microRNA Identification and Target Prediction in the Whitefly (Bemisia tabaci)

Alexis Aleman

University of Southern Mississippi

Follow this and additional works at: https://aquila.usm.edu/honors_theses

Recommended Citation
https://aquila.usm.edu/honors_theses/686

This Honors College Thesis is brought to you for free and open access by the Honors College at The Aquila Digital Community. It has been accepted for inclusion in Honors Theses by an authorized administrator of The Aquila Digital Community. For more information, please contact Joshua.Cromwell@usm.edu.
microRNA Identification and Target Prediction in the Whitefly (*Bemisia tabaci*)

by

Alexis Aleman

A Thesis
Submitted to the Honors College of
The University of Southern Mississippi
in Partial Fulfillment
of Honors Requirements

July 2019
Approved by

_____________________________________________________
Alex Flynt, Ph.D., Thesis Adviser
School of Biological, Environmental, and Earth Sciences

_____________________________________________________
Jake Schaefer, Ph.D., Interim Director
School of Biological, Environmental, and Earth Sciences

_____________________________________________________
Ellen Weinauer, Ph.D., Dean
Honors College
Abstract

RNA interference, referred to as RNAi, is a biological phenomenon whereby knock-down of gene expression can be achieved through the use of RNA molecules, including small interfering RNAs (siRNAs) and microRNAs (miRNAs). The use of miRNAs is an endogenous pathway that results in the degradation of the miRNA strand’s complementary messenger RNA (mRNA), preventing the translation of the mRNA and therefore the production of the proteins necessary for gene expression. While RNAi is a biological phenomenon that occurs naturally, it is also a method that can be used and manipulated in the laboratory to try to control the expression of genes associated with various issues. One such problem that has been studied using RNAi is pest control. One of the most destructive pests to crops globally is the whitefly, Bemisia tabaci. RNAi has produced the best results in B. tabaci when double-stranded RNAs (dsRNAs) are fed to the fly. In order to better understand miRNA evolution and to examine the process of miRNA-mediated mRNA degradation via RNAi in the whitefly, a miRDeep analysis was performed on the genome of Bemisia tabaci to identify miRNAs. Doing so allowed for the identification of 41 novel miRNAs and a total of 177 confidently assigned miRNAs in the whitefly. Of the novel miRNAs discovered, 83% were monocistronic, and the 177 confidently assigned miRNAs were predicted to regulate 51% of the mRNAs in the whitefly, consistent with the findings of previous studies. The ten mRNAs that are targeted the most by miRNAs in the whitefly play serious roles in medicine, including coding for a brain tumor protein, pointing to the clinical significance of further investigating the RNAi pathways in these genes. In addition, of the 51% of the whitefly’s mRNAs that were predicted to be regulated by miRNAs, 4% were discovered to be
targeted by 100 or more miRNAs, indicating numerous RNAi pathways that can be further studied and analyzed in order to better control the expression of these genes.

**Key words:** Whitefly, *Bemisia tabaci*, RNAi, miRNA, mRNA, dsRNA
Dedication

To my husband, Jonathan Stockman, who has supported me through the many obstacles and trials I faced during this process. Without your unwavering belief in me as a scientist and researcher, I do not know where I would be today.

To my family members, especially Mom, Mawmaw, and Pawpaw, for never ceasing to believe in me and for encouraging me when I lost hope. To my best friends, Savannah and Cory, for constantly telling me that I am a phenomenal scientist and that I can achieve anything with the right amount of determination and resilience. If not for your love, friendship, and support, I would not have accomplished this great achievement.
Acknowledgments

I would like to thank my thesis advisor, Dr. Alex Flynt, for his mentorship during the process of completing this study. You have taught me so much and have continually shown me patience and understanding. This work would not have been possible had it not been for your guidance and support. Thank you for all you have seen me through, and for continuing to believe in the completion of this project.

Additionally, I want to specially thank the faculty of the Honors College. Throughout my four years at The University of Southern Mississippi, their support, resources, and guidance have molded me into the scholar and woman that I am today.

I also would like to thank Dr. Mac Alford of The University of Southern Mississippi for his assistance in creating Fig. 4, a phylogenetic tree of the shared miRNAs between the whitefly and its relatives, by allowing me to use his WinClada software. Your aid came after I had become overwhelmed trying to map so many miRNAs by hand and I greatly appreciate your help.
# Table of Contents

List of Tables ........................................................................................................................................... ix
List of Figures .............................................................................................................................................. x
List of Abbreviations ............................................................................................................................... xi
Chapter 1: Introduction ............................................................................................................................. 1
Chapter 2: Methodology ........................................................................................................................... 3
  2.1 miRDeep Analysis ............................................................................................................................. 4
  2.2 Classification as Polycistronic or Monocistronic .......................................................................... 6
  2.3 Discovering Shared miRNAs with Related Species ..................................................................... 6
  2.4 Target Prediction ............................................................................................................................. 6
  2.5 dsRNA Feeding and DEseq ........................................................................................................... 7
  2.6 Calculating the Percent of Whitefly mRNAs that Are Regulated by miRNAs ............................. 7
Chapter 3: Results ..................................................................................................................................... 8
Chapter 4: Conclusions ............................................................................................................................ 18
References .................................................................................................................................................. 20
List of Tables

Table 1. Classification of Potential miRNAs from miRDeep Analysis..........................8
Table 2. Net Number of miRNAs Shared Between Related Orders and the Whitefly ......12
Table 3. Top Ten Genes Targeted by Whitefly miRNAs ...............................................16
List of Figures

Figure 1. Scoring Criteria for Evaluating Potential miRNA Candidates..........................5
Figure 2. Arrangement of Novel miRNAs in the *Bemisia tabaci* Genome ......................9
Figure 3. The Seven *Bemisia tabaci* Polycistronic miRNAs......................................10
Figure 4. Phylogenetic Tree of *Bemisia tabaci* and Relatives with Shared miRNAs ......11
Figure 5. Target Prediction of Known miRNAs...........................................................13
Figure 6. Target Prediction of Novel miRNAs ..............................................................14
Figure 7. DEseq After dsRNA Feeding .........................................................................15
Figure 8. Distribution of Whitefly mRNA Target Site Values .......................................16
Figure 9. Percent of mRNAs Regulated by miRNAs ....................................................17
Figure 10. Percent of mRNAs Targeted by 100 or More miRNAs ..............................17
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGO</td>
<td>Argonaute family protein</td>
</tr>
<tr>
<td>Bps</td>
<td>base pairs</td>
</tr>
<tr>
<td>dsRNA</td>
<td>double-stranded RNA</td>
</tr>
<tr>
<td>miR-#</td>
<td>micro RNA (followed by number, e.g., miR-1120)</td>
</tr>
<tr>
<td>miRNA</td>
<td>micro RNA</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>pre-miRNA</td>
<td>precursor miRNA</td>
</tr>
<tr>
<td>pri-miRNA</td>
<td>primary miRNA</td>
</tr>
<tr>
<td>RISC</td>
<td>RNA-induced silencing complex</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>siRNA</td>
<td>small interfering RNA</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

MicroRNAs are small, non-coding, highly conserved ribonucleic acids with a length of about 22 nucleotides that have been found to play a variety of significant roles in molecular biology. For example, many microRNAs (miRNAs) have been linked to development. Bantam, a gene in *Drosophila melanogaster*, encodes a 21-nucleotide miRNA that triggers cell proliferation and inhibits apoptosis (Brennecke, Hipfner, Stark, Russell, and Cohen, 2003). Various miRNAs have also been tied to homeostasis and cartilage development in mammals, including miR-140, whose loss resulted in the development of osteoarthritis in mice (Miyaki et al., 2010). Still other miRNAs have been found to play roles in the development of larvae and neurons and the differentiation of hematopoietic stem cells (He and Hannon, 2004). In addition, miRNAs play a crucial role in the cell’s stress response, such as when p53 triggers the transcription of various miRNAs that halt growth and induce apoptosis when the cell’s DNA is damaged (Leung and Sharp, 2010). Due to all of their critical functions, the pathways involving miRNAs and the mechanisms that offer ways to manipulate them have been vastly studied.

One biological phenomenon in which miRNAs are involved that has been extensively examined is RNAi, or RNA interference. RNAi is a gene silencing pathway that occurs posttranscriptionally and uses various kinds of small ribonucleic acid molecules, including small interfering RNAs (siRNAs) and miRNAs, to degrade messenger RNA, mRNA (Lam, Chow, Zhang, and Leung, 2015). The RNAi pathway involving miRNAs begins when primary miRNA (pri-miRNA) transcripts are cleaved by the microprocessor complex, consisting of Drosha (a ribonuclease [RNase] III enzyme) and a double-stranded RNA (dsRNA) binding protein named DGCR8 (Gregory et al.,
The microprocessor complex cleaves the pri-miRNA transcripts into precursor miRNA (pre-miRNA) hairpins in the nucleus. These hairpins are then processed in the cytoplasm where they are cleaved by Dicer, another RNase III enzyme, into miRNA duplexes. One of the strands of the duplex is loaded into the RNA-induced silencing complex (RISC), a ribonucleoprotein complex consisting of members of the Argonaute (AGO) protein family. The RISC then brings the miRNA strand to bind to its complementary mRNA target, which occurs when the miRNA seed region (nucleotides 2-8) complementary base pair to the 3’ untranslated region (UTR) of the mRNA (Flynt and Lai, 2008). Upon binding, the miRNA initiates degradation of the mRNA. The other strand of the miRNA duplex, the star or passenger strand, is destroyed (Kim and Rossi, 2008). In this fashion, miRNAs play a critical role in downregulating expression of certain genes by degrading the mRNAs that would be translated into the specific proteins encoded by those genes. Thus, a better understanding of endogenous miRNA pathways will lead to better laboratory applications of RNA interference.

In order to analyze RNAi pathways in a laboratory setting, RNAi can be induced by feeding RNA molecules, such as dsRNAs, to the organism being studied. Many RNAi pathways have been analyzed in various species, but this method of gene silencing has not proven to be totally effective. Many RNAi studies on insects, especially whiteflies and other sap-feeding insects, have produced only modest knock-down of gene expression. Several organisms contain RNases that can degrade the dsRNAs or other small RNAs fed to them. In order to counteract this dilemma, some scientists have discovered that the effects of RNAi can be enhanced by using dsRNAs that specifically target the RNAi-suppressing RNases (Luo et al., 2017). In this way, feeding an organism
dsRNAs serves two purposes, to allow for RNAi to occur and to combat the effects of any RNAi-suppressing genes.

One area of research where RNAi has been applied and studied is pest control. One of the most important crop pests and threats to global food security today is the whitefly, *Bemisia tabaci*, which belongs to a group of plant-feeding pests called psyllids (Taning, Andrade, Hunter, Christiaens, and Smagghe, 2016). In fact, *B. tabaci* is one of the 100 worst invasive species in the world (Chen et al., 2016). The whitefly spreads many viruses that restrict the production of significant crops around the globe (Navas-Castillo, Fiallo-Olivé, and Sánchez-Campos, 2011). As a result of the severe detrimental impact that *Bemisia tabaci* has on food security, many scientists have investigated RNAi pathways that offer potential ways to control this pest.

Studying miRNA pathways in an organism allows scientists to identify novel miRNAs, to assess the expression levels of miRNA loci, and to learn about miRNA evolution and processing in that organism (Mohammed et al., 2018). This study seeks to accomplish all of these tasks in *B. tabaci*, to understand the evolution of miRNAs in the species and its relatives and to gain knowledge of the endogenous miRNAs that could lead to better manipulation and use of RNAi pathways in this organism.

**Chapter 2: Methodology**

To begin the miRDeep analysis of the whitefly, the mature sequences of the miRNAs for all of the species in miRBase for the subphyla Chelicerata, Crustacea, and Hexapoda were downloaded in an unaligned fasta format. Of the organisms present in miRBase for subphylum Hexapoda, all were included except those of the Drosophila genus. All organisms present in miRBase for subphyla Chelicerata and Crustacea were
included. Conducting the miRDeep command produced a html sheet with the results of the miRNA analysis of the whitefly. The resulting potential miRNA molecules were analyzed using a list of six criteria, including length, the presence of a star strand, good hairpin structure with a central bulge, expression level, precision of processing the 5’ end, and the presence of two nucleotide three-prime overhangs. An example of a good match for each criterion, as well as an explanation of evaluating each criterion, is given in Fig. 1.

2.1 miRDeep Analysis

After evaluating all six criteria, each potential miRNA molecule was given a score out of six for how well it reflected these standards. Based on their scores, the candidates were placed into one of five categories. Those with a score of six were placed in a confident category. Candidates with a score of five were placed in a high probability category, those with a score of three or four were placed in a middle probability category, and those with a score of one or two were placed in a low probability category. Any data entry that either failed to meet any of the six criteria or that had no star present was categorized as non-miRNA. We excluded potential candidates that lacked a star strand because the presence of both the star and mature strands is critical to indicate that there was a miRNA duplex. After identifying the confident miRNAs, those with a homolog in a related species were categorized as known, while those without a homolog in another species were categorized as novel.
Figure 1. Scoring Criteria for Evaluating Potential miRNA Candidates. (A) Each miRNA candidate was analyzed to see if it had a length of about 22 nucleotides. To do this, the list of sequences for the mature strand was perused to find the strand with the largest occurrence value (the true sequence), and then the bases of that sequence were counted. (B) Each miRNA candidate also had to have a star strand present, which could be evaluated by viewing the star read count at the bottom of the data. If a star was not present, the candidate was marked as not miRNA, since a star strand is critical in the identification of miRNA since the star strand and mature strand must be present from the miRNA duplex. (C) Expression was the third criterion, and candidates with over ten reads met this category. (D) The fourth criterion was precision of processing the 5’ end. This was measured by finding the mature strand sequences with the largest occurrence values and ensuring their 5’ end included the same bases. Even a difference in one base is significant enough to change the processing of the 5’ end, so even a difference that slight was considered a failure to meet this criterion. (E) The fifth criterion for identification as miRNA was structure, which included a good structure with a central bulge. Panel E shows one such example of good miRNA structure. (F) The last criterion used to evaluate potential miRNA candidates was the presence of two nucleotide 3’ overhangs. This was done by examining the two 3’ ends of the miRNA and ensuring that there were two nucleotides hanging over, as seen by the red U and A and the purple U and C in panel F.
2.2 Classification as Polycistronic or Monocistronic

After identifying which miRNAs were classified as confident and novel, the novel miRNAs were further analyzed to determine if they were monocistronic or polycistronic. To do this, all other data entries with the same scaffold number as a novel miRNA were located in the miRDeep analysis results, and the location of all of these scaffolds in the genome was analyzed. If the novel miRNA scaffold had any other scaffolds that were within 20,000 base pairs (bps) of its location, then that miRNA was classified as polycistronic. If the novel miRNA did not have any other miRNA within 20,000 bps of its location, then that miRNA was classified as monocistronic.

2.3 Discovering Shared miRNAs with Related Species

Since one of the key purposes of this study was to examine if miRNA evolves similarly amongst related species, a phylogenetic tree detailing the number of shared miRNAs between the whitefly and its related species was constructed. To do this, each known miRNA was looked up in miRBase to determine which of the other species in the study also have the miRNA. The results were then organized by grouping the related organisms by order, and the orders and their respective number of shared miRNAs were graphed on the phylogenetic tree. Any miRNAs that have been lost over time were also mapped onto the tree.

2.4 Target Prediction

To further examine the role of miRNAs in whitefly biology, the TargetScan algorithm was used. To utilize this program, the 3’ UTR sequences from the whitefly database were combined with the miRNA calls from the miRDeep analysis. The 3’ UTR sequences were downloaded since miRNAs function by base pairing with the 3’ UTR of
their target mRNAs in order to repress translation or initiate degradation (Olena and Patton, 2009). TargetScan functions by searching the 3’ UTR sequences for segments that Watson-Crick complementarily base pair to nucleotides 2-8 of the miRNAs (Lewis, Shih, Jones-Rhoades, Bartel, and Burge, 2003). The 6mers were deleted and not used in this algorithm, due to their low probability of being regulatory molecules. The algorithm then predicted any mRNAs that were being targeted by the miRNAs.

2.5 dsRNA Feeding and DEseq

Next, to ensure that feeding whiteflies dsRNAs does not impede the normal functioning of the endogenous miRNAs, we performed an adapter ligation-mediated PCR analysis using a Bioo Scientific® kit. The Bioo kit first ligates a 4N adenylated adapter to the 3’ end of the miRNA and then removes and inactivates any excess of the 3’ adapter. Then a 4N adenylated adapter is ligated to the 5’ end of the miRNA and reverse transcription occurs on the 5’ and 3’ adapter ligated miRNA to produce the 1st strand synthesis product. This product then undergoes PCR to be amplified and then undergoes a size selection. The PCR product is then entered into an Illumina® flow cell, where sequencing occurs.

2.6 Calculating the Percent of Whitefly mRNAs that Are Regulated by miRNAs

Lastly, in order to quantify how much of the whitefly’s genome is under regulation by miRNAs, we divided the total of predicted miRNAs by the total whitefly gene number (15,664) to calculate the percent of whitefly mRNAs that are predicted to be regulated by miRNAs. Then we further analyzed the mRNAs predicted to be under control of miRNAs by investigating how many of those mRNAs are targeted by 100 or
more miRNAs. This number was divided by the total whitefly gene number to give the percent of mRNAs predicted to be targeted by 100 or more miRNAs.

**Chapter 3: Results**

The evaluation of the potential miRNA molecules from the miRDeep analysis allowed for the establishment of four categories. The first category, known, includes all miRNA molecules that had a homolog, or presence of the same gene, in another species. The second category, novel, includes all miRNAs that lacked a homolog in another species. The third category, candidate, includes all miRNA entries that had a criteria score of one to five, and the fourth category, confident, includes all the miRNA molecules that had a criteria score of six (including both known and novel miRNAs). The summary of these findings is reported in Table 1. Of the 334 total candidates, 53% were confidently assigned, and of those 177 confidently assigned miRNAs, 23% were novel.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known</td>
<td>136</td>
</tr>
<tr>
<td>Novel</td>
<td>41</td>
</tr>
<tr>
<td>Candidate</td>
<td>72</td>
</tr>
<tr>
<td>Confident</td>
<td>177</td>
</tr>
<tr>
<td>Total</td>
<td>334</td>
</tr>
</tbody>
</table>

**Table 1. Classification of Potential miRNAs from miRDeep Analysis.** The number of miRNAs that were analyzed that are known, novel, candidate, and confident, respectively, are quantified and the total number of miRNAs that were analyzed is given. Note that the reason the confident and candidate numbers do not equal the total is because the data entries that failed to meet any scoring criteria or lacked a star were thrown out.

Analyzing the novel miRNAs further to label them as polycistronic or monocistronic led to the identification of thirty-four monocistronic miRNAs (83%) and seven polycistronic miRNAs (17%). The results of this analysis can be seen in Fig. 2. Our low percentage of polycistronic miRNAs is inconsistent with the predictions of other experiments. A study done by Truscott, Islam, and Frolov in 2016 predicted that about 40% of miRNAs in *Drosophila melanogaster* are polycistronic. Another study predicted
that 40% of the miRNAs in humans are polycistronic (Wienholds and Plasterk, 2005). Our data shows a percentage that is less than half of that value, most likely the result of the whitefly genome being poorly assembled, causing polycistronic RNAs to appear to be further apart and thus identified as monocistronic, while they may in fact be polycistronic. Each of the seven predicted polycistronic miRNAs is displayed in Fig. 3, showing the location of their loci and the distance (in bps) between them. Notably, one of the discovered polycistronic miRNAs is antisense (panel C), which occurs as the result of bidirectional transcription, like the iab-4 miRNA in Drosophila melanogaster (Stark et al., 2008). This is why the horizontal bars in panel C overlap, because transcription occurs on both the positive and negative strands simultaneously.

**Figure 2. Arrangement of Novel miRNAs in the Bemisia tabaci Genome.** The loci locations of the forty-one novel miRNAs were analyzed and thirty-four were predicted to be monocistronic (83%) whereas seven were predicted to be polycistronic (17%). The miRNAs predicted to be monocistronic were labeled as such because they had no other miRNAs within 20,000 bps of their location in the genome, and the miRNAs predicted to be polycistronic were labeled as such because they had another miRNA within 20,000 bps of their location in the genome.
Figure 3. The Seven *Bemisia tabaci* Polycistronic miRNAs. The seven polycistronic miRNAs are shown above. In each panel, the shaded region displays the miRNA calls made after the miRDeep analysis, with a + and blue mark signifying the presence of a miRNA. The white region under it shows the expression levels of each miRNA. On the more zoomed in panels, such as C and D, even the two arms of the miRNA can be seen, further validating the call that the candidate is a true miRNA. (A) Panel A displays scaffold 328, where a novel miRNA was identified from bps 1,375,070-1,375,123 and its partner miRNA was identified from bps 1,368,178-1,368,227. They are 6,843 bp apart. (B) Here, both miRNAs shown are novel. They are in scaffold 365, with one located from bps 771,652-771,700 and the other from bps 789,744-789,792. They are 18,044 bp apart. (C) An antisense miRNA is shown here, resulting from bidirectional transcription, which is why the two miRNAs overlap. One runs from bp 628-737 (the novel miRNA) and the other from bp 684-739. (D) Here scaffold 52 is shown, with miRNAs at bps 6,769,792-6,769,852 (the novel miRNA) and 6,769,969-6,770,025. They are 117 bp apart. (E) Scaffold 650 is shown, with miRNAs at bps 1,465,108-1,465,170 and 1,466,885-1,466,945 (the novel miRNA). They are 1,715 bp apart. (F) Scaffold 3080 is shown here, with miRNAs at bps 186,777-186,837 (the novel miRNA) and 189,211-189,276. They are 2,374 bps apart.
Analyzing the presence of the confident miRNAs in other related species on miRBase led to the identification of shared miRNAs between the whitefly and its relatives. The numbers of shared miRNAs and lost miRNAs were organized by order. The resulting phylogenetic tree is shown in Fig. 4. The number of shared miRNAs is given at each node, as well as the number of any miRNAs that have been lost. A summary of the net number of miRNAs shared or lost between the whitefly and each of the seven orders is shown in Table 2. The order with the most shared miRNAs with the whitefly was found to be order Hymenoptera. This makes sense because the whitefly belongs to this order, so Hymenoptera includes all of the whitefly’s miRNAs, which are included in the total shared number. The groups with the next largest number of shared miRNAs are the Lepidoptera and Diptera, which also makes sense because these orders are the whitefly’s closest related orders on the tree, so they should have the next highest number of shared miRNAs, evolutionarily speaking.

![Figure 4. Phylogenetic Tree of Bemisia tabaci and Relatives with Shared miRNAs](image)

Searching for the known miRNAs found in *Bemisia tabaci* on miRBase allowed for the construction of this phylogenetic tree detailing the number of shared miRNAs between these orders of organisms and *Bemisia tabaci*. The number followed by + is the number of miRNAs shared with the whitefly, while the number followed by – is the number of miRNAs lost over time.
<table>
<thead>
<tr>
<th>Order</th>
<th>Net Number of miRNAs Shared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Araneae</td>
<td>26</td>
</tr>
<tr>
<td>Ixodida</td>
<td>15</td>
</tr>
<tr>
<td>Trombidiformes</td>
<td>19</td>
</tr>
<tr>
<td>Notostraca</td>
<td>33</td>
</tr>
<tr>
<td>Cladocera</td>
<td>19</td>
</tr>
<tr>
<td>Orthoptera</td>
<td>6</td>
</tr>
<tr>
<td>Hemiptera</td>
<td>29</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>56</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>60</td>
</tr>
<tr>
<td>Lepidoptera</td>
<td>57</td>
</tr>
<tr>
<td>Diptera</td>
<td>57</td>
</tr>
</tbody>
</table>

Table 2. Net Number of miRNAs Shared Between Related Orders and the Whitefly. In order to see the big picture instead of tracing individual miRNAs like in Fig. 4, a table was made with the net number of miRNAs shared between the orders and the whitefly, after taking into account the numbers of miRNAs lost. The order with the most shared miRNAs is Hymenoptera, the whitefly’s own order, followed closely by Lepidoptera and Diptera, the two orders most closely related to the whitefly.

The target scan algorithm led to the quantification of the number of genes under control of the 136 known miRNA that were found in the whitefly. Their names and the number of genes they control are shown in Fig. 5., with the ancient miRNAs shown in panel B and all other arthropod miRNAs shown in panel A. The majority of miRNAs in both groups target around the same number of genes, with an average of around 2,000 genes. However, in both cases, there are some miRNAs that target well above that number, such as miR-6006-3p in panel A, which targets just under 18,000 genes. Three other arthropod miRNAs, miR-8499b, miR-315-5p, and miR-8496, all target over 8,000 genes. Further research into how these specific miRNAs function and the pathways that they use to regulate gene expression could lead to mechanisms that control a significantly large number of genes.
Figure 5. Target Prediction of Known miRNAs. The number of genes under control by each known miRNA is shown above. Panel A depicts the number of genes under control by arthropod miRNA, while panel B depicts the number of genes under control by ancient, i.e., discovered long ago, miRNA. Note that the scale between the two panels changes, with panel B reaching a maximum value that is only about a fourth of the scale of panel A.
The target scan algorithm also led to the quantification of the number of genes under control of the 41 novel miRNAs that were discovered in the whitefly. Their names and the number of genes they are predicted to control are shown in Fig. 6.

![Figure 6. Target Prediction of Novel miRNAs.](image)

Interestingly, the novel miRNAs target, on average, about the same number of genes as the known miRNAs, around 2,000 genes. However, as is the case for the known miRNAs, some of the novel miRNAs target well over that number. For instance, the tenth novel miRNA targets just over 18,000 genes, which is fascinatingly close to the number of genes controlled by the most regulatory known miRNA, miR-6006-3p (just under 18,000 genes). Another of the novel miRNAs, the 31st, targets about 10,000 genes, and three others, n7, n11, and n9, all target around 4,000 genes. Since these five mentioned newly identified miRNAs all target such a high number of genes, further research into the
presence of these miRNAs in other insects and the mechanisms by which they regulate gene expression could lead to fascinating finds. If these miRNAs are identified in other insects and are found to be just as highly regulatory, these miRNAs may play critical roles in gene expression in insects.

The DEseq analysis results, shown in Fig. 7, reveal that feeding the whitefly dsRNAs does not impede the ability of the miRNAs to function normally, indicating that the RISC complex is not saturated with dsRNAs to a point where miRNAs cannot bind. This validates the use of dsRNAs in the laboratory, as dsRNAs are fed to organisms in the lab in order to start the process of RNAi.

Figure 7. DEseq After dsRNA Feeding. The DEseq analysis, performed using a BioO kit to perform adapter ligation-mediated PCR, revealed that the feeding of dsRNAs to the whitefly does not alter the activity of the miRNAs.

The target scan analysis allowed for the quantification of the number of target sites in each mRNA that are targeted by miRNAs in the whitefly. The results are displayed in Fig. 8., showing that 5,658 of the mRNAs, 71% of them, have between one and fifty target sites. There are some mRNAs that have many more target sites, as shown by the two mRNA that have between 351 and 400 sites, one mRNA that has between 301 and 350 sites, and eight mRNA that have between 251 and 300 sites. From this data, a
A table of the top ten mRNAs (the mRNAs with the highest number of predicted target sites) was constructed with their names, number of target sites, and functions. This data is shown in Table 3. The mRNA with the largest number of predicted target sites was predicted to have 379 target sites and codes for a brain tumor protein.

![Distribution of Gene Regulation](image)

**Figure 8. Distribution of Whitefly mRNA Target Site Values.** The number of target sites was measured for each of the mRNAs, and the number of mRNAs with a certain number of target sites was graphed.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th># TARGET SITES</th>
<th>FUNCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bta13978</td>
<td>379</td>
<td>Brain tumor protein</td>
</tr>
<tr>
<td>Bta09866</td>
<td>369</td>
<td>Actin-dependent regulator of chromatin subfamily D member 1</td>
</tr>
<tr>
<td>Bta02153</td>
<td>329</td>
<td>F-box only protein 33</td>
</tr>
<tr>
<td>Bta13682</td>
<td>296</td>
<td>Zinc finger protein 710</td>
</tr>
<tr>
<td>Bta10500</td>
<td>290</td>
<td>Serine/threonine-protein phosphatase</td>
</tr>
<tr>
<td>Bta14082</td>
<td>277</td>
<td>Trinucleotide repeat-containing gene 6A protein</td>
</tr>
<tr>
<td>Bta10499</td>
<td>264</td>
<td>Cyclic AMP-responsive element-binding protein 3-like protein 4</td>
</tr>
<tr>
<td>Bta02501</td>
<td>260</td>
<td>RNA polymerase II elongation factor ELL</td>
</tr>
<tr>
<td>Bta10421</td>
<td>256</td>
<td>E3 ubiquitin-protein ligase CBL, putative</td>
</tr>
<tr>
<td>Bta00482</td>
<td>255</td>
<td>RNA-binding motif, single-stranded-interacting protein 1</td>
</tr>
</tbody>
</table>

**Table 3. Top Ten Genes Targeted by Whitefly miRNAs.** The target prediction analysis results were sorted by number of target sites in order to find the top ten genes with the largest number of target sites. These genes are predicted to be regulated the most by the miRNAs in the whitefly. Their number of target sites and functions are given.
Lastly, in order to quantify the role that miRNAs play in regulating mRNAs in the whitefly, the percent of mRNAs predicted to be regulated by miRNAs was calculated and found to be 51%, which is consistent with previous predictions (Li, Flynt, Kim, Solnica-Krezel, and Patton, 2008). Of these mRNAs targeted by miRNAs, the percent targeted by 100 or more miRNAs was calculated to be 4%, indicating there are numerous RNAi pathways that can be manipulated in order to regulate the expression of the genes these mRNAs encode. The results of these two calculations can be seen in Fig. 9 and 10.

**Figure 9. Percent of mRNAs Regulated by miRNAs.** The percent regulated value was found by dividing the number of three no header miRNAs we discovered by the total number in the whitefly genome. We found that 51% of the mRNAs in the whitefly are regulated by miRNAs, which is consistent with previous findings.

**Figure 10. Percent of mRNAs Targeted by 100 or More miRNAs.** The percent of mRNAs targeted by 100 or more miRNAs was found by dividing the number of mRNAs targeted by 100 or more miRNAs by the total number of mRNAs targeted by miRNAs in the whitefly. 4% of the mRNAs in the whitefly are targeted by 100 or more miRNAs.
Chapter 4: Conclusions

Of the 177 miRNA molecules that were categorized as confident and analyzed in this study, 23% were newly discovered, opening the door for future study into how these miRNAs and their target mRNAs can be manipulated or controlled via RNAi. From the analysis of shared miRNAs between the whitefly and its related species, an interesting discovery was made of a large number of miRNAs that have been lost in the order Orthoptera. This raises questions about whether certain species of animals gain or lose miRNAs at different rates, since the other orders have lost, at most, half of this number. Since currently there are no studies that explain this phenomenon in the Orthopterans, future research investigating gene regulation in this order of insects may be able to offer insight into how this massive loss of miRNAs has occurred.

The target prediction analysis revealed that both the known and novel miRNAs target a similar number of genes, indicating that the newly predicted miRNAs play just as large of a role in the whitefly as the already discovered miRNAs. Thus, further research into the novel miRNAs may lead to a great number of discoveries of important genes that can be regulated using RNAi.

The top mRNAs (those that have the highest number of miRNA target sites) analyzed in this study code for various important proteins, ranging from a brain tumor protein, a Zinc finger protein, and a ubiquitin-protein ligase. These findings indicate that the RNAi pathways that regulate these specific proteins could lead to potential breakthroughs in certain diseases involving these proteins.

After analyzing the miRNAs predicted in the miRDeep analysis and their complimentary mRNAs, it was predicted that 51% of the mRNAs in the whitefly are
being regulated by miRNAs. Further, 4% of these mRNAs were predicted to be targeted by 100 or more miRNAs, indicating that there are a vast number of RNAi pathways that can be studied to learn more about the regulation of these specific mRNAs. Synthetic miRNAs can be designed for the miRNAs that expressed highly in our dataset, in order to use them in RNAi pathways to regulate the many mRNAs they control.

This study led to the identification of 41 novel miRNAs, and further research into which genes they control and the RNAi pathways they are involved in may reveal interesting finds. Future studies can reference Fig. 6 to identify the novel miRNAs that exhibit the most regulatory activity and then investigate which genes these highly active miRNAs regulate as well as the pathways and mechanism by which they control gene expression. In addition, the tracing of the loss and gain of miRNAs between the whitefly and related species raises interesting questions about the rate of evolution among these species. Since no study to date has attempted to explain evolutionary rates of gain and loss of miRNAs, future research into these orders of insects and the mechanisms by which these insects have lost or gained miRNAs may offer insight. Another question that remains to be answered is if other orders of animals have shown this phenomenon as well. We have predicted which miRNAs in the whitefly exhibit the most regulatory activity, as well as the number of target sites in the mRNAs they control, setting the ground for further research and manipulation of these RNAs for the identification of more effective RNAi pathways in the whitefly.
References


