family *Palaemonidae* in the world (De Grave *et al*, 2008; Wowor *et al*, 2009). Adult males and adult female prawns have been reported with a total body length of up to 33 cm and 29 cm respectively (New, 2002). Formerly, named as *Palaemon carcinus, Palaeon whitei*, *Palaemon rosenbergii*, *Palaemon spinipes* and *Palaemon dacqueti* (Holthuis, 1980; New, 2002), *M.rosenbergii* is commonly known as 'giant river prawn', 'giant Malaysia prawn', 'giant freshwater prawn', 'giant long-armed prawn' and 'cherabin' (New, 2002). In Malaysia, *M.rosenbergii* is known as 'udang galah' (Holthuis, 1980). The giant river prawn is distributed naturally from western Pacific islands to Southeast Asia (Holthuis, 1980; New, 2002). It is distributed in diverse localities in rivers, lakes, swamps, irrigation ditches, canals and ponds (Nandlal and Pickering, 2005; New, 2002).

Macrobrachium rosenebrgii (Figure 2.1) can be distinguished from other species at the genus level by major physical characteristics such as the long rostrum which develops at the tips of the head and has 11 to 14 upper teeth and 8 to 14 lower teeth (Nandlal and Pickering, 2005; New, 2002). The tip of the telson extends till the end of the uropods (Nandlal and Pickering, 2005; New, 2002). It also can be identified by its

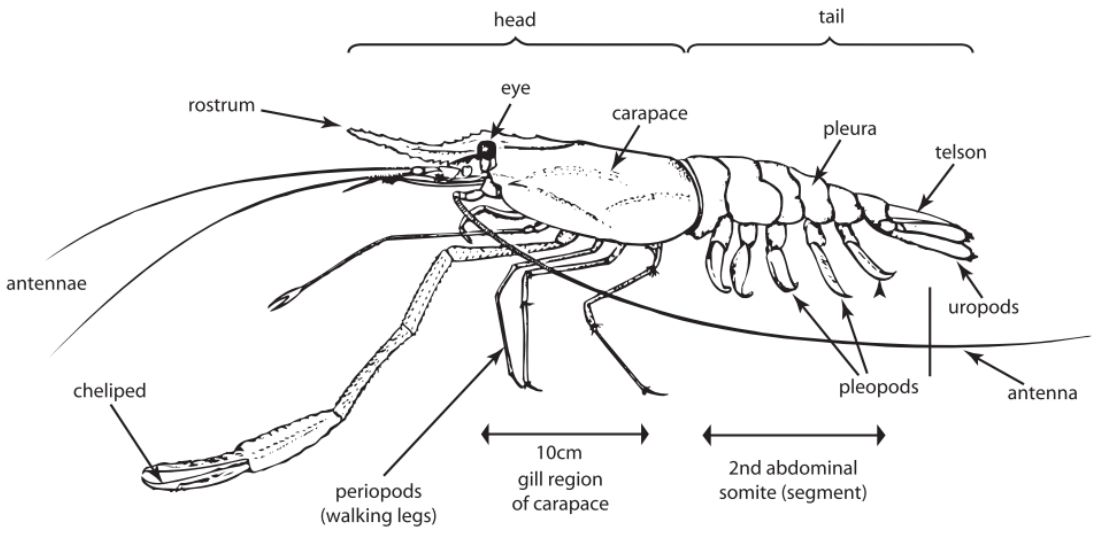

Figure 2.1 External anatomy of the freshwater prawn, *Macrobrachium rosenbergii* (Adopted from Nandlal and Pickering, 2005).

long legs (chelipeds) which are covered with spines and tubercles (Nandlal and Pickering, 2005; New, 2002) and the movable finger of the legs which is covered with spongy fur (New, 2002). The body of freshwater prawn consists of head (cephalothorax) and tail (abdomen) which is divided into 20 segments (abdominal somite) (Nandlal and Pickering, 2005; New, 2002). There are 14 segments in the head which is covered by carapace and another 6 segments are located in the abdomen. The head supports the features such as stalked eyes, antennae (tactile and sensory perception), mandible (cutting and grinding food), maxillae (food handling) and pereiopods (capturing food, walking). In the tail (abdomen), each of the first five segments has a pair of swimming legs (pleopods) and the sixth segment of the abdominal somite features uropods and telson (Nandlal and Pickering, 2005; New, 2002). The body colour of the freshwater prawn ranges from dark green to grayish blue. Freshwater prawns can also be distinguished from marine shrimp by looking at the pleura in the second abdominal somite. In freshwater prawn the second pleuron overlaps with both first and third

pleuron whereas in marine prawn the second pleuron overlaps only the third pleuron abdominal somite (Nandlal and Pickering, 2005; New, 2002).

Before the successful development of breeding and rearing methods, *M. rosenbergii* was widely caught in the rivers (Holthius, 1980). The breeding and rearing method of *M.rosenbergii* was introduced in 1960s. The *M.rosenbergii* was imported into Hawaii from Malaysia in 1965 for research purposes. Ling (1969) discovered that the prawn larvae require brackish water for development and survival. Following the pioneer work of Ling (1969), a mass rearing technique for commercial-scale production of postlarvae was developed by Fujimura and Okamoto (1972). This mass production method has been introduced into almost every continent for farming purposes (New, 2002). It was reported by Food Agriculture Organization (FAO, 2002) of the United Nations that the production of farmed *M.rosenbergii* is expected to exceed 400, 000 tonnes by 2010.

2.1.1 Problems associated with production of *M.rosenbergii*

Macrobrachium rosenbergii is an alternative species for farming worldwide. This freshwater prawn species has a high growth rate and have high market price. However, diseases are causing substantial economic loss and restraining the growth of the aquaculture industry. The major occurrences of disease problems affecting the production of *Macrobrachium rosenbergii* were due to lack of water treatment, poor husbandry, overcrowding, poor sanitation, and insufficient quarantine procedures (New, 2002). The reported viruses infecting *M.rosenbergii* are Macrobrachium hepatopancreatic parvo like virus (MHPV) (Anderson *et al.,* 1990), infectious hypodermal and hematopoietic necrosis virus (IHHNV) (Hsieh *et al*., 2006; Hazreen

Nita *et al*., 2011; Kua *et al*., 2011), white spot syndrome virus (WSSV) (Sahul Hameed *et al*., 2000), *Macrobrachium rosenbergii* nodavirus (MrNV) (Sahul Hameed *et al*., 2004) and extra small virus-like particle (XSV) (Sahul Hameed *et al*., 2004). Among these, IHHNV is associated with high mortalities of up to 80% at postlarvae and juvenile stages in *M. rosenbergii* (Arcier *et al*., 1999; Hsieh *et al*., 2006). The causative agents of white tail disease such as MrNV and XSV have resulted in 100% mortality of *M. rosenbergii* in hatcheries and farms (Sahul Hameed *et al*., 2004). In Malaysia, the wild *M.rosenbergii* was found infected with IHHNV in Rubana River, Perak (Hazreen Nita *et al*., 2011) whereas Parvo-like virus and bacteria was found infecting the wild giant freshwater prawn in Damak Sea of Rejang River in Kuching, Sarawak (Kua *et al*., 2011).

2.2 Small non coding RNA

Small non-coding RNAs are characterized as functional RNA that do not code for protein. This class of RNAs includes transfer RNA (tRNA), ribosomal RNA (rRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), microRNA (miRNA), small interfering RNA (siRNA) and Piwi-interacting RNA (piRNA) (Katia *et al*., 2011).

Current research has implicated miRNAs, single-stranded RNAs of approximately 20 to 22nt in length as important participator in posttranscriptional gene regulation in both animals and plants (Bartel, 2009; Carthew and Sontheimer, 2009; reviewed by Krol *et al*., 2010). In the genome context, miRNA are located within the introns of protein-coding transcription unit or within either intron or exon of non-coding transcription unit or in messenger like non-coding RNA (mlncRNA) transcripts (Rodriguez *et al.*, 2004; Kim *et al*., 2009; Olena *et al*., 2010). Single miRNA can target

different messenger RNA (mRNA) transcripts and at the same time, more than one miRNA gene can target the same protein-coding transcripts (Kim *et al*., 2009; Peter, 2010).

Another class of non-coding RNAs, siRNAs (19-21nt in length), are similar in size to miRNAs and both are involved in posttranscriptional regulation of gene expression. siRNAs differ from miRNA in their biogenesis i.e. siRNA are derived from Dicer cleavage of long double-stranded RNAs (dsRNA) (Elbashir *et al*., 2001; Carthew and Sontheimer, 2009; Ghildiyal and Zamore, 2009) whereas miRNA is generated by Dicer from hairpin like transcript (precursor miRNA) that are produced by Drosha from its primary transcript transcribed from one strand of the genome loci (Carthew and Sontheimer, 2009). These small RNAs i.e. miRNA and siRNA regulate other RNA molecules depending on the degree of base-pairing complementarity between the small RNA and its target which leads to translational inhibition typically by miRNA (imperfect complementarity) or mRNA cleavage typically by miRNA and siRNA (perfect or near perfect complementarity) (Carthew and Sontheimer, 2009). SiRNAs can be derived from transposon transcripts, repetitive elements and sense-antisense transcript pairs (Ghildiyal and Zamore, 2009; Hao *et al.,* 2010).

Small RNA sequencing studies have also contributed to the discovery of a novel class of small RNAs, termed piRNAs which are approximately, 24 to 31 nucleotides in length that are clearly distinct in their size from miRNA (Kim *et al*., 2009). They are processed from single-stranded precursor RNAs and form complexes with Piwi protein which is crucial for self renewal of germ line stem cells (Cox *et al.,* 1998; Cox *et al.,* 2000; Szakmary *et al.,* 2005) and for mobility control of the transposon (Kalmykova *et al.,* 2005) as review by Siomi *et al*., (2011). The piRNA is highly abundant in germline

cells (Lau *et al*., 2009; Lau, 2010). The piRNAs are involved in the repression of retrotransposon and control of non-transposable element in the germline (Malone and Hannon, 2009; Lu and Clark, 2010).

2.3 MicroRNA biogenensis

Metazoan miRNA genes (Figure 2.2) are transcribed either as single genes or in clusters by RNA polymerase II into long primary miRNA transcripts that contain a 5'-end cap structure and a poly A-tail sequence (Lee *et al*., 2004; Cai *et al*., 2004; reviewed by Olena *et al*., 2010). Thus, miRNA genes are commonly encoded as primary mRNA-like sequences (Zhao *et al.*, 2008). The length of animal primary miRNA transcripts is between hundreds of nucleotides and tens of kilobases and contains imperfectly base-paired hairpin structures (Kim *et al*., 2009). Hairpin structures within the primary miRNA gene transcript are recognized and cleaved by the RNase III endonuclease, Drosha with its cofactor i.e. DiGeorge syndrome critical region gene 8, DGCR8 (*Homo sapiens*) or Pasha (*Drosophilla melanogaster* and *Caenorhabditis elegans*) in the nucleus to produce approximately 70 to 80 nucleotide miRNA precursors. DGCR8/Pasha helps Drosha in substrate recognition (Denli *et al.* 2004; Gregory *et al.* 2004; Han *et al.* 2004; Landthaler *et al.* 2004; reviewed by Olena *et al*., 2010). Following nuclear processing, precursor miRNA hairpins are exported from nucleus to cytoplasm by nuclear transport receptor complex, exportin-5 with its cofactor, Ran-GTP (Yi *et al.*, 2003; Bohnsack *et al.* 2004; Lund *et al.*, 2004; Zeng and Cullen, 2004; reviewed by Olena *et al*., 2010). In the cytoplasm, GTP is hydrolysed to GDP and the precursor miRNAs are released from the export complex. Zheng and Cullen (2004) reported that the structural motif of short stem with dinucleotide 3' overhangs which is

located in the precursor miRNA stem loop might serve as a signature motif that is recognized by exportin-5. Later, in the cytoplasm another RNase III endonuclease, Dicer with its cofactor i.e. Loquacious (*Drosophilla melanogaster*) or trans-activator RNA-binding protein, TRBP (*Homo sapiens*) processes the precursor miRNAs thereby releasing three products of largely invariant lengths (Chendrimada *et al.,* 2005; Forstemann *et al.,* 2005; Saito *et al.,* 2005; reviewed by Olena *et al*., 2010). One of these is the loop of the hairpin which is degraded as a by-product. The two other products form a duplex whereby the unstable star strand (miRNA*) is typically degraded, whereas the mature miRNA strand is incorporated into effector complexes i.e. miRNA-containing ribonucleoprotein complex (miRNP) or miRISC (miRNA-containing RNA-induced silencing complex) (Olena *et al*., 2010; Kim *et al*., 2009). The mature miRNA sequence functions by guiding the effector complex to target mRNA transcripts by partial sequence complementary. The seed region which is approximately six nucleotides starting at position two from the 5' end of the mature sequence (Kim *et al*., 2009) is particularly important for target recognition. The effector complex regulates the target mRNA transcript through inhibition of translation or decreasing its stability (Olena *et al*., 2010).

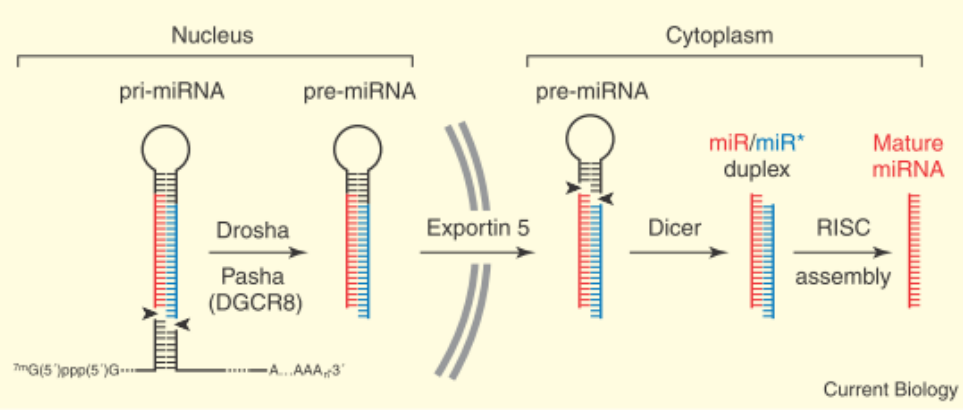

Figure 2.2 The miRNA biogenesis pathway in animal (Adopted from Tomari and Zamore, 2004).

2.4 Mechanism of microRNA function

MiRNAs are a naturally occurring RNA are extensively characterized in a multitude of organisms. In animals, miRNA can manifest as translational repression of the target mRNA and destabilization of mRNA (Olena *et al*., 2010). miRNA associated with effector complexes (miRNP or miRISC) can repress target mRNA transcripts (Olena *et al*., 2010). In animals, miRNAs are often found to inhibit translation due to partial sequence complementarity within 3' untranslated regions (UTR). Early evidence of miRNA-mediated translational repression came from studies of the nematode, *Caenorhabditis elegans* where the number of targeted mRNA molecules (*lin-14* and *lin-28*) remained the same but the abundance of protein that regulates developmental timing of embryonic events encoded by the targeted mRNA was remarkably reduced (Olsen and Ambros, 1999; Seggerson *et al.*, 2000). Animal miRNAs have partial complementarity to the mRNA target and have the ability to repress translation by directly interfering with the translation initiation factors or by disrupting the poly-A tail function (Humphreys *et al.*, 2005). Through transfecting *CXCR4* miRNA into HeLa cells, Humphreys *et al*. (2005) reported that the *CXCR4* miRNA repressed the translation initiation of *Renilla*-lucifrase mRNA by inhibiting the initiation factor 4E/cap and poly-A tail functions (reviewed by Fabian *et al*., 2010).

In contrast to miRNA in plants, which are homologous to their target mRNAs, miRNA-directed cleavage of mRNA targets in animals is less common than translational inhibition. However, some examples of cleavage are reported, such as miR-196 which mediates cleavage of *HOXB8* mRNA in mouse embryos (Yekta *et al.*, 2004; Hornstein *et al.*, 2005; reviewed by Olena *et al*., 2010). A few studies have found that most of the target mRNAs which substantially undergo translational repression by miRNAs, will

also show miRNA destabilization (Baek *et al.,* 2008; Selbach *et al.* 2008). miRNA that mediate destabilizing of their target mRNAs cause mRNA de-adenylation and degradation. Through transfecting miRNA (*miR-1* and *miR-124*) into cells that normally did not express this miRNA, Lim *et al.,* (2005) reported that the presence of such miRNA was found to reduce the abundance of mRNAs which contained the target sites for the miRNA (reviewed by Fabian *et al*., 2010). In addition, Bagga *et al.,* (2005) demonstrated that several miRNA targets of *let-7* and *lin-4* in *Caenorhabditis elegans* were reduced in abundance in an miRNA-dependent manner (reviewed by Fabian *et al*., 2010). miRNA is associated with P-bodies which are enriched with decapping enzyme and exonucleases that are involved in mRNA degradation (Franks and Lykke-Andersen, 2008; reviewed by Krol *et al*., (2010). When the miRNA-targeted mRNAs sequester to P-bodies, they undergo deadenylation, decapping by Dcp1/Dcp2 enzymes and degradation by Xrn via 5'to 3' exonuclease activity (Chatterjee and Grosshans, 2009; reviewed by Krol *et al*., 2010).

2.5 MicroRNA expression and function in animals

miRNA have diverse expression patterns and thus their profiles may help to elucidate their potential roles (Liu *et al*., 2010a). The basis is that, if a miRNA is highly expressed in a tissue or cell type or at a specific developmental stage, it might plausibly play a regulatory role in that particular tissue or cell or developmental stage. The abundance of potential miRNA targets suggest that miRNAs are likely to be involved in broad spectrum of biological roles such as development timing, carcinorgenesis, and virus-host interaction (Skalsky and Cullen, 2010; Watanabe and Kanai, 2011).

2.5.1 MicroRNAs mediate development timing

In *Caenorhabditis elegans,* miRNA of lin-4 and let-7 play important roles in post-transcriptional regulation of a set of transcription factors such as LIN-14, HBL-1 and LIN-29 that control larval developmental transition (first larval stage, L1 to final larval stage, L4) (Ambros, 2011). The progression of first larval stage (L1) to second larval stage (L2), involved the downregulation of transcription factor LIN-14 by lin-4 miRNA. The downregulation of HBL-1 by let-7 family facilitate the transition of second larval stage (L2) to third larval stage (L3) (Abbott *et al*., 2005).The let-7 family includes mir-48, mir-84 and mir-241(Abbott *et al*., 2005). The activity of the let-7 family is mediated by TRIM/NHL protein NHL-2 (Hammell *et al*., 2009). The transition from L2 to L3 also involves the negative regulation of LIN-28 by lin-4 miRNA. Progressions from final larval stage (L4) to adult stage also involve downregulation of LIN-41 and thereby cause the upregulation of the LIN-29. The activity of let-7 is repressed by HBL-1 to ensure that the upregulation of let-7 occurs only after completion of the ealier steps. The downregulation of nuclear hormone receptor moulting factor NHR-23 and NHR-25 by let -7 family miRNAs help to cease the larval moulting cycle (Hayes *et al*., 2006).

2.5.2 MicroRNAs mediate carcinogenesis

It is common for genes controlling developmental process such as cell proliferation and differentiation to be associated with tumorigenesis (reviewed by Lewis *et al*., 2010). For example, the miR 17-92 cluster plays important roles to ensure proper development of the heart and lung as well as foetal and adult B cell lymphoma through the regulation of pro-apoptotic gene *Bim* (Koralov *et al*., 2008; Ventura *et al*., 2008). The experimental support for a direct contribution of miR 17-92 cluster to tumorigenesis

comes from mouse experiments. The over-expression of miR 17-92 cluster in a murine B-cell lymphoma model results in acceleration of tumour development (He *et al*., 2005). The orthologs of lin-4 which is the miR-125 also contribute to leukemogenesis as supported from murine models. The over expression of miR-125 in transplanted mouse fetal liver caused elevated white blood cell such as neutrophils and monocytes that eventually will cause acute lymphoblastic leukemia (Bousquet *et al*., 2010).

2.5.3 Interaction of host and viral miRNA

Virus-encoded miRNAs or miRNA-like molecules have been identified in several mammalian viruses, including the herpesvirus family, such as Epstein-Barr virus (EBV) (Zhu *et al*., 2009; Seto *et al*., 2010), Kaposi sarcoma-associated herpesvirus (KSHV) (Cai et al., 2005; Chandriani *et al*., 2010; Lin *et al*., 2010), and human cytomegalovirus (HCMV) (Dhuruvasan *et al*., 2011), polyomavirusses such as simian virus 40 (SV40) (Sullivan *et al*., 2005) and murine Polyomavirus (muPyV) (Sullivan *et al*., 2009), adenoviruses such as human adenovirus (Xu *et al*., 2007) and ascoviruses such as *Helliothis virescens* ascovirus (Hussain *et al*., 2008). Virus miRNAs are involved in regulating latent-lytic switch, modulate immune response and support viral replication by promoting cell survival, proliferation and differentiation (Skalsky and Cullen, 2010). The modulation of the viral and cellular mRNA by the virus miRNA was to establish a host environment favorable to completion of viral life cycle (Skalsky and Cullen, 2010). miRNAs from polyomaviruses have been shown to direct cleavage of early transcriptomes that encode the T antigen (reviewed by Grundhoff and Sullivan, 2011). For example, the SV40 miRNAs target the viral T-antigen mRNAs and mediate viral T-antigen mRNA cleavage late in infection (Sullivan *et al*., 2005). This reduces the

expression of the viral T antigens, which makes the infected cells less susceptible to killing by T-antigen-specific cytotoxic T cells (Sullivan *et al*., 2005), hence increasing the chances of successful infection (Sullivan *et al*., 2005). Endogenous cellular miRNAs can mediate antiviral defense. The endogenous miR-32 prevents the accumulation of the primate foamy virus type 1 (PFV-1) in human cells by targeting the PFV-1 genome for translation inhibition (Lecellier *et al*., 2005). However, some viruses overcome the miRNA that regulate the antiviral defense by encoding protein that suppress RNA silencing via sequestering small RNA to disrupt the key proteins involved in miRNA processing (Haasnoot *et al*., 2007). The Tas, which is a PFV-1-encoded protein, was able to suppress the miRNA that mediates inhibition of PFV-1 (Lecellier *et al*., 2005).

2.5.4 MicroRNAs involved in innate immunity of crustacean

The innate immune system, including cell-based immunity and humoral immunity, is the first line of defense against the infection of pathogens in animals (Yang *et al*., 2012). Until recently, very little had been reported for miRNAs related to immunity in crustaceans. A study by Yang *et al*., (2012) characterized miRNAs of marine shrimp (*Marsupenaeus japonicas*, taxon: Crustacean) and their targets using small RNA sequencing and degradome sequencing. The degradome sequencing analysis indicated that the target genes of miR-1 and let-7 were endonuclease-reverse transcriptase and transmembrane protein 14C-like indicating that miR-1 may be involved in the regulation of phagocytosis, apoptosis and pro-phenoloxidase system by targeting the endonuclease reverse transcriptase gene. Host miRNAs that respond to DNA virus infection have also been characterized through small RNA sequencing of marine shrimp (*Marsupenaeus japonicus)* challenged by white spot syndrome virus

(WSSV) at different times post-infection (Huang et al., 2012). Of 63 host miRNAs identified in the study, 31 were differentially expressed in response to WSSV infection. Among the differentially expressed miRNAs, miR-1, miR-7 and miR-34 are highly conserved in shrimp, fly and human and mediate similar pathways i.e. apoptosis and MAPK pathways. For example, the predicted target genes for miR-7 include mitogen-activated protein kinases (MAPK) and WSSV early gene (wsv477). Whilst only predictive, these reports suggest that marine shrimps may inhibit WSSV infection by targeting viral transcripts with host miRNAs.

2.6 Approaches in identification of microRNA

Identification and analysis of miRNAs enhances our understanding of the important roles that small RNAs play in complex regulatory networks. In general, methods for identifying miRNAs include cloning, computational and deep sequencing approaches.

Conventional experimental approaches to miRNA discovery have relied on cloning and Sanger sequencing protocols. Even though this method is useful for identification of miRNA genes such as the founding members, *lin-4* and *let-7* in *Caenorhabditis elegans* (Lee *et al.,* 1993, Reinhart *et al.,* 2000), this method is limited to the detection of highly expressed miRNAs and is both laborious and time-consuming. miRNA are small in size, and have significant variation in their expression levels in different tissues and under different environmental conditions. This attribute makes it difficult to identify miRNAs using traditional approaches (Wang *et al.,* 2009).

Complementary approaches involving miRNA discovery by computational predictions, have been developed to complement conventional approaches. The

principles of the computational approaches are based on the major attributes of miRNAs including their ability to form appropriate stem-loop structures, high evolutionary conservation from species to species and high minimal free energy of hybridization (Zhang *et al.,* 2006). There are several computational programs designed for predicting potential miRNA using different algorithms and strategies. Homology based approaches use sequences and secondary structure filters to predict orthologous miRNA families by searching nucleotide databases i.e. expressed sequence tag (EST) and genome sequence survey (GSS) using BLAST (Altschul *et al*., 1997). Examples of these computational prediction programs are miRAlign (Wang *et al*., 2005) and microHARVESTER (Dezulian *et al*., 2005) which was developed to predict animal (e.g. *Anopheles gambiae*; Wang *et al*., 2005) and plant (e.g. *Arabidopsis thaliana*; Adai *et al*., 2005) miRNA genes respectively. Gene finding approaches which complement the homology based approaches can be used for an entire genome search. The gene finding tool aims to identify the conserved genomic regions and assess whether the sequences that are evolutionary conserved are able to fold into an appropriate hairpin structure, with a secondary structure prediction program such as Mfold (Zuker *et al*., 2003). Examples of these computational prediction programs are miRseeker (Lai *et al*., 2003) and miRscan (Lim *et al*., 2003).

Small non-coding RNA (ncRNA) discovery and profiling is one of the applications of next generation sequencing. The longest reads, 250 to 300 nucleotides are obtained by the 454 Life Sciences/Roche technologies but each sequencing reaction yields less reads up to 500,000 (Moxon *et al*,. 2008; Morozova *et al.*, 2008). In contrast, Solexa/Illumina platform allows for the generation of one to three million short reads of 36 nucleotides each (Glazov *et al.,* 2008; Morozova *et al.,* 2008). To date, most studies

have used Solexa/Illumina and 454 Life Sciences/Roche technologies to discover new and different non coding RNA classes in an array of organisms including human (Morin *et al*., 2008), mouse (Barbiarz *et al*., 2008), chicken (Rathjen *et al*., 2009), nematode (Kato *et al*., 2009), amphioxus (Chen *et al*., 2009), cotton (Ruan *et al*., 2009), silkworm (Liu *et al*., 2010b) and sea urchin (Wei *et al.,* 2011). Deep parallel sequencing provides a good tool for miRNA studies. To date several web-server tools i.e. computational identification of miRNA, CID-miRNA (Tyagi *et al*., 2008), miRCat (Moxon *et al*., 2008), miRanalyzer (Hackenberg *et al*., 2009), miRTools (Zhu *et al*., 2010) and standalone tools e.g. miRExpress (Wang *et al*., 2009) and miRDeep2 (Friedlander *et al.*, 2011) have been developed for application of massive data analysis for conserved and novel miRNA discovery or expression profiling. In addition, there are a few miRNA discovery studies reported in species such as rice (Sunkar *et al.* 2008), maize (Zhang *et al.*, 2009), human (Langenberger *et al.*, 2010) that largely rely on in-house scripts. Overall, availability of a bioinformatics pipeline has to be taken into consideration for the application of massive data analysis.

2.7 MicroRNA target prediction

miRNAs are present in most metazoans and are essential regulators of various biological functions. Due to the difficulty of experimental identification of miRNA targets, there was sudden increase in computational target prediction algorithms as shown in Table 2.1. These computational approaches rely mainly on miRNAs complementarity to their target mRNAs. The general criteria for target prediction include 1) strong complementarity of the seed region at 2 to 7 or 8 nucleotides of 5' miRNA sequences to the 3'UTRs of target mRNA, 2) conservation of miRNA binding

site, and 3) minimum free-energy (MFE) of miRNA/mRNA target duplex. However, each algorithm can result in both false positive and false negative prediction (Li *et al*., 2010). Hence, to make a more reliable miRNA target prediction, more than one algorithm is often used.

Table 2.1 Computational target prediction algorithms

Tool	Principle	References
miRanda	Use training set of reference species such as human, mouse or rat to determine the optimal weighted alignment emphasizing seed.	John *et al*., 2004
RNAhybrid	Use training set of 3'UTR sequences and dinucleotide frequency of the 3'UTR sequences from reference species such as nematode and fly to determine the seed match and minimum free energy.	Rehmsmeier *et al*., 2004
TargetScan/ TargetScanS	Use training set of miRNA that is conserved in multiple organisms (human, mouse, worm, fly) and a set of orthologous 3'UTR sequences from the genes in these organisms to established rule for seed which is positioned at nucleotide 2 to7 in the 5'end of the miRNAs.	Lewis *et al*., 2003; 2005
PicTar	Use hidden markov model (HMM): (1) to identify perfect complementarities between seed region of the miRNA and 3'UTR sequences; (2) to identify perfect complementarities interrupted by at most one nucleotide bulge, mismatch or G:U wobbles.	Krek *et al*., 2005
miTarget	Use support vector machine, SVM to learn target rules automatically from positive and negative examples of validated miRNA-mRNA target.	Kim *et al*., 2006
TargetSpy	Use support vector machine, SVM to learn target rules automatically from positive and negative examples of validated target site (Argonautes binding site).	Sturm *et al*., 2010

Computational predictions of miRNA targets were not consistent across different algorithms and usually yield hundreds of potential miRNA target genes (Subramaniam *et al*., 2005; Gusev, 2008; Zhu *et al*., 2010). Hence, there is a clear need for data reduction methods which would allow prioritizing the list of targets of a large group of co-expressed miRNAs as well as to find biological functions that are most significantly affected by multiple miRNAs. For this reason, gene set enrichment analysis would be more informative than gene by gene search when considering large number of miRNA target genes (Subramaniam *et al*., 2005; Gusev, 2008). This statistical analysis could be used to determine the gene ontology (GO) term or pathways that are enriched with miRNA target genes (Gusev, 2008).

It is also important to have computational methods that are able to combine mRNA and miRNA expression profiles for accurate and reliable target prediction as well as to minimize false positives and to detect the functional miRNA targets under a specific biological condition. This integrative approach is based on the principle that an inverse expression relationship should be held between specific miRNA and mRNA (Dai and Zhou, 2010). This approach is suitable for mRNA targets regulated by miRNA via degradation but is not suitable for mRNA targets regulated by miRNA via translational repression (Dai and Zhou, 2010). Examples of algorithms or software tools developed to interpret miRNA expression data, to infer the regulation of miRNA from the transcriptome profiles and to combine mRNA and miRNA expression profiles for target prediction include GenMIR++ (Huang *et al*., 2007), miRGator (Nam *et al*., 2008), Sigterm (Creighton *et al*., 2008), MIR (Cheng and Li, 2008), TopKCEMC (Lin and Ding, 2009), CORNA (Wu and Watson, 2009), GeneSet2miRNA (Antonov *et al*., 2009),

microRNA and mRNA integrated analysis (MMIA) (Nam *et al*., 2009) and MiRonTop (Le Brigand *et al*., 2010).