GLOBAL ANALYSIS OF MICRORNA SPECIES IN THE GALL MIDGE MAYETIOLA DESTRUCTOR

by

CHEN DU

B.S., Nanjing Agriculture University, 2013

A THESIS

Submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Entomology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2015

Approved by:

Co-Major Professor
R. Jeff Whitworth

Approved by:

Co-Major Professor
Ming-Shun Chen

Copyright

Abstract

MicroRNA (miRNA) plays a role in nearly all the biological pathways and therefore may provide opportunities to develop new means to combat the Hessian fly, *Mayetiola destructor*, a destructive pest of wheat. This study presents a comprehensive analysis of miRNA species via deep-sequencing samples from Hessian fly second instar larvae, pupae and adults. A total of 921 unique miRNA species were identified from approximately 30 million sequence reads. Among the 921 miRNA species, only 22 were conserved among Hessian fly and other insect species, and 242 miRNA species were unique to Hessian fly, the remaining 657 share certain sequence similarities with pre-miRNA genes identified from various insect species. The abundance of the 921 miRNA species based on sequence reads varies greatly among the three analyzed stages, with 20 exclusively expressed in adults, two exclusively expressed in pupae and two exclusively expressed in second instar larvae. For miRNA species expressed in all stages, 722 were with reads lower than 10. The abundance of the remaining 199 miRNA species varied from zero to more than eight-fold differences among different stages. Putative miRNA-encoding genes were analyzed for each miRNA species. A single putative gene was identified for 594 miRNA species. Two putative genes were identified for 138 miRNA species. Three or more putative genes were identified for 86 miRNA species. The three largest families had 14, 23 and 34 putative coding genes, respectively. No gene was identified for the remaining 103 miRNA species. In addition, 1516 putative target genes were identified for 490 miRNA species based on known criteria for miRNA targets. The putative target genes are involved in a wide range of processes from nutrient metabolism to encoding effector proteins. Analysis of the expression patterns of miRNA and pre-miRNA for the miRNA family PC-5p-67443, which contains 91 genes, revealed that miRNA and pre-miRNA were expressed differently in different

developmental stages, suggesting that different isogenes are regulated by different mechanisms, or pre-miRNAs had other functions in addition to as an intermediate for miRNA biogenesis. The large set of miRNA species identified here provides a foundation for future research on miRNA functions in Hessian fly and for comparative studies in other species. The differential expression patterns between a pre-miRNA and its encoded mature miRNA in a multigene family is an initial step toward understanding the functional significance of isogenes in dramatically expanded miRNA families.

Table of Contents

List of Figures

List of Tables

Acknowledgements

My deepest gratitude goes first and foremost to my co-major advisors Dr. Ming-shun Chen and Dr. R. Jeff Whitworth, for their constant encouragement and guidance to help me well plan and develop my research. Especially Dr. Chen, he walked me through all the stages of my writing of this thesis. Without his consistent and illuminating instruction, this thesis could not have reached its present form.

I would also like to express my deepest gratitude to Dr. Gregory Ragland, for being my committee member and always willing to give his time generously and provide constructive suggestions to help me improve my master's study.

I would like to offer my special thanks to Dr. Xuming Liu, for his concerns for my professional study as well as personal developmental during my research. I am also grateful to Shauna Dendy for preparing me with the Hessian flies, and for teaching me how to increase the insect properly.

I would like to extend my great appreciation to all the people at the main office, Evelyn Kennedy, Ruth Nihart, Sharon Schroll, and Sherry Figge, for always being there to help with my master program.

I am also thankful to all the people in my lab, Sandra Garces, Xiaoyan Cheng, Zhuhong Wang, Hongmu Ai, and Zainab Abdalhussin Ali, for providing me with all the help I need when doing experiments and always support me. And I would like to thank all my friends in the department of Entomology for their support.

Finally I would like to give my special thanks to my parents for providing me the opportunity to study here.

Dedication

I would like to dedicate this thesis to my family. My parents devoted their love and faith on me and always trusted me on my career. Their unconditional supports made me as good as the person I wanted to be. My aunt, who led to the career of Agriculture especially Entomology, and always guides me toward a successful scientist and a good person.

Introduction

Researches on the function and biological implications of microRNA (miRNA) have been explored in various biological fields since its discovery. The reason that miRNA attracts such wide attention of biologists from different fields is that miRNA is involved in nearly all biological processes through regulating the expression of protein-encoding genes. Insects are the most numerous types of animals on earth and affect human life in various ways. A proportion of insect species are pests of plants, and cause damage to crops in agriculture. Control of plant pests is always a challenge because insects can acquire resistance to chemical insecticides and produce virulent biotypes that can overcome plant defense. Studies on miRNAs in insects could result in discovery of new means, or improve existing means, to more efficiently control damage caused by insect pests. Host plant resistance and plant – insect interactions are important research areas for biologists in agriculture. Knowledge on the mechanisms of host plant resistance to parasitic pests could be translated into new measures for pest control and therefore increase food production. A better understanding of plant – insect interactions can also enrich our knowledge on fundamental biology on interspecies interaction, communication, adaptation, and co-evolution. The gall midge, *Mayetiola destructor*, the so called Hessian fly, is one of the most destructive pests of wheat, and causes serious damage almost every year in major wheat production regions all over the world, including North Africa, Europe, Western Asia, Central Asia, North America, and New Zealand. We hypothesize that miRNA species play a critical role in wheat – Hessian fly interaction and therefore affect wheat resistance to the Hessian fly. Even though there are tremendous advances in understanding the functions of miRNA species in various

biological fields, research on the role of miRNA in plant – insect interactions is just beginning and very little is known about the specific roles of miRNA species in host plant resistance. The long-term goal of this project is to use Hessian fly as a model to explore the molecular functions of miRNA species in host plant resistance and plant – pest interactions. As the first step, this study was to identify and initially characterize miRNA species in differential developmental stages of Hessian fly through a genome wide analysis.

Literature Review

miRNAs

Discovery of miRNAs: miRNAs are groups of small (18 to 24 nucleotides) non-protein coding RNA species that regulate gene expression post-transcriptionally in Eukaryotes (Sun et al., 2010). miRNAs were first discovered by Rosalind Lee et al (1993) when they found a 22 nucleotides-long noncoding transcript, *Lin-4*, which could down regulate the expression of the protein-coding gene *Lin14* in first instar larvae of *Caenorhabditis elegans* (Lee et al., 1993). At that time it was only regarded as a small piece of RNA. Seven years later Frank Slack et al (2000) discovered another 21 nucleotide long small RNA (named miRNA *Let-7* later) in *C. elegans* larvae, which could negatively regulate the expression of the *LIN-41* gene (Slack et al., 2000). These two findings began to bring attention to these small RNA transcripts, and the term "microRNA" for this type of small transcript was officially defined in the year 2001 by three independent research groups. Meanwhile, a system to name miRNA species and their corresponding genes was also established, with miR-X for individual miRNAs and *mir-X* for their corresponding genes, respectively (Lagos-Quintana, 2001; Lau, 2001; Lee, 2001). Since then, studies on miRNAs have expanded dramatically. According to miRbase, over 24,000 miRNA loci from 206 species that produce over 30,000 mature miRNAs have been identified (Griffiths-Jones, 2014), and the number of miRNA reported increases rapidly within time.

Origin and evolution of miRNA genes: miRNA genes have multiple evolutionary origins with different genetic mechanisms. In most plant species, one way

to form a new miRNA gene is through inverted duplication of protein-coding genes. Duplications can occur in a complete or a partial protein-coding gene sequence with or without promoter region in either head-to-head or tail-to-tail direction. Reverse transcription of the newly duplicated locus would generate foldback transcripts. A Dicer-like enzyme (DCL) assists the foldback transcript to adapt to miRNA biogenesis and induces the formation of a newborn miRNA gene with mature hairpin structure (Allen et al., 2004).

Transposable elements can also contribute to new miRNA genes arising as observed in *Arabidopsis* and rice (Piriyapongsa et al., 2007). Transposable elements, especially a DNA-type element named miniature inverted-repeat transposable elements (MITE), which retain the terminal inverted repeats (TIRs) but do not contain the internal open reading frame (ORF). The transcription of inverted repeats of MITE would result in the formation of hairpin structure, which is the basic requirement of pre-miRNA formation. Then, with further processing by Dicer, the folded hairpin structure would yield a nascent miRNA, resulting in the formation of a new miRNA gene (Piriyapongsa et al., 2008).

A miRNA species can be encoded by either single or multiple genes. For those miRNAs with multiple coding genes, the multiple coding genes were regarded as a miRNA gene family. For most arthropod species, the miRNA gene families are much smaller than those of plant species (Li et al., 2007; Nozawa et al., 2010). The miRNA origin for Arthropods is quite different from plants. Novel miRNAs may be generated from random formation of a hairpin structure from accumulated nucleotide mutations of non-miRNA transcripts such as rRNAs, tRNAs, snRNA and snRNAs. Each of these

pathways for the formation of new-born miRNAs have different efficiency, as has been demonstrated in *Drosophila* (Lu et al., 2008).

miRNA genes can also arise from duplication and diversification of newly duplicated existing miRNA genes, a phenomenon relatively common in vertebrates. For example, the human mir17 microRNA cluster including miR-17, miR-18, miR-19b, miR-20, miR-25, miR-92, miR-106a, and miR-106b, was formed through a series of local duplications as well as duplications of the entire clusters, followed by nucleotide substitutions and rearrangements of the newly duplicated genes (Tanzer et al., 2004). The mutated gene or gene fragment becomes a new functional gene if it can form a complementary duplex with a new protein gene target.

miRNAs biogenesis: miRNA biogenesis can be through different mechanisms. Most miRNA species encoded by independent genes are produced through the so-called canonical biogenesis pathway, which has three different stages. First, miRNA transcripts are transcribed as the so-called primary miRNAs (Pri-miRNAs) in the nucleus, which are usually hundreds to thousands of nucleotides long with several hairpin structures. These Pri-miRNAs are then modified by a microprocessor complex, which consists of proteins with multiple functions. Among the components of a microprocessor complex, one of the most important proteins is DGCR8, a conserved protein composed of two double-strand RNA-binding domains that recognize pri-miRNA targets (Du et al., 2005). Another important component is the RNase III enzyme Drosha, which cleaves pri-miRNA into a ~70 nt long RNA fragment with a mature stem-loop structure, called precursor-miRNA (Pre-miRNA) (Lucas et al., 2013). Pre-miRNAs are then transported to cytoplasm via a double-stranded miRNAs (dsRNA)-binding protein, called exportin 5,

which specifically carries a pre-miRNA and exports it to cytoplasm (Bohnsack et al., 2004). In cytoplasm, another ribonuclease called Dicer cleaves the pre-miRNAs to generate a complementary miRNA: miRNA* duplex. This duplex is then loaded onto the so called RNA-induced silencing complex (RISC) for further regulation (Bohnsack et al., 2004). The miRNA strand guides RISC to complimentarily pairing to a target mRNA, while miRNA* strand would be degraded except under some specific circumstances when they can also display a regulatory function (Okamura et al., 2008).

miRNAs can be encoded either by an independent gene or derived from an intron of a protein-encoding gene. miRNAs produced from introns of a gene are called intronic miRNAs. The biogenesis pathways of intronic miRNAs have two different mechanisms. One is canonical intronic miRNA biogenesis through Drosha and Dicer cleavage, as described in the previous paragraph, except that heterogeneous nuclear RNA (hnRNA) serves as pri-miRNA (Ramalingam et al., 2014). Another mechanism is through direct intronic splicing, with intron mimic the structure of pre-miRNA. The pre-miRNA product from this process is called "mirtron". Mirtron pathway of intronic miRNAs was first identified from *D. melanogaster* and *C. elegans*, with 14 mirtrons identified from *D. melanogaster* and four identified from *C. elegans* (Ruby et al., 2007). Upon mRNA splicing in nucleus, the spliced mRNA exons are ligated, and the intron was separated with the mRNA transcript to form a lariat structure. The mirtron lariat is further processed by the debranching enzyme and folded into a mirtron with a stem-loop structure observed in a pre-miRNA. The mirton was then transported to cytoplasm by Expotin-5 for further Dicer cleavage to generate mature miRNAs (Kim et al., 2007).

Mechanism for miRNA functions: Biochemically, miRNAs function through complementary base pairing with their regulatory targets. During this process, a so called "miRNA seed", which is a seven nucleotide region ranging from the second to the eighth nucleotide of the mature miRNA 5' region, initiates the pairing to the 3'-untranslated region (3'UTR) of a target mRNA through so called "Watson-Crick seed match". This seed match is also the key factor for manipulating miRNA target recognition and pairing (Lewis et al., 2003; BARTEL et al., 2004). Complete pairing mostly happens in plant species and leads to translational repression whereas incomplete pairing with a few base mismatches is more likely to occur in animals and results in target mRNA cleavage (Goodall et al., 2013).

Because miRNAs regulate the levels of transcripts encoding for various proteins, miRNAs can affect diverse biological processes such as cell differentiation and development, proliferation and apoptosis, tumor genesis, immune response, metabolism, as well as host-parasite interactions in general (Cai et al., 2009).

miRNAs in insects: Since the discovery of the first two miRNAs in *C. elegans* (Slack, 2000), studies on miRNAs have been undertaken on various insect species. According to miRBase, there are a total of 3824 mature miRNAs identified from 26 insect species, representing 12.6% of all the miRNAs in the database. The most extensively studied insect species for miRNAs are *D. melanogaster* (Nozawa et al., 2010), *Manduca sexta* (Zhang et al., 2014), *Bombyx mori* (Cao et al., 2008), *Aphis gossypii* (Rebijith, 2014), *Tribolium castaneum* (Singh et al., 2008) and *Aedes aegypti* (Li et al., 2009).

Functional analysis of insect miRNAs reveals various important involvements in insect developmental events. *D. melanogaster* is the first and one of the most extensively studied insect species with numerous families of miRNAs having been identified. miRNAs in *D. melanogaster* have been found to perform functions in aging, apoptosis, cell growth and proliferation, signal transduction, metabolism regulation, neurodegeneration and stress response (Behura et al., 2007). In social insects, species-specific miRNAs perform specific regulatory functions in regulating social behavior as found in honeybees (Liu et al., 2012). In pea aphids, miRNAs participate in manipulating reproduction modes (Legeai et al., 2010). In silkworm, miRNAs are involved in embryogenesis and metamorphosis as well as in regulating silkworm fibroin productions (Cao et al., 2008; Zhang et al., 2009). miRNAs are also involved in signal transduction in *Aphis gossypii* (Rebijith et al., 2014). A recently identified mosquito-specific miRNA, Aae-miR-2904 was shown to reduce flavivirus infection during anti-viral response, indicating that miRNAs function as an immune regulation factor in insects (Hussain et al., 2014).

Insect miRNAs regulate insect development and those miRNAs themselves are also developmentally regulated. For instances in *B.mori*, the conserved miRNA species miR-1000 is highly expressed in eggs, but only at low level in the adult. The expression of miR-1268 is also predominant in the egg and larval stages but much lower in the pupal and adult stages. Besides, miR-iab-4-3p is highly expressed in the egg and larval stages but have very low expression in other stages (Huang et al., 2010). In *M. sexta*, 30 miRNA species such as mse-miR-8 and mse-miR-71 were highly expressed in embryo,

and 16 miRNA species were found to have dominant expression level in the adult stage including mse-miR-100 and mse-miR-277 (Zhang et al., 2012).

The Hessian fly

The gall midge family Cecidomyiidae (Diptera): Cecidomyiidae is one of the largest families of Diptera. Phylogenetic analysis indicates that four subfamilies of Cecidomyiidae exist. Among those subfamilies, Catotrichinae, Lestreminnae and Porricondylinae are ancient families that are parasitic on saprophytic fungi, referred to as fungal gnats. The fourth subfamily is Cecidomyiinae, commonly known as Gall midges. The Cecidomyiinae is the largest subfamily with 565 genera and 4819 known species, representing about 80% of all known cecidomyiids (Gagné et al., 2014). Almost all the insect species within this subfamily are important plant gall-inducing pests such as the rice gall midge *Orseolia oryzae*, orange wheat blossom midge *Sitodiplosis mosellana*, and the Hessian fly *Mayetiola destructor* (say) (Harris et al., 2003). Gall midges as a group have a wide range of host plants including wheat, barley, rice, soybean, pine needle, alfalfa, etc. (Weis et al., 1988). However, individual gall midge species usually only live on a single host or very narrow range of host plants. Gall midges induce the formation of galls on various tissues such as buds, stems, leaves, flowers and fruit.

Hessian fly distribution and damage: Hessian fly is one of the most destructive pests in the plant-feeding gall midge subfamily (Schumann et al., 1976), and is widely distributed in North Africa, Europe, Western Asia, Central Asia, North America, and the Pacific ocean island of New Zealand (Stuart et al, 2012). Hessian fly is believed to have been first brought to the United States in 1779 by Hessian soldiers

(Bailey et al., 2000). Since it was first discovered in Long Island, New York, Hessian fly has spread to nearly every wheat-cultivating state including Alabama, Mississippi, Indiana, Kansas, Idaho, Pennsylvania, Washington, etc. (Ratcliffe et al., 2000). Annual grain loss caused by Hessian fly can reach one-third of wheat yield in outbreak areas (Lhaloui et al.,1992). In 1999, 5 to 10% of total wheat production in the United States was lost due to the infestation of this midge (Buntin et al., 1999). At present, Hessian fly remains one of the most devastating pests of wheat in the U. S. (Ratcliffe et al., 2000).

Hessian fly life cycle: In the warm and humid Southern region of the US, Hessian fly usually has three to six generations a year. But in the Northern region, the weather is cooler and less humid, and Hessian fly has only two generations per year (Flanders et al., 2008).

Under laboratory conditions, a complete Hessian fly life cycle takes about 28 days at 20 ℃, going through egg, three instars of larvae, pupae and adults. Under field conditions, Hessian fly's lifespan varies based on temperatures (the cooler, the longer). Hessian fly adult is a non-feeding stage and causes no damage. A female Hessian fly adult lives for 3-4 days and can oviposit 50-400 eggs (Stuart et al., 2012). Female flies deposit tiny reddish eggs in the parallel grooves of the upper surface of younger wheat leaves. It takes around four days at 20 ℃ for eggs to hatch into first instar larvae and then 12-24 hours for reddish larvae or "maggots" to crawl down to the base part of a wheat plant between a leaf sheath and the stem to establish a permanent feeding site. Within seven days at 20 ℃, larvae molt and turn from reddish to transparent white, and this stage is the second instar. Then after another five days at 20 ℃, mature larvae enter into the puparium stage, during which process the outer skin of larvae begins to harden

into a dark brown, protective husk, called "flaxseed". This dormancy stage lasts for about three months at 4 °C and can be terminated by a period of 10 days at room temperature. Once dormancy is broken, the adults emerge from the puparia, then mate and deposit eggs (Stuart et al., 2008).

Hessian fly feeding mechanism: Among the developmental stages of the overall life cycle of Hessian fly, the first instar is the most damaging stage and determines the fate of both the Hessian fly and the attacked plant. In a susceptible host plant, a single Hessian fly larva can effectively manipulate the whole plant through inhibiting host plant growth, suppressing plant defense, and inducing the formation of nutritive cells at the feeding site. The infested tiller dies after the larva finishes development and no longer needs nutrients. In a resistant plant, a larva can feed in a wheat plant initially, but dies within three to five days due to plant antibiosis defense, whereas the plant grows normally after some initial growth deficit. A first instar larva feeds by sucking up plant juices through mouthparts and at the same time, the larva injects saliva through its mandibles, which are connected to a pair of salivary glands, into leaf sheath cells during feeding. Effector proteins, also known as Secretory Salivary Gland Proteins (SSGPs) in Hessian fly saliva, are likely to induce changes of normal cells and convert them into nutritive cells, which form a so-called nutrient sink at the feeding site. Nutrients that are produced during plant photosynthesis flow to the feeding site and are ingested by the Hessian fly larva.

It appears that Hessian fly larvae can also enhance photosynthesis by manipulating a higher concentration of chlorophyll in susceptible plants. Due to the increase of chlorophyll, the leaves of infested plants become dark green instead of the

more normal light green. Another important manipulation of host plants by a virulent Hessian fly larva is to irreversibly inhibit plant growth. The inhibition in plant growth is thought to be due to reprogrammed cellular activity and lack of nutrients (Stuart et al., 2012). However, the fact that the plant's growth cannot be restored after five days of continuous feeding by a larva suggests that the overall biological and physiological pathways in the infested plant have changed irreversibly. Thus, this formation of nutritive cells and a nutrient sink, the inhibition of wheat growth, and the enhancement of photosynthesis, effectively converts the whole plant into a "gall", creating a favorable environment for a larva to live and develop.

Hessian fly management: Several tactics for Hessian fly management have been applied in the field to control Hessian fly damage. Those tactics include chemical control, biological control, cultural measures, and the use of resistant wheat varieties (Royer et al., 2009).

Chemical control includes seed treatment with insecticides such as imidacloprid and thiamethoxam, and field-applications of insecticides for fall Hessian fly infestation. Seed treatment is usually expensive and the effect of seed treatment only lasts for about 30 days (Lamiri et al., 2001). Biological control through releasing natural enemies is an environmental friendly tactic to suppress Hessian fly populations. For example, the release of parasitic wasps *Homoporus destructor* and *Eupelmus allynii* has been found to effectively kill up to 80% of Hessian fly in Texas (Morgan et al., 2005). However this method was only found in very limited areas.

Cultural practices such as crop rotation that interrupts Hessian fly host from wheat to a non-host plant species, can effectively disrupt continuous availability of host plants to local Hessian fly populations. However, crop rotation in large areas is hard to coordinate. In addition, wheat-related grass species could serve as alternative hosts. Also delayed planting, the so called fly-free date (Now more appropriately called the best pest management planting date), would potentially ensure wheat fields escape from Hessian fly infestation (Duyn et al., 2004). But the fly-free date varies from year to year due to temperature fluctuation. Destruction of volunteer wheat can also be important for Hessian fly control. To be effective, volunteer wheat plants need to be destroyed at least two weeks before planted wheat germinates.

Since Hessian fly stays in wheat stubble as a puparium, destruction of stubble has been regarded as one of the most effective ways to fight Hessian fly. Burning wheat stubble is easy to achieve and can kill Hessian fly flaxseed in stubble above the soil surface. However, flaxseeds below ground can survive. In addition, burning stubble destroys nutrients in wheat residues that are otherwise available for the next wheat crop. Another way to deal with stubble is plowing. Infested stubble can be buried into deep soil for four or more inches, and this practice may wipe out all the Hessian fly flaxseed (Ormerod et al., 1887). However, no-till practice is popular at present due to other economic and environmental benefits, and therefore the destruction of wheat stubble for Hessian fly control is currently not as widely utilized as could be.

Currently, the most effective method to control Hessian fly is the deployment of resistant wheat cultivars (Garcés-Carrera et al., 2014). Deployment of Hessian fly-resistant wheat cultivars adapted to the Great Plains region dates back to the 1920s, and

historical data also suggests that in the hard and soft winter wheat regions in the Great Plains including Kansas, Nebraska and Missouri, 60 Hessian fly-resistant cultivars were successfully developed from 1950 to 1983, and the yield loss caused by Hessian fly was reduced to 1% (Ratcliffe, 2013). This historical data provides the promise of success for future study of resistant wheat cultivars and their deployment.

Host plant resistance: Hessian flies interact with wheat following a typical gene-for-gene model between wheat resistance (R) gene and Hessian fly avirulence (*Avr*) gene. Two types of interactions exist, incompatible interaction and compatible interaction. During incompatible interaction, a wheat plant contains a dominant R gene that can specifically activate or initiates the biotype-specific *Avr* gene in Hessian fly. The reaction between the R gene and *Avr* gene then results in host defense response and inhibition of Hessian fly development. During compatible interactions then, host plants do not contain an *R* gene, or Hessian fly does not carry the corresponding *Avr* gene, thus Hessian fly can successfully grow and develop (Thakur et al., 2007). So far, 35 resistant genes, namely *H1-3, h4, H5-H34* and *Hdic* have been identified from common wheat, durum wheat, rye and goat grass (Subramanyam et al., 2015). These genes, except for *h4,* are either dominant or semi-dominant with antibiotic effects on first instar larvae. On the other hand, four of the genes (*H13, H21, H26* and *Hdic*) confer high levels of resistance to Hessian flies in the Great Plain region.

However, using host plant resistance to combat Hessian fly remains a challenge despite its obvious advantages in terms of economy, ecology and environment. Hessian fly has developed genetic mechanisms during the long course of coevolution with wheat to change biotype compositions (Sardesai et al., 2005). It only takes a few years for a

resistant cultivar to lose effectiveness because of the occurrence of new virulent biotypes in the field (Gould et al., 1998). Due to the lack of durability of resistant wheat cultivars, efforts are needed to improve the durability of resistant wheat or design complete new strategies to combat the Hessian fly pest.

Hessian fly miRNA studies: Hessian fly miRNAs were initially analyzed in 2013, and a total number of 273 miRNA species were identified from first instar larvae. The expression patterns of some of the identified miRNA species were regulated differently between larvae feeding on resistant vs susceptible plants (Khajuria et al., 2013). Based on this initial observation, we hypothesize that miRNAs are involved in Hessian fly-wheat interactions. A more comprehensive study of miRNA species in Hessian fly will help reveal how miRNAs regulate genes that are involved directly or indirectly in Hessian fly virulence and host plant resistance.

Objectives and Rationale

miRNA species were initially analyzed from first instar larvae of Hessian fly, and discovered that some families of miRNA species are unique to this gall midge. Several families of Hessian fly-unique miRNA species are dramatically expanded in the Hessian fly genome. In addition, many miRNA species that are unique to Hessian fly are differentially regulated in insects feeding on resistant and susceptible wheat plants. These findings indicate that miRNA species are likely to play important roles in wheat – Hessian fly interactions. Since miRNAs are involved in diverse physiological and biochemical processes, identification and characterization of miRNAs may also reveal

new findings that can result in alternative strategies for Hessian fly control. The objectives of this research are:

1. To conduct a comprehensive identification and analysis of miRNA species in the Hessian fly genome through deep-sequencing RNA samples from different developmental stages including second instar larvae, pupae, and adults. The previous analysis had only analyzed miRNA species that are expressed in the first instar (3-day-old larvae). Since expression of miRNA species is strictly expressed at different stages and tissues, there are likely more miRNA species that are expressed in other developmental stages.

2. To identify potential target genes of selected miRNA species. The expressions of miRNA species were negatively correlated with the expression of target transcripts, which allows the identification of potential target genes of each miRNA species. The long-term goal of the research is to identify those miRNA species that regulate the expression of genes that play a role in wheat – Hessian fly interaction, or play other critical roles in Hessian fly development.

3. To analyze the expression patterns of different miRNA species in different developmental stages of the Hessian fly. Several families of miRNA species have been greatly expanded in the Hessian fly genome, a phenomenon that has not been observed in any other species. For example, the family of miRNA PC-5p-67443 contains 91 apparently functional genes according to sequence conservation. These 91 genes encode an identical or very similar miRNA, which is presumable to have the same function because of the identical sequence.

4. To analyze the transcripts of pre-miRNA and their corresponding mature miRNAs of members in an expanded family. This objective is to determine if there is any similarity or differences between the expression pattern of pre-miRNAs and their corresponding mature miRNAs. Due to the unique expansion of the PC-5p-67443 miRNA family, the comparison between expression patterns of pre-miRNA and mature miRNA can help determine whether pre-miRNAs also perform regulatory functions.

Materials and Methods

Hessian fly population and Greenhouse increase

The Hessian flies used in this study were the "Great Plains" biotype (GP), which is virulent to wheat varieties that carry no resistant genes, but cannot survive on wheat varieties carrying R genes *H3, H5, H6,* and *H7H8* and most other R genes (Ratcliffe, 2000b). The Hessian fly population was originally collected in Scott County, Kansas (Chen, 2009). Since collection, the Hessian fly population has been continuously maintained as a laboratory colony by infesting the susceptible winter wheat cultivar "Karl 92" in greenhouse.

During colony maintenance, Karl 92 seedlings were infested with Hessian fly adults in greenhouse at 20°C ~ 22°C. After 16 to 18 days, about 80% of Hessian flies reached the puparium stage, infested plants were collected and placed in boxes covered by plastic film, then stored in cold room at 4 °C for at least two months to break diapause before use.

Wheat cultivars and Hessian fly infestation

The wheat cultivar used in this study was the susceptible winter wheat cultivar Newton. Plant germination and fly emergence were carefully coordinated so that freshly emerged flies were available to infest wheat seedlings at the right stage. To achieve that, fly stocks were taken out from a cold room four days ahead of planting, since it usually takes 10 days for adult flies to emerge (17 ℃ ~ 20 ℃) and seven days for a plant to reach the 1.5-leaf stage (20 ℃). Wheat seeds were planted in a 10-cm-diameter pot with 20 seeds per pot and maintained in a growth chamber at 20 ℃, with a photoperiod of 14:10h (L: D). Each day, six pots were planted and the planting was continued for one week. Plants were infested with female adult flies, which have a reddish and bigger abdomen than male flies, at 1.5-leaf stage with 30 flies per pot (1.5 flies per plant) to ensure each plant had enough insect samples. Infested pots were covered with transparent screens to ensure infestation within the specified pot, as well as to protect other plants from accidental infestation.

Hessian fly sample collection and total RNA isolation

For Hessian fly sample collection, one sample represent insects collected from a specific developmental stage. Two groups of Hessian fly samples were prepared for this project. The first group of samples were used for Hessian fly miRNA deep sequencing. Insects were collected at three different developmental stages: seven-day-old larvae (representing second instar larval stage), pupae, and female adults. One sample for each developmental stage was collected following three steps: 1) dissect the base part of wheat stem where insects were attached; 2) soak the dissected stem with insects in a 2.0 ml

centrifuge tube filled with 1.5 ml sterilized water; 3) remove the water after all the insects were collected into the tube. For each tube, approximately 100 mg insects were collected.

The second group of Hessian fly samples were used for Real-time PCR analysis. Insect samples were collected at six different developmental stages: one-day-old larvae and three-day-old larvae (representing first instar larval stage), five-day-old larvae and seven-day-old larvae (representing second instar larval stage), pupae, and female adults. For sample preparation, three samples for each developmental stage (three biological replicates) were collected at the same time from the plants infested in the same day (2 pots/replicate). The samples were collected following three steps: 1) dissect the base part of wheat stem where insects were attached; 2) soak the dissected stem with insects in a 2.0 ml centrifuge tube filled with 1.5 ml sterilized water; 3) remove the water after all the insects were collected into the tube. For each tube, approximately 200 mg insects were collected.

A total of 21 insect samples were collected, total RNAs were extracted from collected samples using RNAzol®RT following the manufacture's protocol (Molecular Research Center, Inc. Cincinnati, OH). RNA samples are analyzed using the NanoDrop 2000c spectrophotometers (Thermo Fisher Scientific Inc. Waltham, MA) to determine the quality and quantity. Analyzed total RNA samples were stored in -80 °C refrigerators for further research use.

miRNA deep sequencing and data analysis

For miRNA deep sequencing, total RNA samples from Hessian fly seven-day-old larvae, pupae and adults were shipped to a commercial sequencing company (LC

Sciences, LCC. Houston, TX) for deep-sequencing. The RNA samples were first processed to generate cDNA libraries. The cDNA libraries were then further processed for cluster generation on Illumina's Cluster Station and were sequenced on Illumina GAIIx following vendor's instructions.

The software ACGT101-miR v4.2 (LC Sciences) was used for sequence data analysis, a series of digital filters (LC Sciences) were employed to remove various un-mappable sequencing reads from obtained raw data, the un-mappable reads included low quality reads, copy number less than three and length less than 15 nucleotides, and other junk sequences including sequence with 80% or more of the same base (either A, C, G or T), or reads with only A,C or G,T. The remaining mappable reads were further classified to groups.

Identification of putative miRNA coding genes

Two steps were followed to identify putative genes encoding miRNAs in the Hessian fly genome. Firstly, miRNA sequences were used to search the Hessian fly genome sequence database using Megablast (http://blast.ncbi.nlm.nih.gov/) to identify matched genomic sequences. Those sequences with two or less mismatches for sequence length equal to or less than 19 nucleotides and three or less mismatches for sequence length more than 19 nucleotides were taken as candidate genes for further analysis. For each candidate gene, a ~140 nucleotide long fragment containing miRNA encoding region and ~60 nucleotides flanking sequences at both 5' and 3' ends was extracted from the database for further analysis.

For the second step, the potential matches obtained from step one were analyzed

using RNA Mfold (http://mfold.rit.albany.edu/?q=mfold) to determine if the putative pre-miRNA coding region of a matched genomic sequence could form a stem-loop structure, which is the prerequisite of miRNA biogenesis. Only those sequences that can form a complete loop structure were kept as putative miRNA-coding genes.

Primer design and synthesis

Gene-specific primers were designed for different coding genes of the largest miRNA family PC-5p-67443. Primers were designed for specific detection of both mature miRNAs and pre-miRNAs. All primers were designed based on the nucleotide differences among the sequences of the isogenes in the PC-5p-67443 family, and then analyzed with OligoAnalyzer 3.1 (https://www.idtdna.com/calc/analyzer) to ensure the specificity of each primer for each isogene. For mature miRNAs, forward primers were designed at miRNA encoding region with up to four flanking nucleotides at either 5' or 3' end. Reverse primers were designed at the complementary region of the miRNA-coding region following the same criteria. The resulting miRNA PCR fragment is 60 to 90 nucleotides long. For pre-miRNA, forward primers were designed to cover the 5' flanking sequence of putative miRNA-coding region and reverse primers were designed to cover the 3' flanking sequence. The primers were synthesized by Integrated DNA Technology (https://www.idtdna.com/site), and the resulting pre-miRNA PCR fragment was approximately 120 nucleotides long.

Reverse transcription

For reverse transcription and the following Real-Time PCR, total RNAs isolated from one-day-old larvae, three-day-old larvae, five-day-old larvae, seven-day-old larvae, pupae, and adults were analyzed. Approximately 2,500 ng total RNA of each sample was

used for reverse transcription. For miRNAs, Mir-X^{TM} miRNA First Strand Synthesis Kit from TaKaRa Biotech Company was used to convert miRNAs into cDNAs to enable specific RNAs to be quantified by real-time PCR following the manufacture's protocol (Clontech laboratories. Inc. Mountain View, CA). Total RNA sample was added into a 0.2 ml RNase-free tube together with mRQ enzyme and mRQ Buffer, which is a complex containing active reagents necessary for polyadenylation and reverse transcription of RNA template (content was not listed due to manufacturer's proprietary). The mixture was incubated for 1 hour at 37 ℃ for reverse transcription of RNA to cDNA. The reaction was terminated by incubating at 85 ℃ for 5 minutes. The solution containing the resulting cDNA product was adjusted to 100 μl and stored at -20 ℃ for further experiments.

For pre-miRNA reverse transcription, the SuperScript® III First-Strand Synthesis kit was used following the manufacturer's protocol (Invitrogen by Thermo Fisher Scientific Inc. Waltham, MA). The whole process included two steps. First, the mixture of total RNA sample with a reverse primer and dNTPs in DEPC-treated water was incubated at 65 ℃ for 5 min to denature RNA secondary structures, followed by a chill on ice for primer annealing. Second, the annealed solution was mixed with RT buffer, Mg^{2+}, DTT, RNaseOUT and Reverse transcriptase and was then incubated at 50 ℃ for 50 minutes for cDNA synthesis. The reaction was terminated by incubation at 85 ℃ for 5 minutes. The cDNA products were stored at -20 ℃ for further experiments.

Real-time PCR analysis

Followed by reverse transcription, the synthesized cDNAs were used for Real-time PCR analysis. For miRNAs, the SYBR® Advantage® qPCR Premix kit was used

following manufacturer's protocol (Clontech laboratories. Inc. Mountain View, CA). Two sets of reactions were prepared for each sample. For qPCR reaction, the mixture of synthesized cDNA, SYBR advantage premix, ROX Dye, miRNA-specific forward primer and mRQ 3' primer with a total volume of 25 µl was prepared. A reference control was carried out with primers specific to the U6 small nucleolar RNA (snoRNA). For the U6 qPCR reaction, the components were the same except for the primers were substituted with U6 forward and reverse primers. The PCR program used was the comparative CT program. RT-PCR cycling parameters included 95 ℃ for 20 sec for enzyme activation, followed by 40 cycles of 95 ℃ for 3 sec for template denaturation and 60 ℃ for 30 sec for annealing and extension. At the end of the last cycle, a melting curve stage starts at 95 ℃ for 15 sec to denature the sample, followed by 60 ℃ for 1 min to annealing, and then continue to 95 ℃ for 15 sec for the high resolution melting. SYBR green dye was involved in this stage to serve as a signal for eliminating the possibility of primer-dimer formation.

For pre-miRNAs, the SYBR® Green Master Mix was used following manufacturer's protocol (Life technologies by Thermo Fisher Scientific Inc. Waltham, MA). The synthesized cDNA template of pre-miRNA was mixed with SYBR green PCR master mix together with specific forward and reverse primers. An internal control with the Hessian fly18s rRNA was included and was prepared in the same way except with a pair of primers specific to the 18S rRNA transcript. RT-PCR reactions were run under comparative CT program. The cycling parameters included 95 ℃ for 20 sec for enzyme activation, followed by 40 cycles of 95 ℃ for 3 sec for template denaturation and 60 ℃ for 30 sec for annealing and extension. At the end of the last cycle, a melting curve stage

started at 95 °C for 15 sec to denature the sample, followed by 60 °C for 1 min to annealing, and then continued to 95 °C for 15 sec for the high resolution melting. SYBR green dye was involved in this stage to serve as a signal for eliminating the possibility of primer-dimer formation.

For Real-time PCR, two technical replicates (repeats) were assigned to each of the three biological replicates for each developmental stage. Real-time PCR was performed with the StepOnePlus® Real-Time PCR System (Applied Biosystems, Thermo Fisher Scientific Inc. Waltham, MA). All experiments were performed with parameters under default settings suggested by the Real-Time PCR system.

Target gene identification and analysis

Three steps were carried out to obtain putative target genes of miRNA species. First, the 921 Hessian fly miRNA species were used to search against the Hessian fly Official Gene Set transcript database released on May 2012 (http://agripestbase.org/hessianfly/?q=blast). Sequences with more than 60% similarity within the miRNA encoding regions were selected for further analysis. Second, among the selected genomic sequences, only those with pairings located at the 3'-untranslated region were kept for the next step analysis. The sequences with complementary pairings which occurred in other regions were discarded. Finally, the expression patterns of all miRNA species among different developmental stages were analyzed based on the numbers of sequence reads from the deep sequencing data. The expression patterns of miRNA were compared with the expression patterns of protein-coding genes that were determined in a separate project. Those protein-coding genes with expression trends negatively correlated with a miRNA expression were taken as putative target genes for

that miRNA species. Genes satisfying these three conditions were considered putative target genes for a specific miRNA species.

Results

Identification of miRNA species from different developmental stages of Hessian fly via deep sequencing

In order to characterize miRNA species from the Hessian fly genome and their possible involvement in Hessian fly-wheat interaction, we isolated total RNAs from seven-day larvae, pupae and adults of the insect. The RNA samples were sent for sequencing through a commercial contract (see Methods). A total of 34,165,186 raw reads were obtained through sequencing a combined cDNA library with different barcodes for samples obtained from seven-day larvae, pupae, and adults, respectively. The raw reads were roughly evenly distributed among the three developmental stages, with 10,438,827 reads from the larval sample, 10,625,454 reads from the pupal sample, and 13,100,905 reads from the adult sample. There are three types of low quality reads in the raw sequence reads: 1) sequences with no 3'adaptor; 2) sequences with <15 nucleotides after removal of the 3' adaptor; and 3) the so called junk reads, which contained 80% or more of the same base (either A, C, G or T), or reads with only A,C or G,T. Among the three types of low quality reads, 15.1% were reads with <15 nucleotides. Reads without 3'adaptor and junk reads were each less than 1%. For some reason, the low quality reads were more common in the samples from the larval and pupal stages, especially the reads <15 nucleotides long, which represented 22.2% and 19.3% of the total raw reads from the larval and pupal samples, respectively, but represented only 6.0% of the total raw reads from the adult sample. After removing the low quality reads,

28,897,698 mappable reads, or 84.6% of the total reads, were maintained for further analysis (Table 1). A combination of more total raw reads and fewer percentages of low quality reads resulted in more mappable reads in the adult sample, which was 12,252,748, or 42.4% of total mappable reads from all three developmental stages. The mappable reads for the larval and pupal samples were 8,102,257 (28%) and 8,542,693 (29.6%), respectively.

The size of mappable reads ranged from 15 to 45 nucleotides (nt), as shown in Figure 1A. The reads with 30 nucleotides represented approximately 35% of mappable reads, the highest among all the groups. In comparison, the second highest group of reads was 31 nucleotides long, representing only 7% of total mappable reads, nearly 5-fold less than the group with 30 nucleotides. Reads with 29 nucleotides was the third highest, with 5% of total mappable reads. Together, reads with 29 to 31 nucleotides represented nearly half of the total mappable reads. For the remaining reads, 40% of the total mappable reads were distributed among 15 to 29 nucleotide groups with each group containing 2-4%. The remaining 13% of the total mappable reads belonged to the 32 to 45 nucleotide groups with each group of less than 1-3% of total reads.

The total 28,897,698 mappable reads were mapped to various public available databases including miRbase, mRNA, Rfam, and repbase. As expected, the largest proportion (43.8%) of the mappable reads represented various ribosomal RNA (rRNA), with the most abundant group of 29 – 31 nucleotides, which were mapped to the 5.8S ribosomal RNA. Other non-miRNA reads included 15.9% reads representing other small RNAs including tRNAs, snRNAs and snoRNAs, 5.9% reads representing repeated sequences in Repbase, and 32.3% representing other Rfam RNAs. After removing all the

non-miRNA reads, the remaining 677,789.33 mappable reads were kept for further analysis. The distribution of the 677,789.33 miRNA reads is shown in Figure 1B. The majority of the miRNA reads were 21 – 23 nucleotides, consistent with the size of majority of miRNAs reported in other species.

These unique miRNA reads were further assembled into 921 miRNA clusters, defined as 921 miRNA species. The 921 unique miRNA species were divided into groups according to their sizes, which ranged from 15 to 31, and 34 nucleotides (Figure 1B, Table 2, Supplemental Table S1). The most noticeable feature of these groups was the disproportion between the number of miRNA species and the number of unique reads contained in each group. The group of miRNA species with 15 nucleotides represented 253 different miRNA species, or 27.47% of the 921 identified miRNA species, the largest among all groups. However, the total unique reads contained in the 15 nucleotide group were only 36,145.41 or 5.29% of the total unique reads, or 143 unique reads per miRNA species on average. The disproportion between the percentage of unique miRNA species and the percentage of unique reads contained in that group was found in other miRNA groups as well. For example, the second largest group of unique miRNA species, the 22 nucleotide group, contained 102 different miRNA species, or 11.07% of the 921 identified miRNA species, but it had 214,929.59 or 31.71% of total unique reads, or 2107 unique reads per miRNA species on average. Similarly, the 21 nucleotide miRNA group consisted of only 69 or 7.49% of the identified miRNA species. However, the 21 nucleotide group contained a total of 150,822.05 or 22.25% of total reads, or 2186 reads per miRNA species on average, representing the most abundant group of miRNA species in terms of average unique reads per miRNA species. In summary, groups with 5% or

more miRNA species included the groups with 15, 16, 17, 18, 21, 22, and 23 nucleotides long, which together represented 75.55% of the 921 identified miRNA species; whereas the groups with 5% or more unique reads were 15, 21, 22, 23, and 24 nucleotides long, which together represented 87.43% of total unique reads. Other groups, especially the groups with 19, 20, 25 to 34 nucleotides, were minor groups with less than 4.00% of either total miRNA species, or total unique reads.

Classification of identified miRNA species

The 921 unique miRNA species were classified into four groups based on their sequence similarity to known miRNA species in miRBase (Supplemental Table S2). Group 1 contained 22 miRNA species that are identical to known miRNA species characterized from other insect species but novel found in Hessian fly. The name, sequence, miRNA origins and references of this 22 miRNA species are given in Table 3. Group 2 contained 73 miRNA species that are unique to Hessian fly but identified previously in our lab. Group 3 contained 657 miRNA species that share certain sequence similarities with pre-miRNAs of other insect species based on genomic sequences. Group 4 contained 169 miRNA species identified from this project for the first time, and are also unique to Hessian fly, with no sequence similarity to any known miRNA species in public databases.

Identification of putative miRNA-encoding genes from a Hessian fly genome draft sequence

From the first step of putative gene identification through BLAST search, a total of 3500 matched genomic sequences were identified for 918 miRNA species, while the remaining three miRNA species have no match identified (Table 4, Supplemental Method

1). Among the matched sequences, 3280 were considered to potentially contain full-length genes since they had extra sequences at least 65 nucleotides flanking the matched region at both 5' and 3' ends. There were 107 identified sequences that were considered to contain potential coding genes but were truncated at the 5' end, and 110 identified sequences were potentially miRNA-coding genes, but truncated at the 3'-end. The majority (406 or 44.1%) of the miRNA species had one matched sequence identified in the Hessian fly genome sequence. One hundred and sixty-nine (18.3%) miRNA species had two matched sequences. The remaining 343 (37.2%) miRNA species had three or more matched sequences, and among them, 17 miRNA species had more than 20 (20 ~ 108) matched sequences.

From the second step Mfold analysis of the 3500 matches, 1252 matched sequences were found to be able to form a typical stem-loop structure in the region surrounding the miRNA-matched region, and these sequences were considered to be putative miRNA-coding genes. The distribution of the putative miRNA-coding genes is shown in Table 5. The majority (594) of miRNA species have only one putative coding gene (Supplemental Method 2-Group 2). One hundred and thirty-eight miRNA species had two putative coding genes (Supplemental Method 2-Group 3). Eighty-six miRNA species had three or more putative coding genes (Supplemental Method 2). The remaining 103 miRNA species had matches identified through BLAST search but all those matches cannot form a putative stem-loop structure, and therefore no gene has been identified yet (Supplemental Method 2-Group 8). Among miRNA species that had multiple genes, PC-5p-107087_9, mde-mir-2505-p3, and mde-mir-7-p5_1ss2TG, contained 14, 23, and 34 putative coding genes, respectively.

In addition to the above analysis, sequence alignments were performed for those with multiple putative coding genes to determine their evolutionary relationship and functionality (Supplemental Table S3). Altogether 12 miRNA families with more than five coding genes per family were further analyzed with sequence alignments to show relationships between the miRNA coding region and flanking region (Figure 2, Supplemental Table S3). The multiple coding genes within each expanded miRNA family had more than 79% sequence similarities in miRNA encoding regions, with the majority (75%) of them identical, while the 5' and 3' flanking regions had more than 30% sequence identity, indicating that the multiple genes for those miRNA families are likely functional considering that sequences with less than 15 similarities were more likely to be random sequences.

Estimation of the abundance of the identified miRNA species in different developmental stages of Hessian fly

Assuming the sequence reads of each miRNA species were proportional to their abundance in each RNA sample, the abundance of the identified miRNA species in different developmental stages of Hessian fly was estimated based on the distribution of sequence reads among these developmental stages. The average reads per miRNA species were 1571. There were 39 miRNA species with sequence reads more than the average, 160 miRNA species with sequence reads less than the average but more than 10, and 722 miRNA species with sequence reads less than 10 when the data from larvae, pupae, and adults were pooled (All stages) (Table 6, Supplemental Table S4). When the data from different stages were analyzed separately, the overall pattern of miRNA read distribution was similar among larval (Supplemental Table S5), pupal (Supplemental

Table S6) and adult (Supplemental Table S7) stages with 2.6% to 4.2% of total miRNA species with more than average reads, 11.5% to 17.4% of miRNA species with less than average but more than 10 reads per miRNA, and 78.4% to 85.9% of total miRNA species with less than 10 reads.

To further characterize changes in expression of miRNA species in different developmental stages of Hessian fly, we compared the changes in the number of sequence reads between samples from different stages, including a comparison among all the stages, and comparisons between Larva-Pupa, Pupa-Adult, and Larva-Adult stages. As shown previously in Table 6, altogether 722 miRNA species have less than 10 reads in all three stages, and these low abundance miRNA species were excluded from further analysis. The variation in sequence reads for the remaining 199 miRNA species is shown in Table 7. When samples from all three developmental stages were compared (Supplemental Table S8), nearly half (94 or 47.23%) of the 199 miRNA species exhibited >8 fold variation in the number of sequence reads among samples from the three developmental stages, compared with only five (or 2.51%) miRNA species that did not show any significant change in sequence reads among these samples. The remaining 100 miRNA species exhibited 1.5 to 8 fold variation among the samples from the three developmental stages. When comparison was made between samples from two different developmental stages, the Larva-Adult comparison resulted in the greatest variation (Supplemental Table S9), with 81 (or 40.7%) miRNA species exhibiting more than eight fold variation, compared with 24 (or 12.06%) miRNA species that did not show significant variation. The Larva-Pupa comparison resulted in the smallest variation (Supplemental Table S10), with 56 (or 28.14%) miRNA species that did not show

significant changes. The Pupa-Adult comparison revealed that 23 miRNA species showed no significant variation while 77 miRNA species exhibited more than 8-fold variation (Supplemental Table S11).

Identification of putative target genes regulated by specific miRNA species

Based on current knowledge, gene regulation by miRNAs requires complementary pairing between a target mRNA in the 3'-untranslated region and a miRNA in the conserved seed region. In addition, the expression levels of a target gene and the regulatory miRNA are negatively correlated as a result of the miRNA action. Accordingly, we followed two steps to identify potential target genes that are regulated by specific miRNA species. First, we searched the Hessian fly database containing all predicted transcripts based on the Hessian fly draft genome sequence. Transcripts with at least 70% sequence identity to a specific miRNA were extracted from the database for further analysis. From this step, a total of 10041 transcripts that have matches with one of the 921 Hessian fly miRNA species and located at the 3'-untranslated region were identified. Second, these matched transcripts were subjected to comparison of expression dynamics between miRNA abundance based on sequence reads and mRNA levels based on RNAseq data from another project. Through data analysis, 1516 unique target transcripts were found to have negative correlation between the abundance of a miRNA species and its putative target mRNA, and therefore, were taken as potential target genes. These 1516 putative target transcripts were obtained from 490 miRNA species, while the remaining miRNA species had no qualified target transcripts identified (Supplemental Table S12). The negative correlations were categorized into nine types based on miRNA expression pattern in larvae, pupae and adults (Table 8, Figure 3). The first type is that

miRNA species were exclusively detected in larvae, which in turn, the transcript levels of the candidate target genes were lowest in larvae (Figure 3-1a, 1b). Forty-eight target genes for 30 miRNA species were classified into this category, with an average of two target genes per miRNA. The second type is that miRNA species were only detected in pupae, whereas the transcript levels of the candidate target genes were lowest in pupae (Figure 3-2a, 2b). Forty-four target genes for 30 miRNA species were classified into this group, with an average of 1.5 targets per miRNA. The third type is that miRNA species were only detected in adults, whereas the transcript levels of putative target genes were lowest in adults (Figure 3-3a, 3b). This category contained 1194 target genes for 315 miRNA species with an average of four targets per miRNA. The fourth category is that miRNA species were highly expressed in larvae, but have much lower expression in pupae, while the transcript levels of the target genes were highest in pupae but lowest in larvae (Figure 3-4a, 4b). Thirteen target genes for 9 miRNA species were classified into this group with an average of 1.5 target genes per miRNA. The fifth category is miRNA species were highly expressed in pupae while lowest in adults, whereas the expression levels of target genes were highest in adults and lowest in pupae (Figure 3-5a, 5b). Ten target genes for nine miRNA species belonged to this group with an average of one target per miRNA. miRNA species in the sixth category had the highest expression in pupae and lowest expression in larvae, while in turn, their target genes had highest expression in larvae and lowest expression in pupae (Figure 3-6a, 6b). Twenty-one putative genes for fourteen miRNA species were in this category with an average of two target genes per miRNA. The seventh category is that the expressions of miRNA species were highest in adults and lowest in pupae, whereas the corresponding target genes were highly

expressed in pupae but much lower in adults (Figure 3-7a, 7b). Nineteen target genes for fifteen miRNA species were in this category with an average of one target per miRNA. The eighth category of miRNA species had a decreasing expression tendency among larvae, pupae, and adults, while their target genes had an ascending tendency (Figure 3-8a, 8b). This category contained 104 target genes for 39 miRNA species with an average of three targets per miRNA. The ninth category is miRNA species with a decreasing expression pattern among the three developmental stages, while the putative targets have an increasing tendency (Figure 3-9a, 9b). This category had 63 target genes for 29 miRNA species, with an average of two target genes per miRNA, respectively.

Functions of the putative target genes

The functions of the 1516 putative target genes were further annotated based on BLAST searches against the Genbank database. The specific functions for each target gene are given in Supplemental Table S13-S15 and are summarized in Table 9. The putative target genes were classified into three categories: (1) genes with functions unknown (Supplemental Table S13), including 162 genes with functions not well defined, 380 genes encoding hypothetical proteins, and 217 genes encoding unique protein for Hessian fly but no sequence similarity with sequences in the Genbank; (2) genes encoding Secreted Salivary Gland Proteins (SSGPs) in Hessian fly, which function as putative effectors (Supplemental Table S14); and (3) genes encoding proteins with known functions (Supplemental Table S15). The putative target genes with known functions were further divided into seven groups, including 208 genes encoding regulatory proteins, 118 genes encoding transporters, 104 genes encoding proteins with functions in nutrient metabolism, 98 genes encoding structural and adhesion proteins, 95

genes encoding proteins involve in protein metabolism, 41 genes encoding proteins with functions in RNA metabolism, and 24 genes encoding proteins with functions in redox and detoxification, respectively.

Expression of selected isoforms of the miRNA family PC-5p-67443 in different developmental stages of Hessian fly

A previous study of first instar Hessian fly miRNA species found that some miRNA gene families were dramatically expanded. The largest miRNA gene family among the expanded genes is PC-5p-67443, which has 91 different encoding genes. The genes encoding different isoforms in the PC-5p-67443 family were named as PC-5p-67443-1 to PC-5p-67443-91, respectively. Among these genes, 65 encode an identical miRNA isoform, whereas the other 26 genes encode miRNA isoforms with at least one residue difference (Figure 4). Among the 26 miRNA isoforms, 17 have one nucleotide difference, five have two nucleotides difference, and four have three nucleotides difference.

Based on the miRNA sequence alignment, three sets of primers specific to three coding genes were designed, with each targeting miR-PC-5p-67443-72, miR-PC-5p-67443-74, or miR-PC-5p-67443-80 (Figure 5B). Through Real-Time PCR with the gene-specific primers, the expression patterns of the three iso-genes were determined with samples from six Hessian fly developmental stages including one-day-old, three-day-old, five-day-old and seven-day-old larvae, pupae, and adults. The abundance of the three iso-miRNAs was very different as reflected in the differences in the coordinate scales (Figure 6). The scale in the graph for miR-PC-5p-67443-72 is 0 to 20, the highest compared to the scale 0-3.5 for miR-PC-5p-67443-74 and the scale 0-2.5 for miR-PC-5p-

67443-80. In addition to the difference in the overall levels of expression, the patterns of expression were also different for the three genes among the larval stages examined. For miR-PC-5p-67443-72, the expression level was the highest in five-day-old larvae, followed by seven-day-old larvae. The expression levels in one-day-old and three-day-old larvae were the lowest. For miR-PC-5p-67443-74, the expression level was again the highest in five-day-old larvae, but followed by one-day-old and then seven-day-old larvae. For miR-PC-5p-67443-80, however, the gene expression level was the highest in one-day-old larvae, followed by three-day-old larvae, and then five-day-old larvae; and the seven-day-old larvae exhibited the lowest expression level. Despite the different expression patterns associated with different genes in different developmental stages, all three genes exhibited similar expression patterns in pupae and adults, with the highest expression level in the pupal stage and the lowest expression in the adult stage.

Expression of selected pre-miRNA isoforms of the miRNA family PC-5p-67443 in different Hessian fly developmental stages

To determine if the levels of pre-miRNAs from different genes were different, we further characterized the expression patterns of pre-miRNA species encoded by different genes. Pre-miRNAs are produced in the nucleus and transported to cytoplasm, where pre-miRNAs are processed to form mature miRNAs (Figure 7). As mentioned previously, even though the 91 miR-PC-5p-67443 genes encode nearly identical mature miRNA, each of the 91 genes encodes a different pre-miRNA. Thus, the expression levels of different pre-miRNAs derived from different genes would reveal the relative contribution to the mature miRNA from different genes. Primers specific for pre-miRNA species derived from the coding genes pre-mir-21 and pre-mir-59, as well as their corresponding

mature miRNA coding genes miR-PC-5p-67443-21 and miR-PC-5p-67443-59 were designed and synthesized based on nucleotide differences among sequences (Figure 8).

Using these specific primers, the expression patterns of the mature miRNAs, miR-PC-5p-67443-21 and miR-PC-5p-67443-59, as well as the corresponding pre-miRNAs, pre-mir-21 and pre-mir-59, were analyzed through Real Time-PCR (Figure 9). The expression patterns of the mature miRNAs miR-PC-5p-67443-21 and miR-PC-5p-67443-59 were quite different from each other (Figure 9a). Firstly, the overall RNA level of miR-PC-5p-67443-21 was much higher than that of miR-PC-5p-67443-59 based on the big difference in the ordinate scale of the figure. Secondly, even though both miR-PC-5p-67443-21 and miR-PC-5p-67443-59 exhibited the highest levels of expression in the pupal stage, the relative levels of miR-PC-5p-67443-21 in other developmental stages, including one-day-old to seven-day-old larvae and adults, were 8-fold lower than that observed in the pupal stage, whereas the relative expression levels for miRNA miR-PC-5p-67443-59 were only 2-fold lower in those respective stages compared with the pupal stage.

The expression patterns for the pre-miRNAs pre-mir-21 and pre-mir-59 were surprisingly different, both relative to each other and with respect to their corresponding mature miRNA counterparts. The overall expression levels of pre-mir-59 were 25-fold to more than 200-fold greater than the expression levels of pre-mir-21 in most developmental stages (Figure 9b). The expression level of the pre-mir-21 in one-day larvae was the highest, in contrast to the relatively low level of the corresponding mature miRNA miR-PC-5p-67443-21 in the same developmental stage of Hessian fly; and this was also in contrast to the relatively low level of the pre-miRNA pre-mir-59 in the same

insect sample. The expression levels of the pre-miRNA pre-mir-59 varied widely from different developmental stages, with the highest level in five-day-old larvae, followed by three-day-old larvae and adults. Much lower expression levels were detected in seven-day-old larvae and pupae, and the expression level of pre-mir-59 was barely detectable in one-day-old larvae.

Discussion

miRNA species in Hessian fly

In this study, a total of 921 unique miRNA species were identified from three developmental stages, including seven-day-old larvae, pupae, and adults of Hessian fly. A previous study identified 273 miRNA species from Hessian fly first instar larvae. Together, 1121 unique miRNA species have been identified through deep-sequencing. Among the 1121 miRNA species, 768 are identical or very similar to miRNA species reported from other insect species. The remaining 353 miRNA species are unique to Hessian fly and have not been reported from any other insect species. From literature, one of the most extensively studied insect species in terms of the number of identified miRNA species to date is *Laodephax striatellus*, with 501 miRNA species identified from first to fifth instar female and male nymphs (Zhou et al., 2014). Among the 501 miRNA species, 274 are novel miRNA species and 227 are conserved miRNA species among Arthropoda species. Of the 227 conserved miRNA species, 38 were highly conserved among insect species. Another extensively studied insect species on miRNA is *B. mori*. A total of 563 mature miRNA species from *B. mori* have been published in miRBase. The identification of 1121 miRNA species therefore places the Hessian fly on the top of most extensively identified miRNA species among insects. This should provide a solid

foundation for studying the functions of miRNA species in Hessian fly biology, and for comparative research among insect species.

The dominant form of Hessian fly miRNA species

As shown in Figure 1A, the dominant forms of total mappable reads were 29 – 31 nucleotides long. Further analysis revealed that sequences with 29 – 31 nucleotides were mainly derived from the 5.8S ribosomal RNA. For example, the most abundant read within the 30-nucleotide group, TGTATAGACCACACATGGTTGAGGGTTGTA, was derived from 5.8S ribosomal RNA. This result is expected since the 5.8S ribosomal RNA is the most abundant small RNA in eukaryotic cells. After removing sequence reads from small rRNAs, snRNAs, snoRNAs, and tRNAs, the distribution of miRNA reads was quite different from total reads (Figure 1B). The predominant forms of miRNA reads were 21 – 24 nucleotides.

Size distribution of miRNA species identified from *Locusta migratoria* is from 11 to 30 nucleotides, with 22 and 28 nucleotide length as the dominant forms, depending on different locust biotypes (Wei et al., 2009). The major miRNA species from *Blattella germanica* also have 21 to 23 nucleotides, with the 22 nucleotide reads nearly four-fold more than the other two forms (Cristino et al., 2011). Previous Hessian fly miRNA study indicated that the dominant miRNA species identified from first instar larvae are with 26 nucleotides (Khajuria et al., 2013), which is somewhat different from other insects, suggesting that the abundances of different miRNA species vary greatly among different insect species and between the first instar larvae and other developmental stages of Hessian fly.

This is not surprising considering the fact that the biology of first instar Hessian fly larvae are quite different from other types of insect species and from other developmental stages of Hessian fly. The first instar larvae express the greatest proportion of unique genes that are required for manipulation of host plants (Chen et al., 2008; Chen et al., 2010; Zhao et al., 2015). The unique physiology in first instar Hessian fly larvae may need expression of a unique set of miRNA species. Further research is needed to reveal the biochemical basis for the observed difference in the abundance of different miRNA species among different developmental stages of this insect.

Another interesting characteristic of Hessian fly miRNA is the disproportion between the number of miRNA species and their relative abundance. Though unique reads with 21 and 22 nucleotides are the most abundant, each with 22.25% and 31.71% of total unique reads, respectively. The miRNA species with 21 and 22 nucleotides were only 7.49% and 11.07% of total miRNA species, suggesting that a few miRNA species were highly abundant. On the other hand, miRNA species with 15 nucleotides represented 27.47 % of total miRNA species, but only 2.48% of total reads, suggesting that each miRNA species had much fewer reads on average than the 21 and 22 miRNA group. It is understandable that the 21 – 22 nucleotide group was the most abundant group since they are the major forms in other organisms as well (Cristino et al., 2011). However, the large number and low abundance of miRNA species with 15 nucleotides, to a lower degree of the 16 and 17 nucleotides as well, observed in Hessian fly was somehow unique to Hessian fly. The origins and biological significance of the short and low abundance miRNA species remain to be determined. One possibility for the origin of short miRNA species was that the short miRNA species might be the degradation

products of longer miRNA species. However, sequence alignment analysis of the short miRNA species with longer miRNA species revealed that their sequences are different, even though some of the short miRNA sequences share certain sequence similarities with the longer miRNA sequences. For example miRNA dwi-mir-308-p5_1ss8TC (15 nucleotides) and dme-mir-4945-p5_1ss3AG (21 nucleotides) have 45.8% similarity, miRNA tca-miR-927a-3p_L-1_1ss14TC (15 nucleotides) share nearly 60% sequence similarity with dme-miR-927-3p_L-1_1ss21AG (21 nucleotides) and dme-miR-927-3p_1ss21AG (22 nucleotides). The differences in miRNA sequences indicate that the miRNA species with different sizes were produced for specific functions, not due to degradation of longer miRNA species.

Expansion of Hessian fly miRNA families

One unique characteristic of Hessian fly miRNA species is that some miRNA families are dramatically expanded in the Hessian fly genome. This phenomenon was initially observed in several miRNA families from the first instar larvae (Khajuria el al., 2013). In this study, a more systematic analysis of Hessian fly miRNA species was carried out with samples from multiple developmental stages of Hessian fly. Among the 921 miRNA species, 232 were found to have multiple coding genes. The three largest miRNA families are PC-5p-107087_9, mde-mir-2505-p3, and mde-mir-7-p5_1ss2TG, which contained 14, 23, 34 putative coding genes, respectively (Table 5). The miRNA encoding sequences of these three miRNA families, in each case, are identical, but the 5' and 3' end flanking sequences are more diversified, with 33% to 77% sequence similarity, higher than sequence similarity among unrelated sequences (Figure 2). This suggests that the multiple genes in each miRNA family are likely functional and each isogene may

have specific regulatory significance. It is interesting to note that all expanded families are unique to Hessian fly. The 22 miRNA species that are also conserved in other insect species are all single gene for each miRNA except for the miRNA aga-mir-124-p5, which have two coding genes that have 88% sequence similarity at miRNA encoding region and 32% to 50% sequence similarity at 5' end and 3'end flanking region, respectively. This observation supports the notion that the expanded miRNA families play specific roles in biological processes that are unique to Hessian fly.

Multiple genes for a miRNA family have been reported from other species. Previous miRNA studies indicated that miRNA families with larger size usually have relatively low abundance in plant species while those with smaller family size have higher abundance in animal species (Li et al., 2007). The largest miRNA family identified previously is the miRNA family miR395 in rice, which contains 24 perfectly matched coding genes (GUDDETI, 2005). The miR395 family has also been identified in the maize genome, which contains 16 different coding genes (Zhang et al., 2009). For insect species, the largest miRNA family is *B. mori* miRNA family bmo-miR-466e, which has 16 isogenes (Yu et al., 2008). Another study of *T. castaneum* miRNA identified only eight miRNA families with three or more isogenes and the largest families mir-2 and mir-467 both have six isogenes, respectively (Luo et al., 2008). Compared with the largest miRNA family PC-5p-67443 identified from Hessian fly first instar larvae, which contains 91 isogenes, as well as the largest miRNA family mde-mir-7-p5_1ss2TG identified from Hessian fly seven-day-old larvae, pupae, and adults, apparently the Hessian fly genome contains more expanded families and the largest number of isogenes in the expanded families.

Developmental regulation of Hessian fly miRNA species

The numbers of sequence reads for many different miRNA species varied greatly at the same developmental stage, and the numbers of sequence reads for the same miRNA species varied significantly among different developmental stages. Assuming that the numbers of sequence reads were proportional to the abundance of miRNA species, these observations suggest that the expression of different miRNA species varied among different developmental stages of the insect. Some miRNA species were only detected in a particular stage. For example miRNA mde-mir-375-2-p5 and PC-5p-92404_11 were only found in the seven-day-old larval stage; PC-3p-91476_11 was only found in the pupal stage; and mde-miR-193-3p was only found in the adult stage. Some miRNA species were found in two of the developmental stages. For example, miRNA mde-mir-3874-p5_1ss11TG was detected in larvae and pupae but not from adults; PC-5p-12029_129 was detected in pupae and adults but absent in larvae. miRNA mde-mir-2791-p5_1ss7TG were detected in larvae and adults but absent in pupae. These miRNA species that are exclusively expressed in a single developmental stage of the insect are nearly 30 % of the total analyzed miRNA species. These stage-specific miRNAs may be critical for developmental transition from one stage to the next, and therefore could be targets, for the development of novel pest control measures. Even for those miRNA species detected from all stages, their abundance in terms of reads varies from two-fold to more than eight-fold (Table 8, Supplemental Table S7). Only five miRNA species, mde-miR-965-3p_R-1_1ss9TG, mde-miR-317-3p_1ss21-T, mde-miR-993-5p_R-1, mde-miR-252-5p_R+1, and mde-miR-2779_L-3R-2_1ss9CT, exhibited no significant differences

in abundance among all three developmental stages of the insect. These results indicate that each developmental stage of Hessian fly expresses a unique set of miRNA species.

Developmental regulation of miRNA species was also found in many other insect species. For example, in *D. melanogaster*, miR-3 to miR-7 are exclusively expressed during embryogenesis, while miR-1, miR-2, and miR-8 to miR-13 are expressed at all developmental stages with certain levels of abundance variation. miR-1, miR-8 and miR-12 are highly expressed during different larval stages, whereas the expression levels of miR-9 and miR-11 decrease dramatically during the adult stage (Lagos-Quintana, 2001). In *C. elegans*, the first identified miRNA species *lin-4* is closely regulated with the timing of developmental transitions (Lee et al., 1993). All of these reports suggest that a great number of miRNA species are developmentally regulated in a wide range of different organisms.

In addition to developmental regulation, many miRNA species are also tissue-specific. Again in *D.melanogaster*, miR-1 and miR-8 to miR-11 are detected from different developmental stages, but have shown no expression in S2 cell, which derives from late stage embryos. On the other hand, miR-3 to miR-6, though highly expressed during embryogenesis, exhibited no expression in S2 cell (Lagos-Quintana, 2001). For Hessian fly, some of the miRNA species identified from current study may be expressed only in certain tissues, such as salivary glands, fat body, mid gut, and Malpighian tubes. Considering the salivary glands are the critical tissue for Hessian fly first instar larvae to attack plants (Chen et al., 2008). Tissue-specific expression of the identified miRNA species is the next priority for the next phase of study for this project.

Putative target genes of different miRNA species

The availability of the Hessian fly genome sequence and RNAseq data with samples from different developmental stages allowed us to identify protein-coding genes that are likely to be targets regulated by a specific miRNA species. Based on the knowledge that miRNA forms complementary pairing at the 3'-untranslated region of its target mRNA and that the abundance of miRNA is negatively related to its target mRNA, a total of 1516 putative target genes have been identified. Identification of these putative target genes provides a base for future studies to examine the functions of specific miRNA species.

An interesting fact revealed from the target identification is that since some of the Hessian fly miRNA families are dramatically expanded, we would expect that those miRNA families might have multiple targets that perform different regulatory function. However, our results suggested a different pattern. The most expanded miRNA family identified was mde-mir-7-p5_1ss2TG, which had 34 different coding genes, but only one qualified target gene identified, which functions as a transporter. Another significantly expanded miRNA family, PC-5p-107087_9, contained 14 different coding genes and also had one target gene identified, which functions as a regulatory protein in signal transduction. In contrast, the second largest miRNA family, mde-mir-2505-p3, had no qualified target gene identified. Interestingly, some of the single-gene miRNA families have multiple target genes. For example, miRNA mde-miR-9b-5p_R-1_1ss10AG, had three different target genes, each with different functions. Thus, the expansion of miRNA families does not necessarily mean expanded target genes.

The functions of miRNA target genes embrace a wide range of physiological processes, including regulation of gene expression, transport, and protein metabolism. Except for those target genes with known functions, the functions for the majority of miRNA target genes remain to be determined. The genome-wide search in this project is the first comprehensive analysis of putative target genes for Hessian fly miRNA species identified from different developmental stages, and the database generated from this project would be an important resource for further study of Hessian fly miRNA interaction with targets, and should lead to elucidating the regulation mechanism of Hessian fly miRNA species.

miRNA iso-genes in an expanded family are differentially regulated

The dramatic expansion of some miRNA families is a unique phenomenon observed in the Hessian fly genome. To examine the potential functional significance of each isogene in an expanded family, real-time PCR was carried out to determine the expression patterns of different isogenes in different developmental stages. Thus, it has been determined that different isogenes have unique expression patterns among different developmental stages (Figure 6). Considering the fact that some iso-miRNA species has one or two residues different among them (Figure 5A), the different expression pattern among these isogenes could suggest that different genes have different target genes and therefore perform different functions. Since the miR-PC-5p-67443 family has 91 isogenes and only three of them were examined in this research, further analysis on the other isogenes may discover more useful information on the biological significance of this expanded miRNA family.

Differential expression of pre-miRNAs the same family

Pre-miRNA is an intermediate during miRNA biogenesis and how pre-miRNAs are processed by the Dicer complex determines the characteristics of mature miRNA (Huang et al., 2011). Previous research on pre-miRNA and mature miRNA suggested a positive correlation between the expression of pre-miRNAs and their corresponding mature miRNAs, but the regulation mechanism of pre-miRNA and mature miRNA are different (Gan et al., 2013; Schmittgen et al., 2004).

In this study, the expression levels of isogenes miR-PC-5p-67443-21 and miR-PC-5p-67443-59 were found to be lowest in the adult stage and first instar larval stage but highest in the pupal stage (Figure 9a), consistent with the expression pattern of miR-PC-5p-67443-72 and miR-PC-5p-67443-74 (Figure 6). This provides further evidence for the differential regulatory functions of different isogenes through pairing with different target genes. However, the expression levels of pre-mir-21 and pre-mir-59 were not proportional to their corresponding mature miRNA species miR-PC-5p-67443-21 and miR-PC-5p-67443-59. In fact, the pre-miRNA and mature miRNA exhibited totally different expression patterns. The highest expression level of pre-mir-21 was in one-day-old larvae, followed by a dramatic decrease during late stages. In comparison, the highest expression level of miR-PC-67443-21 was in pupae while the lowest level was in one-day-old larvae. The highest expression level of pre-mir-59 was in five-day-old larvae with much lower level in pupae, compared with the highest expression level of the mature miRNA miR-PC-5p-67443-59 in pupae with lower level in five-day larvae. The different expression patterns between pre-miRNAs and the corresponding mature

miRNAs are quite interesting and the biological significance of this phenomenon remains to be determined.

Conclusion

This project is the first systematic and comprehensive study of Hessian fly miRNA species. We identified 921 unique miRNA species from three Hessian fly developmental stages: second instar larvae, pupae and adults. The 921 miRNA species represent the largest number of miRNA species so far identified among any insect species. The most abundant forms of Hessian fly miRNA species are 21 to 23 nucleotides long. The abundance of miRNA species varied greatly among different developmental stages, suggesting that each developmental stage of the insect may have a unique set of miRNA expressed and thus perform specific functions. Some miRNA species were found to have up to 34 different genes encoding an identified miRNA, which is consistent with earlier findings that many miRNA families are dramatically expanded. A genome-wide search was conducted and 1516 putative target genes for 490 Hessian fly miRNA species were identified. The results suggested that each miRNA species might have several different target genes with different functions. The number of target genes for a specific miRNA family, however, is not proportional to the number of isogenes in each family. Single-gene families can have multiple target genes whereas a family with multiple isogenes may have only one target gene. The expression profiling of the largest Hessian fly miRNA family PC-5p-67443 suggested that the pre-miRNA and mature miRNA encoded by isogenes within an expanded miRNA family have different expression patterns among different developmental stages of Hessian fly.

This identification of a large number of miRNA species through a systematic analysis provides a good foundation for future analysis of miRNA function in Hessian fly. The initial estimation of expression patterns of miRNA species in different developmental stages provides a rough reference to identify candidate miRNA with regulatory roles in different Hessian fly developmental stages. The identification of candidate target genes for specific miRNA species also provides useful information for future characterization of miRNA functions. Finally, the methodology established and initial results obtained on determining the differential expression of isoforms in expanded miRNA families should be very useful for further analysis on functions of those expanded miRNA families.

Figures and Tables

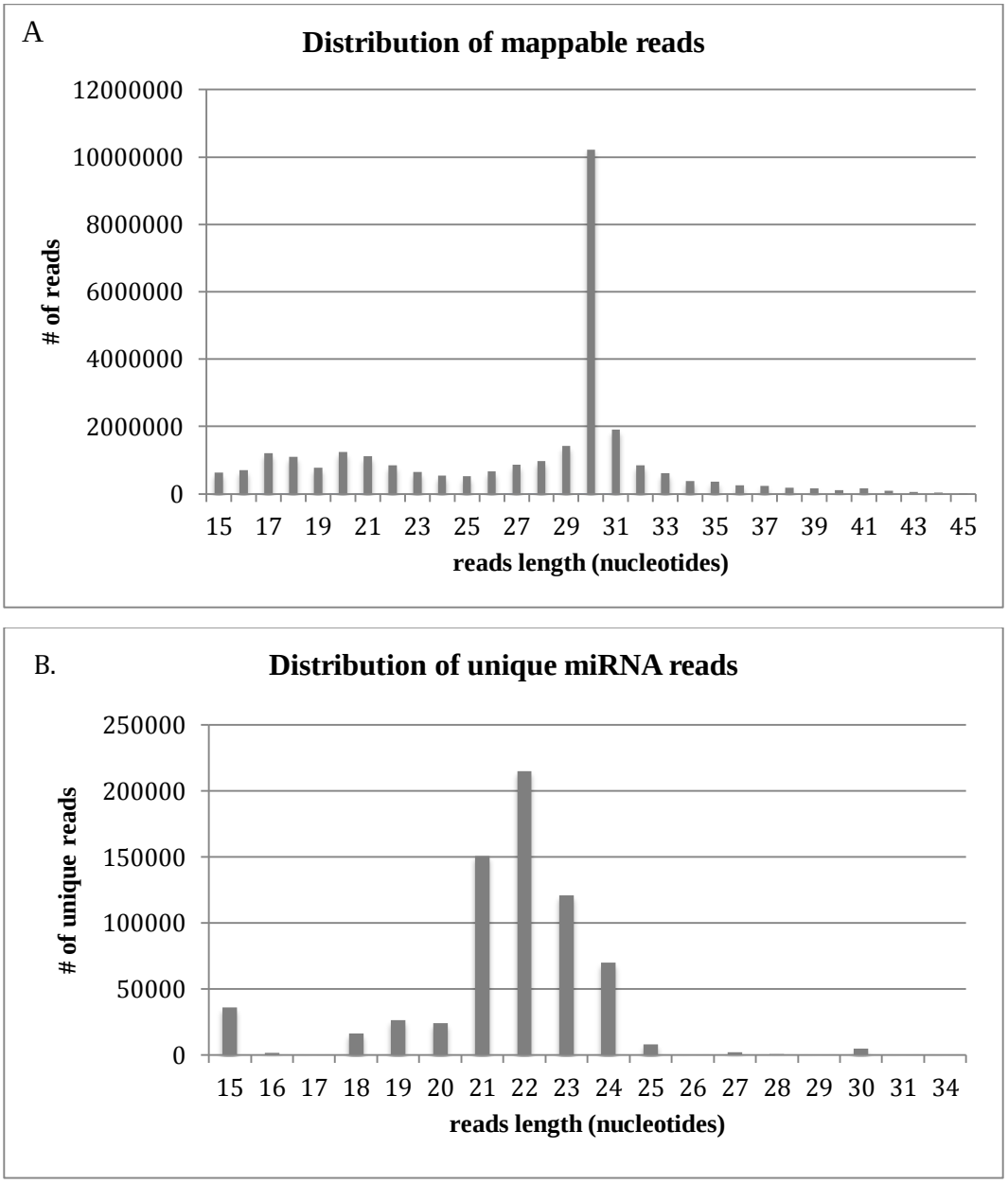


Figure 1. Size distribution of miRNA reads identified by deep sequencing of samples obtained from seven-day larvae, pupae and adults. A. Size distribution of mappable reads. B. Size distribution of unique miRNA reads.

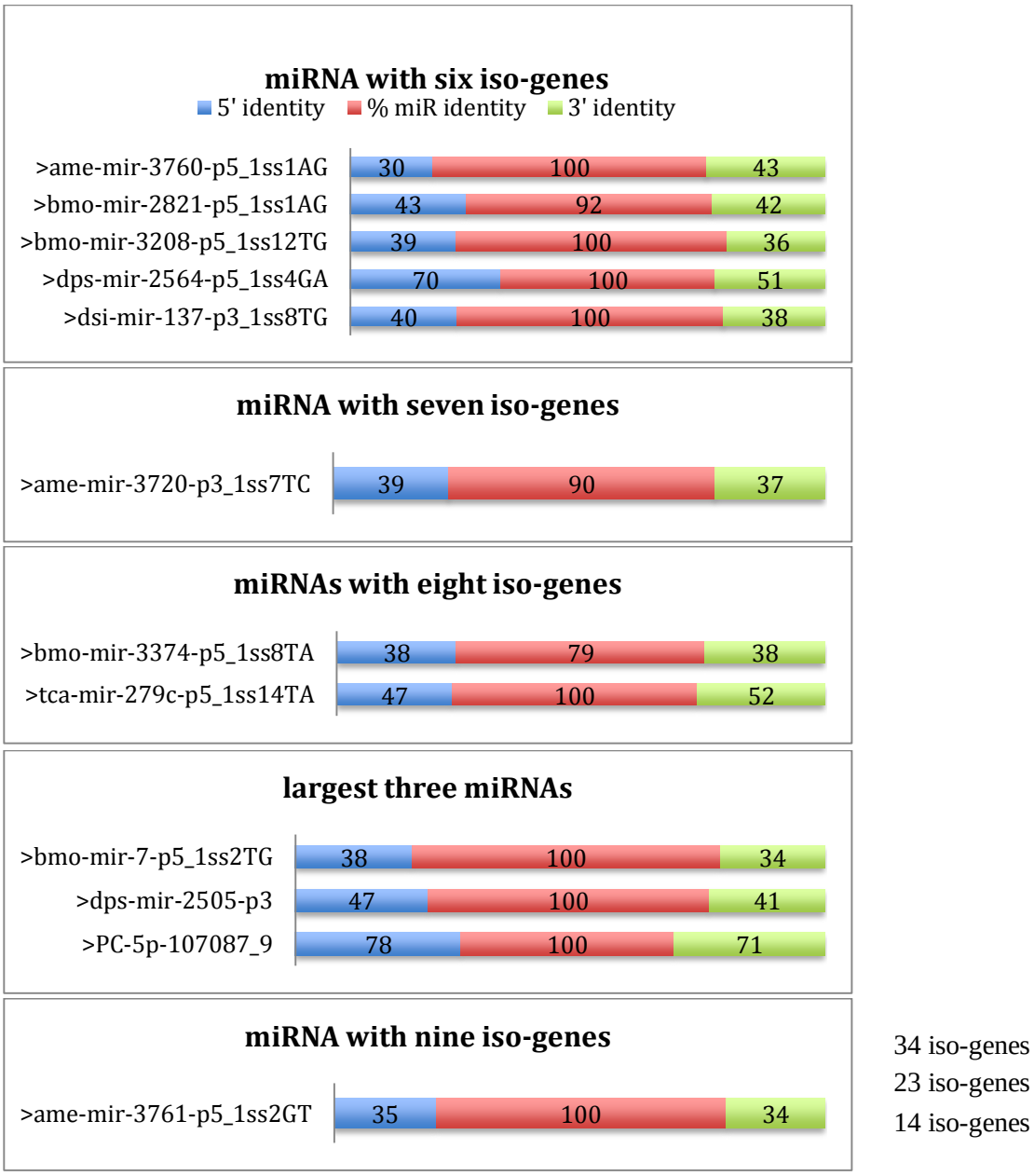
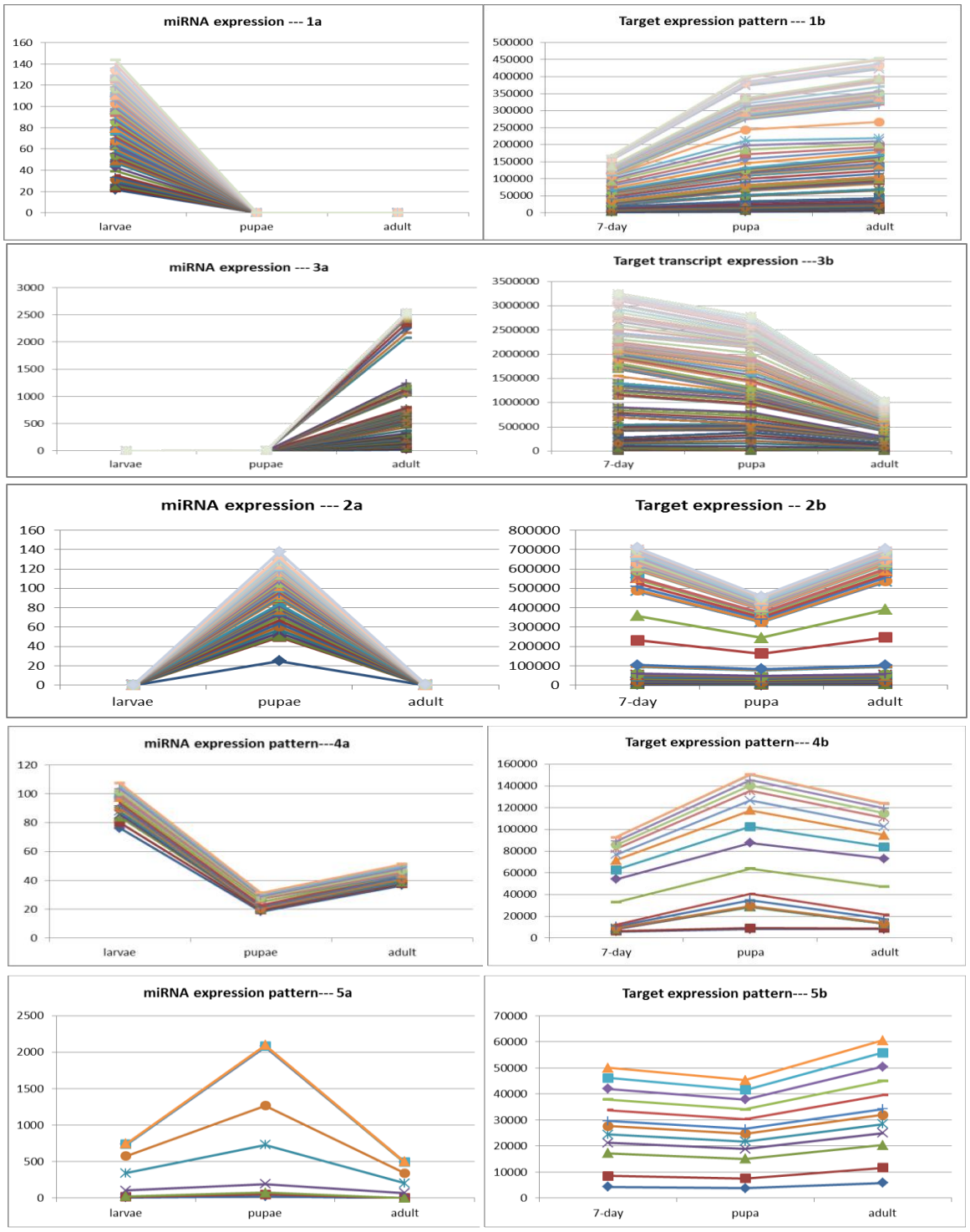


Figure 2. **Sequence alignment of isogenes in the expanded miRNA families.** The alignments was divided to the 5' flanking region (blue) and 3' flanking region (green), both are 65~81 nucleotides long, as well as mature miRNA coding region (red), which is 15 to 22 nucleotides long. Numbers on each bar indicate the sequence identity among all the coding genes in a miRNA family. (Supplemental Table S3)


miRNA expression --- 1a
160
140
120
100
80
60
40
20
0
larvae
pupae
adult
Target expression pattern --- 1b
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
7-day
pupa
adult
miRNA expression --- 3a
3000
2500
2000
1500
1000
500
0
larvae
pupae
adult
Target transcript expression ---3b
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
7-day
pupa
adult
miRNA expression --- 2a
160
140
120
100
80
60
40
20
0
larvae
pupae
adult
Target expression -- 2b
800000
700000
600000
500000
400000
300000
200000
100000
0
7-day
pupa
adult
miRNA expression pattern---4a
120
100
80
60
40
20
0
larvae
pupae
adult
Target expression pattern--- 4b
160000
140000
120000
100000
80000
60000
40000
20000
0
7-day
pupa
adult
miRNA expression pattern--- 5a
2500
2000
1500
1000
500
0
larvae
pupae
adult
Target expression pattern--- 5b
70000
60000
50000
40000
30000
20000
10000
0
7-day
pupa
adult

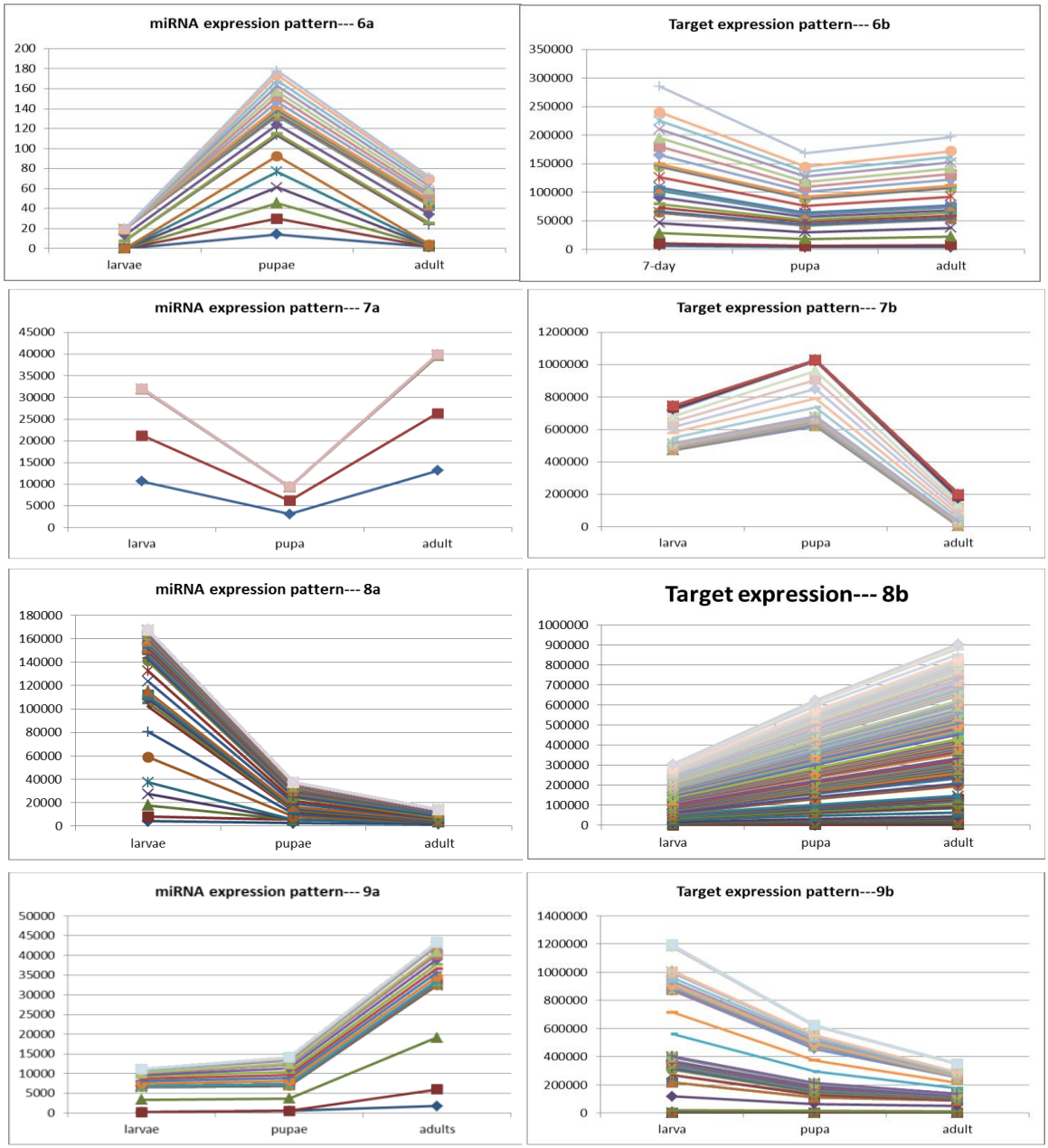


Figure 3. Nine categories of negative correlation between miRNA abundance and the expression levels of their putative target transcripts. a: miRNA abundance in different developmental stages. b: Expression levels of putative target genes. (Supplemental Table S12)

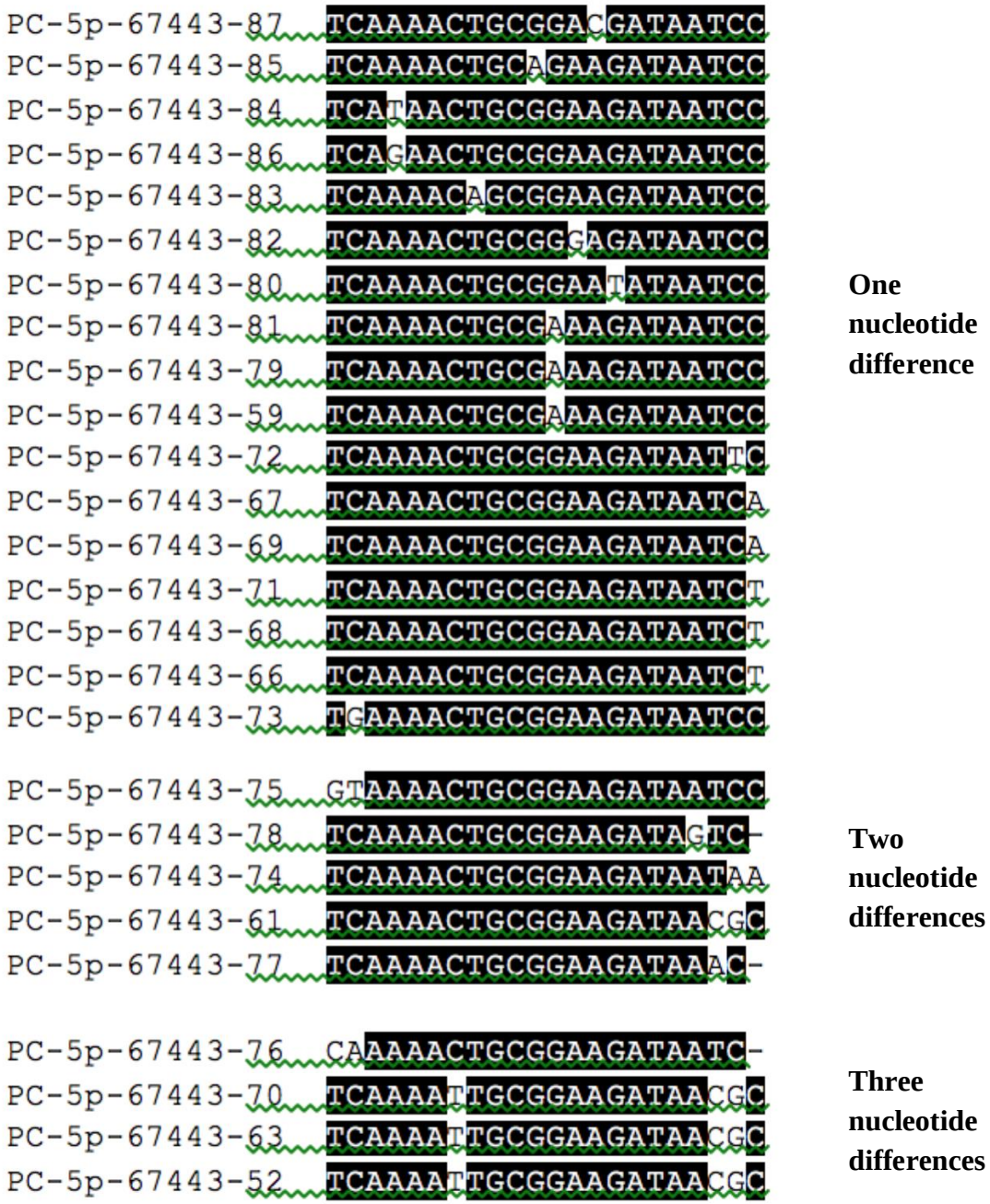


Figure 4. Sequence alignment of the miRNA coding region of multiple genes in the PC-5p-67443 family. Nucleotide residues that differed among different genes were marked by white, and the shadowed regions are conserved sequences. A nucleotide deletion was indicated by the symbol '-'.

A.

```
PC-5p-67443-72      AGAATGAAAATGGTCCTTTTTCAAAACTGCGGAAGATAATTCCAGCATAAAATAAATGC
PC-5p-67443-74      AGAATGAAAATGGTCCTTTTTCAAAACTGCGGAAGATAATAACAGCATAAAATAAATGC
PC-5p-67443-80      AGAATGAAAATGGTTCTTTTTCAAAACTGCGGAATATAATCCCAGCATAAAATAAATGC
                    ************** ****************** *****  *****************
```

B.

miR-PC-5p-67443-72

```
                                  A   C              UCCA    AAA
UGAAUCGCUGAGAAUAAAUAUGGGACUUUUC AAA UGCGGAAGAUAAU    GCAU   \
ACUUAGCGACUCUUAUUUAUACUCUGAAAAG UUU ACGCCUUCUAUUG    CGUA   A
                                  C   A              ----    AAU
```

miR-PC-5p-67443-74

```
                                      A   C              AACA    AAA
UUUCAAUUGAAUGGCACAGAAUGAAAAUGGUCCUUUUC AAA UGCGGAAGAUAAU    GCAU   \
AAAGUUAACUUACCGUGUCUUACUUUUACCAGGAAAAG UUU ACGCCUUCUAUUG    CGUA   A
                                      C   A              ----    AAU
```

miR-PC-5p-67443-80

```
                                    A   C       U      CCCA    AAA
AUAUGAAUGGCACAGAAUGAAAAUGGUUCUUUUC AAA UGCGGAA AUAAU    GCAU   \
UAUACUUACCGUGUCUUACUUUUACCAGGAAAAG UUU ACGCCUU UAUUG    CGUA   A
                                    C   U       C      ----    AAU
```

C.

Encoding gene	Forward primer	Reverse primer
miR-72	AAACTGCGGAAGATAATT	CGCAATTTCGAAAAGTCTCA
miR-74	TCAAAACTGCGGAAGATAATAA	GCAATTTCGAAAAGGACCA
miR-80	TCAAAACTGCGGAAT	GTTATCTTCCGCATTTTCGA

Figure 5. Design of primers specific to three isogenes in the PC-5p-67443 miRNA family. A. Alignment of a fragment of the miR-72, miR-73 and miR-80 encoding genes. The shadow region shows mature miRNA sequences. The nucleotides marked in red are the varied nucleotides. B. Stem-loop structures of the pre-miRNAs of miR-72, miR-73 and miR-80. The sequence marked in red indicates mature miRNA sequence. Underlined sequences indicate the gene-specific primers. Nucleotides in yellow shades indicate the nucleotide variations. C. Specific primers to miRNA-encoding genes.

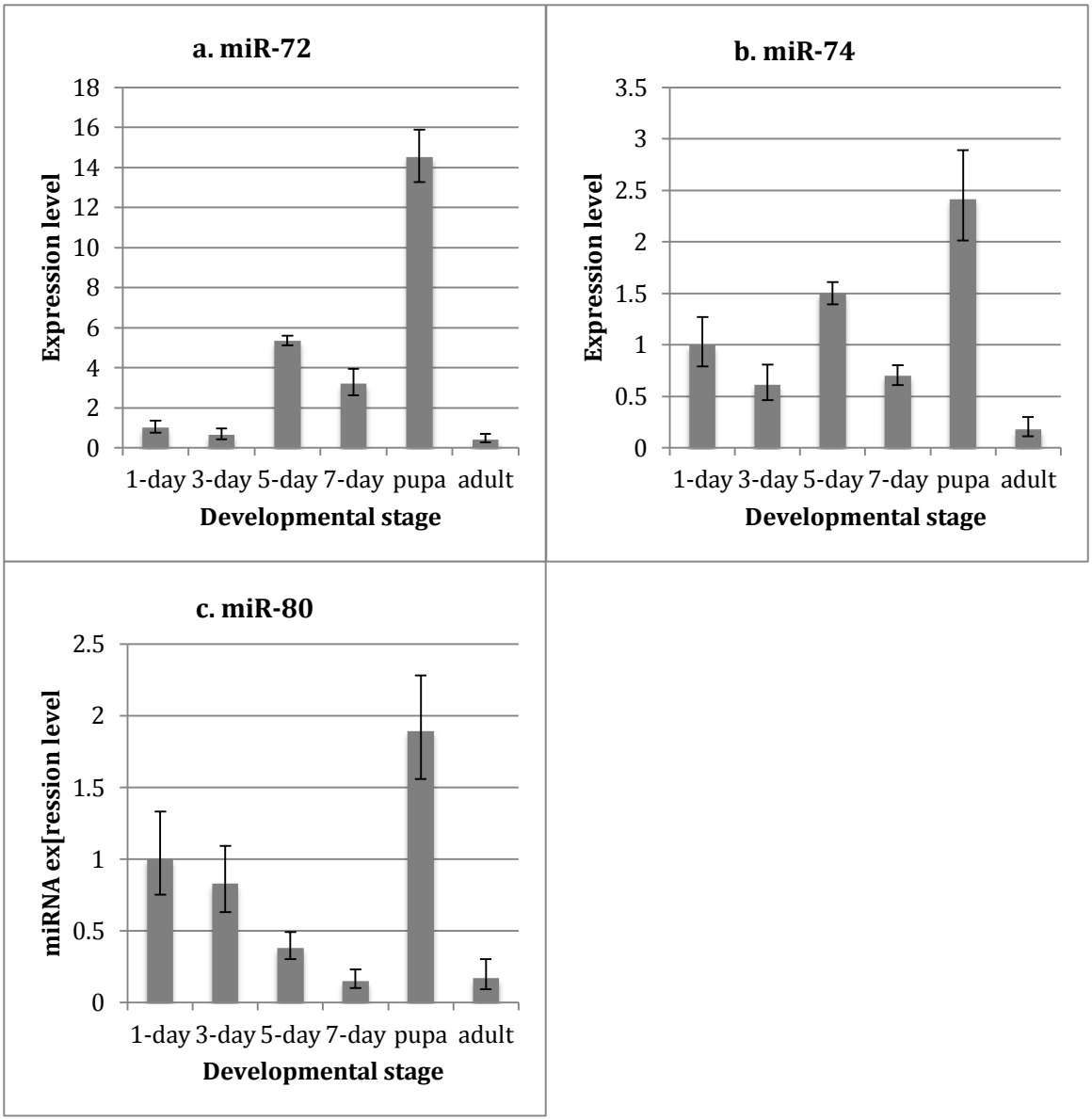


Figure 6. Differential expression patterns among miRNAs miR-72, miR-74 and miR-80 from the PC-5p-67443 miRNA family. The expression patterns of the three miRNA coding genes miR-72, miR-74, and miR-80 were evaluated through real-time PCR. A primer pair specific to U6 small nucleolar RNA was used as an endogenous control. miRNA expressions were evaluated with RNA samples collected from one-day-old, three-day-old, five-day-old, seven-day-old larvae, pupae, and adults.

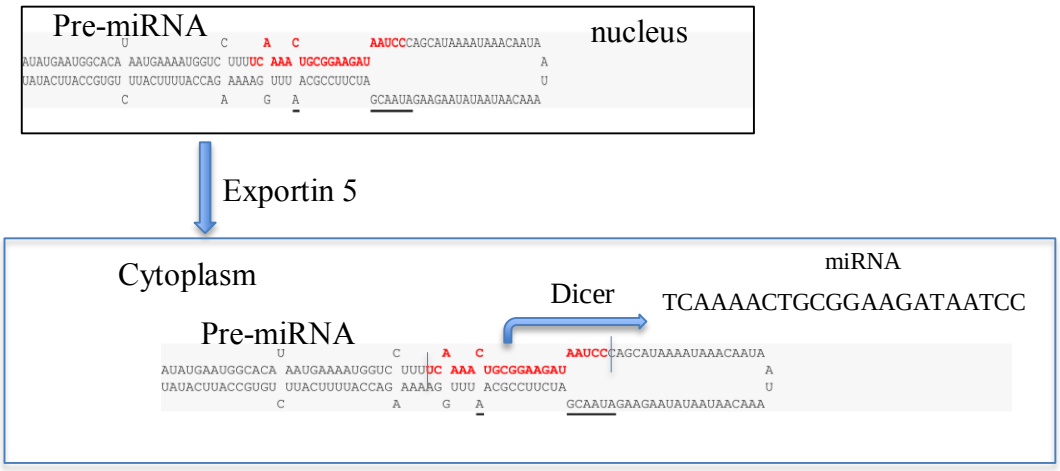
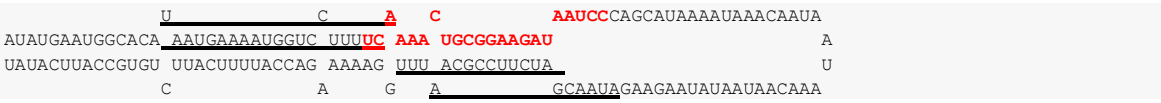
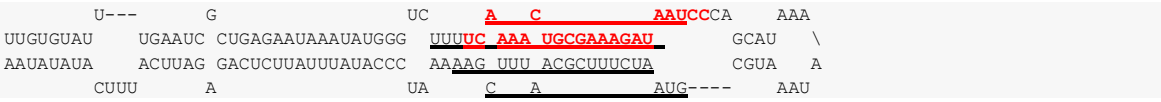


Figure 7. Processing of a pre-miRNA to produce a mature miRNA. A Pre-miRNA is ~140 nucleotides long with a typical stem-loop structure. Pre-miRNAs are produced in nucleus and are transported to cytoplasm by the protein exportin 5. A pre-miRNA is further processed in cytoplasm by dicer, resulting in a mature miRNA.

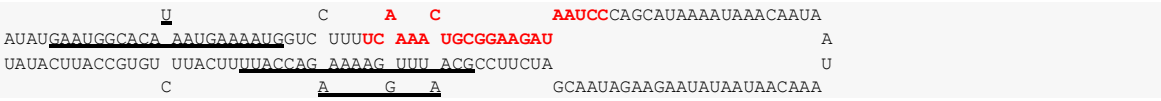
A-1. miR-21



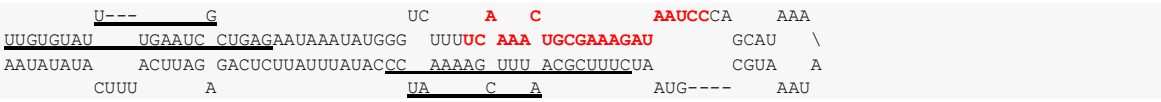
A-2. miR-59



B-1. pre-mir-21



B-2. pre-mir-59

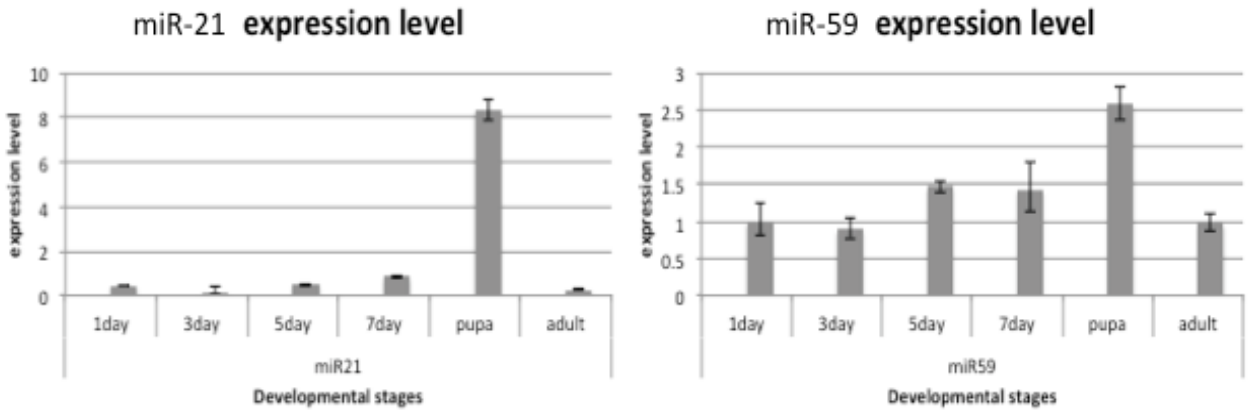


C.

	Encoding genes	Forward primer	Reverse primer
miRNA	miR-21	TAATGAAAATGGTCCTTTCA	ATAACGATCTTCCGCAATTT
	miR-59	TTTTCAAAACTGCGAAAGATAAT	GTAATCTTTCGCAATTTCGAA
Pre-miRNA	Pre-mir-21	GAATGGCACATAATGAAAATG	GCAATTTGGAAAAAGACCATT
	Pre-mir-59	TTGTGTATTTGAATCGCTGAG	CTTTCGCAATTTCGAAAAATCC

Figure 8. Design of primers specific to selected pre-miRNAs and miRNAs. A-1 and A-2 show the miRNA-specific primers for miR-21 and miR-59. B-1 and B-2 show the pre-miRNA primers for mir-21 and mir-59. Sequence in red shows mature miRNA region, underlined sequences indicated the primers. C. Specific primers for miRNAs and pre-miRNAs.

A.



B.

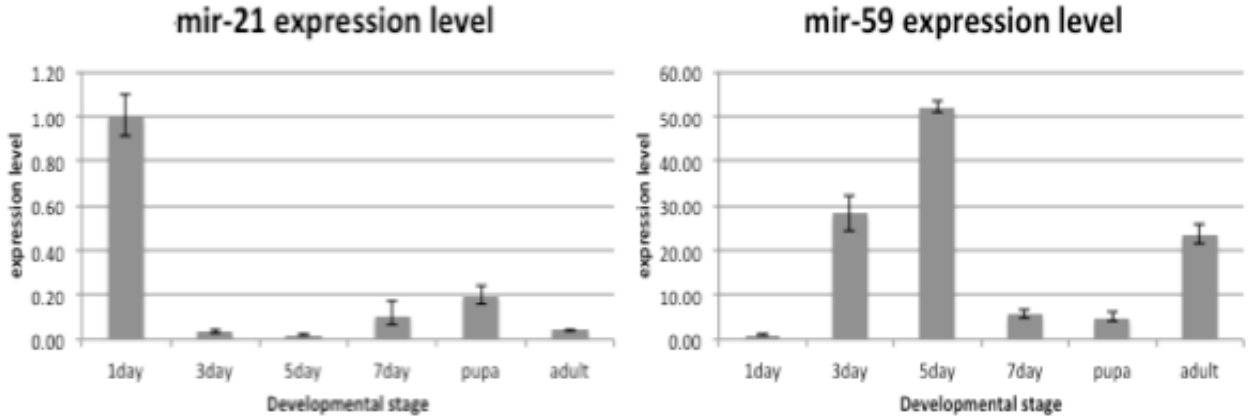


Figure 9. Differential expression patterns of the miRNA and pre-miRNA encoded by isoforms in the PC-5p-67443 miRNA family. A. miR-21 and miR-59 expression pattern. B. mir-21 (pre-miR-21) and mir-59 (pre-miR-59) expression pattern.

Table 1. **Distribution of Hessian fly miRNA sequence reads among samples from seven-day larvae, pupae and adults.**

	Larva	Pupa	Adult	Total
# Raw reads	10,438,827 (100%)	10,625,454 (100%)	13,100,905 (100%)	34,165,186 (100%)
# Reads without 3' adaptor	10,463 (0.1%)	35,223 (0.2%)	20,029 (0.2%)	65,715 (0.2%)
# Reads with length <15	2,322,558 (22.2%)	2,042,528 (19.3%)	789,295 (6.0%)	5,154,381 (15.1%)
# Junk reads	3,549 (0.0%)	5,010 (0.0%)	38,833 (0.3%)	47,392 (0.1%)
#mappable reads	8,102,257 (77.6%)	8,542,693 (80.4%)	12,252,748 (93.5%)	28,897,698 (84.6%)

Table 2. **Size distribution of unique miRNA species and the corresponding reads contained in each category of miNRA species.** (Supplemental Table S1)

miRNA Length	# of unique Reads	Percentage of unique reads (%)	# of miRNA	Percentage of miRNA species (%)
15	36,145.41	5.29	253	27.47
16	1,641.40	0.24	86	9.33
17	349.37	0.04	69	7.49
18	16,499.73	2.43	51	5.53
19	26,543.97	3.92	30	3.26
20	24,015.76	3.62	14	1.52
21	150,822.05	22.25	69	7.49
22	214,929.59	31.71	102	11.07
23	120,786.02	17.82	66	7.17
24	70,209.98	10.36	41	4.45
25	7,921.33	1.17	25	2.71
26	252.34	0.04	35	3.80
27	2,171.27	0.32	28	3.04
28	631.18	0.09	23	2.50
29	131.69	0	18	1.95
30	4,737.20	0.7	9	1.00
31	0.54	0	1	0.11
34	0.5	0	1	0.11
Total	677,789.33	100	921	100.00

Table 3. **Hessian fly miRNA species identical to miRNA species identified from other insects.** (Supplemental Table S2)

Hessian fly seq name	miR_name	miRNA seq	Origins	Reference
mde-miR-281-5p	tca-miR-281	AAGAGAGCTATCCGTCGACAGT	*Tribolium castaneum*	(Nagaraju, 2008)
mde-miR-9a-5p	dme-miR-9a	TCTTTGGTTATCTAGCTGTATGA	*Drosophila melanogaster*	(Lagos-Quintana, 2001)
mde-miR-305-5p_R+1	dme-miR-305	ATTGTACTTCATCAGGTGCTCTGG	*Drosophila melanogaster*	(Lai, 2003)
mde-miR-10_R+1	dan-miR-10	ACCCTGTAGATCCGAATTTGTT	*Drosophila ananassae*	(Clark et al., 2007)
mde-miR-184a	api-miR-184a	TGGACGGAGAACTGATAAGGGC	*Acyrthosiphon pisum*	(Sathyamurthy, 2009)
mde-miR-8-5p	aae-miR-8	CATCTTACCGGGCAGCATTAGA	*Aedes aegypti*	(Li, 2009)
mde-miR-8-3p	aae-miR-8	TAATACTGTCAGGTAAAGATGTC	*Aedes aegypti*	(S. Li et al., 2009)
mde-miR-137-3p_L-1R+1	tca-miR-137	TATTGCTTGAGAATACACGTAG	*Tribolium castaneum*	(Nagaraju, 2008)
mde-miR-210_L+1R-1	dvi-miR-210	CTTGTGCGTGTGACAGCGGCT	*Drosophila virilis*	(Clark et al., 2007)
mde-miR-iab-4-5p	dme-miR-iab	ACGTATACTGAATGTATCCTGA	*Drosophila ananassae*	(Gagen, 2001)
mde-miR-305-3p_R-1_2ss7AG13AG	dme-miR-305	CGGCACGTGTTGGAGTACACTC	*Drosophila melanogaster*	(Lai et al., 2003)
mde-mir-79-p3	cqu-mir-79	TAAAGCTAGATTACCAAAGCAT	*Culex quinquefasciatus*	(Skalsky, 2010)
mde-miR-10-3p_R-1	cqu-miR-10	CAAATTCGGTTCTAGAGAGGTT	*Culex quinquefasciatus*	(Skalsky et al., 2010)
mde-miR-124_R-3	aga-miR-124	TAAGGCACGCGGTGAATGCC	*Anopheles gambiae*	(Sathyamurthy, Ganesh; Swamy, 2009)
mde-miR-252-5p_L+1	aae-miR-252	CTAAGTACTAGTGCCGCAGGAG	*Aedes aegypti*	(Li, 2009)
mde-miR-281-3p_L+1_1ss11GT	tca-miR-281	CTGTCATGGATTTGCTCTCTTT	*Tribolium castaneum*	(Nagaraju, 2008)
mde-miR-210-5p_R+1_2ss10AC20AT	lmi-miR-210	AGCTGCTGGCCACTGCACATGAT	*Locust*	(Wei, 2009)
mde-miR-iab-8-5p_R-3	dme-miR-iab	TTACGTATACTGAAGGTATA	*Drosophila melanogaster*	(Ruby, 2007)
mde-miR-9a-3p_L+1_1ss13TG	dme-miR-9a	ATAAAGCTAGCTGACCGAAGTTA	*Drosophila melanogaster*	(Lagos-Quintana et al., 2001)
mde-miR-79_R-8	cqu-miR-79	GCTTTGGCGCTTTAG	*Culex quinquefasciatus*	(Skalsky et al., 2010)
mde-mir-124-p5	aga-mir-124	ACTGTAGGCCTGTATG	*Anopheles gambiae*	(Sathyamurthy, Ganesh; Swamy, 2009)
mde-miR-252-3p_1ss21GA	aae-miR-252	CTGCTGCCCAAGTGCTTATCA	*Aedes aegypti*	(S. Li et al., 2009)

Table 4. **Distribution of sequence matches from the miRNA BLAST analysis against the Hessian fly draft genome sequence.** (Supplemental Method 1)

Number of Matches	Number of miRNA family
0	3
1	406
2	169
3	115
4	62
5	51
6	26
7	17
8	15
9	9
10	7
11	1
12	6
13	5
14	4
15	2
16	3
19	2
20	1
23	1
25	2
27	1
28	1
32	1
37	1
40	1
48	1
70	1
71	1
82	1
100	3
106	1
108	1

Table 5. **Distribution of putative miRNA-coding genes.** (Supplemental Method 2)

Number of identified Putative gene	Number of miRNA family
0	103
1	594
2	138
3	54
4	13
5	7
6	5
7	1
8	2
9	1
14	1
23	1
34	1

Table 6. **Distribution of miRNA sequence reads in seven-day larvae, pupae and adults.** (Supplemental Table S4-S7)

miRNA category	All stages	Larva	Pupa	Adult
	# of species (percentage)	# of species (percentage)	# of species (percentage)	# of species (percentage)
With reads more than average	39 (4.23)	24 (2.61)	24 (2.61)	27 (2.93)
With reads less than average	160 (17.37)	106 (11.51)	116 (12.60)	133 (14.44)
With reads less than 10	722 (78.39)	791 (85.88)	781 (84.80)	761 (82.63)
Total	921 (100.00)	921 (100.00)	921 (100.00)	921 (100.00)

Table 7. Changes in miRNA abundance among samples from seven-day larvae, pupae, and adults. Abundance variances were calculated based on the miRNA reads distribution among the three stages.

Stage compared	Fold change										
	<1.5	1.5	2	2.5	3	4	5	6	7	8	>8
All three stages	5	6	20	13	20	13	9	5	9	5	94
Larva-Pupa	56	21	35	13	12	15	8	3	4	0	32
Pupa-Adult	23	30	15	8	14	10	11	3	3	5	77
Larva-Adult	24	12	25	9	17	10	5	4	9	3	81

Table 8. **Categories of negative correlations of expression levels between miRNA species and their putative target transcripts.** L: Larvae, P: Pupae, A: Adults. (Supplemental Table S12)

Category	#miRNAs	# target genes	miRNA expression	Target expression
1	30	49	Only in L	Lowest in L
2	30	44	Only in P	Lowest in P
3	315	1194	Only in A	Lowest in A
4	9	12	L>A>P	P>A>L
5	9	10	P>L>A	A>L>P
6	14	20	P>A>L	L>A>P
7	15	19	A>L>P	P>L>A
8	39	104	L>P>A	L>P>A
9	29	64	A>P>L	A>P>L
Total	490	1516		

Table 9. **The functions of putative targets.** (Supplemental Table S13-S15)

Group	# Putative targets	Functional category
Target genes with unknown function	380	Hypothetical proteins
	217	Proteins with no matches to any sequences in Genbank
	142	Proteins with functions unknown or hypothetical proteins
	20	Proteins with functions not well defined
SSGPs	69	Secreted Salivary Gland Proteins (putative effectors)
Target genes with known function	95	Proteins involved in Nutrient Metabolism
	104	Proteins involved in Protein Metabolism
	41	Proteins involved in RNA activity
	208	Regulatory Proteins
	118	Transporters and Proteins Involved in Transport
	24	Proteins involved in Redox and Detoxification
	98	Structural Proteins and Proteins involved in Cell Adhesion

References

Allen, E., Xie, Z., Gustafson, A. M., Sung, G.-H., Spatafora, J. W., Carrington, J. C. (2004). Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat Genet*, *36*(12), 1282–1290.

Bailey, M. L. B. and W. C. (2000). Hessian Fly Management on Wheat. *Mu Extension, University Of Missouri-Columbia*.

Bartel, D. (2004). MicroRNAsGenomics, Biogenesis, Mechanism, and Function. *Cell*, *116*(2), 281–297. doi:10.1016/S0092-8674(04)00045-5

Behura, S. K. (2007). Insect microRNAs: Structure, function and evolution. *Insect Biochemistry and Molecular Biology*, *37*(1), 3–9. doi:10.1016/j.ibmb.2006.10.006

Bohnsack, M. T., Czaplinski, K., Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA (New York, N.Y.)*, *10*, 185–191. doi:10.1261/rna.5167604

Buntin, G. D. (1999). Hessian Fly (Diptera: Cecidomyiidae) Injury and Loss of Winter Wheat Grain Yield and Quality. *Journal of Economic Entomology*, *92*(5), 1190–1197.

Cai, Y., Yu, X., Hu, S., Yu, J. (2009). A Brief Review on the Mechanisms of miRNA Regulation. *Genomics, Proteomics and Bioinformatics*, *7*(4), 147–154. doi:10.1016/S1672-0229(08)60044-3

Cao, J., Tong, C., Wu, X., Lv, J., Yang, Z., Jin, Y. (2008). Identification of conserved microRNAs in *Bombyx mori* (silkworm) and regulation of fibroin L chain production by microRNAs in heterologous system. *Insect Biochemistry and Molecular Biology*, *38*(12), 1066–1071. doi:10.1016/j.ibmb.2008.09.008

Chen, M.-S., Echegaray, E., Whitworth, R. J., Wang, H., Sloderbeck, P. E., Knutson, A., Royer, T. a. (2009). Virulence analysis of Hessian fly populations from Texas, Oklahoma, and Kansas. *Journal of Economic Entomology*, *102*(2), 774–780. doi:10.1603/029.102.0239

Chen, M.-S., Liu, X., Yang, Z., Zhao, H., Shukle, R. H., Stuart, J. J., Hulbert, S. (2010). Unusual conservation among genes encoding small secreted salivary gland proteins from a gall midge. *BMC Evolutionary Biology*, *10*(1), 296. doi:10.1186/1471-2148-10-296

Chen, M.-S., Zhao, H.-X., Zhu, Y. C., Scheffler, B., Liu, X., Liu, X., Stuart, J. J. (2008). Analysis of transcripts and proteins expressed in the salivary glands of Hessian fly (*Mayetiola destructor*) larvae. *Journal of Insect Physiology*, *54*(1), 1–16. doi:10.1016/j.jinsphys.2007.07.007

Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., MacCallum, I. (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature*, *450*(7167), 203–218. doi:10.1038/nature06341

Cristino, A. S., Tanaka, E. D., Rubio, M., Piulachs, M. D., Belles, X. (2011). Deep sequencing of organ- and stage-specific microrna s in the evolutionarily basal insect *Blattella germanica* (L.) (Dictyoptera, Blattellidae). *PLoS ONE*, *6*(4). doi:10.1371/journal.pone.0019350

Du, T., Zamore, P. D. (2005). microPrimer: the biogenesis and function of microRNA. *Development (Cambridge, England)*, *132*(21), 4645–52. doi:10.1242/dev.02070

Flanders KL, Buntin GD, M. P. (2008). Biology and Management of Hessian Fly in the Southeast. Ala. Coop. Ext. Serv. Bull. ANR-1069. 4 Pp.

Gagné, R.J. and M. Jaschhof. 2014. A Catalog of the Cecidomyiidae (Diptera) of the World. 3rd Edition. Digital version 2.

Gan, L., Denecke, B. (2013). Profiling Pre-MicroRNA and Mature MicroRNA Expressions Using a Single Microarray and Avoiding Separate Sample Preparation. Microarrays, 2(1), 24–33. doi:10.3390/microarrays2010024

Goodall, E. F., Heath, P. R., Bandmann, O., Kirby, J., Shaw, P. J. (2013). Neuronal dark matter: the emerging role of microRNAs in neurodegeneration. *Frontiers in Cellular Neuroscience*, *7*, 178. doi:10.3389/fncel.2013.00178

Gould, F. (1998). Sustainability of Transgenic Insecticidal Cultivars: Integrating Pest Genetics and Ecology. *Annual Review of Entomology*, *43*(1), 701–726. doi:10.1146/annurev.ento.43.1.701

GUDDETI, S., ZHANG, D. C., LI, A. L., LESEBERG, C. H., KANG, H., LI, X. G., MAO, L. (2005). Molecular evolution of the rice miR395 gene family. *Cell Res*, *15*(8), 631–638.

Harris, M. O., Stuart, J. J., Mohan, M., Nair, S., Lamb, R. J., Rohfritsch, O. (2003). GRASSES AND GALL MIDGES: Plant Defense and Insect Adaptation. *Annual Review of Entomology*, *48*(1), 549–577. doi:10.1146/annurev.ento.48.091801.112559

Huang, Y., Shen, X. J., Zou, Q., Wang, S. P., Tang, S. M., Zhang, G. Z. (2011). Biological functions of microRNAs: a review. *Journal of Physiology and Biochemistry March 2011, Volume 67, Issue 1, Pp 129-139.*

Huang, Y., Zou, Q., Tang, S., Wang, L., Shen, X. (2010). Computational identification and characteristics of novel microRNAs from the silkworm (*Bombyx mori L.*). *Molecular Biology Reports*, *37*(7), 3171–3176. doi:10.1007/s11033-009-9897-4

Hussain, M., Asgari, S. (2014). MicroRNAs as mediators of insect host-pathogen interactions and immunity. *Journal of Insect Physiology*, *70C*, 151–158. doi:10.1016/j.jinsphys.2014.08.003

John Van Duyn, Steve Bambara, R. W. (2004). The Hessian Fly: A Pest of Wheat in North Carolina. *North Carolina Cooperative Extension Service North Carolina State University*.

Khajuria, C., Williams, C. E., El Bouhssini, M., Whitworth, R. J., Richards, S., Stuart, J. J., Chen, M.-S. (2013). Deep sequencing and genome-wide analysis reveals the expansion of MicroRNA genes in the gall midge *Mayetiola destructor*. *BMC Genomics*, *14*(1), 187. doi:10.1186/1471-2164-14-187

Kim, Y.-K., Kim, V. N. (2007). Processing of intronic microRNAs. *The EMBO Journal*, *26*(3), 775–783. doi:10.1038/sj.emboj.7601512

Kozomara, A., Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, *42*(Database issue), D68–73. doi:10.1093/nar/gkt1181

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T. (2001). Identification of Novel Genes Coding for Small Expressed RNAs. *Science , 294* (5543), 853–858. doi:10.1126/science.1064921

Lai, E. C., Tomancak, P., Williams, R. W., Rubin, G. M. (2003). Computational identification of Drosophila microRNA genes. *Genome Biology*, *4*(7), R42. doi:10.1186/gb-2003-4-7-r42

Lamiri, A., Lhaloui, S., Benjilali, B., Berrada, M. (2001). Insecticidal effects of essential oils against Hessian fly, *Mayetiola destructor* (Say). *Field Crops Research*, *71*(1), 9–15. doi:10.1016/S0378-4290(01)00139-3

Lau, N. C., Lim, L. P., Weinstein, E. G., Bartel, D. P. (2001). An Abundant Class of Tiny RNAs with Probable Regulatory Roles in *Caenorhabditis elegans*. *Science , 294* (5543), 858–862. doi:10.1126/science.1065062

Lee, R. C., Ambros, V. (2001). An Extensive Class of Small RNAs in *Caenorhabditis elegans*. *Science , 294* (5543), 862–864. doi:10.1126/science.1065329

Lee, R. C., Feinbaum, R. L., Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, *75*(5), 843–854. doi:10.1016/0092-8674(93)90529-Y

Legeai, F., Rizk, G., Walsh, T., Edwards, O., Gordon, K., Lavenier, D., Jaubert-Possamai, S. (2010). Bioinformatic prediction, deep sequencing of microRNAs and expression analysis during phenotypic plasticity in the pea aphid, *Acyrthosiphon pisum*. *BMC Genomics*, *11*(1), 281. doi:10.1186/1471-2164-11-281

Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P., Burge, C. B. (2003). Prediction of Mammalian MicroRNA Targets. *Cell*, *115*, 787–798. doi:10.1016/S0092-8674(03)01018-3

Lhaloui, S.; Buschman, L.; Bouhssini, M. E.; Amri, A.; Hatchett, J.; Keith, D.; Starks, K.; Houssaini, K. E. (1992). Infestations of *Mayetiola spp*. Diptera, Cecidomyiidae in bread wheat, durum wheat and barley: results of five annual surveys in the major cereal growing regions of Morocco. *Al Awamia (77): 21-54.*

Li, A., Mao, L. (2007). Evolution of plant microRNA gene families. *Cell Research, 17*, 212–218. doi:10.1038/sj.cr.7310113

Li, S., Mead, E. A., Liang, S., Tu, Z. (2009). Direct sequencing and expression analysis of a large number of miRNAs in *Aedes aegypti* and a multi-species survey of novel mosquito miRNAs. *BMC Genomics*, *10*, 581. doi:10.1186/1471-2164-10-581

Liu, F., Peng, W., Li, Z., Li, W., Li, L., Pan, J., Su, S. (2012). Next-generation small RNA sequencing for microRNAs profiling in *Apis mellifera*: Comparison between nurses and foragers. *Insect Molecular Biology*, *21*(March), 297–303. doi:10.1111/j.1365-2583.2012.01135.x

Lu, J., Shen, Y., Wu, Q., Kumar, S., He, B., Shi, S., Wu, C.-I. (2008). The birth and death of microRNA genes in Drosophila. *Nature Genetics*, *40*(3), 351–355. doi:10.1038/ng.73

Lucas, K., Raikhel, A. S. (2013). Insect microRNAs: biogenesis, expression profiling and biological functions. *Insect Biochemistry and Molecular Biology*, *43*(1), 24–38. doi:10.1016/j.ibmb.2012.10.009

Luo, Q., Zhou, Q., Yu, X., Lin, H., Hu, S., Yu, J. (2008). Genome-wide mapping of conserved microRNAs and their host transcripts in *Tribolium castaneum*. *Journal of Genetics and Genomics = Yi Chuan Xue Bao*, *35*(6), 349–55. doi:10.1016/S1673-8527(08)60051-X

Mattick, J. S., Gagen, M. J. (2001). The Evolution of Controlled Multitasked Gene Networks: The Role of Introns and Other Noncoding RNAs in the Development of Complex Organisms. *Molecular Biology and Evolution*, *18*(9), 1611–1630. doi:10.1093/oxfordjournals.molbev.a003951

Morgan, G. (2005). Hessian Fly In Texas Wheat. Retrieved from: http://lubbock.tamu.edu/files/2015/06/Hessian_Fly.pdf

Nozawa, M., Miura, S., Nei, M. (2010). Origins and evolution of microRNA genes in Drosophila species. *Genome Biology and Evolution*, *2*, 180–189. doi:10.1093/gbe/evq009

Okamura, K., Phillips, M. D., Tyler, D. M., Duan, H., Chou, Y., Lai, E. C. (2008). The regulatory activity of microRNA species has substantial influence on microRNA and 3' UTR evolution. *Nat Struct Mol Biol*, *15*(4), 354–363.

Ormerod, E. a. (1887). *The Hessian Fly*. *Nature* (Vol. 36). doi:10.1038/036439a0

Piriyapongsa, J., Jordan, I. K. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA*, *14*(5), 814–821. doi:10.1261/rna.916708

Piriyapongsa, J., Mariño-Ramírez, L., Jordan, I. K. (2007). Origin and Evolution of Human microRNAs From Transposable Elements. *Genetics* , *176* (2), 1323–1337. doi:10.1534/genetics.107.072553

Ramalingam, P., Palanichamy, J. K., Singh, A., Das, P., Bhagat, M., Kassab, M. A., Chattopadhyay, P. (2014). Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing. *RNA*, *20*(1), 76–87. doi:10.1261/rna.041814.113

Ratcliffe, R. H., Cambron, S. E., Flanders, K. L., Bosque-Perez, N. a, Clement, S. L., Ohm, H. W. (2000). Biotype composition of Hessian fly (Diptera: Cecidomyiidae) populations from the southeastern, midwestern, and northwestern United States and virulence to resistance genes in wheat. *Journal of Economic Entomology*, *93*, 1319–1328. doi:10.1603/0022-0493-93.4.1319

Ratcliffe, R. H., Cambron, S. E., Flanders, K. L., Bosque-Perez, N. A., Clement, S. L., Ohm, H. W. (2000). Biotype Composition of Hessian Fly (Diptera: Cecidomyiidae) Populations from the Southeastern, Midwestern, and Northwestern United States and Virulence to Resistance Genes in Wheat. *Journal of Economic Entomology*, *93*(4), 1319–1328.

Rebijith, K. B., Asokan, R., Krishna, V., Hande, R. H., Kumar, N. K. K., Ramamurthy, V. V. (2014). In Silico Prediction and Characterization of MicroRNAs from *Aphis gossypii* (Hemiptera: Aphididae). *Annals of the Entomological Society of America*, *107*(2), 521–531. doi:10.1603/AN12158

Ruby, J. G., Jan, C. H., Bartel, D. P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature*, *448*(7149), 83–86. doi:10.1038/nature05983

Ruby, J. G., Stark, A., Johnston, W. K., Kellis, M., Bartel, D. P., Lai, E. C. (2007). Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome Research*, *17*(12), 1850–1864. doi:10.1101/gr.6597907

Sandra Garcés-Carrera, Allen Knutson, Haiyan Wang, Kristopher L. Giles, Fangneng Huang, R. Jeffrey Whitworth, C. Michael Smith, M.-S. C. (2014). Virulence and Biotype Analyses of Hessian Fly (Diptera: Cecidomyiidae) Populations From Texas, Louisiana, and Oklahoma. *Journal of Economic Entomology*, *107(1):417*.

Sardesai, N., Nemacheck, J. A., Subramanyam, S., Williams, C. E. (2005). Identification and mapping of H32, a new wheat gene conferring resistance to Hessian fly. *Theoretical and Applied Genetics*, *111*(6), 1167–1173. doi:10.1007/s00122-005-0048-6

Sathyamurthy, Ganesh; Swamy, N. R. (2009). Computational Identification of MicroRNA Homologs from *Acyrthosiphon pisum* (Pea Aphid). *Journal of Computational Intelligence in Bioinformatics, Vol. 2*(Issue 3), p109–119.

Schmittgen, T. D., Jiang, J., Liu, Q., Yang, L. (2004). A high-throughput method to monitor the expression of microRNA precursors. *Nucleic Acids Research*, *32*(4), e43. doi:10.1093/nar/gnh040

Schumann, H. (1976). Mamaev, B. M. (1975): Evolution of gall forming insects—gall midges (Übersetzung von Crozy, A.; Mamaev, B. M. (1968): Evolyutsiya galloobrazuyushchikh nasekomykh — gallits. — Leningrad). — Brit. Libr., Lending Div., Boston Spa, Wetherby, West Yorkshire, E. *Deutsche Entomologische Zeitschrift*, *23*(4-5), 411. doi:10.1002/mmnd.19760230420

Singh, J., Nagaraju, J. (2008). In silico prediction and characterization of microRNAs from red flour beetle (*Tribolium castaneum*). *Insect Molecular Biology*, *17*(4), 427–436.

Skalsky, R. L., Vanlandingham, D. L., Scholle, F., Higgs, S., Cullen, B. R. (2010). Identification of microRNAs expressed in two mosquito vectors, *Aedes albopictus* and *Culex quinquefasciatus*. *BMC Genomics*, *11*(1), 119. doi:10.1186/1471-2164-11-119

Slack, F. J., Basson, M., Liu, Z., Ambros, V., Horvitz, H. R., Ruvkun, G. (2000). The lin-41 RBCC Gene Acts in the *C. elegans* Heterochronic Pathway between the let-7 Regulatory RNA and the LIN-29 Transcription Factor. *Molecular Cell*, *5*(4), 659–669. doi:10.1016/S1097-2765(00)80245-2

Stuart, J. J., Chen, M.-S., Marion O. Harris. (2008). Hessian Fly. USDA Agricultural Research Service- Lincoln, Nebraska. USDA/UNL Faculty, http://digitalcommons.unl.edu/usdaarsfacpub/397.

Stuart, J. J., Chen, M.-S., Shukle, R., Harris, M. O. (2012). Gall midges (Hessian flies) as plant pathogens. *Annual Review of Phytopathology*, *50*, 339–57. doi:10.1146/annurev-phyto-072910-095255

Subramanyam, S., Sardesai, N., Minocha, S. C., Zheng, C., Shukle, R. H., Williams, C. E. (2015). Hessian fly larval feeding triggers enhanced polyamine levels in susceptible but not resistant wheat. *BMC Plant Biology*, *15*(1), 3. doi:10.1186/s12870-014-0396-y

Sun, W., Julie Li, Y.-S., Huang, H.-D., Shyy, J. Y.-J., Chien, S. (2010). microRNA: A Master Regulator of Cellular Processes for Bioengineering Systems. *Annual Review of Biomedical Engineering*, *12*(1), 1–27. doi:10.1146/annurev-bioeng-070909-105314

Thakur, R. P. (2007). Host Plant Resistant to Diseases: Potential and Limitations. *Indian Journal of Plant Protection, 35 (1). Pp. 17-21.*

Tom A. Royer; Jeff Edwards; Kristopher L.Giles. (2009). Hessian Fly Management in Oklahoma Winter Wheat. *Division of Agricultural Science and Natural Resources, Oklahoma State University*.

Wei, Y., Chen, S., Yang, P., Ma, Z., Kang, L. (2009). Characterization and comparative profiling of the small RNA transcriptomes in two phases of locust. *Genome Biology*, *10*(1), R6–R6. doi:10.1186/gb-2009-10-1-r6

Weis, a. (1988). Reactive Plant Tissue Sites And The Population Biology Of Gall Makers. *Annual Review of Entomology*, *33*(Figure 1), 467–486. doi:10.1146/annurev.ento.33.1.467

Yu, X., Zhou, Q., Li, S.-C., Luo, Q., Cai, Y., Lin, W., Yu, J. (2008). The Silkworm (*Bombyx mori*) microRNAs and Their Expressions in Multiple Developmental Stages. *PLoS ONE*, *3*(8), e2997.

Zhang, L., Chia, J.-M., Kumari, S., Stein, J. C., Liu, Z., Narechania, A., Ware, D. (2009). A Genome-Wide Characterization of MicroRNA Genes in Maize. *PLoS Genet*, *5*(11), e1000716.

Zhang, X., Zheng, Y., Jagadeeswaran, G., Ren, R., Sunkar, R., Jiang, H. (2012). Identification and developmental profiling of conserved and novel microRNAs in *Manduca sexta*. *Insect Biochemistry and Molecular Biology*, *42*(6), 381–395. doi:10.1016/j.ibmb.2012.01.006

Zhang, Y., Zhou, X., Ge, X., Jiang, J., Li, M., Jia, S., Huang, Y. (2009). Insect-Specific microRNA Involved in the Development of the Silkworm *Bombyx mori*. *PloS One*, *4*(3), e4677. doi:10.1371/journal.pone.0004677

Zhao, C., Escalante, L. N., Chen, H., Benatti, T. R., Qu, J., Chellapilla, S., Richards, S. (2015). A massive expansion of effector genes underlies gall-formation in the wheat pest *Mayetiola destructor*. *Current Biology : CB*, *25*(5), 613–20. doi:10.1016/j.cub.2014.12.057

Zhou, G., Wang, T., Lou, Y., Cheng, J., Zhang, H., Xu, J.-H. (2014). Identification and Characterization of MicroRNAs in Small Brown Planthopper (*Laodephax striatellus*) by Next-Generation Sequencing. *PLoS ONE*, *9*(7), e103041.