Spring 5-2019

Fate of Ingested RNA in the Two-Spotted Spider Mite

Obrie D. Scarbrough
University of Southern Mississippi

The University of Southern Mississippi

Fate of ingested RNA in the two-spotted spider mite

by

Obrie Scarbrough

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

May 2019

Approved by:

Alex Flynt, Ph.D., Thesis Adviser
Professor of Biological Sciences

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

Ellen Weinauer, Ph.D., Dean
Honors College

Abstract

RNA interference, or RNAi, is a gene regulation mechanism that uses small noncoding RNAs (sRNAs) to silence the expression of certain genes. The application of RNAi has been extended to insect pest control. The two-spotted spider mite *Tetranychus urticae* is a persistent agricultural pest that tends to develop pesticide resistance at an alarming rate, making it a perfect candidate for RNAi technology development. It was hypothesized that unique sRNAs could be isolated from RNA soaked spider mites, and new synthetic RNAs could be synthesized to elicit greater knockdown than was achieved in previous studies. To perform this research, a small RNA column filtration method that allows for enrichment of the organism's endogenous functional RNAi effectors, which are small RNAs from 18 to 31 nucleotides long, was used on mites soaked in different RNA molecules and fed on four diverse species of plants. After sequencing the column-extracted RNA fraction, there was a predominance of 18-21 nt reads with a 1-2 nt offset. This is consistent with RNA processing by an enzyme called Dicer and is observed in other arthropods. However, antisense reads mapped to regions outside of the double-stranded RNA (dsRNA) showed a clear "T" at the 5' end and an "A" at the 10^{th} base from the 5' end, which is suggestive of a different processing method of RNAi that is not dependent on Dicer and not seen in insects. It was determined that the majority of the recovered RNA fragments were derived from plant chloroplast RNA, which is slower to degrade during mite digestion of plant materials. Common reoccurring structures from the plant chloroplast RNA data collected from mite specimens were then curated. Synthetic RNA structures were created and fed to the mites in an attempt to elicit a superior inhibition of target gene expression. Most synthetic RNAs fed to mites

exhibited knockdown equal to or greater than actin-targeted dsRNA fed to mites, indicating that standard RNAi strategies could be improved. This work highlights the changing nature of RNAi during animal evolution and the value of detailed analysis in a particular species to formulate an optimal strategy for controlling gene expression with this technology.

Keywords: RNAi, sRNAs, *Tetranychus urticae*, column filtration, Dicer, chloroplast.

Acknowledgements

I would like to thank Dr. Alex Flynt, my thesis advisor, for giving me an opportunity to earn hands-on research experience in my field, and a place to belong in the Department of Biological Sciences. He always answered my many questions, usually more than once, and encouraged me to continue on in my research, even after my tenth failed attempt at qPCR. I could not imagine a better advisor and mentor to perform my research under, and I would recommend him to anybody interested in Molecular Biology, Bioinformatics, or research in general.

I would like to thank my first graduate student, Mosharrof Mondal, for laying the ground work for this study and showing me how to perform multiple different research techniques. After Mosharrof Mondal became Mosharrof Mondal, Ph. D., and continued to a post-doctoral position, Jacob Oche Peter stepped in and helped me to complete the rest of my thesis. Good luck Jay on completing the remainder of your Doctoral program!

Table of Contents

List of Figures …….……viii

List of Abbreviations……..ix

Chapter 1: Introduction……1

Chapter 2: Literature Review……...……2

Chapter 3: Methodology……..5

3.1: Mite care ……..…….…..5

3.2: RNA preparation…..…..…….…..6

3.3: Soaking of mites …..…….……..6

3.4: Column filtration …..…….……..7

3.5: TRIzol® total RNA extraction…….……8

3.6: RT- qPCR……..…...…9

3.7: TrueSeq Small RNA Library Prep Kit……..…9

3.8: Bioinformatics pipelines and software used……...10

3.9: Chloroplast genome accession numbers…….11

3.10: Data visualization……..12

Chapter 4: Results……..12

Chapter 5: Discussion ……23

References……..26

List of Figures

Figure 1. Main pipeline used to quantify RNA sequencing data……..10

Figure 2. Comparison of miRNA length sequences obtained from column purified RNA and from total RNA extraction……13

Figure 3. RNA extracted from GFP and actin dsRNA soaked mites …….…..14

Figure 4. dsRNA expression in actin and GFP soaked mites……...15

Figure 5. Plant based data derived from RNA extracted from mites fed on *P. vulgaris, Arabidopsis,* and *Brachypodium*…….....……...17

Figure 6. Size distribution and expression rates for reads occurring in *P. vulgaris, Arabidopsis, Brachypodium,* and moss……...……..19

Figure 7. Expression of the synthetic structures……...……...21

List of Abbreviations

cDNA	complementary DNA
ddH_2O	double distilled water
dsRNA	double stranded RNA
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
GFP	green fluorescent protein
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IGV	Integrated Genome Viewer
KOAc	potassium acetate
ML	master loci
miRNA	micro RNA
mRNA	messenger RNA
nt	nucleotide
PBS	phosphate buffered saline
PMSF	phenylmethylsulfonyl fluoride
piRNA	PIWI interacting RNA
RT-qPCR	real time quantitative polymerase chain reaction
RNA	ribonucleic acid
Rdrp	RNA dependent RNA polymerase
RISC	RNA inducing silencing complex
RNAi	RNA interference
rpm	rotations per minute

siRNA	small interfering RNA
sRNA	small exogenous noncoding RNA
rRNA	ribosomal RNA
tRNA	transfer RNA

Chapter 1: Introduction

RNA interference, or RNAi, is a gene regulation mechanism where small exogenous noncoding RNAs (sRNAs) form complexes with complementary transcripts, thereby silencing the expression of certain genes[1]. The silencing molecules used in RNAi pathways in eukaryotic systems vary by clade and can be broadly separated into three major classes: microRNA (miRNA), small-interfering RNA (siRNA) and PIWI interacting RNA (piRNA)[1]. Across eukaryotic phyla, the function of each class of small RNA molecules existing in one clade or species may vary from those existing in others. In addition to the variation in type of sRNAs found in various clades, sRNAs also have major features that can be used to distinguish one sRNA type from another. These differences may range anywhere from type of protein they associate with, length of RNA, 5' U-bias, 5' mono- or tri-phosphorylation, 3' terminal 2'-O-methylation, and biogenesis, to type of transcript they ultimately target[2-5]. As a general rule of thumb, sRNAs usually either hybridize in a complementary fashion to generate dsRNAs with the target transcript, thereby preventing its translation, or sRNAs guide proteins of the Argonaute family to target and cleave transcripts in a sequence specific manner. Both pathways bring about post-transcriptional gene regulation[6].

The two-spotted spider mite, *T. urticae*, is a successful arthropod that can have a substantial economic effect. Spider mites are an exceptionally polyphagous agricultural pest that can feed on as many as 1100 plant species and develop a resistance to common pesticides in one year's time[6]. Their reproduction is favored by dry, warm climates, which suggests that the two-spotted spider mite will become an ever-increasing burden

on agriculture through climate change, and it is questionable if conventional pesticides will be able to maintain control over their increased reproductive rates[6].

The aim of this research was multifaceted. First, the goal was to obtain an RNA purification method that would filter out all RNA fragments that were smaller than 18 nucleotides and larger than 31 nucleotides[26]. Once a successful method was found, it was used to process and analyze the sRNA sequencing data obtained from mites reared on multiple plant species. Then, Next Generation sequencing and different computational pathways were used to find any reoccurring structures that were not being broken down by Dicer. If the RNA structure could survive Dicer processing inside the mite, then it might be used to elicit some sort of knockdown in gene expression levels that would result in a population reduction. With this data, synthetic RNAs were created to later feed to the two-spotted spider mite in an attempt to generate a greater knockdown in expression levels than reported in previous studies[20, 21]. Long term applications of this research could lead to a successful but unconventional method for spider mite population control through the use of RNAi.

Chapter 2: Literature Review

miRNAs are generally processed from single stranded hairpin-like precursor transcripts, first in the nucleus and then in the cytoplasm, by two RNAse III enzymes: Drosha and Dicer proteins, respectively, into approximately 22 nt long mature ssRNAs. This class of sRNAs is well conserved in the majority of eukaryotes and known to be found associated with the AGO subfamily of the Argonaute proteins, directing them to

target and regulate messenger RNAs (mRNAs)[7] and subsequent gene expression via mRNA cleavage or translation inhibition. A second class of sRNAs, siRNAs, is found to be approximately 21 nt long and generated either endogenously from mRNA, transposons, repetitive regions of the genome, inverted repeat transcripts, or exogenously either from viral RNAs or a transgene[5]. In addition to regulating mRNAs, siRNAs have been shown to act as genome defenders targeting and silencing the activities of transposable elements. This group, unlike miRNAs, bypasses Drosha processing and can be recognized by their 2 nt 3' overhangs typical of Dicer processing[8]. Additionally, in some organisms such as *Caenorhabditis elegans* and *Arabidopsis thaliana* the protein RNA dependent RNA polymerases (Rdrp) can amplify siRNA production due to the ability of this protein to generate either long or short (Dicer-independent) dsRNA from ssRNA templates[9-11]. siRNAs made directly by Dicer cleavage of endogenous, exogenous, and viral dsRNA can be distinguished from Rdrp-associated siRNAs, as Rdrp is known to tri-phosphosphorylate the 5' end of their product[10,12].

The piRNA class sRNAs are between 24 and 31 nt in length and was mostly thought to be gonad specific until the recent discovery of somatic piRNAs in arthropods[13-15]. The length of this sRNA class is one way to distinguish them from their counterparts. In addition to this characteristic, their hallmark is the presence of a 5' U-bias, a usually 3' terminal 2'-O-methylation, and their role in to silencing transposable elements. piRNAs are associated with PIWI proteins, unlike miRNAs and siRNAs that load into AGO proteins to form the RNA inducing silencing complex (RISC). Two mechanisms produce piRNAs: a primary pathway that converts licensed transcripts, and an amplification loop called the ping pong amplification cycle. Amplification proceeds

through collaborative cleavage of partner piRNAs, such as AGO3 and Aubergine in *Drosophila*[16,17].

In insects, antiviral pathways exploit the use of a long dsRNAs strategy to inhibit the replication of RNA and some DNA viruses[18]. Efficiency of RNAi varies widely between insects which is reflected in the poor conservation sequence conservation of pathway factors found in these groups. In *Drosophila*, for example, antiviral RNAi is mainly triggered by viral dsRNA molecules and is critical to suppress viral infection[9].

The use of long dsRNA to elicit RNAi has been particularly successful in the control of the western corn rootworm beetle, *Diabrotica virgifera virgifera*[19]. Companies have introduced commercial products that use RNAi technology to reduce the survival rate beetle larva by inserting invert repeat transcript expressing transgenes into the maize genome[19]. A recent study on *Tetranychus urticae* attempted to elicit RNAi using various methods to deliver dsRNA[20]. Although successful, it showed low efficiency in comparison to what was seen with *D. virgifera virgifera*. These differences, emerging from the divergent nature of RNAi pathways between chelicerates and insects, is the rationale for this study which seeks to look into the biogenesis of exogenously-derived siRNAs in spider mites.

Unlike the well-studied *Drosophila* system, spider mites possess Rdrp as well as significant siRNA producing loci[15]. The two-spotted spider mite possesses a unique genomic feature called siRNA Master Loci (ML) that appear to operate upstream of piRNA production, which is opposite from piRNA-based genome surveillance mechanisms in *C. elegans*[15]. This unusual combination of RNAi factors in spider mites make their RNAi pathway different from most arthropods.

Spider mites have the capacity to metabolize a wide range of exogenous transcripts from the plants that they eat [21]. Therefore, it is possible that these mites can ingest some transcripts more efficiently than others [22,23]. Similarly, the RNAi machinery of the animal might metabolize certain RNA conformations better than others. For any effective RNAi approach to control spider mite populations by producing the RNAi trigger in the host plant, the first strategy would be to ensure that target animals ingest the effector molecule efficiently. In a previous study it was determined that soaking in long dsRNA was a successful option for modest induction of RNAi in the spider mite[20].

Chapter 3: Methodology

Mite care

Mites were reared on *Phaseolus vulgaris* plants throughout the entirety of their life cycle, except for the experiments discussed in Figures 5 and 6. For those steps of the study, separate colonies were reared on each plant discussed: *Arabidopsis thaliana*, *Brachypodium distachyon*, *Phaseolus vulgaris,* and *Selaginella uncinata.* During extraction, mites were dislodged by gentle tapping from the plant leaves into an Eppendorf tube, both juveniles and adults were kept from this leaf extraction and used in the protocols described below. All samples consisted of approximately 0.1 mL of mites by volume determined by markings on a microfuge tube.

RNA preparation

For the initial experiment, two dsRNAs were generated from cloned fragments of *T. urticae* actin and of Green Fluorescent Protein (GFP) from *Aequorea victoria* (accession numbers CAEY01002033.1 and FJ172221.1, respectively). The positive control, actin RNA, and negative control, GFP RNA, were both approximately 350 nt long and created using the MEGAscript™ *in vitro* transcription kit (Ambion). Templates for dsRNA synthesis were PCR products amplified with primers possessing T7 RNA polymerase promoter sequences on their 5' ends. Following *in vitro* transcription, lithium chloride precipitation was performed to purify products, and the resulting RNAs were annealed through heating to 95 ^{o}C followed by cooling to 4 ^{o}C.

Four synthetic structures were designed from RNA folds present in chloroplast transcripts that were observed to yield small RNAs. Templates for these structures were created by annealing long, approximately 100 nt, oligonucleotides. The resultant dsDNAs encoded a T7 promoter at the 5' end followed by the sequence designed to mirror the chloroplast RNA folds. Short synthetic RNAs were synthesized using the MEGAshortscript™ *in vitro* transcription kit (Ambion). The synthesis products were purified by phenol chloroform extraction, and ethanol precipitation.

Soaking of mites

The soaking of the mites was done by collecting specimen from *P. vulgaris* plants into a 1.5 mL microfuge tube. Then 200 µL of a 160 ng/µL solution, consisting of dsRNA and nuclease-free water, was placed into the Eppendorf tube. The mites were

then soaked overnight (15-16 hours) and rinsed from the microfuge tube using approximately 2 mL of a PBS, pH 7.4, and 0.1% Tween 20 solution. Afterwards, they were collected on a paper towel and allowed to partially dry before being placed onto a *P. vulgaris* leaf inside of a large petri dish at room temperature. After five days, mites were washed from the petri dishes and collected in one tube, which was flash frozen and stored at -80 ^{o}C.

Column filtration

Column purification of sRNAs used a 1 mL HiTrap Q FF column (GE Lifesciences) was used for RNA extraction[24]. Columns were equilibrated following manufacturer instructions. First, 5 mL of start buffer (20 mM HEPES-KOH, pH 7.9) was applied and passed through the column (1 mL/minute). Then 5 mL of elution buffer (20 mM HEPES-KOH, pH 7.9, 1 M NaCl), and 10 mL start buffer (20 mM HEPES-KOH, pH 7.9, 100 mM KOAc) were applied sequentially.

Bulk harvested animals, across all developmental stages, were washed several times with PBS, pH 7.4, followed by flash freezing and grinding with a mortar and pestle. One mL chilled binding buffer (20 mM HEPES-KOH pH 7.9, 100 mM KOAc, 0.2 mM EDTA, 1.5 mL $MgCl_2$, 10% glycerol, 0.2% PMSF, 1 mM DTT, 1X Roche EDTA-free protease inhibitor cocktail) was mixed with the pulverized animals. The lysate was then clarified by centrifuging for 30 minutes, 13,000 rpm at 12^{o}C. Clear lysate was applied to the column and passed at a speed of 1 mL/minute. The column was washed with 1 mL binding buffer followed by applying 2 mL elution buffer (binding buffer with 300 mM KOAc). The Pass-through and eluted products were collected in a 15 mL conical tube on

ice. An equal volume (approximately 3.5 ml) of acid phenol-chloroform was added to the tube, and the tube was rocked at room temperature for 10 minutes. Next, the content of the 15 mL tube was distributed equally between 1.5 mL Eppendorf centrifuge tubes. The tubes containing the newly divided product were then centrifuged for 15 minutes, 13,000 rpm, at 4^{o}C. The aqueous layer of each sample was transferred to new Eppendorf tubes, and 600 µL of isopropanol was added to each. The solutions were vortexed and placed in -20^{o}C storage overnight. They were then centrifuged for 15 minutes, 13,000 rpm, at 4^{o}C. The supernatant was disposed of and 600 µL of 70% ethanol was added to each tube, and the samples centrifuged again for 15 minutes, 13,000 rpm, at 4^{o}C. Approximately 400 µL of the supernatant was removed. Each tube was then vortexed, the contents were combined into a single Eppendorf tube, and centrifuged for 15 minutes, 13,000 rpm, at 4^{o}C. The final supernatant was removed, the pellet was resuspended in 30 µL of ddH_2O, and the sample was stored at -80^{o}C. The above protocol was adapted from that of Lau *et al.* [24].

TRIzol® total RNA extraction

1mL of TRIzol® Reagent was added to a 1.5 mL Eppendorf tube containing approximately 0.1 mL *T. urticae* sample and homogenized. The sample was rocked for 10 minutes at room temperature, then centrifuged for 7 minutes, 13,000 rpm, at room temperature. The supernatant was transferred to a new tube and 200 µL of chloroform was added before rocking at room temperature for 7 minutes. The sample was centrifuged for 15 minutes, 13,000 rpm, at 4^{o}C. The aqueous layer was transferred to a new tube and 600 µL of isopropanol was added. The solution was placed in -20^{o}C

storage overnight. It was then centrifuged for 15 minutes, 13,000 rpm, at 4^{o}C. The supernatant was disposed of, 600 µL of 70% ethanol was added and the tube centrifuged for 15 minutes, 13,000 rpm, at 4^{o}C. The supernatant was removed, and the pellet was resuspended in 30 µL of ddH_2O. This is the standard manufacturer's protocol.

RT-qPCR

100 ng/ µL of column purified RNA from spider mite samples was used in cDNA synthesis (Thermo Fisher T-7 kit) using random hexamer primers. The following cDNAs were then used in qPCR assays with SYBR Green real-time PCR master mix (Thermo Fisher), and the manufacturer's protocol was used. In the qPCR assay, actin transcripts were quantified using a primer pair targeted to its mRNA. Expression of actin mRNA was normalized by assessing levels of 18S ribosomal RNA with an appropriate primer pair. Primer sequences were taken from prior publications[15,20].

TrueSeq Small RNA Library Prep Kit

RNA obtained from both column extraction and total RNA extraction was sent for sequencing to the University of Mississippi Medical Center Molecular and Genomics Core Facility. The genomics facility uses an Illumina TrueSeq Small RNA Library Prep Kit, a form of Next Generation Illumina DNA sequencing. The first step of the process is to ligate adapters onto the RNA fragments, and then reverse transcription is performed on the RNA to create cDNA and amplify the reads. Once the cDNA is made, PCR adds the adapter required for the Illumina sequencing platform. Libraries of the cloned RNAs were then sequenced on an Illumina NextSeq500 machine using a single-read 50 base

pair protocol. Illumina sequencing uses fluorescent reversible terminator nucleotides to determine identity of bases in immobilized DNA fragments.

Bioinformatics pipelines and software

For this research, several bioinformatics programs and pipelines were used to include: Fastx toolkit[25], Bowtie[26], Samtools[27], Bedtools[28], and RNAfold[29]. A large majority of the figures were created using R, a statistical programming language, through the ggplot2[30], Seqlogo[31], qqman[32], Pheatmap[33], and Sushi[34] packages.

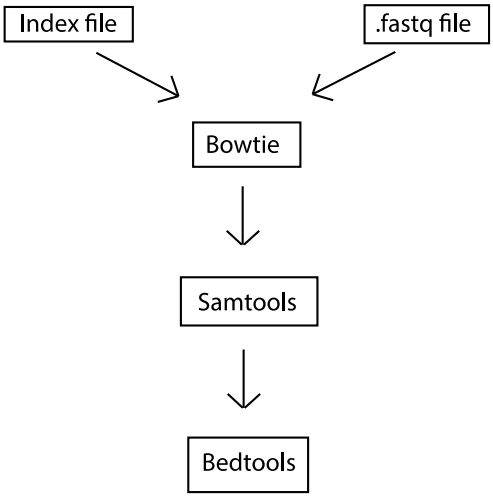

Figure 1- Main pipeline used to analyze RNA sequencing data

Adapter sequences were removed from raw sequencing data files ("clipping") using the Fastx toolkit. As shown in Figure 1, Bowtie was used to align the data with the index file of the chloroplast genome, nucleotide eu196765.1, for the *P. vulgaris* on which the mites were fed, and the parameters "-v 0, –no-unal, and -S" were used. The first two parameters given Bowtie were to only output data that aligned perfectly with the

chloroplast genome for the plant, so any data that did not coincide with these presets was not output to a file. The last parameter instructed Bowtie that the new file should be output in a sequence alignment (SAM) format. A SAM file is a text file that contains sequencing alignment data, which is the format needed to perform the following step in the pipeline. Next, the SAM file was processed through Samtools into a binary alignment map (BAM). BAM files can be loaded in genome browsers like the Integrated Genomic Viewer (IGV) alongside ".bedgraph" files to view the genomic alignment data. The next step in the pipeline was to use Bedtools genomecov, Bedtools merge, and Bedtools multicov to create ".bedgraph" files, merge them, and lastly create expression count tables that categorically list the size of the loci and the number of reads per locus.

RNAfold was used to create a two-dimensional figure of each locus with higher than average expression rates to determine the structures that survived the Dicer pathway. To determine this, Bedtools getfasta was used on the previously created bed files to create a ".txt" file. The text file of the data could then be used in RNAfold to output individual ".png" files of each locus listed in the text file.

Chloroplast genome accession numbers

P. vulgaris- nucleotide eu196765.1

A. thaliana- AP000423

B. distachyon- NC_016131.3

S. uncinata- AY723749

Data visualization

Figure 2A and B was created from data recovered from a Bioanalyzer instrument, which can assess the size, quantity, integrity, and purity of DNA, RNA, and proteins. The RT-qPCR results from Figure 7B was the gene expression graph generated by CFXMaestro, the program that interfaces the RT-qPCR machine. Figures 2C, 2D, 4A, 5A, 5E, 5F, 6E, 6F, 6G, 7C, and 7D were created in the Apple Inc. program "Numbers" and ggplot R package. Figures 3, 6A-D, 7E, and 7F were created using bedtools plot, which created the plot density data, and Sushi to generate the graphs. Figures 4B and C were created using pheatmap. The Manhattan plots in Figure 5C and nucleotide biases graphs in Figures 4E, 4F, and 5B were all created using qqman R package. Figures were laid out in Adobe Illustrator.

Chapter 4: Results

To assess the fate of dsRNAs ingested by spider mites, a sequencing approach for capturing small RNA was used to uncover abundance along with biogenesis patterns. In order to reduce contamination of degradation fragments, a Hi Trap QFF chromatography method was used to enrich for Ago/PIWI associated RNAs as previously described (Fig 2A). Lengths of RNAs isolated from fractionated spider mite lysates showed a single

peak at the sizes expected for small regulatory RNAs (Fig 2A). In contrast, RNAs recovered with the TRIzol® method showed a heterogeneous collection of RNA species (Fig 2B). Moreover, comparison of small RNAs sequenced from each method showed column isolated RNAs to have a bi-modal size distribution consistent with enrichment of target transcripts such as miRNAs, siRNAs (20-22 nt) and piRNAs (25-30 nt) (Fig 2C). In the TRIzol® sample, more reads were seen at small sizes. Mapping of reads to the *T. urticae* genome showed significantly more miRNA derived reads compared to non-coding RNAs (rRNA & tRNA) (Fig 2D).

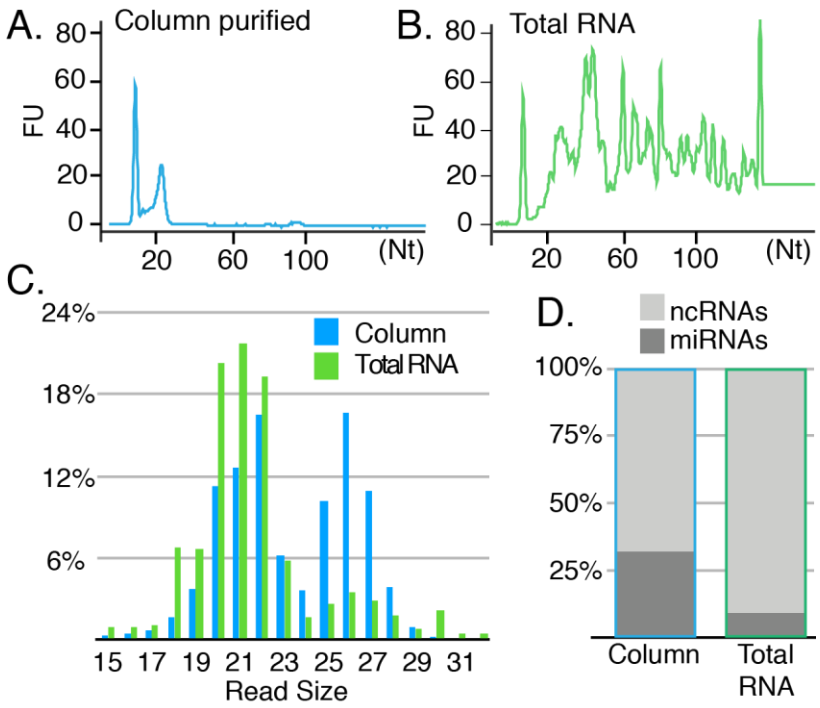

Figure 2- Comparison of miRNA length sequences obtained from column purified RNA vs Total RNA extraction. (A) The Hi Trap QFF Column method for RNA extraction had RNAs that were between 18 and 26 nucleotides long. (B) The TRIzol® Total RNA extraction method had the same RNAs that were kept by the column method, but it also had a high percentage of 30-130 nucleotide sequences. (C) The different percentages of read sizes for the sequencing results obtained with the two RNA extraction methods. (D) When compared to the reference libraries for *T. urticae*, the column extracted RNA was made up of approximately 30% miRNAs, while the Total RNA extraction method had 5% coding miRNAs, with the rest being composed of non-coding RNAs.

Using this chromatography approach, the production of small RNAs derived from ingested long synthetic dsRNAs was investigated. dsRNAs were generated identical to those targeted to actin that were used to determine techniques appropriate for RNAi experiments in *T. urticae*. The GFP sequence derived dsRNAs were also synthesized to examine processing of non-targeted dsRNA. Animals were subjected to "soaking"

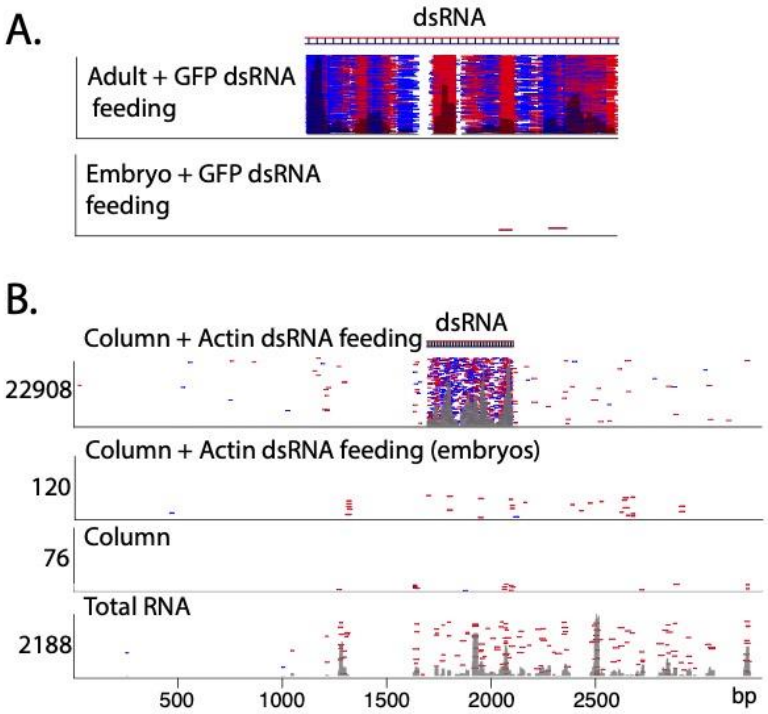

Figure 3- RNA extracted from GFP and actin dsRNA-soaked mites. (A-B) *T. urticae* was soaked in actin and GFP dsRNA solutions, and their RNA was extracted using the Column method. The reads that aligned to the double stranded target regions were then charted and compared to the same position of the non-soaked mites' genome.

experiments as described and subjected to Hi Trap column isolation of small RNAs and small RNA sequencing. Alignment of reads to target mRNAs for actin and GFP showed a considerable accumulation of RNAs at regions corresponding to the dsRNA sequence (Fig 3A and B). In RNAs extracted either by column or TRIzol® without the mites being

soaked in dsRNA, this accumulation was absent with nearly no alignments in column purified sample and apparent degradation fragments in the TRIzol® reagent library (Fig 3B). Although *C. elegans* and other species have shown that RNAi performed in the adult can pass to the germline and subsequent generations[15], *T. urticae* does not seem to exhibit the same pathways. Throughout the 5-day period of the mites being kept in the petri dish post dsRNA soaking, there was a number of embryos laid, but not any noteworthy dsRNA fragments were apparent in the RNA extracted from embryos of either actin or GFP dsRNA-soaked mites (Fig 3).

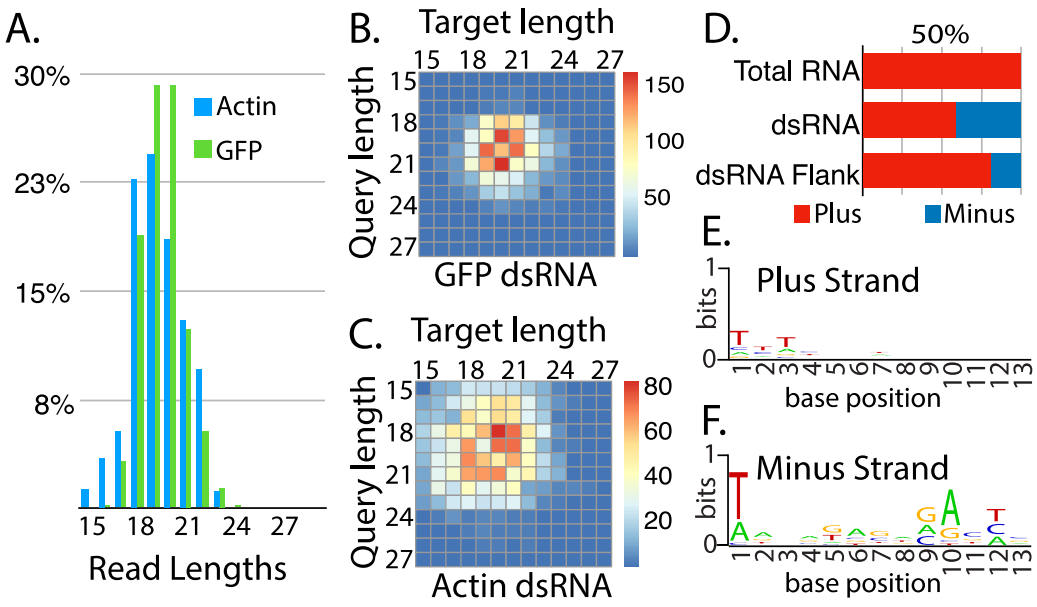

Figure 4- dsRNA expression in actin and GFP soaked mites. (A) The percentage of the different read lengths for both actin and GFP dsRNA-soaked mites. (B and C) Of the reads recovered from the column, there was a high number of 2-nucleotide overhangs with the most common being query lengths of 18-20 nucleotides. (D) Of the RNAs recovered, the Total RNA extraction method yielded all sense strand RNAs, but the column extracted dsRNA-soaked RNA yielded mostly equal amounts of sense and anti-sense reads in the dsRNA region shown in Figure 2A. (E and F) Of the sense and anti-sense reads collected, there was an observable amount of nucleotide biases for T and A at position 1 and 10, respectively.

Distribution of read sizes for small RNAs derived from GFP and actin dsRNAs showed predominantly 18-21nt length reads, consistent with Dicer class products observed in insects and nematodes, however, there was a tendency of actin dsRNA derived reads to be shorter, suggesting that when a target RNA is present, transcript fragments may be gated into small RNA pathways (Fig 4A). This is consistent with the presence of Rdrp homologs in the spider mite genome, which could be involved in converting cleavage products into new, small RNAs. To further investigate biogenesis, abundance of small RNA pairs exhibiting 2 nt overhangs indicative of Dicer processing were assessed (Fig 4B and C). Small RNAs of 15-27 nt were used to query target pairs, also 15-27 nt long, that exhibited an overlap of 2 nt less than the size of query RNA. The number of reads for each combination was then plotted on a heatmap showing a matrix of small RNA pair length combinations that are likely produced by Dicer. Interestingly, the most common pairs were offset by 1-2 nt, for example in GFP dsRNA derived reads 21/20 nt and 19/20 nt pairs were some of the most common combinations. Duplex asymmetry could be the result of post Dicer modifications. Furthermore, actin small RNAs showed significantly more diversity in pair lengths compared to GFP siRNAs, indicating that the presence of a target has significant influence on sRNA length modifications. Asymmetry may be the result of base removal by a nibbler-like activity or extension by Rdrp. Indeed, in actin RNAs there is a bias for query RNAs to be paired with longer target RNAs, suggesting nucleotide addition by Rdrp may have occurred when the mRNA target was present.

Rdrp activity was also evident in the presence of antisense reads mapping to regions outside of the dsRNA region, which are not present in the TRIzol® extracted

library (Fig 4D). Investigation of sequence biases in flanking reads found very little base preference in sense reads, however, in the antisense reads, a clear "T" was seen at the 5' end and an "A" at the 10th base from the 5' end, which is suggestive of piRNA ping pong biogenesis (Fig 4E and F).

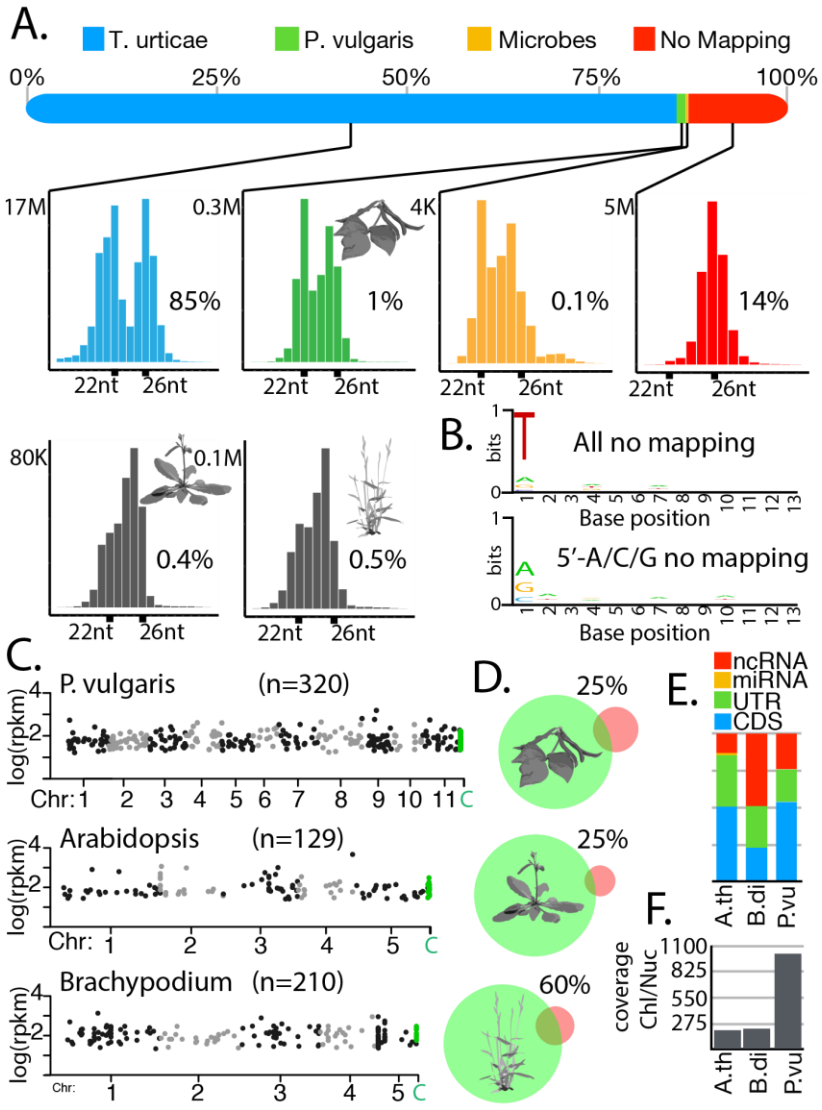

Figure 5- Plant based data derived from RNA extracted from mites fed on *P. vulgaris*, *Arabidopsis*, and *Brachypodium*. (A) The sequencing data was referenced to libraries for *T. urticae*, the plant that the mite was fed on, and common microbes that are associated with the spider mites and the plants. (B) This sequencing data also showed the nucleotide biases discussed earlier. (C) A Manhattan plot of the data that correlated with the plant RNA sequence and the number of loci called. While chromosomal data is present, there is also plant chloroplast RNA maintained. (D) The sequencing data recovered was referenced against 1000 high expressing loci from public libraries. (E) Of the data recovered for each plant, it was mostly miRNAs and un-transcribed regions. However, there is a very small percentage of miRNAs recovered for *A. thaliana*. (F) The ratio of chloroplast to nuclear coverage.

To gain deeper insight into the structural and sequence requirements for entry of ingested RNAs to gain access to RNAi pathways, the fate of transcripts derived from diet was investigated. Spider mites can feed on a variety of plants and are typically raised on bean (*P. vulgaris*) plants in the laboratory. In a sRNA sequencing library constructed from column purified RNAs obtained from animals raised on bean plants, approximately 85% of reads mapped to the spider mite nuclear genome and approximately 1% mapped to bean sequences (Fig 5A). Reads were also mapped to a collection of microorganism sequences, which accounted for only 0.1% of library reads. For RNAs from each origin, a bimodal distribution of read sizes was seen, suggesting a mix of siRNA and piRNA class small RNAs were generated. In addition to bean plants, mites were also raised on *A. thaliana* and *B. distachyon*. The remaining 14% of reads were defined as un-mapping, so it was assumed that these un-mapping reads aligned to areas of the spider mite or plant genome that have not been fully sequenced yet. This set of sequencing data also showed the nucleotide biases discussed earlier (Fig 5B). The data that correlated with the plant RNA sequence and the number of loci called is shown in Figure 5C, while chromosomal data is present, there is also plant chloroplast RNA maintained. The sequencing data recovered was referenced against 1000 high expressing loci from public libraries (Fig 5D). Most the data recovered for each plant corresponded to miRNAs and un-transcribed regions (Fig 5E). However, a very small percentage of miRNAs was recovered for *A. thaliana*. The coverage for the chloroplast and nuclear genome was then mapped and the ratio was shown (Fig 5F). This shows that, in comparison to the plants' nuclear genome, a vastly larger amount of residual chloroplast genomic data is remaining in the mite after ingestion of the plant.

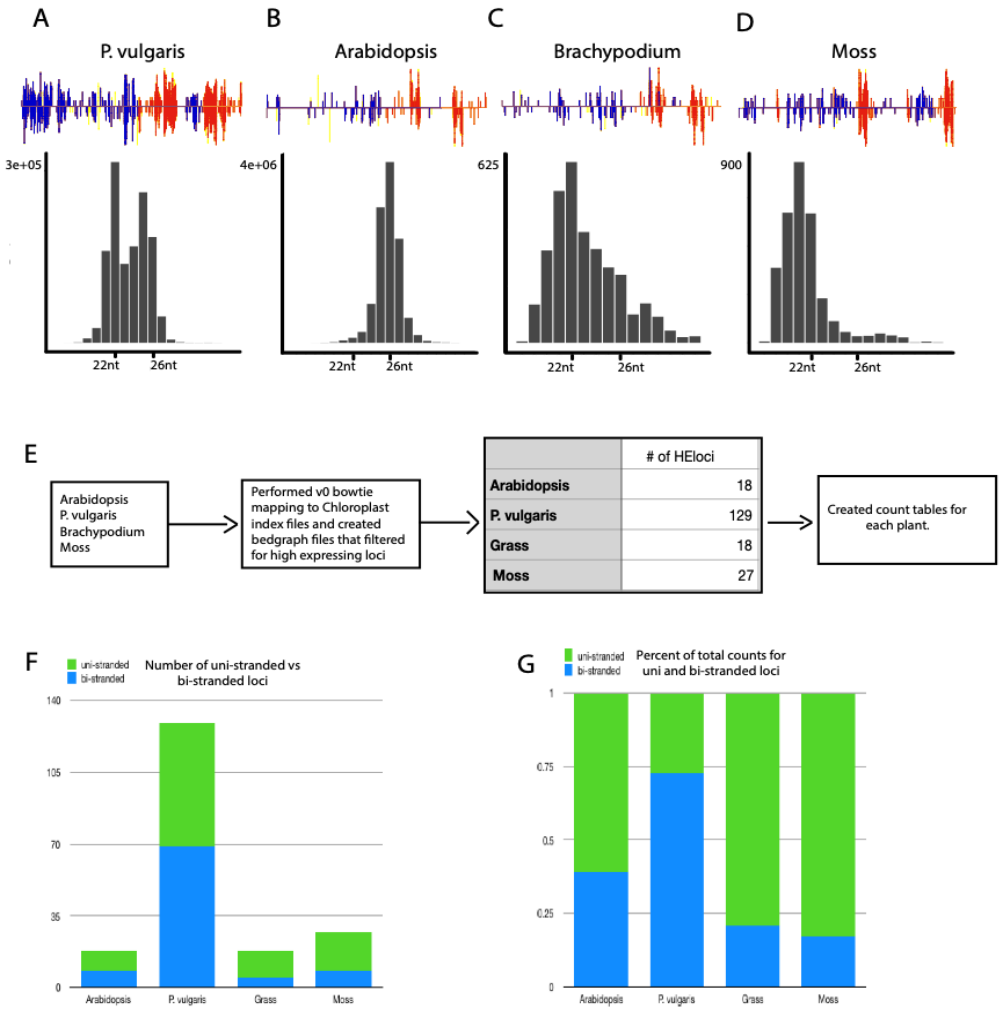

Figure 6- Size distribution and expression rates for reads occurring in *P. vulgaris, Arabidopsis, Brachypodium,* and moss. (A-D) Size distribution for reads occurring in *P. vulgaris, Arabidopsis, Brachypodium,* and moss. The top graphs are read density plots showing the unique (blue) reads and residual (red) reads. The residual reads are found across all four of the plant species. (E) Flowchart depicting the process for selection of synthetic RNAs. (F) Reads recovered were designated as uni-stranded or bi-stranded, and the resulting data was graph. (G) The uni-stranded and bi-stranded loci data was quantified for the number of reads expressed for each type of structure.

Of the plant data recovered, exponentially more reads corresponded to *P. vulgaris* and *A. thaliana* nucleic acids, but all of the data showed RNA fragments from 18-26 nt, with the occasional longer fragment. *Phaseolus vulgaris* had a bimodal distribution of

reads, which suggests a mix of siRNA and piRNA class small RNAs were generated, as previously discussed in Figure 5. The lower expression rates shown for *Brachypodium* and moss could have many explanations, but a lower feeding rate for mites fed on these plants was observed and noted throughout the extent of the research (Fig 6A-D). The top plots of Figures 6A-D are read density plots, and they show the expression of reads from across the genome. The blue peaks account for unique reads only expressed in the chloroplast genome of each individual plant, and the red peaks are present across all four of the plant species and seem to be ribosomal RNA. The plus and minus strands of both the red and blue peaks coincide to the smaller reads (approximately 22 nt long), and these reads are most likely siRNAs. Although small in number for most of the plants, the yellow peaks are longer (26 nt long) and could be some sort of artifact mapping (Fig 6A-D). There was an overall higher expression for the smaller segments, and the larger fragments seemed to be confined to small regions. For the remaining Figures, a bioinformatic pipeline was created and used to quantify the high expressed loci and their read expressions. The information gained from the plant study led to the development of synthetic RNAs that were based on frequently observed, high expressing structures (Fig 6E). These structures were derived from the uni-stranded loci with higher than average expression levels, since RNA fragments that survive Dicer are single stranded. To better determine which loci were expressed the most, count tables were created that quantified the expression rates for each locus (Fig 6F and G). These loci were rendered in a two-dimensional form using RNAfold, and the resulting images were used to determine which structures survived degradation in the mite.

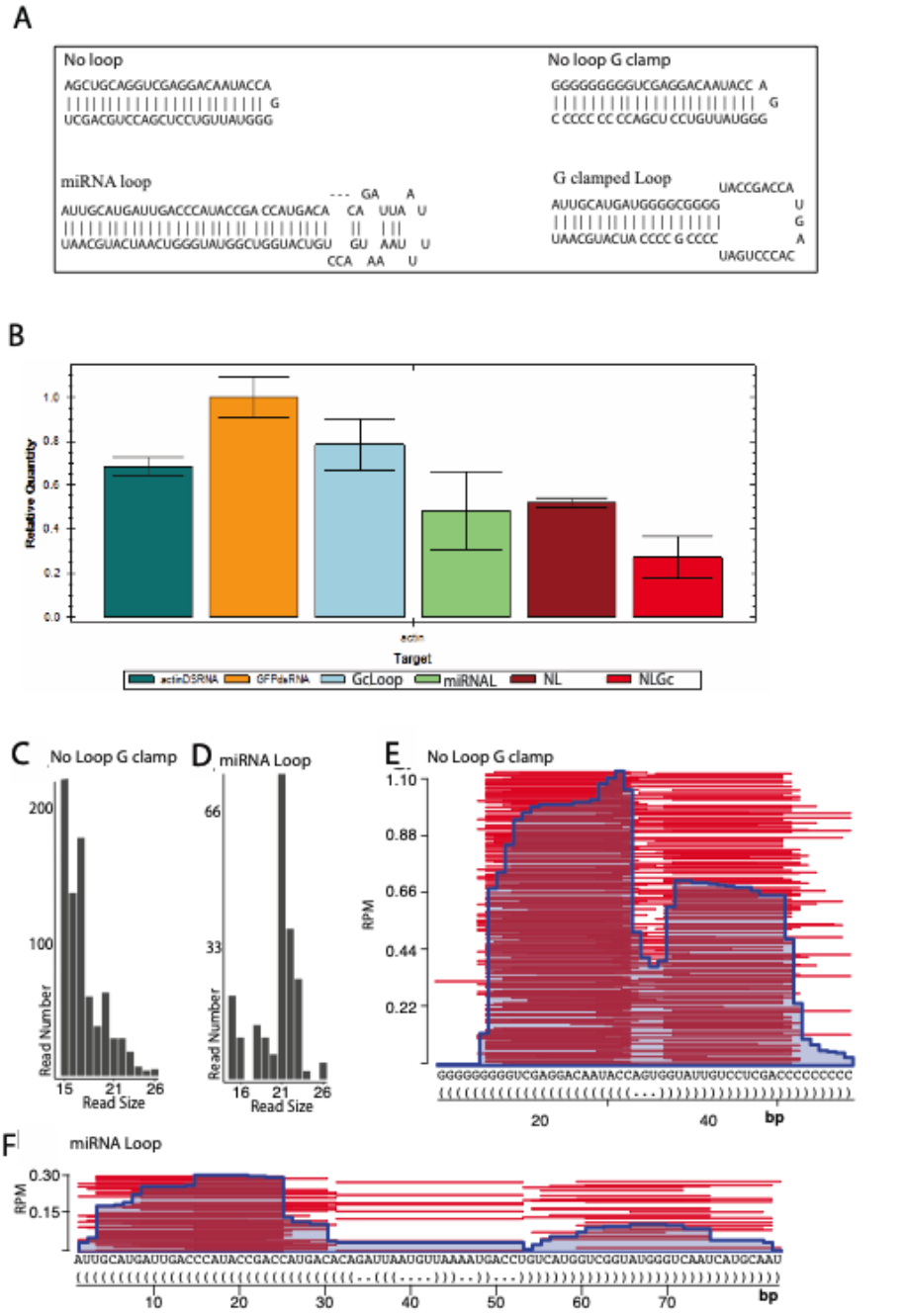

Figure 7- Treatment with the synthetic structured RNAs. (A) Two-dimensional structure of the 4 RNAs that were synthesized. (B) RNA that was extracted using the Total RNA method was used in qPCR assays with actin and GFP dsRNA-soaked mites for positive and negative controls, respectively. (C-D) Number of reads per read size for two selected synthetic RNAs after column extraction. (E-F) Read density plot for the two selected structures.

The synthetic structures created (Fig 7A) using data shown in Figure 6 were fed to the mites, and the subsequent RNA that was extracted using the total RNA method was

used in qPCR assays (Fig 7B). The actin and GFP dsRNA created earlier in the experiment were used as positive and negative controls, respectively. While all of the synthetic structures had greater knockdown than the actin dsRNA, "miRNA loop" and "No loop G clamp" were the structures yielding the highest overall knockdown in gene expression, so they were chosen to proceed to the next phase of the experiment. However, it was noted that there was higher variability in the knockdown efficiency for these RNAs. It was hypothesized that "No loop G clamp" would have greater survivability in the mite than the regular "No loop" structure due to the strength of the G-C bonds in the center of the structure, and "miRNA loop" was structured after a commonly occurring miRNA structure in the mite RNA (Fig 6). "G clamped loop" was designed to assess how well the actin-based loop would survive with G-C bonds directly surrounding the loop.

New mites were soaked in the two selected structures, and after column method RNA extraction and sequencing, the read sizes and expression rates were determined. "No loop G clamp"-soaked mite RNA had higher read numbers and smaller fragments compared to the "miRNA loop"-soaked mite RNA. "No loop G clamp" also generated a large amount of sRNA fragments, but both structures resulted in similar amounts of mature RNA fragments (approximately 21 nt) (Fig 7C and D). Read density plots were then created from the data, and "No loop G clamp" RNA had approximately 80 million more reads than "miRNA loop" RNA (Fig 7E and F). "No loop G clamp" seemed to be cut by Dicer in the center of the hairpin, but due to the structure of "miRNA loop", it was processed into three fragments. However, for both of the structures, the first

fragment at the 5’ end showed greater accumulation compared to sRNAs arising from other segments of the structures (Fig 7E and F).

Chapter 5: Discussion

This study sought to assess the fate of exogenous, ingested RNAs in the two-spotted spider mite to improve research approaches and promote prospects for addressing spider mite infestation with RNAi technology. By using a sRNA column filtration method that enriches for functional sRNAs in conjunction with high throughput sequencing, processing of ingested actin and GFP dsRNA was evaluated[20]. While an RNAi response was evident in adults, none of the embryos from RNA soaked mites exhibited any signs of that knockdown in gene expression. The lack of knockdown in subsequent spider mite generations supports that RNAi in *T. urticae* does not occur transgenerationally, so that successive generations do not inherit the effects. It was also noted that specific sRNA pair length combinations were offset by 1-2 nucleotides. This duplex asymmetry could be the result of post Dicer modifications. The actin dsRNA also had a significantly higher diversity in pair lengths compared to the GFP dsRNA-soaked mites. This means that the presence of a target could have a significant influence on RNA modification, specifically extension by Rdrp. There was also strong evidence of piRNA ping pong biogenesis when looking at the antisense reads of the column extracted actin RNA. This is consistent with prior work that implicated siRNA-class Dicer products as upstream triggers of piRNA production[12,15,16,17]. Together, this data suggests that long dsRNA has an activity different from those found in insect systems.

To further investigate structural and sequence requirements for exogenous RNAs to enter mite RNAi pathways, RNAs introduced through ingested plant material were characterized. Four different species of plants were examined to determine if there was a common theme, i.e., reoccurring RNA structures being recovered. It was discovered that the chloroplast genome of the different plant species was overrepresented in the RNA data sequenced from the mites. This suggests that RNAs originating in the chloroplast of plants were able to survive the Dicer pathway and evade degradation for a longer period of time, and from this chloroplast RNA data, read density plots were created to identify RNA structures with the highest levels of accumulation. These single stranded sRNAs from each set of plant fed mites were compared, and four synthetic structures were created.

The RT-qPCR results for mites soaked in each of these synthetic structures all showed a knockdown in target gene expression equal to or greater than the positive control, mites soaked in the actin dsRNA. Lastly, a read density map was created for the RNA data to help determine which nucleotide components were responsible for the greater knockdown capabilities. The structure with the greatest knockdown had a larger number of 15-18 nt fragments, but the same amount of 21 nt reads was discovered for both structures. siRNAs are also 21 nt long and are believed to be one of the main effectors in spider mite RNAi. This suggests that the synthetic structures were broken down by Dicer into functioning siRNAs, and this could have contributed to the greater knockdown in gene expression.

While these experiments produced many important insights, further studies will be needed to determine if there is a way to promote RNAi in the mite's gonad, so that

future generations inherit changes in gene expression. Also, the study could be expanded to more agricultural crops on a larger scale to determine if there are other reoccurring RNA structures. That knowledge could serve to better inform future researchers on which structures to use in RNAi technology. If the goal of research into spider mite RNAi pathways is to control their spread as agricultural pests, more studies should be directed towards genetically engineering the chloroplast genome of crops to harbor synthetic structures that result in knockdown in spider mite gene expression but will not affect other species.

References

1 Okamura K., Lai, E. C. (2008). Endogenous small interfering RNAs in animals. *Nat Rev Mol Cell Biol*, 9, 673-678, doi:10.1038/nrm2479.

2 Stein, C.B., Genzor, P., Mitra, S., Elchert, A.R., Ipsaro, J.J., Sobti, S., Su, Y., Hammell, M., Joshua-Tor, L., Haase, A.D. (2019). Decoding the 5' nucleotide bias of PIWI-interacting RNAs. *Nat Comm*, 10.1038/s41467-019-08803-z.

3 Pinzón, N., Bertrand, S., Subirana, L., Busseau, I., Escrivá, H., Seitz, H. (2019). Functional lability of RNA-dependent RNA polymerases in animals. *PLOS Gen*, 15, e1007915, doi:10.1371/journal.pgen.1007915.

4 Wang, N., Qu, S., Sun, W., Zeng, Z., Liang, H., Zhang, C.Y., Chen, X., Zen, K. (2018). Direct quantification of 3’ terminal 2'-O-methylation of small RNAs by RT-qPCR. *RNA*, 24(11):1520-1529 doi: 10.1261/rna.065144.117.

5 Guleria, P., Mahajan, M., Bhardwaj, J., Yadav, S. K. (2011). Plant small RNAs: biogenesis, mode of action and their roles in abiotic stresses. *GPB*, 9, 183-199, doi:10.1016/S1672-0229(11)60022-3.

6 Grbić, M., Van Leeuwen, T., Clark, R. M., Rombauts, S., Rouzé, P., Grbić, V., Osborne, E. J., Dermauw, W., Ngoc, P. C., Ortego, F., Hernández-Crespo, P., Diaz, I., Martinez, M., Navajas, M., Sucena, É., Magalhães, S., Nagy, L., Pace, R. M., Djuranović, S., Smagghe, G., Iga, M., Christiaens, O., Veenstra, J. A., Ewer, J., Villalobos, R. M., Hutter, J. L., Hudson, S. D., Velez, M., Yi, S. V., Zeng, J., Pires-daSilva, A., Roch, F., Cazaux, M., Navarro, M., Zhurov, V., Acevedo, G., Bjelica, A., Fawcett, J. A., Bonnet, E., Martens, C., Baele, G., Wissler, L., Sanchez-Rodriguez, A., Tirry, L., Blais, C., Demeestere,

K., Henz, S. R., Gregory, T. R., Mathieu, J., Verdon, L., Farinelli, L., Schmutz, J., Lindquist, E., Feyereisen, R., Van de Peer, Y. (2011). The genome of Tetranychus urticae reveals herbivorous pest adaptations. *Nat,* 479, 487-492, doi:10.1038/nature10640 (2011).

7 Kim, V. N., Han, J., Fau - Siomi, M. C., Siomi, M. C. (2009). Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol*, 10(2): 126-39. doi:10.1038/nrm2632.

8 Antoniewski, C. (2014). Computing siRNA and piRNA overlap signatures. *M Mol Biol*, 1173:135-46. doi: 10.1007/978-1-4939-0931-5_1.

9 Karlikow, M., Goic, B., Saleh, M.C. (2014). RNAi and antiviral defense in Drosophila: Setting up a systemic immune response. *DCI*, 42, 85-92, doi:10.1016/j.dci.2013.05.004.

10 Marker, S., Le Mouel, A., Meyer, E., Simon, M. (2010). Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in *Paramecium tetraurelia*. *Nucleic Acids Res*, 38, 4092-4107. doi:10.1093/nar/gkq131.

11 Younis, A., Siddique, M. I., Kim, C.K., Lim, K. B. (2014). RNA interference (RNAi) induced gene silencing: a promising approach of hi-tech plant breeding. *Int J Biol Sci*, 10, 1150-1158, doi:10.7150/ijbs.10452.

12 Tsai, H. Y., Chen, C. C., Conte, D. Jr., Moresco, J. J., Chaves, D. A., Mitani, S., Yates, Jr., Tsai. M.D., Mello, C. C. (2015). A ribonuclease coordinates siRNA amplification and mRNA cleavage during RNAi. *Cell,* 160, 407-419, doi:10.1016/j.cell.2015.01.010.

13 Lewis, S. H., Quarles, K. A., Yang, Y., Tanguy, M., Frezal, L., Smith, S. A, Sharma, P. P., Cordaux, R., Gilbert, C., Giraud, I., Collins, D. H., Zamore, P. D., Miska, E. A., Sarkies, P., Jiggins, F. M. (2018). Pan-arthropod analysis reveals somatic piRNAs as an ancestral defense against transposable elements. *Nat Ecol Evol*, 2, 174-181, doi:10.1038/s41559-017-0403-4.

14 Zamore, P. D. (2010). Somatic piRNA biogenesis. *EMBO J*, 29, 3219-3221, doi:10.1038/emboj.2010.232.

15 Mondal, M., Mansfield, K., Flynt, A. (2018). siRNAs and piRNAs collaborate for transposon control in the two-spotted spider mite. *RNA*, 24, 899–907. doi:10.1261/rna.065839.118.

16 Weick, E.M., Miska, E. A., (2014). piRNAs: from biogenesis to function. *Dev*, 141, 3458, doi:10.1242/dev.094037.

17 Czech, B., Hannon, G. J. (2016). One loop to rule them all: the ping-pong cycle and piRNA-Ggided silencing. *Trends Biochem Sci*, 41, 324-337, doi:10.1016/j.tibs.2015.12.008.

18 Gammon, D. B., Mello, C. C. (2015). RNA interference-mediated antiviral defense in insects. *Current Opinion in Insect Science*, 8, 111-120, doi:10.1016/j.cois.2015.01.006.

19 Niu, X., Kassa, A., Hu, X., Robeson, J., McMahon, M., Richtman, N. M., Steimel, J. P, Kernodle, B. M., Crane, V. C., Sandahl, G., Ritland, J. L., Presnail, J. K., Lu, A. L., Wu, G. (2017). Control of western corn rootworm (*Diabrotica virgifera virgifera*) reproduction through plant-mediated RNA interference. *Sci Rep*, *7*(1), 12591. doi:10.1038/s41598-017-12638-3.

20 Suzuki, T., Nunes, M. A., España, M. U., Namin, H. H., Jin, P., Bensoussan, N., Zhurov, V., Rahman, T., De Clercq, R., Hilson, P., Grbic, V., Grbic, M. (2017). RNAi-based reverse genetics in the chelicerate model *Tetranychus urticae*: a comparative analysis of five methods for gene silencing. *PloS One*, *12*(7), e0180654. doi:10.1371/journal.pone.0180654.

21 Khila, A., Grbić, M. (2007). Gene silencing in the spider mite *Tetranychus urticae*: dsRNA and siRNA parental silencing of the Distal-less gene. *Dev Genes Evol*, 217:241-251.

22 Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., Gascuel, O. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36, W465-469, doi:10.1093/nar/gkn180.

23 Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nat,* 391, 806-811, doi:10.1038/35888.

24 Lau, N. C., Seto, A. C., Kim, J., Kuramochi-Miyagawa, S., Nakona, T., Bartel, D. P., Kingston, R. E. (2006). Characterization of the piRNA complex from rat testes, *Science*, 21;313(5785):363-7. doi:10.1126/science.1130164.

25 Gordon, A. (2009). Fastx toolkit- (http://hannonlab.cshl.edu/fastx_toolkit/).

26 Langmead, B., Trapnell, C., Pop, M., Salzberg, S. L., (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10(3):R25. doi: 10.1186/gb-2009-10-3-r25.

27 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup. (2009). The sequence Aagnment/map format and SAMtools. *Bioinformatics*, 15; 25(16):207809. doi: 10.1093/bioinformatics/btp352.

28 Quinlan, A. R., Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, *26*(6), 841–842. doi:10.1093/bioinformatics/btq033.

29 Lorenz, R., Bernhart, S. H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., Hofacker, I. L. (2011). ViennaRNA package 2.0. *Algorithms for molecular biology: AMB, 6*, 26. doi:10.1186/1748-7188-6-2.

30 Wickham, H. (2016). ggplot2: elegant graphics for data analysis. *Springer-Verlag New York*, 2016.

31 Schneider, T. D., Stephens, R. M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*, 25; 18(20): 6097-6100. Doi:10.1093/nar/18.20.6097.

32 Turner, T. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software*, 3(25), 731, https://doi.org/10.21105/joss.00731.

33 Kolde, R. (2012). Pheatmap: pretty heatmaps. https://cran.r-project.org/web/packages/pheatmap/index.html.

34 Phanstiel, D. H. Boyle, A. P., Araya, C. L., Snyder, M. P. (2014). Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-

quality multi- panel figures. *Bioinformatics*, 30(19):2808-10. doi: 10.1093/bioinformatics/btu379.