Title
The piRNA System in Aedes aegypti

Author
Han, Michael

Publication Date
2017

UNIVERSITY OF CALIFORNIA
RIVERSIDE

The piRNA System in *Aedes aegypti*

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Cell, Molecular, and Developmental Biology

by

Michael Eric Han

September 2017

Dissertation Committee:
Dr. Peter W. Atkinson, Chairperson
Dr. Anupama Dahanukar
Dr. Ted Karginov

The Dissertation of Michael Eric Han is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

I would like to thank my adviser Dr. Peter Atkinson for his invaluable assistance and guidance through the challenges of a complex project. I am grateful for his efforts in making me a better scientist, writer, and presenter. I would also like to thank my committee members Dr. Anupama Dahanukar and Dr. Ted Karginov for their discussions and advice on my progress over the years.

I would also like to thank members of the Atkinson laboratory. For bioinformatic advice, Peter Arensburger and especially Patrick Schreiner, who helped me on a wide range of topics, from drawing tables in R to navigating the biocluster. On the benchwork side, Jennifer Wright and especially Robert Hice, who taught me much of what I know about molecular biology. For their companionship and many talks about science, Presha Shah and Lee Doss.

I would also like to thank Crystal Pontrello for her advice on insect dissections, and Pinky Kain for her advice on confocal microscopy.

Always thankful to my family for their unconditional support and care.

ABSTRACT OF THE DISSERTATION

The piRNA System in *Aedes aegypti*

by

Michael Eric Han

Doctor of Philosophy, Graduate Program in Cell, Molecular, and Developmental Biology
University of California, Riverside, September 2017
Dr. Peter Atkinson, Chairperson

The aim of the research presented in this thesis is to examine the piRNA pathway in *Aedes aegypti,* with an emphasis on understanding the role of the pathway in the soma. Chapter one reviews the piRNA pathway's role in transposon regulation as well as transposon-independent roles, such as sex-determination in *Bombyx mori.* In addition, preliminary research from the Atkinson laboratory showed an expansion in the number and expression domain of the PIWI family in *Aedes aegypti* compared to a model Dipteran organism, *Drosophila melanogaster.* Chapter two introduces research I performed that showed the somatic expression of an important PIWI gene, Ago 3, in somatic ovarian follicular cells and larval gastric caecum. Piwi 2 was found to have a germline localization. In addition, an Ago 3 RNAi knockdown line (M14) exhibited a phenotype of larval mortality. Chapter three focuses on a new, more stringent method of annotating piRNA clusters in *Ae. aegypti* from different types of mosquito sRNA libraries, including both somatic and germline tissue. Two fairly distinct sets of piRNA

clusters were discovered, one in the soma and one in the germline. Somatic clusters produced piRNA against predominately *gypsy* elements; somatic piRNA bore strong U1 signatures but weaker A10 signatures, and also bore less hallmarks of the piRNA ping-pong amplification loop. In contrast, germline clusters produced piRNA against a more varied set of transposons, and germline piRNA had both strong U1 and A10 signatures. Germline libraries also had larger quantities of transposon-derived piRNA. Chapter four examines the effect of Ago 3 knockdown in mosquito larvae. Modest decreases in U1 and A10 signatures were seen in piRNA sequenced from Ago 3 knockdown mosquitoes; in addition, the relative percent of piRNA mapping against transposons declined from wild-type and control conditions. A global decrease in mRNA mapping to transposons was also detected. Together, these data show that somatic piRNAs exist in *Ae. aegypti.* These piRNA play a role in transposon defense, but based on comparison with germline piRNA, somatic piRNA may also play a role in different pathways, such as gene regulation or viral defense.

TABLE OF CONTENTS

Chapter 1:
Introduction……...1

1.1 The Argonuate family of proteins……...1
1.2 Introduction to the piRNA pathway in *D. melanogaster*……...……...4
1.3 Genome-transposon conflict……6
1.4 Germline specification and development……..8
1.5 Maternal mRNA degradation……..9
1.6 A role in epigenetics and canalization for PIWI proteins…….....10
1.7 PIWI and post-transcriptional silencing……..13
1.8 PIWI and transcriptional silencing……...14
1.9 The piRNA pathway in *Ae. aegypti*……...……16
1.10 Somatic piRNA in *Ae. aegypti*……...18
1.11 Sex regulation in *Bombyx mori*……..19
1.12 piRNA in mammals……...20
1.13 Somatic piRNA in mammals – stem cells and cancer…...……...21
1.14 *Ae. aegypti* as a potential model for piRNA pathways……..22
1.15 References……...24

Chapter 2: Localization of PIWI family proteins in *Ae. aegypti*……31
2.1 Introduction……..31
2.2 Materials and methods……36
2.3 Results……41
2.4 Discussion……47
2.5 References……54
2.6 Figures and tables……57

Chapter 3: Characteristics of piRNA clusters in *Ae. aegypti*……..69
3.1 Introduction……..69
3.2 Materials and methods……79
3.3 Results……86
3.4 Discussion……96
3.5 References……...107
3.6 Figures and tables……..111

Chapter 4: A somatic role for Ago 3 in *Ae. aegypti*……130
4.1 Introduction……130
4.2 Materials and methods……..134
4.3 Results……..141

4.4 Discussion……...149
4.5 References……...160
4.6 Figures and tables……..162

List of tables

Table 2.1 Primers for PIWI qPCR……..……57
Table 2.2 General Leica SP5 confocal microscopy settings……58
Table 2.3 Leica SP5 confocal laser settings for Hoechst DNA stain detection……....59
Table 2.4 Leica SP5 confocal laser settings for Alexa555 detection……..59
Table 2.5 M14 line developmental delay and mortality……60
Table 2.6 Ago 3 enrichment in larval tissues……...60
Table 2.7 PIWI expression in larval gastric caecum……60
Table 3.1 U1 and A10 biases of transposon mapping piRNA……..…111
Table 3.2 U1 and A10 biases of *Ae. aegypti* piRNA……112
Table 3.3 Statistics on proTRAC called piRNA clusters……...113
Table 3.4 proTRAC clusters called from larval gastric caecum libraries……...114
Table 3.5 proTRAC clusters called from 4th instar larval libraries……..114
Table 3.6 proTRAC clusters called from ovarian libraries……...115
Table 3.7 proTRAC clusters called from early embryonic libraries……...115
Table 3.8 proTRAC clusters called from Ago 3 pulldown libraries……..…..116
Table 3.9 proTRAC clusters called from Piwi 2 pulldown libraries……..116
Table 3.10 Summary of proTRAC cluster calls……117
Table 3.11 Overlap between cluster calls in different tissues in base pairs...……...117
Table 3.12 Overlap between cluster calls in different tissues by percent……..118
Table 3.13 proTRAC called split unidirectional clusters……118
Table 4.1 Primers for transposon qPCR……162
Table 4.2 Reads mapping to Ago3-SB transformation plasmid……..163
Table 4.3 Differentially expressed genes in female larvae upon induction……...…...163
Table 4.4 Differentially expressed genes between male and female larvae……..164
Table 4.5 Differentially expressed genes in male larvae upon induction……..164
Table 4.6 Differentially expressed transposons in male larvae upon induction……..165
Table 4.7 Differentially piRNA targeted transposons in male larvae upon induction…166

List of figures

Figure 2.1 Ago3-SB transformation plasmid……...61
Figure 2.2 Ago 3 and Piwi 2 Western blot……62
Figure 2.3 Ago 3 qPCR in ovaries and larval tissues……..63
Figure 2.4 PIWI qPCR in larval gastric caecum……,……..64
Figure 2.5A PIWI immunohistochemistry in ovaries……,……65
Figure 2.5B Zoom of Figure 2.5A……,……66
Figure 2.6 PIWI immunohistochemistry in testis…….……..66
Figure 2.7 PIWI immunohistochemistry in larval tissues……67
Figure 3.1 Seqlogo of transposon mapping sRNA reads……119
Figure 3.2 Genome occupancy of proTRAC called piRNA clusters….……..120
Figure 3.3 Cluster origin of sequenced sRNA……...120
Figure 3.4 5′ ten nucleotide overlap library Z-scores…….……..……121
Figure 3.5 Nucleotide distributions of gastric caecum sRNA reads…….……...……..122
Figure 3.6 Average percent identify of called clusters between tissues……123
Figure 3.7 Shared top clusters between tissues……123
Figure 3.8 Sense/antisense orientation of cluster features……...124
Figure 3.9 Somatic cluster supercont 1.478……124
Figure 3.10 sRNA mapping to split cluster supercont1.1……125
Figure 3.11 First half of split cluster supercont1.1……126
Figure 3.12 Second half of split cluster supercont1.1……127
Figure 3.13 sRNA mapping to cluster supercont1.192……127
Figure 3.14 Germline cluster supercont1.192……128
Figure 3.15 Types of *Ae. aegypti* piRNA clusters……...……129
Figure 4.1 M14 line male GFP phenotypes……...…..…167
Figure 4.2A M14 mosquito transformation plasmid integration sites……...166
Figure 4.2B EGFP PCR in M14 mosquitoes……169
Figure 4.3 qPCR of Ago 3 and a transposon panel in M14 mosquitoes……170
Figure 4.4 qPCR of Ago 3 and a transposon panel in M14 mosquitoes……171
Figure 4.5 qPCR of Ago 3 and a transposon panel in Orlando mosquitoes……...…...172
Figure 4.6 Ago 3 knockdown in male M14 mosquitoes……...173
Figure 4.7 Log-fold change and expression of transposons in female larvae……...174
Figure 4.8 Log-fold change and expression of transposons in male larvae……..……....175
Figure 4.9 Log-fold change and expression of transposon mapping sRNA……176

Chapter 1 – Introduction to Argonaute and the piRNA pathway

1.1 The Argonaute family of proteins

Argonaute proteins are a highly conserved class of small RNA binding proteins present in eukaryotes that have been split into three clades – AGO, a family which includes AGO1 from *Arabidopsis thaliana* and are the only Argonaute clade found in plants; PIWI, a family which includes Piwi from *D. melanogaster* and whose members are predominately found in the germline of animals; and WAGO, a family found in *Caenorhabditis elegans* (Aravin et al., 2007a; Tolia and Joshua-Tor, 2007; Yigit et al., 2006).

The functional domains of Argonaute proteins are the PAZ (Piwi-Argonaute-Zwille), MID (middle) and PIWI domains. The PAZ domain binds to the 3' end of a sRNA, and the MID domain binds to the 5' end; the PIWI domain functions as an endonuclease that can cleave the RNA complement of a bound sRNA guide strand (Jinek and Doudna, 2009). Although all Argonaute proteins have a PIWI domain, the catalytic endonuclease activity is not always conserved or necessary for the role of the protein in effecting RNAi. In humans, only AGO2 in the AGO clade is catalytically active (Liu, 2004). In *D. melanogaster,* Piwi in the PIWI clade has endonuclease activity in vitro, but this activity is not required for the function of the protein; mutation of the catalytic domain was not found to result in any mutant phenotypes (Darricarrère et al., 2013). Piwi, which is localized in the nucleus, is instead thought to play a role in transcriptional repression of transposons through the formation of heterochromatin (Sienski et al., 2012; Yu et al., 2015).

The AGO clade is responsible for the classical, Drosha and Dicer associated RNAi pathway (He and Hannon, 2004). This RNAi pathway involves two classes of small RNA guide molecules – microRNA (miRNA) and small interfering RNA (siRNA). The first known miRNA gene, *lin-4,* was originally identified in *C. elegans* as a regulator of *lin-14*. This regulation was necessary for the establishment of a protein gradient required for proper cell lineage development (Wightman et al., 1993). *Lin-14* contains in its 3' UTR short sequence elements complementary to *lin-4* RNA. These sequences were necessary for its regulation by *lin-4*, leading to a model, later confirmed, where *lin-4* RNA directly regulates *lin-14* translation through an anti-sense RNA-RNA interaction (Lee et al., 1993; Wightman et al., 1993).

Similarly, the siRNA RNAi pathway was first observed in in *C. elegans*, where experimental introductions of double stranded RNA (dsRNA) were found to have strong, specific repressive effects on the expression of a myofilament protein, UNC-22. Introductions of solely sense or solely anti-sense RNA to *unc-22* were much less effective than the dsRNA molecules (Fire et al., 1998). Endogenous, rather than introduced, siRNA were first described in tobacco plants. 25 nucleotide long antisense RNA species to transgenes were detected – both to transgenes with homology to endogenous genes, as well as the gene for green fluorescent protein (GFP) (Hamilton, 1999).

miRNA have their origin in specific precursor genes that code for hairpin loops within the genome, and typically target endogenous genes for regulation (Lagos-Quintana et al., 2001). siRNA, however, often derive from dsRNA of foreign origin, such

as from viruses, and thus can function as a viral defense pathway (Hannon, 2002). Both pathways depend on a family of RNAse-III enzymes, Dicer, for the maturation of small 20-25 nucleotide anti-sense guide RNA molecules from their respective origins. These processed RNA molecules then bind with Argonaute proteins to form complementation-guided RNA-induced silencing complexes (RISCs) to mediate post-transcriptional silencing (He and Hannon, 2004). Targeting of the guide RNA to its target message can either occur with mismatches at several sites, often within untranslated regions (UTR). The binding of multiple RISCs to the UTR inhibits or completely blocks translation. This is often the mechanism of action for miRNA genes found in animals (He and Hannon, 2004). Alternatively, the guide RNA can form a perfect complement with a specific site on its target and cause the catalytic cleavage and destruction of the message. This is often the mechanism of action for siRNAs and plant miRNAs (Hannon, 2002). The prime determinant of which of these two pathways is favors seems to be the degree of complementarity of the guide RNA to its target, rather than the siRNA or miRNA origin of the guide RNA, again pointing to the similarity of the RISC of both RNAi pathways. For instance, mammalian miR-196 cleaves its target like a plant miRNA or a siRNA, and siRNA engineered imperfectly to the UTR sites of genes can cause translational repression rather than the traditional siRNA cleavage (Doench, 2003; Yekta, 2004).

1.2 Introduction to the piRNA pathway in *D. melanogaster*

In animals, there are three small RNA silencing pathways - the miRNA pathway, the sRNA pathway, and the piwi-interacting, or piRNA pathway. Unlike the other two pathways, which are ubiquitous throughout the organism, in its most studied model organism, *D. melanogaster,* the piRNA silencing system has been most studied and is most active in the germline and associated soma, such as the nurse cells. The primary function of the piRNA pathway is the suppression of transposon activity, but recent studies have opened the possiblity of other functions as well (Seto et al., 2007).

Similar to the two other RNA silencing pathways, the piRNA pathway is mediated in part through the formation of RISC. RISCs include an Argonaute family protein, which catalyzes RNA silencing through mechanisms such as RNA cleavage, and a bound small RNA, which guides the complex to its targets through complementation. Since the function of a cell is ultimately determined by gene expression and translation, RNA silencing pathways have the potential to regulate every cellular process (Tomari and Zamore, 2005). The piRNA pathway is associated with a subfamily of the Argonaute proteins called PIWI. In *D. melanogaster*, the three proteins in the family are Piwi, Argonaute 3 (Ago3), and Aubergine (Aub), the last two of which are only expressed in the fly germline (Vagin et al., 2006). In contrast to Aub and Ago3, Piwi is also expressed in the somatic support cells of the ovary. There is little support for the expression of these genes in any other cells of the adult fly. In addition to different expression patterns when compared to siRNA and miRNA mediators, which are not restricted to the

germline or germline-adjacent, the three PIWI family proteins also bind longer small RNAs. PiRNAs are on average 24-30 nucleotides long (Brennecke et al., 2007).

The source of most piRNAs are genomic loci, such as *flamenco*, which are made up of truncated, inactivated transposon fragments. Some clusters are germline specific, some are only expressed in the somatic support cells of the germline, and some are expressed in both. A single long transcript is synthesized from these loci, which is further processed into piRNA. It is thought that one way *D. melanogaster* gains control over new transposons is when they transposition into a piRNA locus, after which the transposons are subject to piRNA silencing (Malone et al., 2009; Robert et al., 2001).

Novel piRNAs may be generated to invading transposons and spread throughout the length of the element through a mechanism called piRNA phasing, and then increased in abundance through a mechanism called ping-pong amplification. These novel piRNA are produced through the initial cleavage of a transposon transcript, which generates new small RNAs in a stepwise fashion across the length of the transcript from the initial site of cleavage. These new piRNA can then amplified by interactions of Ago3 with Aub. Ago3 binds piRNAs in the sense orientation to its target loci, while Aub binds piRNAs in the antisense orientation; the two thus produce substrates for each other's binding after cleavage of target transcripts (Han et al., 2015).

1.3 Genome – transposon conflict

Transposons are selfish genetic elements present within genomes with the ability to replicate and transposition themselves throughout the host genome. Successful transposons are especially active in the germline of their hosts, which allows for propagation of the element down generations through vertical transmission (Senti and Brennecke, 2010). Uncontrolled transposons can have devastating effects on the reproductive fitness of an organism, as movement of the transposon can disrupt key genes and damage the host genome as it replicates itself.

This phenomenon of transposon mediated reproductive damage is called hybrid dysgenesis, and was discovered in *D. melanogaster* when naïve, or reactive, females produced sterile progeny when mated with inducer flies that contained a new transposon. Fertile offspring resulted when the male was naïve and the female introduced the new element. As sterility only occurred when the new transposon was introduced paternally, it became apparent that the suppression of transposons in the developing embryo is mediated through factors transmitted maternally (Engels and Preston, 1980).

Although the technology was lacking at the time, these factors were later discovered to be piRNAs. piRNAs are present in the embryo before zygotic transcription, indicating that they are deposited by the mother. When a novel transposon is introduced into female flies, they have few piRNA targeted against the element. There is high concordance in the piRNA population of the maternal ovaries and

its derived embryos, and this continues even after the embryo has developed into an adult. Despite being genetically identical to reactive male/inducer female crosses, reactive female/inducer male progeny had fewer piRNA targeted against the induced transposon element (Brennecke et al., 2008). This difference in piRNA populations may have to do with the ping-pong or positive feedback mechanism of piRNA biogenesis, which is affected by the initial piRNA population imparted by the mother (Senti and Brennecke, 2010). In this way, maternal deposition of piRNA in the embryo can have lasting effects in the adult organism, explaining why reactive female/inducer male progeny are sterile while the reciprocal cross is not.

Due to their association with PIWI family proteins, piRNAs were already known to affect fertility. PIWI, or P-element induced wimpy testis, mutants have defects in the maintenance of germline stem cells (GSCs). The cells differentiate without self renewing, leading to eventual depletion of the germline (Cox et al., 2000; Lin and Spradling, 1997). It may be that this phenotype is the result of the activity of derepressed transposases, since active transposases cause double stranded breaks in host DNA and can lead to cell cycle arrest (Aravin et al., 2007a). In support of this theory is the finding that insertion or deletion mutants of just one locus from which piRNA are derived, *flamenco,* leads to female sterility as well as transposon derepression (Mével-Ninio et al., 2007). In addition, Aub mutants resemble spindle-class mutants that are defective in mitotic progression due to kinase-dependent mitotic checkpoint activation. Animals mutant for both *Aub*

and the cell-cycle kinases that regulate checkpoint activation have a partially relieved Aub mutant phenotype (Klattenhoff and Theurkauf, 2008; Wilson et al., 1996).

1.4 Germline specification and development

In addition to the requirement of the piRNA pathway for the maintenance of germ cells, evidence has emerged that piRNAs play a role in germ cell specification and development. This role appears to be independent of transposon regulation. In *Drosophila,* cells are induced to take a PGC fate through the inheritance of cytoplasmic factors that are present in high concentrations in the germ plasm, or nuage, located in the posterior pole of the embryo (Houston and King, 2000). These factors are products of maternally expressed genes, and are often proteins or RNA products that are incorporated into the PGCs. Examples of these factors are oskar, which is deposited as an mRNA and is necessary to form the nauge, and a suite of genes called *grandchildless* (Ephrussi and Lehmann, 1992; Mahowald, 2001).

Grandchildless mutants have sterile progeny; examples of *grandchildless* mutants are *tudor* and *aub* (Boswell and Mahowald, 1985; Harris and Macdonald, 2001). Aub contains symmetrically dimethylated arginine (sDMAs) which bind tudor. In wild type flies, both Aub and A3 are localized to the nuage. In *tudor* mutants, Aub is no longer localized to the nuage, suggesting that Aub or the complex of Aub and Tudor in the nuage is necessary for germ cell specification. The mechanism of this function is unclear, but one possibility is that Aub mediated piRNA silencing of specific mRNAs is required

for cells to take a germ cell fate. It seems unlikely that transposon regulation plays a role in germ cell specification, as *tudor* mutants did not alter Piwi proteins or piRNA populations in female ovaries. Transposon de-repression was only lost for one transposon (Kirino et al., 2010). The precise mechanism for the sterility seen in Tudor and Aub mutants may lie elsewhere.

1.5 Maternal mRNA degradation

Another recently described transposon regulation independent function of piRNA is the deadenylation and degradation of maternal mRNA in the *Drosophila* embryo. During the maternal-to-zygotic transition, some maternal mRNA are degraded through the actions of the RNA-binding protein Smaug and the deadenylase CCR4 (Semotok et al., 2005). One such maternal mRNA is transcribed from *nanos (nos),* a gene which codes for a morphogen with highest concentration in the posterior pole of the fly. Nos is deposited throughout the embryo, and only through translational repression and degradation are maternal *nos* transcripts restricted to the posterior of the embryo. Without this repression, head and thorax segmentation does not proceed properly (Dahanukar and Wharton, 1996).

When embryos mutant for the piRNA pathway were assessed for *nos* deadenylation activity, it was found that mRNA tail shortening was impaired in the mutants, compared to WT embryos. This led to stabilized *nos* mRNA outside of it's WT posterior restricted range, as well as ectopic Nos protein throughout the embryo.

Morphologically, the piRNA pathway mutant embryos had head defects. These mutant phenotypes were not relieved by mutations in the DNA damage checkpoint activator Chk2, indicating that the phenotypes were unrelated to derepression of transposons and the resulting genomic DNA damage and cell cycle arrest, which can also lead to defects in axis specification (Klattenhoff et al., 2007).

Mechanistically, the PIWI family proteins Aub and Ago3 associate with Smaug and CCR4 into complexes, and Aub also associates with *nos* mRNA. There are regions in the *nos* UTR which are capable of complementation with some transposon derived sequences that generate piRNA. This suggests that Aub and Ago3 are complexing with RNA-binding proteins and deadenylases and guiding them to the maternal *nos* mRNA through complementation of piRNAs (Rouget et al., 2010).

A follow up study focusing on maternal mRNA in the early embryo that associate with Aub did not replicate *nos* mRNA as a target of Aub. However, gene expression studies performed through sequencing of Aub mutants did reveal that a small portion of Aub-interacting maternal mRNA undergo Aub-dependent degradation. In addition, these mRNA are often involved in germ cell specification, which has support in other studies as described previously. These result suggest that Aub and piRNA may have a role in maternal gene regulation in *D. melanogaster* embryonic development (Barckmann et al., 2015). However, a clear consenus has yet to be reached.

1.6 A role in epigenetics and canalization for PIWI proteins

Complementing the role of the piRNA pathway in preventing transposon induced genome change, and perhaps related to Piwi's role in chromatin modification, piRNAs have been shown to assist in the process of canalization in *Drosophila.* Canalization, or developmental robustness, is a term for how an evolutionarliy selected pheonotype is protected against environmental factors and other stresses during development that might alter patterns of gene expression and the organism's phenotype (Waddington, 1959). A key protein in canalization is Hsp90, a heat shock protein. When Hsp90 is inactivated by severe organismal stress during development, morphological changes are seen in the embryo. Hsp90 mutants have the same effect. In addition, the morphological changes are inheritable (Rutherford and Lindquist, 1998).

One such change is ectopic outgrowth of bristles in the fly eye, found in a mutant screen. Mutants that caused the phenotype included *trithorax* (TrxG) group genes, which affect epigenetic inheritance of acetylated chromatin and gene expression patterns, as well as Hsp83, a chaperon for Hsp90. The ectopic outgrowth phenotype was more pronounced when these mutations were carried in the mother. When flies were subject to chemical treatment that inactivated Hsp90 and then bred over several generations, the ectopic outgrowth phenotype persisted and even raised in frequency. TrxG mutations are known to often cause histone hypoacetylation, pointing to an epigenetic role in Hsp90's control of canalization. When flies chemically inactivated for Hsp90 were also

treated with histone deacetylase inhibitors, the ectopic overgrowth phenotype was suppressed in successive generations (Sollars et al., 2003).

Building on the finding that the outgrowth phenotype was more severe when the mutations were carried by the mother, studies were done on the role of maternal piRNA deposition and canalization, finding that mutant *piwi* alleles were dominant enhancers of the eye outgrowth phenotype. Piwi was found to associate with Hsp90, and even the loss of one allele of *piwi* was sufficient to cause ectopic outgrowths, though the loss of only one *piwi* allele does not compromise it's function in transposon silencing. Bearing up this observation is the fact that transposon RNA levels do not increase in heterozygous *piwi* mutants (Gangaraju et al., 2011). Given that Piwi associates with Hsp90, which has an epigenetic function in canalization, it seems that Piwi's role in this pathway is guiding epigenetic regulation of endogenous genes, and not just transposons.

Elsewhere in the fly head, a recent study has shown that Aub and Ago3 are expressed in the fly brain, with implications that the tight control of transposons in the brain may play some role in genome rearrangements that leads to neuronal plasticity, in analogy to how domesticated transposons create variability in the production of lymphocytes in the vertebrate adaptive immune response (Perrat et al., 2013).

1.7 PIWI and post-transcriptional silencing

Ago 3, along with Aub, plays a critical role in the post-transcriptional, ping-pong piRNA pathway in *D. melanogaster,* one of two pathways by which piRNA regulate transposon expression in the fly. Similar to the miRNA and siRNA RNAi pathways, post-transcriptional silencing mediated by PIWIs involve the formation of a RISC and transcript cleavage. However, the piRNA-induced silencing complex (piRISC) involve PIWI-clade proteins instead of AGO-clade proteins, and have several biochemical differences. piRISCs bind guide RNAs of longer sizes (23-33 nucleotides), enforce either a U1 or an A10 nucleotide bias, and also involve a different mode of guide RNA loading (Czech and Hannon, 2016).

The loading of guide RNA into piRISCs is thought to involve first the loading of a piRNA intermediate generated from a piRNA cluster (Brennecke et al., 2007). Trimming of this intermediate by either an unknown 3' exonuclease or Zucchini endonucleolytic cleavage results in a guide RNA 23-33 nucleotides in length (Han et al., 2015). The last processing step to form a mature, piRISC bound piRNA is Hen1 methylation of the 3' end (Horwich et al., 2007). Both Aub and Piwi preferentially bind piRNA with a U at position one, a preference first revealed by RNA-sequencing of PIWI pulldowns (Brennecke et al., 2007). The crystal structure of Siwi, the silkworm homologue of Piwi, revealed a specificity loop in the MID domain of Siwi. One of the residues in the loop, Tyr603, repulsed modeled A, G, and C nucleotides, either in a steric or a charge-related fashion. There was no repulsion of U. DmelAgo3 has a bias towards

binding piRNA with an A10 instead of piRNA with a U1; accordingly, the specificity loop in the MID domain of BmAgo3 contains different residues than the loop in Siwi (Matsumoto et al., 2016). The observed A10 bias seen in DmelAgo3, however, may not be an intrinsic feature of Ago3, or a result of the U1 Piwi/Aub bias; instead, Piwi and Aub have a separate preference for an adenine at position one of their target RNA (a t1A), again, regardless of the sequence of their bound guide RNA (Wang et al., 2014). This is supported by the crystal structure of Siwi – similar to human argonuate-2 (hAgo2), Siwi contains a potential t1A binding pocket. The potential pocket in Siwi contains a Ser561; the equivalent residue in DmelAgo3 and BmAgo3 is instead a glutamine, which sterically blocks the pocket. Upon target slicing by the Aub or Siwi associated piRISC, the target RNA bound by the piRISC beginning with a t1A has been processed and matured into a sense piRNA with an A at position 10 (Czech and Hannon, 2016). This sense piRNA can then be bound by Ago3 and enter into the ping-pong cycle, finding anti-sense targets to cleave and create more anti-sense guide piRNA which can be bound by Aub/Piwi (Brennecke et al., 2007).

1.8 PIWI and transcriptional silencing

In contrast to Aub and Ago3, studies have shown that catalytic mutants of Piwi that lack PAZ, or nuclease, domain activity do not cause mutant phenotypes (Darricarrère et al., 2013). Piwi is able to perform its functions without its nuclease

domain. Instead of transcript cleavage, Piwi, which is localized in the nucleus, may be repressing transposon activity by silencing transcription through chromatin methylation. ChIP QPCR data shows that Piwi colocalizes with HP1a, a protein associated with formation of heterochromatin, to piRNA complementary sequences in the genome. A Piwi-piRNA RISC complex is capable of guiding epigenetic factors to specific areas of the genome through complementary hybridization of its bound guide sRNA strand. DNA that is bound to Piwi/HP1a complexes was also found to be enriched in repressive chromatin marks H3K9me2 and H3K9me3 (Huang et al., 2013).

Potential partners that scaffold the sequence specificity of the piRNA-bound Piwi with the general chromatin modification machinery of the cell – proteins such as HP1a – includes proteins such as Asterix and Panoramix. While transcription of the gene to be silenced is ongoing, the Piwi-piRNA complex binds to the nascent transcript and then further recruits these scaffolding proteins. Panoramix in particular may be the key scaffold or adaptor in this process in *D. melanogaster*, though it is not conserved even among Diptera. Repressive chromatin marks are then directed against the genome in the vicinity of transcription, silencing the activity of the piRNA target, whether they be transposable elements or endogenous genes (Yu et al., 2015). In this fashion Piwi can silence transposons, even in the fly soma in the absence of Aub and Ago3.

1.9 The piRNA pathway in *Ae. aegpyti*

In *Ae. aegypti,* less work has been undertaken to understand the piRNA system. The low rate of germline transformation of *Ae. aegypti* , along with the difficulty of rembolizing transposons such as piggyBac or Hermes following integration, suggests that the piRNA system may be even more complex and robust in *Ae. aegypti* compared to *D. melanogaster.* In *D. melanogaster,* elements can be remobilized and have been used as genomic tools, ranging from enhancer trapping to insertional mutagenesis (Häcker et al., 2003; Smith and Atkinson, 2010).

Further evidence for increased complexity of the piRNA pathway in *Ae. aegypti* is the expansion in the gene family. While *D. melanogaster* has three members of the PIWI family, *Ae. aegypti* has seven. Based on sequence similarity, there is one *Ago3* gene and six other *Piwis*. Which of these six Piwi proteins are most closely related in function to *D. melanogaster* Aub or Piwi remains unknown (Arensburger et al., 2011).

An explanation for this gene expansion may be related to the differences in genome structure and transposon load. The *D. melanogaster* genome is relatively small (144 Mb) and has a low transposon load (15.8% of base pairs) when compared to *Ae. aegypti* (Kaminker et al., 2002; Smith et al., 2007). *Ae aegypti* has a much larger genome (1.38 Gb), though the number of protein coding genes is comparable with *D. melanogaster,* and a high transposon load (47% of base pairs) (Sinkins, 2007). These overarching genomic differences may pose challenges that required an expansion in order to police transposon activity. PIWI family genes in other insects tends to support

this; *Acyrthosiphon pisum* has a genome size of 464 Mb with a transposon load of 38% and contains 10 PIWI proteins, and *Culex quinquefasciatus* has a genome size of 540 Mb with a transposon load of 30% and contains 7 PIWI proteins. *Anopheles gambiae,* which is similar to *D. melanogaster* with a genome siz of 286 Mb and a transposon load of 16%, only contains 3 PIWI proteins (Campbell et al., 2008).

The origin and diversity of piRNAs in *Ae. aegypti* also has marked differences compared to *D. melanogaster.* In *D. melanogaster,* using criteria of 5 uniquely mapping piRNA per 5 kilobases of genomic sequence, 81% of uniquely mapping piRNA (which comprised roughly 20% of the total piRNA population sequenced in sRNA libraries in one experiment) and up to 92% of total piRNAs could potentially derive from loci at 142 genomic locations that make up only 3.5% of the fly genome. Of these piRNA, 42-50% match transposon sequences (Brennecke et al., 2007).

In contrast, piRNA clusters in *Ae. aegypti* are more diffuse. Even using a more stringent critera of 10 uniquely mapping piRNa per 5 kilobases of genomic sequence, piRNA clusters were found to occupy 20.5% of the genome. These clusters could be the source of up to 84% of total sequenced piRNA. Unlike clusters in *D. melanogaster,* which are made up of a very high density of transposons, identified clusters in *Ae. aegypti* were not enriched in transposons when compared to the rest of the genome. Although the increase in genome occupancy might be partly explained by the greater transposon load in *Ae. aegypti,* the increase in occupancy is out of proportion with the greater load. In addition, of the piRNA sequenced from *Ae. aegypti,* only about 19% mapped to

transposable elements, less than the 42-50% figure for *D. melanogaster.* piRNAs were found to map against endogenous protein coding genes, as well as viral-derived sequences (Arensburger et al., 2011). This expansion of piRNA generating loci in *Ae. aegypti* thus may not merely be an adaption to transposon load, but an indication that the piRNA pathway is being used for other functions in the mosquito.

1.10 Somatic piRNA in *Ae. aegypti*

Along with an expansion in gene number in the PIWI family, PIWIs in *Ae. aegypti* can also have expanded expression domains. qPCR and RNA-seq experiments in our lab and others has shown that some *Ae. aegypti* PIWI genes are expressed not only in the ovary and testes, but also in the wider soma. Many PIWI genes are expressed in high to moderate levels in developmental stages post-embryonically, the larvae and pupae, and also in adult mosquitoes which have had their germline tissues removed. Most of the PIWIs reach their highest levels of expression in the blood fed ovary, but some, such as *piwi4* and *piwi6,* have their peak in later stage embryos. *piwi7*, is not expressed or is expressed at low levels in the germline, and is only heavily transcribed in the embryo following the maternal-to-zygotic transition. a*go3* is enriched in blood fed ovaries and early embryos, but in other developmental stages has a fairly steady expression of around 15 reads per kilobase of transcript per million mapped reads (RPKM) (Akbari et al., 2013). The purpose of the expanded somatic range of the PIWI family in *Ae. aegypti* is unclear, but, judging by the *D. melanogaster* work, members of the family may be playing

roles in epigenetic regulation, somatic repression of transposons, or regulating development in both embryonic and post-embryonic stages.

Another potential role is viral defense. piRNAs with the hallmark A10-U1 ping-pong signature have been found in *Ae. aegypti* directed against viral sequences (Morazzani et al., 2012). What proteins are involved and in which tissues (germline or somatic support cells) is less understood and further complicated by the large expansion of PIWIs in *Ae. aegypti.* However, these viral piRNA were detected in the wider soma of the adult.

1.11 Sex regulation in *Bombyx mori*

A striking role in embryonic development for piRNAs is found in the silkworm, *B. mori.* Sex determination in the silkworm is controlled by a sex chromosome, W, which contains a feminizing factor (Fujii and Shimada, 2007). However, no protein coding genes leading to feminization could be found on the W chromosome. Instead, the W chromosome is almost entirely composed of retrotransposon and other transposon sequences (Abe et al., 2005). With growing knowledge of the piRNA pathway, small RNAs that regulate transposons, a potential link between the W chromosome, sex determination, and piRNAs was established. Sequencing revealed that the W chromosome was a source of piRNAs enriched in females compared to males, and these piRNAs were derived from a loci on the W chromosome already known to be a critical

sex determining region, although no protein coding gene had ever been identified within (Kawaoka et al., 2011).

Further studies identified a female-specific piRNA precursor gene named *Fem* that targeted a transcription factor, *Masc*. Inhibition of *Fem* derived piRNA led to derepression of *Masc* and the production of male-specific splice factors of *doublesex*, a gene which acts at the end of sex-determination cascades in many insects. Without piRNA mediated inhibition of *Masc,* silkworm embryos would not properly feminize. Therefore, a single piRNA is responsible for primary sex determination in the silkworm embryo (Kiuchi et al., 2014).

1.12 piRNA in mammals

Piwi family proteins and piRNAs are also present in mammals, where they play a similar role in defending and maintaining the germline. In mammals however, the piRNA pathway is more complex, and there may be other effectors that also play a role in the germline. Mice defecient for Piwi family homologues MIWI, MILI, and MIWI2 show defects in transposon regulation in the primordial testis as well as post-meiotic arrest of spermatogenesis; however, no defects are observed in the female germline, despite the presence of MILI and associated piRNAs in the ovary. piRNAs in primordial testis are derived from transposon sequences, and mouse mutants for MIWI proteins have upregulated expression of transposon mRNA (Aravin et al., 2007b).

In contrast to piRNAs from nascent testis, testicular piRNAs in the adult mouse show no sign of ping-pong amplification and are mostly derived from non-transposon intergenic regions and are localized to the germ line only. They appear to play a transposon independent role in meiosis progression, since MIWI mutants cannot complete spermatogenesis and are sterile, but the exact mechanism is unclear (Beyret, 2009; Beyret and Lin, 2011).

1.13 Somatic piRNA in mammals – stem cells and cancer

In addition to the role of mammalian Piwi family homologues in spermatogenesis in adult stem cells in the mouse testis, human hematopoietic stem cells express human homologues of Piwi family genes, HIWI genes (Sharma et al., 2001). It is likely that expression of these HIWI genes is not necessary for stem cell function. Knock out experiments showed that in the absence of HIWI, stem cell function maintenance and differentiation was not effected (Nolde et al., 2013). However, there may be related Argonaute or Piwi class family genes which provide redundancy of function, and it is still possible that in mammals, Piwi family genes play a role not only in germline stem cells, but also stem cells for somatic cells.

Other somatic tissues where HIWI genes are expressed are in cancerous tumors, such as gastric cancer, breast cancer, and renal cancer (Liu et al., 2006; Taubert et al., 2007). It is unknown whether the HIWI genes are acting as oncogenes in these tumors.

Further support for somatic piRNA in mammals comes from a study showing that a small fraction of small RNAs from mammalian species, the mouse and the macaque, contain somatic piRNA. For instance, about 4% of sRNA reads in the mouse pancreas were identified as piRNA. *In situ* hybridization experiments also showed the presence of macaque Piwi gene mRNA in somatic tissue (Yan et al., 2011). However, these results may be due to contamination due to the small fraction of piRNAs identified, may have low sample size issues due to how many reads are in the sequenced libraries, and have yet to be validated (Ross et al., 2014).

1.14 *Ae. aegypti* as a potential model for piRNA pathways

As discussed in previous sections of this introduction, studies are now uncovering the many roles of the piRNA pathway in a variety of different organisms. *Ae. aegypti* may be particularly well suited in answering many outstanding questions about piRNA and PIWI family genes, such as the somatic role of piRNA in epigenetic regulation of the genome, specification of germline cells, and development. Unlike other organisms, *Ae. aegypti* has strong RNA-seq and qPCR support for the somatic expression of an expanded suite of PIWI genes (Akbari et al., 2013), and likely also has somatic piRNA as well. The fact that the majority of piRNA in *Ae. aegypti* do not map to transposon sequences also supports this hypothesis (Arensburger et al., 2011).

Related to the more canonical role of piRNAs in defending against transposons, the large and complex *Ae. aegypti* genome is more similar to mammalian genomes such

as humans, which are also large and have heavy transposon loads. The challenges that such genomes face may be different from the challenges that a genome such as *D. melanogaster* faces. The expanded suite of *Ae. aegypti* PIWIs point to this fact; although there are only four human homologuous PIWIs, the fact that in mammals the PIWI family operates mainly in the male germline suggests that there may be other, more complex transposon defense systems and proteins in the female germline (Aravin et al., 2007b). In addition, the difficulties of germline integration in *Ae. aegypti* and it's ability to rapidly silence integrated transposons and prevent their remobilization makes it an interesting organism to study.

Another potential role of the piRNA pathway in *Ae. aegypti* is the co-opting of the pathway's adaptive ability to amplify piRNAs that match target transcripts, in a similar way to how adaptive immunity works. Studies have already shown that viral-derived piRNA exist, and this may be another function of somatic piRNA in *Ae. aegypti* (Morazzani et al., 2012).

The goal of this dissertation is to elucidate the function of a subset of piRNA pathway proteins in *Ae. aegypti*, focusing mainly on *ago3*. By knocking down *ago3* in the germline and early developmental stages of the mosquito, defects in development, gene regulation, transposon repression, and piRNA populations can be assessed.

1.15 References

Abe, H., Mita, K., Yasukochi, Y., Oshiki, T., and Shimada, T. (2005). Retrotransposable elements on the W chromosome of the silkworm, Bombyx mori. Cytogenet. Genome Res. *110*, 144–151.

Akbari, O.S., Antoshechkin, I., Amrhein, H., Williams, B., Diloreto, R., Sandler, J., and Hay, B.A. (2013). The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. G3 Bethesda Md *3*, 1493–1509.

Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007a). The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race. Science *318*, 761–764.

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007b). Developmentally Regulated piRNA Clusters Implicate MILI in Transposon Control. Science *316*, 744–747.

Arensburger, P., Hice, R.H., Wright, J.A., Craig, N.L., and Atkinson, P.W. (2011). The mosquito Aedes aegypti has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs. BMC Genomics *12*, 606.

Barckmann, B., Pierson, S., Dufourt, J., Papin, C., Armenise, C., Port, F., Grentzinger, T., Chambeyron, S., Baronian, G., Desvignes, J.-P., et al. (2015). Aubergine iCLIP Reveals piRNA-Dependent Decay of mRNAs Involved in Germ Cell Development in the Early Embryo. Cell Rep. *12*, 1205–1216.

Beyret, E. (2009). Function of the Mouse PIWI Proteins and Biogenesis of Their piRNAs in the Male Germline.

Beyret, E., and Lin, H. (2011). Pinpointing the expression of piRNAs and function of the PIWI protein subfamily during spermatogenesis in the mouse. Dev. Biol. *355*, 215–226.

Boswell, R.E., and Mahowald, A.P. (1985). tudor, a gene required for assembly of the germ plasm in Drosophila melanogaster. Cell *43*, 97–104.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell *128*, 1089–1103.

Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. Science *322*, 1387–1392.

Campbell, C.L., Black, W.C., Hess, A.M., and Foy, B.D. (2008). Comparative genomics of small RNA regulatory pathway components in vector mosquitoes. BMC Genomics *9*, 425.

Cox, D.N., Chao, A., and Lin, H. (2000). piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. Dev. Camb. Engl. *127*, 503–514.

Czech, B., and Hannon, G.J. (2016). One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem. Sci. *41*, 324–337.

Dahanukar, A., and Wharton, R.P. (1996). The Nanos gradient in Drosophila embryos is generated by translational regulation. Genes Dev. *10*, 2610–2620.

Darricarrère, N., Liu, N., Watanabe, T., and Lin, H. (2013). Function of Piwi, a nuclear Piwi/Argonaute protein, is independent of its slicer activity. Proc. Natl. Acad. Sci. U. S. A. *110*, 1297–1302.

Doench, J.G. (2003). siRNAs can function as miRNAs. Genes Dev. *17*, 438–442.

Engels, W.R., and Preston, C.R. (1980). Components of Hybrid Dysgenesis in a Wild Population of Drosophila Melanogaster. Genetics *95*, 111–128.

Ephrussi, A., and Lehmann, R. (1992). Induction of germ cell formation by oskar. Nature *358*, 387–392.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). [No Title]. Nature *391*, 806–811.

Fujii, T., and Shimada, T. (2007). Sex determination in the silkworm, Bombyx mori: a female determinant on the W chromosome and the sex-determining gene cascade. Semin. Cell Dev. Biol. *18*, 379–388.

Gangaraju, V.K., Yin, H., Weiner, M.M., Wang, J., Huang, X.A., and Lin, H. (2011). Drosophila Piwi functions in Hsp90-mediated suppression of phenotypic variation. Nat. Genet. *43*, 153–158.

Häcker, U., Nystedt, S., Barmchi, M.P., Horn, C., and Wimmer, E.A. (2003). piggyBac-based insertional mutagenesis in the presence of stably integrated P elements in Drosophila. Proc. Natl. Acad. Sci. *100*, 7720–7725.

Hamilton, A.J. (1999). A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. Science *286*, 950–952.

Han, B.W., Wang, W., Li, C., Weng, Z., and Zamore, P.D. (2015). piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science *348*, 817–821.

Hannon, G.J. (2002). RNA interference. Nature *418*, 244–251.

Harris, A.N., and Macdonald, P.M. (2001). Aubergine encodes a Drosophila polar granule component required for pole cell formation and related to eIF2C. Dev. Camb. Engl. *128*, 2823–2832.

He, L., and Hannon, G.J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. Nat. Rev. Genet. *5*, 522–531.

Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The Drosophila RNA Methyltransferase, DmHen1, Modifies Germline piRNAs and Single-Stranded siRNAs in RISC. Curr. Biol. *17*, 1265–1272.

Houston, D.W., and King, M.L. (2000). Germ plasm and molecular determinants of germ cell fate. In Current Topics in Developmental Biology, G.P. Schatten, ed. (Academic Press), pp. 155-IN2.

Jinek, M., and Doudna, J.A. (2009). A three-dimensional view of the molecular machinery of RNA interference. Nature *457*, 405–412.

Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., et al. (2002). The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol. *3*, RESEARCH0084.

Kawaoka, S., Kadota, K., Arai, Y., Suzuki, Y., Fujii, T., Abe, H., Yasukochi, Y., Mita, K., Sugano, S., Shimizu, K., et al. (2011). The silkworm W chromosome is a source of female-enriched piRNAs. RNA N. Y. N *17*, 2144–2151.

Kirino, Y., Vourekas, A., Sayed, N., de Lima Alves, F., Thomson, T., Lasko, P., Rappsilber, J., Jongens, T.A., and Mourelatos, Z. (2010). Arginine methylation of Aubergine mediates Tudor binding and germ plasm localization. RNA N. Y. N *16*, 70–78.

Kiuchi, T., Koga, H., Kawamoto, M., Shoji, K., Sakai, H., Arai, Y., Ishihara, G., Kawaoka, S., Sugano, S., Shimada, T., et al. (2014). A single female-specific piRNA is the primary determiner of sex in the silkworm. Nature *509*, 633–636.

Klattenhoff, C., and Theurkauf, W. (2008). Biogenesis and germline functions of piRNAs. Dev. Camb. Engl. *135*, 3–9.

Klattenhoff, C., Bratu, D.P., McGinnis-Schultz, N., Koppetsch, B.S., Cook, H.A., and Theurkauf, W.E. (2007). Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. Dev. Cell *12*, 45–55.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853–858.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

Lin, H., and Spradling, A.C. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. Dev. Camb. Engl. *124*, 2463–2476.

Liu, J. (2004). Argonaute2 Is the Catalytic Engine of Mammalian RNAi. Science *305*, 1437–1441.

Liu, X., Sun, Y., Guo, J., Ma, H., Li, J., Dong, B., Jin, G., Zhang, J., Wu, J., Meng, L., et al. (2006). Expression of hiwi gene in human gastric cancer was associated with proliferation of cancer cells. Int. J. Cancer J. Int. Cancer *118*, 1922–1929.

Mahowald, A.P. (2001). Assembly of the Drosophila germ plasm. B.-I.R. of Cytology, ed. (Academic Press), pp. 187–213.

Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. Cell *137*, 522–535.

Matsumoto, N., Nishimasu, H., Sakakibara, K., Nishida, K.M., Hirano, T., Ishitani, R., Siomi, H., Siomi, M.C., and Nureki, O. (2016). Crystal Structure of Silkworm PIWI-Clade Argonaute Siwi Bound to piRNA. Cell *167*, 484–497.e9.

Mével-Ninio, M., Pelisson, A., Kinder, J., Campos, A.R., and Bucheton, A. (2007). The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. Genetics *175*, 1615–1624.

Morazzani, E.M., Wiley, M.R., Murreddu, M.G., Adelman, Z.N., and Myles, K.M. (2012). Production of Virus-Derived Ping-Pong-Dependent piRNA-like Small RNAs in the Mosquito Soma. PLoS Pathog *8*, e1002470.

Nolde, M.J., Cheng, E.-C., Guo, S., and Lin, H. (2013). Piwi genes are dispensable for normal hematopoiesis in mice. PloS One *8*, e71950.

Perrat, P.N., DasGupta, S., Wang, J., Theurkauf, W., Weng, Z., Rosbash, M., and Waddell, S. (2013). Transposition-Driven Genomic Heterogeneity in the Drosophila Brain. Science *340*, 91–95.

Robert, V., Prud'homme, N., Kim, A., Bucheton, A., and Pélisson, A. (2001). Characterization of the flamenco Region of the Drosophila melanogaster Genome. Genetics *158*, 701–713.

Ross, R.J., Weiner, M.M., and Lin, H. (2014). PIWI proteins and PIWI–interacting RNAs in the soma. Nature *505*, 353–359.

Rouget, C., Papin, C., Boureux, A., Meunier, A.-C., Franco, B., Robine, N., Lai, E.C., Pelisson, A., and Simonelig, M. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature *467*, 1128–1132.

Rutherford, S.L., and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. Nature *396*, 336–342.

Semotok, J.L., Cooperstock, R.L., Pinder, B.D., Vari, H.K., Lipshitz, H.D., and Smibert, C.A. (2005). Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early Drosophila embryo. Curr. Biol. CB *15*, 284–294.

Senti, K.-A., and Brennecke, J. (2010). The piRNA pathway: a fly's perspective on the guardian of the genome. Trends Genet. *26*, 499–509.

Seto, A.G., Kingston, R.E., and Lau, N.C. (2007). The Coming of Age for Piwi Proteins. Mol. Cell *26*, 603–609.

Sharma, A.K., Nelson, M.C., Brandt, J.E., Wessman, M., Mahmud, N., Weller, K.P., and Hoffman, R. (2001). Human CD34(+) stem cells express the hiwi gene, a human homologue of the Drosophila gene piwi. Blood *97*, 426–434.

Sienski, G., Dönertas, D., and Brennecke, J. (2012). Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression. Cell *151*, 964–980.

Sinkins, S. (2007). Genome sequence of Aedes aegypti, a major arbovirus vector. Science *316*, 1718–1723.

Smith, R.C., and Atkinson, P.W. (2010). Mobility properties of the Hermes transposable element in transgenic lines of Aedes aegypti. Genetica *139*, 7–22.

Smith, C.D., Shu, S., Mungall, C.J., and Karpen, G.H. (2007). The Release 5.1 annotation of Drosophila melanogaster heterochromatin. Science *316*, 1586–1591.

Sollars, V., Lu, X., Xiao, L., Wang, X., Garfinkel, M.D., and Ruden, D.M. (2003). Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. Nat. Genet. *33*, 70–74.

Taubert, H., Greither, T., Kaushal, D., Würl, P., Bache, M., Bartel, F., Kehlen, A., Lautenschläger, C., Harris, L., Kraemer, K., et al. (2007). Expression of the stem cell self-renewal gene Hiwi and risk of tumour-related death in patients with soft-tissue sarcoma. Oncogene *26*, 1098–1100.

Tolia, N.H., and Joshua-Tor, L. (2007). Slicer and the Argonautes. Nat. Chem. Biol. *3*, 36–43.

Tomari, Y., and Zamore, P.D. (2005). Perspective: machines for RNAi. Genes Dev. *19*, 517–529.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. Science *313*, 320–324.

Waddington, C.H. (1959). Canalization of development and genetic assimilation of acquired characters. Nature *183*, 1654–1655.

Wang, W., Yoshikawa, M., Han, B.W., Izumi, N., Tomari, Y., Weng, Z., and Zamore, P.D. (2014). The Initial Uridine of Primary piRNAs Does Not Create the Tenth Adenine that Is the Hallmark of Secondary piRNAs. Mol. Cell *56*, 708–716.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855–862.

Wilson, J.E., Connell, J.E., and Macdonald, P.M. (1996). aubergine enhances oskar translation in the Drosophila ovary. Dev. Camb. Engl. *122*, 1631–1639.

Yan, Z., Hu, H.Y., Jiang, X., Maierhofer, V., Neb, E., He, L., Hu, Y., Hu, H., Li, N., Chen, W., et al. (2011). Widespread expression of piRNA-like molecules in somatic tissues. Nucleic Acids Res. *39*, 6596–6607.

Yekta, S. (2004). MicroRNA-Directed Cleavage of HOXB8 mRNA. Science *304*, 594–596.

Yigit, E., Batista, P.J., Bei, Y., Pang, K.M., Chen, C.-C.G., Tolia, N.H., Joshua-Tor, L., Mitani, S., Simard, M.J., and Mello, C.C. (2006). Analysis of the C. elegans Argonaute Family Reveals that Distinct Argonautes Act Sequentially during RNAi. Cell *127*, 747–757.

Yu, Y., Gu, J., Jin, Y., Luo, Y., Preall, J.B., Ma, J., Czech, B., and Hannon, G.J. (2015). Panoramix enforces piRNA-dependent cotranscriptional silencing. Science *350*, 339–342.

Chapter 2 – Localization of PIWI family proteins in *Ae. aegypti*

2.1 Introduction

2.1.1 Localization of PIWI family proteins in the *D. melanogaster* ovary

In *D. melanogaster,* the PIWI family proteins are confined to germline and ovarian follicular tissue (Brennecke et al., 2007). Of the three PIWI proteins, Aubergine (Aub) and Argonaute 3 (Ago3) are confined to the true germline – cells derived from primordial germ cell precursors. Aub and Ago3 are localized in the cytoplasm and enriched in the nuage of nurse cells, a perinuclear structure. Ago3 shows stronger enerichment to nuage than Aub (Brennecke et al., 2007). Aub has also been detected in the pole plasm of the developing occyte; Ago3 has not.

In contrast, Piwi is present in both the true germline and also the follicular support cells that surround the ovary. Piwi is primarily nuclear and thus does not show enrichment in the nuage. Small amounts of Piwi have also been detected in the pole plasm of the developing occyte (Brennecke et al., 2007; Cox et al., 2000). These expression domains inform how the proteins interact with each other and the genome in order to function in suppressing transposon activity.

In the somatic follicular cells of the ovary, the only protein expressed is Piwi. Therefore, piRNA produced in the *D. melanogaster* soma do not undergo the amplification loop known as ping-pong amplification. Instead, piRNAs in the soma are primary piRNA generated from long precursors transcribed from piRNA clusters

present in the genome (Aravin et al., 2007). Not all piRNA clusters are expressed in all tissues – some clusters, such as *flamenco,* are soma-specific, some, such as *42AB,* are germline-specific, and some are transcribed in both (Brennecke et al., 2007). These expression patterns may have coevolved with the different strategies that transposable elements use to replicate in the germline of their hosts. For instance, the major somatic piRNA cluster *flamenco* contains many inactivated fragments of gypsy and ZAM transposable elements (Brennecke et al., 2007). Even before the mechanism of function was known, it was observed that *flamenco* is required for the control of those elements – deletion or disruption of *flamenco* resulted in defects in oogenesis (Mével-Ninio et al., 2007; Robert et al., 2001). If unrepressed, *gypsy,* a retroelement that may have once been a retrovirus, will invade the germline from the somatic support cells (Chalvet et al., 1999). The somatic expression of piRNA proteins in the *D. melanogaster* female then serves the function of protecting the germline from DNA damage (Brennecke et al., 2007).

2.1.2 Localization of PIWI family proteins in the *D. melanogaster* testis

Expression patterns in the *D. melanogaster* testis are similar to expression patterns in the ovary. Ago3 and Aub are both present in the cytoplasm of true germline cells: the germline stem cells, primary gonial cells which descend from the stem cells, and as development proceeds, the protein carries over into early spermatocytes (Gunawardane

et al., 2007; Nishida et al., 2007). Neither Ago 3 or Aub are detected in the somatic support cells of the testis – the post-mitotic hub which serves as a niche for the germline stem cell progentitors at the apical tip of the testis (Gunawardane et al., 2007; Nishida et al., 2007). In contrast, Piwi, which is required for spermatogenesis, is seen only at low levels in the germline stem cells. Instead, Piwi is predominately located in the nucleus in the post-mitotic hub where Ago 3 and Aub are absent, as well as in the somatic cyst cells (Gonzalez et al., 2015; Nishida et al., 2007).

2.1.2 Localization of PIWI family proteins in *Ae. aegypti*

Little is known about the localization of the expanded set of PIWI proteins in *Ae. aegypti*. Based on work done in our own laboratory, as well as RNA-seq experiments done elsewhere (Akbari et al., 2013), the PIWI family members *piwi1, piwi2, piwi3,* and *piwi7* are similar to *DmelAubergine* and *DmelArgonaute 3* in that they are only present at significant levels in germline cells. *AaegArgonaute 3* as well as *piwi4, piwi5,* and *piwi6,* however, have an expanded expression domain and do not appear to be restricted solely to the germline since they are also expressed in larval, pupal, and adult stages. mRNA from these genes can be detected through sequencing or through qPCR of RNA extracted from adult carcass dissections with the germline removed(Akbari et al., 2013; Wright, 2011). In addition, *piwi7* has a unique expression domain that is not seen in *D. melanogaster. piwi7* only has significant expression during embryonic development

(Akbari et al., 2013). In *D. melanogaster,* PIWI-clade genes are expressed maternally in the embryo, but then expression disappears (Gunawardane et al., 2007). The numerical expansion of the PIWI-clade in *Ae. aegypti,* along with the expanded expression domain in both somatic tissue and during embryonic development, suggests a larger role for this gene family in the mosquito, one that is not mostly restricted to preserving genome integrity in the germline (Arensburger et al., 2011; Vodovar et al., 2012).

2.1.3 Insect ovary/testis and salivary glands/gastric caecum

Oogenesis in insects occurs in multiple strings of ovarioles within each ovary (Chapman, 1969). Each ovariole has a distal germarium where germline stem cells reside and produce oocytes. As oocytes leave the germarium they become associated with somatic follicular tissue which covers the occyte in a epithelial sheet. Together this unit – both the developing oocyte and its sister nurse cells, as well as the somatic follicular tissue – is called a follicle. As the oocyte grows, by necessity the follicular sheet grows with it. These follicular epithelial cells play important roles in oocyte maturation; along with contribution from the fat body and the nurse cells, they produce yolk proteins and mRNA and deposit them into the oocyte. In addition, the follicular cells produce the vitelline envelope, ligands and enzymes necessary for proper developmental axis specification, and the chorion (Chapman, 1969).

Spermatogenesis similarly proceeds from a distal germarium that is associated with somatic cells – in Dipteran testes, an apical syncytium, or hub. Spermatogonia arise from the germarium stem cell niche and remain associated with the apical hub through cytoplasmic connections for early development before detaching into cysts, which are spermatogonia associated with two somatic cyst cells. Cyst cells also arise from the hub. The cysts then travel down the testis; the spermatogonia they contain mature to spermatocytes and finally to spermatozoa as the cysts approach the vas deferens (Chapman, 1969).

Larval salivary glands in *Ae. aegypt* are derived from rudiments near the proventriculus and are relatively small compared to the lower lateral caeca of which they are adjacent (Christophers, 1960). The gastric caeca are part of the midgut of the alimentary canal, and are composed of many pouches and folds to increase the surface area of the gut. Cells in the gastric caeca both secretory functions – digestive enzymes – and also absorb nutrients and ions. The gastric caeca are not carried over into the adult stage (Chapman, 1969). In addition, microbiotia such as bacteria are present in the gastric caeca, as well as the wider gut, and are required for larval development (Coon et al., 2014).

2.1.4 Chapter aims

The goal of the research described in this chapter was to gain further insight into the somatic and germline roles of AaegAgo3. *ago3* mutants were made to test the phenotype of *ago3* knockdown in *Ae. aegypti* and compare with the phenotypes of *ago3* mutants in *D. melanogaster,* which have increased transposon activity and reduced fertility (Li et al., 2009). Because of the expanded somatic domain of Ago3, it was possible that a more severe phenotype might be observed. Additionally, Ago3, Piwi2, and Piwi7 antibodies were made to detect the localization of these proteins via immunohistochemistry. Previous RNA-seq and qPCR work done in *Ae. aegypti* suggested that Ago3 was present in the soma of *Ae. aegypti*, but these experiments could not determine the subcellular localization of the proteins (Akbari et al., 2013; Wright, 2011). *Ae. aegypti* PIWI genes – with the exception of Ago3 – bioinformatically cluster separately from *D. melanogaster* Aub and Ago3 (Arensburger et al., 2011; Campbell et al., 2008), so finding a nuclear localization for an *Ae. aegypti* PIWI might provide clues to a functional homologue of *D. melanogaster* Piwi.

2.2 Materials and methods

2.2.1 Mosquito stocks and rearing

Strains used in these experiments were either the Liverpool or Orlando strains of *Ae. aegypti.* The strains are maintained in a University of California, Riverside insectary,

under standard conditions as described in (Munstermann, 1997). Larvae were reared in distilled water and fed on a diet of Milkbone Original Dog Biscuits mixed with wheast in a ratio of 1:2. Adults were fed on a 10% sucrose solution and fed either on mice or in an artifical blood feeder system with bovine or sheep blood through a Parafilm membrane.

Ago 3 RNAi lines were transformed at the University of Maryland Insect Transformation Facility with plasmids made by a Robert Hice, a research associate in the Atkinson laboratory. *Ae. aegypti* Orlando mosquitoes were transformed by microinjection of transformation donor (pBacTet-ago3SB) and a helper plasmid expressing piggyBac (pBac) transposase. These lines express a 372 base pair sequence of Ago 3 in a snapback configuration to form a double-stranded RNA molecule under the control of a tetracycline-induced promoter. The tetracycline reversible trans-activator is under the control of a heat shock promoter, hsp70 (Figure 2.1). To induce RNAi targeted against Ago 3, eggs were hatched under vacuum conditions in one liter of distilled water sterilized by UV light. A 30mg/ml aqueous stock solution of doxycycline hyclate was then added, to create final rearing conditions of 30ug/ml of doxycycline.

2.2.2 Antibodies

Primary antibodies were reared in rabbits and purchased from Open Biosystems. Anti-Argonaute 3 was raised against the peptide sequence GQSVKRNPDEALNDKLFYL and anti-Piwi 2 was raised against RPTFQHPGAEGRAMTHRDASAGRGA. Anti-Piwi7 was raised against EYRPRGGRGGNNQARGNVGGEG. Other antibodies used include

anti-EGFP purchased from Invitrogen. Secondary antibodies include HRP-conjugated goat anti-rabbit from Pierce and Alexa Flour 555 goat anti-rabbit.

2.2.3 Tissue dissections

Ae. aegypti females were blood fed and then ovaries dissected either one day, two days, or three days post blood-feeding, depending on the experiment. Testes were dissected from adult male mosquitoes one-to-two days post eclosion. Larval tissues were dissected from mixed-sex 4th instar larvae. Instar development was determined through the size of the animal as well as time passed since hatching. All tissues were dissected in phosphate-buffered saline using forceps underneath a dissecting microscope and then placed into 4% paraformaldehyde fixative in preparation for staining, or snap-frozen in liquid nitrogen for RNA extractions.

2.2.3 Immunohistochemistry

After dissection, tissues were fixed in 4% paraformaldehyde in a cold room at 4° C for two to three days, with gentle rocking. The tissues were held in place inside a Beem embeeding capsule placed into the well of a 24-well plate which was filled with the paraformaldehyde fixative. After fixation, tissues were washed with 0.5% Triton-X in phosphate-buffered saline solution (PBST) three times for five minutes each. Washes were done with gentle rocking at room temperature. The tissues were then blocked with 5% normal goat serum at 4° C overnight, washed three times with PBST, and then

incubated with primary antibody. The primary PIWI antibody was incubated at a concentration of 5 ng/μl overnight, and then the sample was washed three times with PBST. The secondary Alexa 555 flour antibody was likewise incubated at a concentration of 5 ng/μl overnight, and then the samples were washed three times with PBST. Finally, DNA was stained with Hoescht 33258 at a concentration of 1:4000 for 15 minutes. The samples were then washed with PBST and mounted on glass slides in 70% glycerol.

Tissue samples were imaged on either a Leica SP5 inverted microscope or a Zeiss 510 located at the UCR Genomics core. Excitation of Hoechst 33258 was done using a 405 nm laser line. Excitation of Alexa 555 was done using a 543 nm HeNe laser line. Metadata of exact acquisition parameters are shown in Table 2.2, Table 2.3, and Table 2.4.

2.2.5 Western blotting

Proteins were extracted from tissues or animals snap frozen in liquid nitrogen, placed in a modified SDS-free radioimmunoprecipitation assay buffer (RIPA, 25mM Tris, 150mM NaCl, 0.5% sodium deoxycholate, 1% Triton X-100)containing Roche (04693159001 ROCHE) protease inhibitors to prevent protein degradation, and then homogenized. Bio-Rad (500121) RC/DC protein assays were used to determine the concentration of proteins extracted from each sample for the Western Blot loading. 25 to 50 μg of each sample were loaded into Pierce Precise Tris-HEPES 10% Protein Gels and run according to the manufacturer's specifications. Estimation of molecular weight was done with prestained BLUEstain protein ladder from Goldbio and band locations were

transferred onto the film by hand. After size separation, proteins were blotted onto a nitrocellulouse membrane and blocked with Pierce T20 PBS Superblock for two hours at room temperature. The superblock was then drained, but the blot was not washed. From primary antibody stocks of 1 mg/ml, dilutions of 1:20000 and 1:10000 were made in TBST (50 mM Tris, 150mM NaCL, 0.1% Tween 20). Primary antibody incubation was carried out overnight, with gentle agitation, at 4° C. The blot was then washed for 6 minutes in TBST four times. The secondary antibody used for Western detection was a HRP-conjugated goat anti-rabbit antibody from the Pierce SuperSignal West Dura kit at a dilution of 1:2000, incubated for one hour at room temperature. The blot was then washed as before, 6 minutes in TBST four times, and then a final wash in TBST for 30 minutes. Blots were detected with Pierce SuperSignal West Dura stable peroxide and Luminol enhancer. Blots were exposed to Premium Clear Blue X-ray Film from Bioland for 15-30 minutes, and developed on an AFP Imaging Mini-Medical 90 film processor.

2.2.6 Quantitative PCR

RNA was isolated from mosquitoes using TRIzol® reageant following the manufacturer's protocols. For downstream cDNA synthesis, RNA was treated with Ambion® Turbo-DNA-free to destroy genomic DNA. cDNA was synthesized with New England Biolabs® ProtoScript II. Quantitative PCR was performed on a BioRad MyiQ thermal cycler using iQ SYBR® Green Supermix. Efficiency of amplification was determined through a 1:10 dilution series and a standard curve. Primers are listed in

Table 1, and all reactions had the same cycling conditions. Following an initial denaturation at 95°C for three minutes, reaction temperature was cycled between 95°C, 10 seconds, and 60°C, 30 seconds, for 40 cycles. *Ae. aegypti* 30S ribosomal protein 7 (AAEL009496) was used as a housekeeping gene to control for the amount of starting template cDNA. To determine statistical significance of the qPCR results, experiments were run with three biological replicates and then analyzed with REST 2009 for a total of 2000 iterations (Pfaffl et al., 2002).

2.3 Results

2.3.1 Antibody validation

To verify the specificity of the rabbit antibodies reared against Ago 3 and Piwi 2 peptide fragments against the full protein, protein extracts from blood fed ovaries and from developing embryos were probed in a Western blot (Figure 2.2). Banding was seen at 22 and 70 kDA size ranges – this banding was also present when protein extracts were probed using an anti-EGFP antibody also reared in rabbit, and thus was likely a non-specific interaction common to all rabbit anti-sera. In addition to the non-specific bands, a band between 93 and 130 kDA were detected with Ago 3 and Piwi 2 antibodies in blood-fed ovary protein extracts (Figure 2.2). Ago3 weighs 106.5 kDA and Piwi 2 weighs 99.42 kDA; the band in the Piwi 2 lane may be smaller than the band in the Ago 3 land, but the differences in relative signal made it difficult to analyze with certainty. These bands were not present in the EGFP control, indicating a specific interaction with a

protein of the correct molecular weight for both antibodies. Western blots were repeated three times under the same conditions with similar results.

2.3.2 Localization of PIWI proteins in the ovary

Due to previous work showing PIWI family expression in the *D. melanogaster* germline (Brennecke et al., 2007), immunohistochemistry experiments were first performed in blood fed mosquito ovarian tissue to see if these findings could be replicated in a different system. Figures 2.5A and 2.5B show different expression domains for three *Ae. aegypti* Piwi proteins – AaegAgo 3, Piwi 2, and Piwi 7. There was staining for Piwi 7 in some ovarian germariums, but not others, and some staining in follicular epithetlial cells, but not to a high level. This expression may be background. In contrast, AaegAgo 3 had a similar pattern but higher levels of expression in the cytoplasm of germarium and nurse cells. In addition, strong expression was also seen in the cytoplasm of the somatic follicular epithelial cells. Similar to DmeAgo 3, AaegAgo 3 localization did not appear to be nuclear as it did not colocalize with the DNA stain, but although DmeAgo 3 is cytoplasmic, it is predominately localized to the nuage. This was not the case in *Ae. aegypti,* where it is expressed evenly throughout the cytoplasm. In addition, the localization of AaegAgo 3 to the ovarian follicular epithelial cells is not seen in *D. melanogaster (*Brennecke et al., 2007*)*

Piwi 2 was also detected to a high level in germariums, but less expression was detected in the somatic follicular epithelial cells (Figure 2.5A, 2.5B). Piwi 2 signal did not

colocalize directly to the nuclear stain, but a high level of signal was detected around the boundaries of the nurse cell nuclei – an enrichment of Piwi 2 in the nuage of nurse cells. This is similar to the localization of DmelAub (Brennecke et al., 2007). No *Ae. aegypti* PIWI with similar subcellular localization to to DmelPiwi was seen.

2.3.3 Localization of *Ae. aegypti* PIWI proteins in the testis

Both Ago 3 and Piwi 2 exhibited similar localization in the testis (Figure 2.6). However, higher levels of Piwi 2 signal were detected, suggesting that expression of Piwi 2 is higher. Both Ago 3 and Piwi 2 have a decrease in signal at the apical tip of the testis, where the hub resides, suggesting that neither is expressed strongly in that particular somatic tissue. The signal appeared to be primarily cytoplasmic – antibody signal did not colocalize with the DNA stain. The majority of expression is in the developing germ cells and spermatogonia. In the mature spermatozoa further down the testis, expression of both Ago 3 and Piwi 2 was lost. This germline localization pattern is identical to the localization of DmelAgo 3 and Aub (Gunawardane et al., 2007; Nishida et al., 2007). Unlike in the ovaries, AaegAgo 3 did not appear to have an expanded somatic range in the testis. In *D. melanogaster*, Piwi is localized to the apical hub and somatic cyst cells – however, no somatic localization of *Ae. aegypti* Piwi 2 or Ago 3 was seen in in the testis soma.

2.3.4 Phenotypic effects of Ago3 knockdown

When Ago3 M14 mosquitoes were hatched in water with 30ug/ml of doxycycline hyclate to induce foldback Ago3 RNA production and *ago3* knockdown, developmental delay and male mortality was observed (Table 2.5). Upon induction, *ago3* transcript in doxycycline-induced male M14 larvae was on average 19.6% the level in uninduced male M14 larvae, with a P(H1), or the probability of an alternate hypothesis, value of 0%, as determined by qPCR.

Phenotypically, a subset of larva arrested at the 4th instar stage (instar stage was determined by size) and did not develop further. The only adults that successfully eclosed were females, leading to the identification of the developmentally arrested larva as the male larvae. Compared to wild type mosquitoes reared under similar conditions (30 ug/ml of doxycycline hyclate), M14 mosquitoes developed slower – wild type mosquitoes eclosed within six to seven days, while transgenic mosquitoes took up to three weeks to eclose. In addition, the sex of wild type mosquitoes that reached adulthood was roughly split between males (24) and females (16), while in the M14 mosquitoes, only females (15) eclosed. Males died as 4th instar larvae or as pupae.

This severe phenotype of male mortality in 4th instar larvae upon Ago3 knockdown indicated to a vital biological function for Ago3 at this developmental timepoint. Because at this point in development the germline is confined to primordial germ rudiments, I hypothesized that this important biological function of Ago3 was somatic in scope. Previous immunohistochemistry work demonstrating the expanded

somatic role of Ago3 in the germline, along with RNA-seq data showing Ago3 expression in 4th instar larvae, provided some support for this theory (Akbari et al., 2013).

2.3.5 Expression of Piwi family Genes in *Ae. aegypti* Larvae

To confirm that *ago3* was expressed in 4th instar larvae, qPCR was performed to determine if *ago3* transcript could be detected from larval RNA samples. Based on qPCR data, *ago3* transcript was detected in whole 4th instar larva, though not to the level seen in blood fed ovaries (Figure 2.3). Expression in blood fed ovaries was significantly higher (p-value = 0), and was on average of 203.76 times higher (standard error: 91.25 – 325.189) than expression in whole 4th instar larva.

After confirming the presence of *ago3* expression in the whole larvae, enrichment of *ago3* transcript within specific somatic tissues and organs of the larvae was examined. An enrichment in a certain tissue could provide insight into the role of Ago3 within the larvae. Comparing within the larva itself, most tissues did not show significant differences in *ago3* expression compared to the whole larva (Table 2.6). Expression in the gut was on average 1.977 times the expression in the whole larva (standard error: 0.565 – 6.503), but with p-value = 0.599. Based on the calculated p-value, this result of greater expression in the gut was likely to be detected only by chance. Expression in the head was on average 0.126 times the expression in the whole larva (standard error: 0.020 - 0.772), but with p-value = 0.067. Therefore, a change in expression pattern of *ago3* in the

larval head compared to the whole animal was less likely to be detected by chance than the result from the gut, but still above the 5% p-value cutoff that is commonly used.

In contrast, expression of *ago3* in the salivary glands and gastric caecum was on average 8.653 times the expression in whole larva (standard error: 3.325 – 17.339), and with p-value = 0. Although the expression of *ago3* in the larval salivary glands and gastric caeca was not to the high level seen in blood fed ovaries, the difference in expression between larval tissues was detected to a significant level.

Immunohistochemistry in larval tissues (Figure 2.7) showed results supporting the data from the qPCR experiments. Using a specific Ago 3 antibody, signal was detected in the pouches of the gastric caecum. When using a control EGFP antibody also obtained from rabbits, some background signal was detected, but not to the degree seen with the specific antibody. No signal was detected from the specific Ago 3 antibody in the rest of the larval gut.

In *D. melanogaster,* Aub and Ago 3 partner in a ping-pong amplification loop. Further qPCR experiments were performed in *Ae. aegypti* to check for potential piRNA Piwi partners in the salivary glands and gastric caecum (Table 2.7 and Figure 2.4). Expression of *piwi3* was on average 0.01 times the expression of *ago3* (standard error: 0.003 - 0.041), and with p-value = 0.011. Therefore, *piwi3* was not detected at high expression levels in the salivary gland and gastric caecum. In contrast, *piwi4* (0.991 times, standard error: 0.298 – 3.253), *piwi5* (2.203 times, standard error: 0.816-8.712) and *piwi6* (0.931 times, standard error: 0.370-3.518) all had levels of expression similar to *ago3*.

2.3.6 Results Summary

Confirming previous studies (Akbari et al., 2013; Wright, 2011), an expanded somatic role and localization for AaegAgo 3 was observed both in the ovary as well as in the mosquito larvae. Ago3 was present in both the somatic follicular cells of the ovary as well as the germline; knockdown of Ago 3 caused a somatic mortality phenotype in 4th instar larvae, and Ago 3 protein was detected in the somatic larval gastric caeca and salivary glands. No expanded somatic role was seen in the testis. The localization of Piwi 2 was similar to that of *D. melanogaster* Aub. No *Ae. aegypti* PIWI tested had nuclear localization similar to DmelPiwi.

2.4 Discussion

2.4.1 Germline Expression of Piwi Genes in *Ae. aegypti*

To gain insight into the function of the expanded set of Piwi family proteins in *Ae. aegypti,* the laboratory previously had antibodies generated against Ago 3, Piwi 2, and Piwi 7. The goal was to characterize these proteins and see if their subcellular localization and potential function matched with their orthologues in *D. melanogaster* Piwi, Ago 3, and Aub. In addition, RNAi mutant lines of these three genes were also generated with the same goal of comparing to *D. melanogaster* orthologues, though no viable Piwi 2 line was ever recovered, perhaps indicating that Piwi 2 knockdown may be lethal in *Ae. aegypti.*

AaegAgo 3, the PIWI-clade protein which has been determined to be most similar to DmelAgo 3 based on protein sequence analysis (Arensburger et al., 2011), was found to have a much wider somatic expression domain than has been reported in DmelAgo 3. In immunohistochemistry experiments, AaegAgo 3 was also seen in the somatic follicular epithelial cells; therefore, it is not restricted to the true germline of the ovary as reported in *D. melanogaster* (Senti and Brennecke, 2010). In addition, AaegAgo3 was dispersed throughout the cytoplasm. No evidence was seen for a predominate accumulation of Ago 3 in punctate nuage granules around the nurse cell nuclei, as has been shown in *D. melanogaster* (Brennecke et al., 2007; Webster et al., 2015).

One possible explanation for this difference in subcellular localization may be to relate it to the expanded somatic expression domain of Ago 3 in *Ae aegypti.* Since somatic cells do not have nuage, Ago 3 in *Ae. aegypti* may have a function in regulating transposons or other RNA, such as mRNA or viral RNA, outside of the nuage. This is problematic because Ago 3 is known to bind piRNA in the sense orientation to transposable elements – thus, in the cytoplasm, it cannot directly act to guide complementation and cleaveage of transcripts (Senti and Brennecke, 2010). However, there are still parallels to the function of Ago 3 in *D. melanogaster* nuage. In *D. melanogaster* nuage, Krimper helps Ago 3 partner with Aub to form a ping-pong amplification loop that specifically increases the number of Aub complexes bound to antisense piRNA against active transposable elements; Aub then screens transcripts leaving the nucleus for its transposon targets (Webster et al., 2015). Though the kinetics

of interaction will be less favorable with the protein dispersed throughout the cytoplasm, Ago3 could similarly be working with an unknown partner with a commensurate expression pattern in the cytoplasm – perhaps one of the expanded members of the PIWI family in *Ae. aegypti* that are also known to be expressed in the soma, such as *piwi4, piwi5,* or *piwi6. Supporting* this hypothesis, sRNA of the appropriate length and with the hallmarks of ping-pong amplification – a U1 and an A10 – have been found both in adult mosquitoes as well as somatic cell culture lines, such as Aag2, even though these tissues and cells likely lack germline nuage. These piRNA hallmark sRNA are generated in response to viral infection, such as Dengue or Sindbis, and are dependent on Ago 3, Piwi 5 and to a lesser extent Piwi 6. Knockdown of those Piwi family proteins reduced the production of the viral sRNAs (Miesen et al., 2015, 2016).

In contrast, Piwi 2 more closely matches the expression profile of *D. melanogaster* Aub and Ago 3. High levels of Piwi 2 protein are only in the true germline of the ovary, and enrichment was also seen in an even distribution around the nuage, instead of the punctate granules seen in the distribution of DmelAgo 3 and Aub in nuage. Based on protein homology, Piwi 2 is more closely related to Aub than DmelAgo 3, so, taken together, Piwi 2 may be the counterpart of Aub in *Ae. aegypti (*Campbell et al., 2008). If so, its partner in the nuage remains unknown. Possible candidates include Piwi 1 and Piwi 3, which share similar expression pattern to Piwi 2 through a wide range of mosquito tissues and developmental stages, and also seem to be restricted to the germline (Akbari et al., 2013).

Piwi 7 was not detected to a high level in the blood fed ovary, which matches with RNA-seq data of the *Ae. aegypti* developmental transcriptome (Akbari et al., 2013). The role of Piwi 7 in mid-stage embryos, right after the maternal to zygotic transition, is unknown. Another unknown is the identify of AaegPiwi, if it exists. DmelPiwi is located in the nucleus in both the true germline and the somatic epithelial cells, and functions to repress transposons through chromatin modification (Senti and Brennecke, 2010). AaegAgo3, Piwi 2, and Piwi 7 did not have a nuclear localization, but there may be a functional homologue of DmelPiwi with a nuclear localization among one of the untested PIWI-clade proteins.

2.4.2 Somatic Expression of Piwi Genes in *Ae. aegypti Larvae*

The phenotype of male mortality in the 4th instar larval stage in the Ago 3 knockdown transgenic line led to the possibility that Ago 3 has important functions in the larval soma. PIWI family loss of function mutant alleles in *D. melanogaster* have not been observed to be lethal. Instead, these mutant lines tend to suffer defects in fertility and germ stem cell maintenance, such as in the case of *piwi* (Cox et al., 2000; Lin and Spradling, 1997). Similarly, *ago3* mutants in *D. melanogaster* lose the ability to regulate transposable elements and are sterile (Li et al., 2009).

Testing for the expression of *ago3* in 4th instar larvae, qPCR experiments detected mRNA to a much lower level than in blood fed ovaries. However, though at low levels, *ago3* transcript was available to be detected by PCR. Looking further, a significant

enrichment of *ago3* transcript was seen in combined salivary gland and gastric caecum tissue compared to the whole larval animal; however no significant increase in expression was seen in other larval tissues, such as the head or the hindgut. These findings were supported by immunohistochemistry experiments detecting high levels of Ago 3 staining in the pouches of the gastric caecum, but no apparent staining in the hindgut.

The role of Ago 3 in the somatic tissue of the larval gastric caecum and salivary glands may be related to transposon defense, but typically there is more genetic pressure for transposon control in the germline. Mutations that accumulate in germ cells can be passed on to the next generation and may effect long-term reproductive success; mutations that accumulate in the soma will not be passed on. Transposon defense is not emphasized in the soma unless it also effects the germline; for instance, in *D. melanogaster*, Piwi control of *gypsy* in the somatic follicular cells surrounding the oocyte to prevent invasion (Chalvet et al., 1999).

Another possibility is viral defense. As discussed in the previous section, viral piRNAs have been isolated from the somatic cell lines and adult mosquitoes in response to infection, and knockdown of PIWI genes in these systems is accompanied with a rise in viral titer (Miesen et al., 2015, 2016). Similary, *Ae. aegypti* larva are exposed to viruses in their aquatic habitat in the course of feeding and development. These viruses include both dsDNA viruses, include baculoviruses, densoviruses, and iridoviruses, and a dsRNA virus, cypoviruses (Becnel and White, 2007). Infection by one of these viruses, a

invertebrate iridescent virus (IIV), specifically IIV-6, can cause both a patent, or lethal infection, which involves the appeaerance of characteristic iridescent viral particles in the larva, as well as a covert, sublethal infection (Marina et al., 2003; Williams, 2008). Larva exposed to live IIV-6 took longer to develop, and as adults were less healthy and produced less progeny, suffering a 34% reduction in fertility. Larva exposed to heat or UV killed IIV-6 suffered only a 5-17% reduction in fertility, suggesting that sublethal effects of covert IIV-6 infection depends more on viral replication than presence of a viral cytotoxin (Marina et al., 2003). The experiment also suggests that control of IIV-6 infection is important in both the larval and potentially the adult stage in *Ae. aegypti*. In this case, a potent viral defense mechanism in the gastric caecum, where digestive absorption first takes place, would increase fitness, both by decreasing mortality and also increasing the fertility in sublethally infected mosquitoes.

The larval gastric caecum do not persist past metamorphosis – adult mosquitoes have a simpler diet and do not contain caecae – but the contents, including sRNA, are sloughed off into the body cavity and could be retained (Chapman, 1969). If viral piRNA are generated (as has been seen in cell culture) and retained (Miesen et al., 2015), they could packaged into the developing oocyte along with other maternally deposited piRNA as a form of viral immune memory, along with the transposon memory described in *D. melanogaster*, even in the absence of retro-viral integration of viral genes into the genome. Maternally deposited viral derived piRNA in the soma could then

cleave viral transcripts and potentially act in a ping-pong amplification loop characteristic of piRNA.

2.5 References

Akbari, O.S., Antoshechkin, I., Amrhein, H., Williams, B., Diloreto, R., Sandler, J., and Hay, B.A. (2013). The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. G3 Bethesda Md *3*, 1493–1509.

Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race. Science *318*, 761–764.

Arensburger, P., Hice, R.H., Wright, J.A., Craig, N.L., and Atkinson, P.W. (2011). The mosquito Aedes aegypti has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs. BMC Genomics *12*, 606.

Becnel, J.J., and White, S.E. (2007). MOSQUITO PATHOGENIC VIRUSES—THE LAST 20 YEARS. J. Am. Mosq. Control Assoc. *23*, 36–49.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell *128*, 1089–1103.

Campbell, C.L., Black, W.C., Hess, A.M., and Foy, B.D. (2008). Comparative genomics of small RNA regulatory pathway components in vector mosquitoes. BMC Genomics *9*, 425.

Chalvet, F., Teysset, L., Terzian, C., Prud'homme, N., Santamaria, P., Bucheton, A., and Pélisson, A. (1999). Proviral amplification of the Gypsy endogenous retrovirus of Drosophila melanogaster involves env-independent invasion of the female germline. EMBO J. *18*, 2659–2669.

Chapman, R.F. (1969). The insects: structure and function (New York: American Elsevier Pub. Co.).

Christophers, S.R. (1960). Aëdes aegypti (L.) the yellow fever mosquito; its life history, bionomics, and structure. (Cambridge [England: University Press).

Coon, K.L., Vogel, K.J., Brown, M.R., and Strand, M.R. (2014). Mosquitoes rely on their gut microbiota for development. Mol. Ecol. *23*, 2727–2739.

Cox, D.N., Chao, A., and Lin, H. (2000). piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. Dev. Camb. Engl. *127*, 503–514.

Gonzalez, J., Qi, H., Liu, N., and Lin, H. (2015). Piwi Is a Key Regulator of Both Somatic and Germline Stem Cells in the Drosophila Testis. Cell Rep. *12*, 150–161.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science *315*, 1587–1590.

Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., et al. (2009). Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. Cell *137*, 509–521.

Lin, H., and Spradling, A.C. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. Dev. Camb. Engl. *124*, 2463–2476.

Marina, C.F., Ibarra, J.E., Arredondo-Jiménez, J.I., Fernández-Salas, I., Valle, J., and Williams, T. (2003). Sublethal iridovirus disease of the mosquito Aedes aegypti is due to viral replication not cytotoxicity. Med. Vet. Entomol. *17*, 187–194.

Mével-Ninio, M., Pelisson, A., Kinder, J., Campos, A.R., and Bucheton, A. (2007). The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. Genetics *175*, 1615–1624.

Miesen, P., Girardi, E., and van Rij, R.P. (2015). Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. Nucleic Acids Res. gkv590.

Miesen, P., Ivens, A., Buck, A.H., and van Rij, R.P. (2016). Small RNA Profiling in Dengue Virus 2-Infected Aedes Mosquito Cells Reveals Viral piRNAs and Novel Host miRNAs. PLoS Negl. Trop. Dis. *10*.

Munstermann, L.E. (1997). Care and maintenance of Aedes mosquito colonies. In The Molecular Biology of Insect Disease Vectors, J.M. Crampton, C.B. Beard, and C. Louis, eds. (Springer Netherlands), pp. 13–20.

Nishida, K.M., Saito, K., Mori, T., Kawamura, Y., Nagami-Okada, T., Inagaki, S., Siomi, H., and Siomi, M.C. (2007). Gene silencing mechanisms mediated by Aubergine–piRNA complexes in Drosophila male gonad. RNA *13*, 1911–1922.

Pfaffl, M.W., Horgan, G.W., and Dempfle, L. (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. *30*, e36.

Robert, V., Prud'homme, N., Kim, A., Bucheton, A., and Pélisson, A. (2001). Characterization of the flamenco Region of the Drosophila melanogaster Genome. Genetics *158*, 701–713.

Senti, K.-A., and Brennecke, J. (2010). The piRNA pathway: a fly's perspective on the guardian of the genome. Trends Genet. *26*, 499–509.

Vodovar, N., Bronkhorst, A.W., van Cleef, K.W.R., Miesen, P., Blanc, H., van Rij, R.P., and Saleh, M.-C. (2012). Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. PloS One *7*, e30861.

Webster, A., Li, S., Hur, J.K., Wachsmuth, M., Bois, J.S., Perkins, E.M., Patel, D.J., and Aravin, A.A. (2015). Aub and Ago3 Are Recruited to Nuage through Two Mechanisms to Form a Ping-Pong Complex Assembled by Krimper. Mol. Cell *59*, 564–575.

Williams, T. (2008). Natural invertebrate hosts of iridoviruses (Iridoviridae). Neotrop. Entomol. *37*, 615–632.

Wright, J.A. (2011). Investigating Transposable Elements for Use in Dipteran Systems. Ph.D. University of California, Riverside.

2.6. Figures and tables

Primer	**Sequence**
Real RPS7 For	TCAGTGTACAAGAAGCTGACCGGA
Real RPS7 Rev	TTCCGCGCGCGCTCACTTATTAGATT
A3 qPCR r4 F	CGAAGCAGAAGAGCAACTCC
A3 qPCR r4 R	TTCGTACTCGGAGCACATTC
AeB1 Primer F	AGGAAGAGTTGCCCAGTACC
AeB1 Primer R	CGCATGCTTTCATTTGCCAG
Lian Primer F	GAAGAAGCATGTCCTCTGCG
Lian Primer R	AGCCGACTGACTTCACTCAA
Muta Primer F	GAAGGATTACGACGTTTAGCCC
Muta Primer R	CTAGATCGATCCCCAGCTCATT
Gecko F	GACGGACCTGGTGTAGTGG
Gecko R	GCTAGTTCATCTCGGGACCA
Pogo 12 F	TTTGTCATACACTCGCCCGA
Pogo 12 R	TCCAACCGCCTTTCCATAGT
Copia 174 F	TTTGTTTGCTTCGCGGATGA
Copia 174 R	CAGCCGTTAACAAGGATCCG

Table 2.1 – Primers for qPCR

Scanner Settings	
UV Shutter	1
Visible Shutter	1
ScanMode	xyz
Pinhole [m]	112.7 µm
Pinhole [airy]	1.99
Size-Width	775.0 µm
Size-Height	775.0 µm
Size-Depth	0.0 m
StepSize	1.01 µm
Voxel-Width	757.6 nm
Voxel-Height	757.6 nm
Voxel-Depth	1007.1 nm
Voxel-Volume	578032860.096 nm³
Zoom	1.0
Scan-Direction	1
SequentialMode	0
Frame-Accumulation	1
Frame-Average	4
Line-Average	4
Resolution	8 bits
Channels	2
Format-Width	1024 pixels
Format-Height	1024 pixels
Line-Accumulation	1
Sections	18

Table 2.2 – General Leica SP5 confocal microscopy settings

Sequential Setting Nr.1	
Unknown Filter	SMD1 NF 405/470
Polarization FW	Empty
Target Slider	Target Park
UV Lens Wheel	Lens 20x/0.75
PMT 1 (421nm - 487nm)	Active, Gain: 750, Offset: 0
PMT 2 (532nm - 600nm)	Inactive, Gain: 750, Offset: 0
PMT 3 (604nm - 713nm)	Inactive, Gain: 52, Offset: 0
Transmission	Active, Gain: 158, Offset: 0
Laser Line (405)	Shutter: on, Intensity: 20.0024% %
Laser Line (458)	Shutter: on, Intensity: 0.0000% %
Laser Line (476)	Shutter: on, Intensity: 0.0000% %
Laser Line (488)	Shutter: on, Intensity: 0.0000% %
Laser Line (496)	Shutter: on, Intensity: 0.0000% %
Laser Line (514)	Shutter: on, Intensity: 0.0000% %
Laser Line (543)	Shutter: on, Intensity: 0.0000% %
Laser Line (594)	Shutter: on, Intensity: 0.0000% %
Laser Line (633)	Shutter: on, Intensity: 0.0000% %

Table 2.3 - Leica SP5 confocal laser settings for Hoechst DNA stain detection

Sequential Setting Nr.2	
Unknown Filter	Empty
Polarization FW	NF 488
Target Slider	Target Park
UV Lens Wheel	Lens 20x/0.75
PMT 1 (421nm - 487nm)	Inactive, Gain: 932, Offset: 0
PMT 2 (508nm - 593nm)	Active, Gain: 872, Offset: 0
PMT 3 (605nm - 714nm)	Inactive, Gain: 52, Offset: 0
Transmission	Inactive, Gain: 180, Offset: -2
Laser Line (405)	Shutter: on, Intensity: 0.0000% %
Laser Line (458)	Shutter: on, Intensity: 0.0000% %
Laser Line (476)	Shutter: on, Intensity: 0.0000% %
Laser Line (488)	Shutter: on, Intensity: 20.0024% %
Laser Line (496)	Shutter: on, Intensity: 0.0000% %
Laser Line (514)	Shutter: on, Intensity: 0.0000% %
Laser Line (543)	Shutter: on, Intensity: 0.0000% %
Laser Line (594)	Shutter: on, Intensity: 0.0000% %
Laser Line (633)	Shutter: on, Intensity: 0.0000% %

Table 2.4 - Leica SP5 confocal laser settings for Alexa555 detection

	Hatch Rate	Avg. Time to Pupate	Adult Males	Adult Females	Death - 3rd Instar	Death - 4th Instar	Death - Pupae
A3 M14A Induced	44/100	~21 days	0	15	5	20	4
A3 M14A Control	49/100	~6.5 days	24	16	3	2	4

Table 2.5 – A sex bias phenotype was observed in mosquitoes from the M14 line that were reared in tetracycline. Increased larval death in the 4th instar points to male mortality in this developmental stage.

	Ago3 expression, relative to Whole Larvae	P-value	Significant difference?
Whole Larvae	1x	N/A	No
Larval Gut	1.977x	0.599	No
Larval Head	0.126x	0.067	No
Larval SG & GC	8.653x	0.000	Yes
Blood Fed Ovaries	203.76x	0.000	Yes

Table 2.6 – qPCR determined relative Ago3 expression levels in larval tissues and blood fed ovaries, compared to whole 4th instar larvae.

	Expression, relative to Ago3	P-value	Significant difference?
Ago3	1x	N/A	N/A
Piwi 3	0.01x	0.011	Yes
Piwi 4	0.991x	0.779	No
Piwi 5	2.203x	0.453	No
Piwi 6	0.931x	0.834	No

Table 2.7 – qPCR determined protein expression levels of different PIWI genes relative to Ago3, in larval salivary gland and gastric caecae.

ACAACGTAAAGGAATTGGACGCGGCCAGTTGACCTCGGATGCTGGCAGAGCCACCG GAACATCTGGAGCGGATTCATCGGAATCGGATGACAAACAGTCCAGTATTGGATCTG CCATGCCGTCCTTGAGCGGTCGCGGAAGGGCTCAGTTTATTCAGGCCCTTATCCGTCA ACCGTTCGAACCGGCGCCATCGGTTGTCTCCGACGATTCGTCCTCCATGGTGTCTGCA AGGGTGTCGCAAATTGCCTGCGGCAGAGGAAGATTCATCCAGCAGCTGCTGAACAC GGCTGCGGATGCCGAGAGCATCGAAACACAATCGAACGGAAAACATGACGAACTCT CCGAAGCGGTTTCACAGGTCACCATCGCAAA

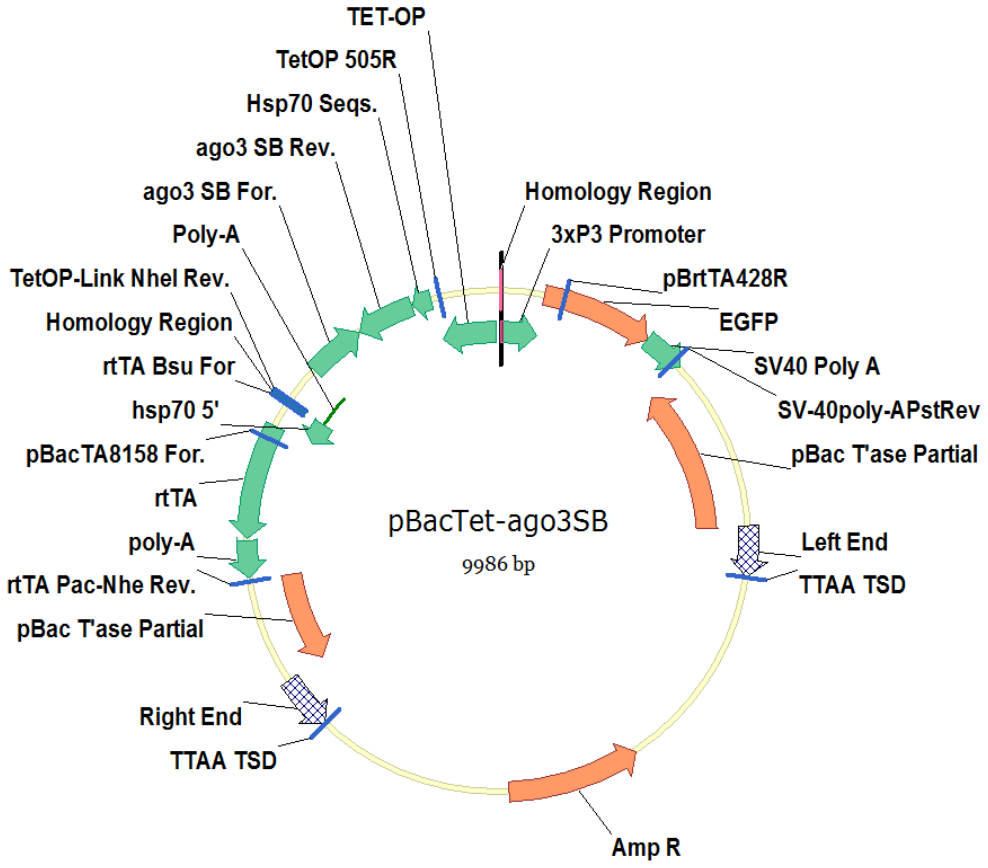

Figure 2.1 – M14 transformation donor plasmid and foldback sequence.

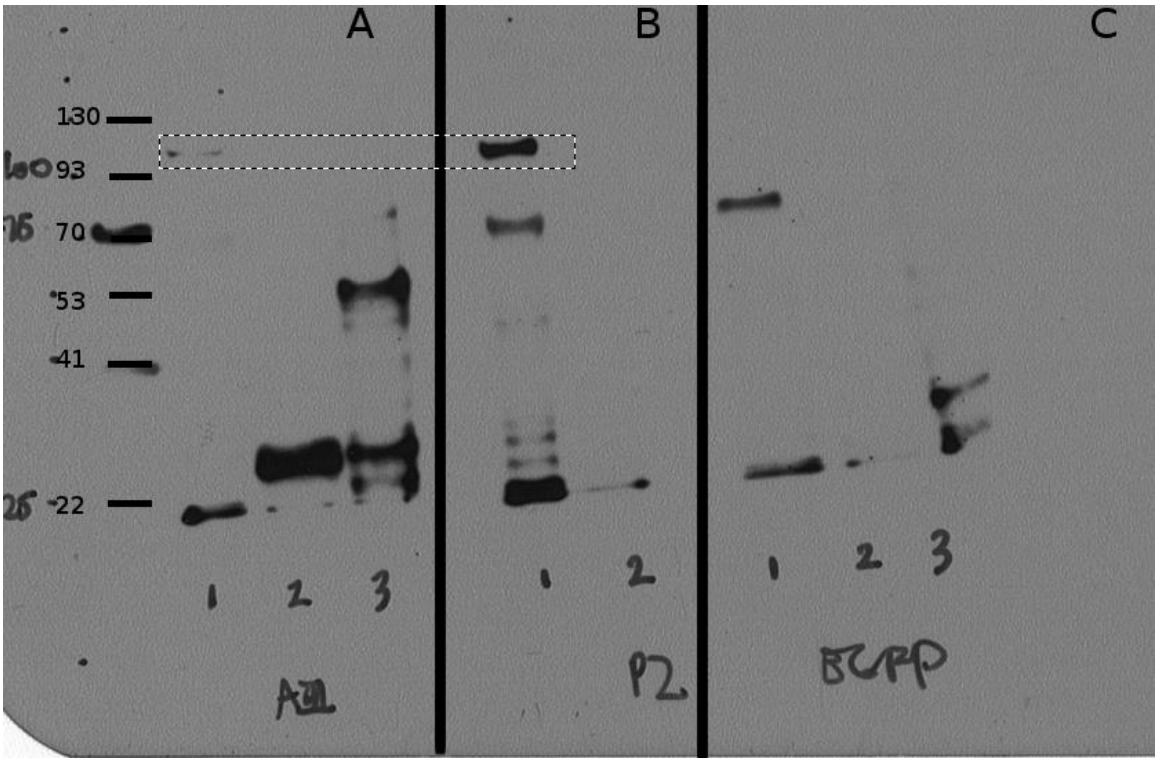

Figure 2.2 – Ago 3 and Piwi 2 Antibody Western Blot. (A) Detection with Ago3 antibody in 1) blood fed ovaries, 2) 4-8 hour embryos, 3) adult whitefly (B) Detection with Piwi 2 antibody in 1) blood fed ovaries and 2) 4-8 hour embryos (C) Detection with control EGFP antibody in 1) blood fed ovaries, 2) 4-8 hour embryos, 3) adult whitefly. Ladder sizes are in kDA.

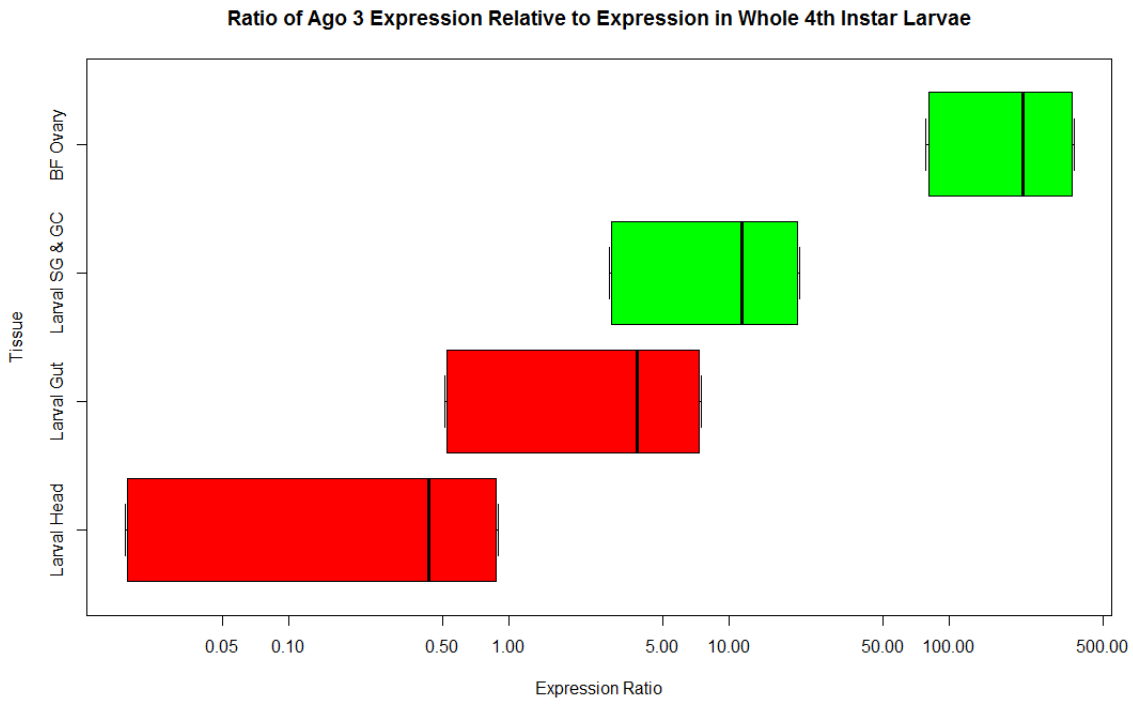

Figure 2.3 - Ratio of Ago3 Expression Relative to Expression in 4th Instar Larvae. Box-and-whisker plot of qPCR data looking at the presence of Ago3 mRNA in various tissues and developmental timepoints. The reference is the mostly somatic tissue of whole 4th instar larvae. Expression of Ago3 is significantly higher in blood ovaries and larval salivary gland and gastric caecum than in whole 4th instar larvae**.** Expression of Ago3 is slightly higher in the gut and lower in the head compared to whole 4th instar larvae, but not to a significant level. X-axis is on a log scale.

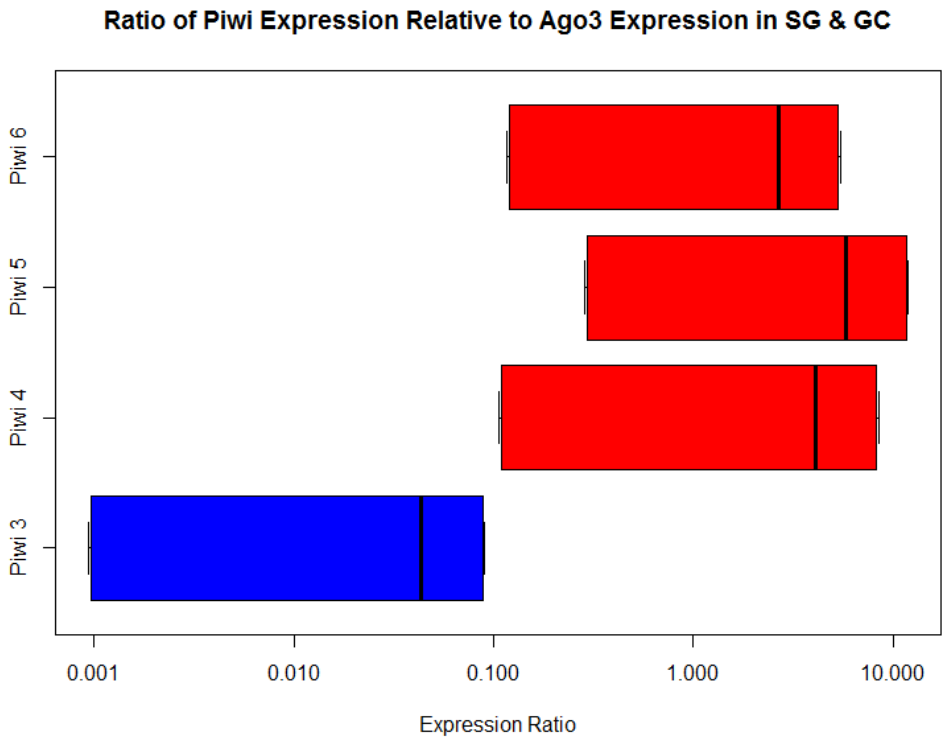

Figure 2.4 - Ratio of Piwi Genes Expression Relative to Expression of Ago3 in Larval Salivary Gland and Gastric Caecum. Box-and-whisker plots of qPCR data looking at the presence of Ago3 mRNA in various tissues and developmental timepoints. The reference is the mostly somatic tissue of whole 4th instar larvae. Expression of Ago3 is significantly higher in blood ovaries and larval salivary gland and gastric caecum than in whole 4th instar larvae, but there is no significant difference in Ago3 expression when compared to the larval gut. X-axis is on a log scale.

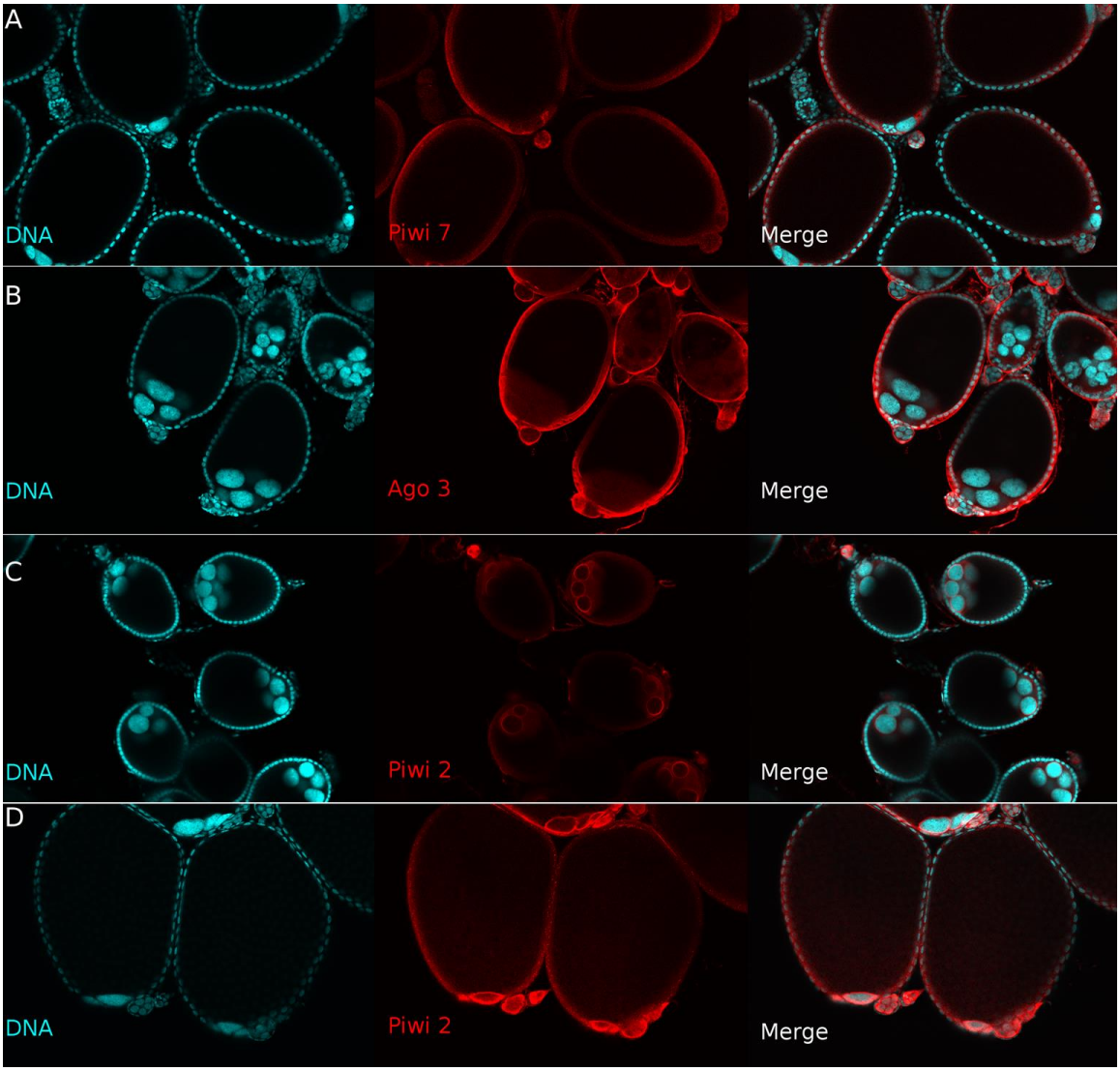

Figure 2.5A - Immunohistochemistry of Ovaries. (A) Signal detected with a Piwi 7 antibody, 2 days post blood feeding (B) Signal detected with a Ago 3 antibody, 2 days post blood feeding (C) Signal detected with a Piwi 2 antibody, 1 day post blood feeding (D) Signal detected with a Piwi 2 antibody, 2 days post blood feeding. Antibody stains are in red and DNA stain is in blue.

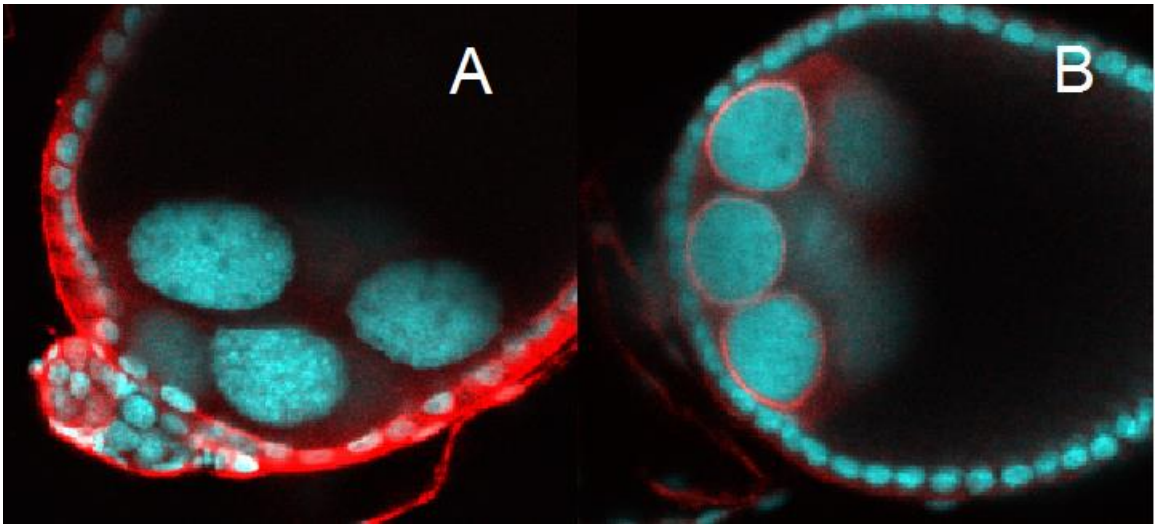

Figure 2.5B - Immunohistochemistry of Ovaries. (A) Zoom of Ago 3 (red) and DNA stain (blue). No colocalization is seen. (B) Zoom of Piwi 2 (red) and DNA stain (blue). No colocalization is seen.

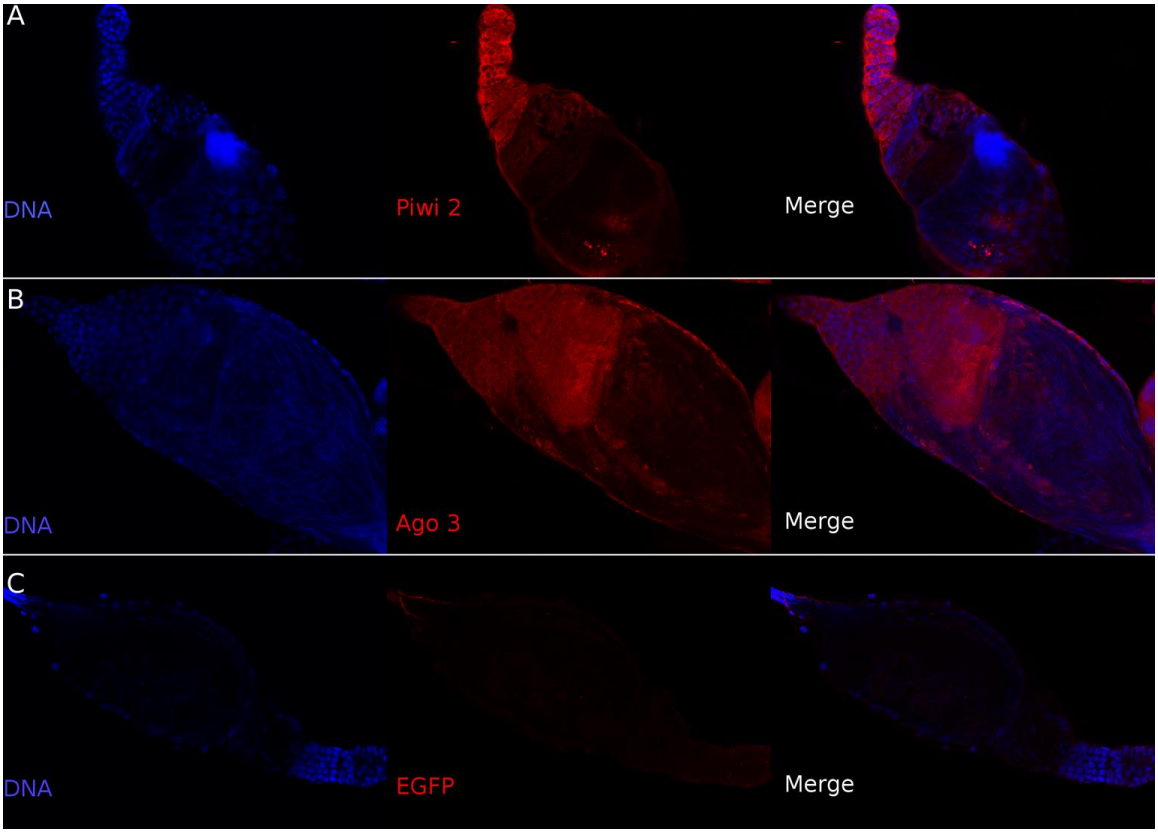

Figure 2.6 - Immunohistochemistry of Testis. (A) Signal detected with a Piwi 2 antibody (B) Signal detected with an Ago3 antibody (C) Signal detected with an EGFP antibody. Antibody stains are in red and DNA stain is in blue.

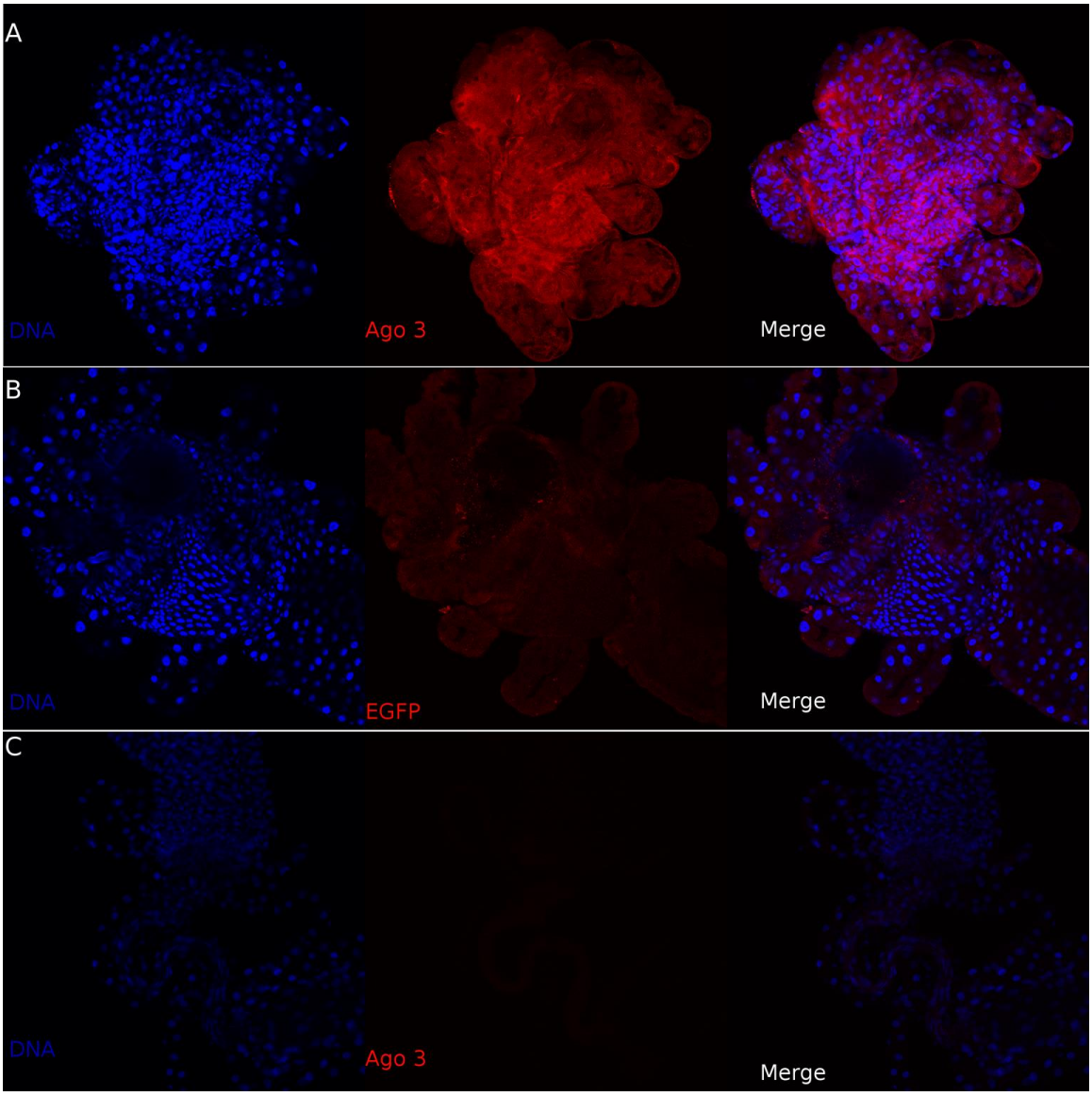

Figure 2.7 - Immunohistochemistry of 4th Instar Larval Tissue. (A) Signal detected with an Ago3 antibody in larval salivary glands and gastric caecum (B) Signal detected with an EGFP antbody in larval salivary glands and gastric caecum (C) Signal detected with an Ago3 antibody in the larval hindgut. Antibody stains are in red and DNA stain is in blue.

Chapter 3 - Characteristics of piRNA clusters in *Aedes aegypti*

3.1 Introduction

3.1.1 *Flamenco* in *Drosophila melanogaster*

Primary piRNA originate from genomic loci known as piRNA clusters. A well characterized piRNA cluster expressed in the somatic support cells of the *D. melanogaster* ovary is the *flamenco* locus, which was first mapped in a mutant stock that had increased mobilization of the *gypsy* element and subsequent deleterious effects on female fertility (Mével-Ninio et al., 2007; Pélisson et al., 1994; Prud'homme et al., 1995). Studies revealed that the effects of *flamenco* on *gypsy* mobilization were restricted to the somatic follicle cells that surrounded the female germline, and that the gene locus played a role in suppressing the expression of a retroviral-like envelope protein; this protein was required for *gypsy* transposition by facilitating the invasion of *gypsy* into the germline (Pélisson et al., 1994). The suppression of this protein was due to a maternal effect factor; *gypsy* transposition permissive mothers produced *gypsy* permissive progeny regardless of the phenotype and genotype of the father (Pélisson et al., 1994). Subsequent mutant lines were found that mobilized the *gypsy*-like elements, *ZAM* and *Idefix;* these also mapped to *flamenco,* but these mutant lines did not also mobilize *gypsy,* suggesting that the locus' ability to control multiple transposable elements is not all-or-nothing – mutant *flamenco* lines exist which are permissive to only a subset of transposons (Coline et al., 2014; Desset et al., 2003).

Sequencing of *flamenco* was difficult because of its location in a highly repetitive, heterochromatic region of the genome, but early surveys of the locus identified a very high density of TEs, with a span of around 180 kilobases containing 104 different element insertions representing 42 different transposon superfamilies, the majority of which were LTR retrotransposons (Quesneville et al., 2005; Smith et al., 2007). The transposon fragments present in *flamenco* were then demonstrated to be the genomic origin of a newly discovered small RNA species, piRNA, which were sequenced from sRNA libraries made from *D. melanogaster* ovaries (Brennecke et al., 2007). Of the piRNA generating loci identified in ovarian tissue, *flamenco* was somewhat unique; 94% of the piRNAs that mapped to it uniquely were partners of only one piRNA protein, Piwi. In addition, *flamenco* produced piRNAs with a very strong U1 signature and anti-sense bias against transposons, consistent with the strongly uni-directional biased orientation of the TEs in the actual genomic locus. Of the 15 top piRNA clusters identified, only 3 had this strongly biased genomic architecture (Brennecke et al., 2007). Based on this and further CHIP-seq pulldown studies with the other *D. melanogaster* PIWI proteins, Ago3 and Aub, it was proposed that *flamenco* generates a single long precursor transcript that is further processed into many anti-sense primary piRNA which bind to Piwi to silence transposons that have inserted into *flamenco*, which tend to be LTR retrotransposons that inserted into the locus. Analysis of these insertions has found that the majority (52.7%) of transposon sequences had 98% identify with reference elements – this lack of sequence divergence suggested that these elements were recent insertions; 12 of these recent

insertions included nine LTR-retrotransposons (*phidippo, blood, stalker4, gedeo, mdg1,* and *pifo)* and three LINE elements (Zanni et al., 2013). No ping-pong amplification or sense-oriented piRNA are known to be involved in this *flamenco* linked mechanism of silencing in the somatic follicular cells (Brennecke et al., 2007; Malone et al., 2009). Instead, heterochromatin recruitment and transcriptional silencing has been found to be involved (Mohn et al., 2014; Yu et al., 2015).

3.1.2 Germline clusters in *D. melanogaster*

In contrast to *flamenco,* the majority of piRNA clusters described and found by Brennecke *et al (*2007) in the *D. melanogaster* ovary do not have the unidirectional orientation that characterizes *flamenco;* instead, they contain a more even distribution on both strands. For instance, 78% of *flamenco* locus sequence was annotated as transposons orientated on the minus strand, compared with 38% of 42AB locus sequence annotated as transposons orientated on the plus strand and 32% on the minus strand (Brennecke et al., 2007). In addition, although the majority of piRNAs that can be mapped to these clusters were anti-sense to the transposons that they silence, a significant fraction were sense, leading to a model of these clusters undergoing dual-stranded precursor transcription, in contrast to the undirectional *flamenco* (Yamanaka et al., 2014). Dual-stranded transcription of these clusters, activated only in the germline of *D. melanogaster* where the full suite of piRNA proteins are expressed, allows for ping-pong amplification of piRNA species against transposable element sequences in both orientations, whether

they derive from the slicing of active transposons or are processed from antisense precursors transcribed from germline clusters (Malone et al., 2009). In fact, knockdown of the sense-piRNA component of the Piwi pathway, Ago3, caused a greater collapse of anti-sense piRNA than sense piRNA, due to disruption of this amplification process (Li et al., 2009).

Unlike *flamenco,* which primarly targets *gypsy*-like elements, germline clusters such as 42AB contain a lesser proportion of *gypsy*-like elements, perhaps because these elements are primarily expressed in the somatic follicular cells and use a viral like nuclear envelope to invade the germline. Instead, TEs that are controlled by insertions into 42AB include long interspersed elements (LINEs) such as the I element, and class II elements such as the P element (Brennecke et al., 2008; Hirakata and Siomi, 2016).

3.1.3 piRNAcClusters in *Anopheles gambiae*

Anopheles gambiae is a related Dipteran species and disease vector more closely related to *Ae. aegypti (*approximately 150 million years of divergence) than *D. melanogaster* (250 million years of divergence) (Gaunt and Miles, 2002; Krzywinski et al., 2006). The sequenced genome size is 278 Mbp, with a transposon content of 16% of the euchromatic genomic sequence and 60% of the heterochromatic genomic sequence, compared to 2% and 8% for *D. melanogaster* (Holt, 2002). The transposon content of *An. gambiae* is higher than found in *D. melanogaster,* but lower than the transposon content

found in *Ae. aegypti,* where transposable elements make up 47% of a large, 1.38 Gbp genome (Nene et al., 2007).

In *An.gambiae,* piRNA populations sequenced from ovarian tissue were found to differ from similar populations in *D. melanogaster.* Depending on mapping methodology, 65.4 – 81.6% of *D. melanogaster* piRNA were found to map to transposable elements; the fraction was smaller in *An. gambiae*, only 23.6 – 39.4% Instead, a greater fraction of *An. gambiae* sRNA libraries mapped to unannotated and genic regions (George et al., 2015). Another difference was the directionality of the piRNA clusters called from the sRNA libraries – in *D. melanogaster* ovaries, the large majority (80%) of top piRNA clusters are bidirectional. In *An. gambiae* ovaries however, based on a piRNA cluster calling methodology of requiring 0.05% of uniquely mapping piRNA to map to a 5 kb region, a greater proportion of unidirectional clusters were observed – 35% (or 1.75x times the *D. melanogaster* fraction) of called piRNA clusters had a strand bias of 70% or more, and 11.2% had a strand bias of 90% or more. Total clusters called were 187 for *An. gambiae* and 155 for *D. melanogaster* (Brennecke et al., 2007; George et al., 2015). Unidirectional clusters are unable to produce piRNA that can undergo ping-pong amplification due to their architecture; in *D. melanogaster*, unidirectional clusters are expressed in the somatic support cells of the ovary, of which the most well characterized example is *flamenco;* this may indicate that there are more somatic piRNA clusters and soma-specific piRNA in the *An. gambiae* ovarian follicular cells.

Examination of the nucleotide biases of *An. gambiae* ovarian piRNA populations revealed that, from 22 million piRNA reads, 79.1% contained a U1 signature and 28.4% contained a A10 signature. Looking at the portion of the library which mapped to transposable elements, for sense-mapping piRNA, 60.9% had a U1 and 54.0% had a A10. For antisense mapping piRNA, 84.3% had a U1 and 27.8% had a A10 (George et al., 2015).

Although the antisense U1 and sense A10 nucleotide bias is conserved in *An. gambiae,* there are also several key differences compared to *D. melanogaster* that may be features of the piRNA pathway in mosquitoes that are absent from the *D. melanogaster* model. A lesser fraction of sequenced piRNA mapping to transposons and a greater fraction mapping to genic regions suggests a greater role for piRNAs in processes unrelated to transposon regulation. This may be related to the greater observed somatic character of the piRNA clusters called in *An. gambiae* and perhaps a larger somatic piRNA pool. Since the sRNA libraries were made from ovaries however, it is not possible to strictly separate somatic and germline piRNAs (George et al., 2015).

3.1.3 piRNA clusters in *Aedes aegypti*

A previous study of piRNA clusters in *Ae. aegypti* focused on sRNA libraries generated from whole body samples (Arensburger et al., 2011). Since these libraries came from the whole organism and mixed both somatic and germline tissue, analysis of these libraries was unable to distinguish between clusters that may be expressed in only

one class of tissue. In addition, analysis of these libraries was unable to determine if different developmental timepoints or tissues in *Ae. aegypti* may have different levels of piRNA cluster transcription or activity, based on the amount of piRNA that could be mapped to genomic locations (Arensburger et al., 2011). The analysis undertaken followed the piRNA cluster calling algorithm used by Brennecke *et al* (2007) to find piRNA clusters in *D. melanogaster,* looking for initial genomic hotspots of mapping piRNA using the criteria of five uniquely mapping piRNA per five kb of genomic sequence, and collapsing the windows that overlapped; in addition, cluster calls within 20 kb of each other were also collapsed, even if there few uniquely mapping piRNA within this interval. This algorithm called 3.5% of the assembled *D. melanogaster* genome as a piRNA cluster, with 92% of the sequenced piRNAs potentially deriving from these called genomic loci Brennecke et al., 2007). For *Ae. aegypti,* a cut-off of greater than ten uniquely mapping piRNA per five kb window was used instead to avoid calling a very high percentage of the genome as a piRNA cluster; in addition, clusters were not collapsed, even if they were within 20 kb of each other, which prevented adjacent called genomic loci from chaining together into large piRNA clusters that contain intervening genomic sequence with low piRNA density and would also increase the genome occupancy of called piRNA clusters. Even with this more stringent criteria, this modified algorithm called 20.6% of the *Ae. aegypti* genome as a piRNA cluster, one-fifth of the total annotated genome and six times higher than the *D. melanogaster* proportion. 84% of the sequenced piRNA could derive from these genomic loci (Arensburger et al., 2011).

Based on the analysis of these libraries and piRNA cluster calls, some characteristics of the *Ae. aegypti* piRNA system were identified. Similar to *D. melanogaster,* the majority of TE-mapping piRNA are antisense – 72%. In addition, both a U1 and an A10 bias was observed in these piRNAs (Arensburger et al., 2011; Brennecke et al., 2007). However, a smaller fraction of the total filtered piRNA library mapped to transposons – 19% – more comparable to the fraction found in *An. gambiae* (23.6 – 39.4%*).* In *D. melanogaster,* transposon mapping piRNAs are the majority (65.4 – 81.6%) instead of the minority (19%). Also similar to *An. gambiae,* more of *Ae. aegypti* piRNAs mapped to genic and unannotated regions, despite the fact that *Ae. aegypti* has a much higher transposon load than both *An. gambiae* and *D. melanogaster* (George et al., 2015). In terms of piRNA cluster calls, differences from *D. melanogaster* include a much higher genome occupancy (almost six times higher, based on this method of piRNA cluster calling) and a much lower overall TE density in called clusters. In *D. melanogaster,* clusters are significantly enriched in TE content, with 70-94% TE occupancy depending on the cluster. In *Ae. aegypti,* TE content in the top 30 called piRNa clusters was found to be 23-42% occupancy, roughly equivalent to the TE load of the actual genome. Like *D. melanogaster* however, many of these clusters, despite a lower overall TE occupancy, had their transposable elements orientated in the same orientation, much like the unidirectional *flamenco* locus (Arensburger et al., 2011; Brennecke et al., 2007).

However, a drawback of this algorithm of calling piRNA clusters was that it utilized a flat cutoff – for instance five or ten piRNAs per five kb genomic window –

which can be somewhat arbitrary. Since potential unique piRNA populations can be very large, ranging from an estimate of 2×10^5 for the mouse to 1.7×10^7 for *Ae. aegypti* and 1.6×10^7 for *D. melanogaster* (Arensburger et al., 2011; Betel et al., 2007), it would take higher read coverages then have been achieved so far to sequence the full pool of *Ae. aegypti* piRNAs. Using a flat, unnormalized measure of piRNA read count density to call genomic loci as piRNA clusters would thus likely call more and more of the genome as a piRNA cluster in a linear fashion simply as a function of sequenced library size, rather than biological relevance. In addition, a method of calling piRNA clusters that works for the genome of one organism, such as *D. melanogaster*, may not work for the genomes of organisms with genomes with different characteristics, such as *Ae. aegypti,*which has a much larger genome with a much higher transposon load (1.38 Gbp; 47% of the genome). In the mouse testes, (2.56 Gbp genome, 37.5% transposon load), clusters covering 2.7 million bp were called based on a criteria of 5 mapping piRNA per 5 kb; however, these clusters were called from a small library of only 77 thousand reads instead of the millions of reads that are common in modern libraries (Chinwalla et al., 2002; Girard et al., 2006).

The criteria for uniquely mapping piRNA can prevent the calling of genomic loci which by chance contain a high concentration of repetitive transposable element sequences as a piRNA cluster – this in part is due to the inability to definitively assign where a non-unique piRNA originated from. Although these non-unique piRNA can be

assigned fractionally to all regions in the genome to which they map, many of these regions will not be the actual origin for a canonical piRNA precursor transcript that is processed into piRNA; instead, many of these sites will be locations where the transposons are actively being silenced. However, completely excluding non-unique piRNA can also mask genomic loci which are in fact true piRNA clusters, perhaps the result of duplication events or simply as a feature of organisms that have a high transposon load, such as *Ae. aegypti* (Assis and Kondrashov, 2009; Girard et al., 2006). Because of the nature of transposons as repetitive sequences, it therefore can be difficult to strike a balance between false-negatives and false-positives when mapping piRNAs to a genome to call piRNA clusters. One method which may be more nuanced than previous methods is to use all mapped reads, not just uniquely mapping, and compute what density of reads count as significant over a uniform distribution of reads over a chromosome or scaffold, thereby controlling for library size. Further analysis takes into account piRNA cluster features, such as U1 and A10 biases and strand asymmetry (Rosenkranz and Zischler, 2012). This method calls much less of the *Ae. aegypti* and *D. melanogaster* genome as piRNA clusters, but the calls may be more biologically significant.

3.1.4 Chapter aims

The goal of the research described in this chapter was to further characterize the unique features of piRNAs and piRNA clusters in *Ae. aegypti.* With the advent of

increasing large sRNA library sets and new information on the expanded somatic expression domain of PIWI family proteins in *Ae. aegypti,* new methods of piRNA cluster calling and annotation that have been developed can be applied to provide a more accurate picture of the genomic loci that produce the densest amount of piRNA. In addition, sequencing of wholly somatic tissues such as the larval gastric caecae and salivary gland provided a unique opportunity to separate somatically expressed piRNA clusters from germline expressed piRNA clusters, which is more difficult to accomplish in mixed-tissue samples such as ovaries which have been the standard for piRNA sequencing thus far.

3.2 Materials and methods

3.2.1 Mosquito stocks

Mosquitoes were reared as described in a chapter 2, section 2.2.1.

3.2.2 Tissue dissections and sample collection

Ovarian and larval tissues were dissected as described in chapter 2, section 2.2.3. For embryo collection, a wet filter paper was placed into a cage containing blood fed females. The females were allowed to lay in darkness for two hours after which the filter paper was removed from the cage and allowed to sit for another two hours at room temperature before the embryos were collected, yielding a sample of 2-4 HR embryos, a

timepoint where zygotic transcription in *Ae. aegypti*, which develops slower than *D. melanogaster*, is low (Biedler et al., 2012).

3.2.3 Ovarian, embryonic, and whole body RNA library construction

Construction of these libraries was performed by former members of the lab, and the method is described in the methods section of Arensburger *et al*., 2011. Briefly, 10-20 μg of Trizol-extracted whole body total RNA was first size selected (16-35bp) from a 15% polyacrylamide/7M urea/TBE gel. The size selected RNA was then run through the Illumina small RNA sample prep kit for 3' and 5' adaptor ligation, with a gel purification and size selection step after each ligation. The ligated RNA was then transcribed into cDNA with Superscript II (Thermo Fisher Scientific), and the library was amplified with Phusion DNA polymerase. Library sequencing was performed on the Illumina Gax2 sequencer at the IIGB Genomics Core (Arensburger et al., 2011).

Analysis and processing of these libraries including adaptor trimming and ribosomal and miRNA removal and genome mapping was performed. I then undertook further analysis of these libraries as described below.

3.2.4 Pull-down library construction

Antibodies used for pull-down library construction were generated by Open Biosystems as described in chapter 2, section 2.2.2. Pull downs were performed by J. Wright, a former student in the laboratory. Briefly, blood-fed ovaries were collected and

proteins were extracted from each sample as described in chapter 2, section 2.2.5. Briefly, each sample was then incubated with antibody in a ratio of 1:10 for Ago3, Piwi2, and the EGFP control, and 1:1 for Piwi7, overnight at 4° C. The homogenate/antibody solution was added to Protein G beads (Thermo Fisher Scientific) and pulled down with a magnet. Pulled down and washed beads were then resuspended in Trizol for RNA extraction. The RNA were then sequenced with the Illumina TruSeq® RNA kit Revision A on the Illumina Genome Analyzer II DNA Sequencer at the IIGB Genomics Core (Wright, 2011).

3.2.5 Mass spectrometry

Mass spectrometry sample preparation immunoprecipitation followed the protocol outlined in chapter 3, section 3.2.2, omitting the final Trizol step to elute the protein bound RNA. Instead, the protein-Protein G bead complex was submitted for mass spectrometry at the IIGB Proteomics Core. The samples were subjected to a chemotrypsin treatment and then spectrometry was performed with a Waters Q-TOF nano-ESI/MS/MS with 2D-nanoAcquity UPCL. For analysis, MatrixScience's Mascot Daemon program was used to analyze the peptide fragments (Wright, 2011).

3.2.6 Quantitative PCR

qPCR was performed as described in chapter 2, section 2.2.6.

3.2.7 Doublesex PCR sexing of larvae

To sex larval RNA samples for downstream qPCR or sequencing analysis, cDNA was synthesized as described in chapter 2, section 2.2.6, and tested for the presence of a 300 base pair female specific splicing transcript of the *Ae. aegypti doublesex* gene, using primers dsx3 and dsx4 as described previously by Salvemini (Salvemini et al., 2011). *doublesex* was chosen because it controls sex differentiation in insects and is known to have sex-specific differential splicing. PCR was performed with Bioneer AccuPower® PCR PreMix Taq polymerase, and products were resolved on a 1.5% TBE-agarose gel stained with ethidium bromide.

3.2.8 Larval library preparation and sequencing

Libraries were prepared with New England Biolab® library preparation kits. For sRNA libraries, the Multiplex Small RNA Prep Kit for Illumina was used. Total RNA was used as input for RNA-seq library poly-A bead enrichment. This RNA was extracted from larval tissue with Trizol as described in chapter 2, section 2.2.6. For sRNA libraries, total RNA was first purified by loading with NEB RNA loading dye into a 15% TBE-urea polyacrylimide gel, and then size separated at 200V for one hour. The gel fraction containing RNA molecules approximately 21-30nt in size was excised and the RNA was eluted into a 0.3M NaCl solution and then used as input for the sRNA enzymatic steps as per the manufacturer's protocol. Eluted RNA was ligated to 3' and 5' Illumina adaptors and then reverse transcribed and amplified.

A 5% TBE gel was used for the final purification and size selection step, and DNA fragments between the sizes of 145 and 160 nucleotides in size were excised, extracted from the gel, and ethanol precipitated. Library samples pooled together in multiples of eight to be sequenced together in a single lane. Sequencing was carried out at the IIGB Genomics Core on a Illumin HiSeq2500. Sequencing runs were single-end and 50 bp long.

3.2.9 sRNA mapping

For these sRNA input libraries, sequencing reads were assigned to different samples by adapter sequence, and then the adapters were trimmed using the command line tool, Cutadapt (Martin, 2011). The Illumina adapter sequence 5'-AGATCGGAAGAGCACACGTCTG-3' was trimmed from the 3' end of all reads, and then only reads between 24 and 31 nucleotides, fitting the size profile of piRNAs, were retained. (Command line: cutadapt -a AGATCGGAAGAGCACACGTCTG -o output.fastq input fastq -m 24 -M 31) In addition, reads with low quality scores were discarded based on the cutadapt quality trimming algorithm. The reads that remained were mapped using Bowtie2 (Langmead and Salzberg, 2012) to a collection of *Ae. aegypti* ribosomal RNA, micro RNA, and snoRNA that were pulled from the non-coding RNA header of a list of *Ae. aegypti* RNA sequences obtained from Vectorbase. Only reads that did not align to one of the above non-coding RNA categories were retained.

A genome annotation file, Aedes-aegypti-Liverpool_BASEFEATURES_Aaegl4.4.gtf was downloaded from Vectorbase to assign reads to genes. To assign reads to transposable elements, a separate file containing known *Ae. aegypti* transposon sequences tabulated previously in the lab by a research associate, Peter Arensburger, was used. To assign reads to viral genomes, a file containing all viral genomic sequences was downloaded off NCBI. Bowtie2 was the mapper used to make these calls.

3.2.10 piRNA profile analysis and cluster calls

To analyze the nucleotide biases of library reads, .fastq files were analyzed by a command-line toolkit called FASTX-Toolkit (Blankenberg et al., 2010). (Command line: fastx_quality_stats -i BC54.fq -o bc54_stats.txt -Q33) Nucleotide bias percentages were then calculated by division from the raw number of reads containing a particular nucleotide compared to the total number of reads. Of most interest were nucleotide biases at position 1, where a U bias is seen in *D. melanogaster*, and position 10, where an A bias is seen in *D. melanogaster.* Fastq files used included the whole libraries (with rRNA, miRNA, and snoRNA removed) as well as the fastq files generated by Bowtie2 containing reads that aligned against transposons, genes, or viruses, to investigate nucleotide biases in just a subset of library reads.

To determine the magnitude of ping-pong overlap (a measure of secondary piRNA abundance), or 10 bp of sense/antisense alignment at the 5' end of two different

sRNA reads, a next generation sequencing (NGS) toolbox perl script (Rosenkranz et al., 2015) was used to determine how many standard deviations above background the number of 10 nt overlapping complementary pairs were, by comparing the numbers of reads with a 10 bp overlap with the number of reads with overlaps of other sizes, from 1 to 9 bp and from 11 to 20 bp. From this number, a ping-pong Z-score was calculated. A Z-score standard deviation value above 1.649 corresponds to a p-value of less than 0.05; a Z-score above 2.362 corresponds to a p-value of less than 0.01 (Rosenkranz et al., 2015). A separate script developed by a graduate student in the lab, Patrick Schreiner, was used to identify specific reads that exhibited a ping-pong overlap for further analysis.

To call potential piRNA clusters in libraries made from different tissues types of *Ae. aegypti (*germline, somatic, and developmental stages), a more stringent method than looking at raw piRNA number mapping per one kb of genomic sequence was used, proTRAC. Reads were first collapsed and the mapped to *Ae. aegypti* genome with the software SNRAmapper to produce the input for ProTRAC (Rosenkranz and Zischler, 2012). ProTRAC uses the sliding window method of calling, but also normalizes for library size and enforces a requirement for typical piRNA characteristics before calling a particular genomic location as a piRNA cluster (Rosenkranz and Zischler, 2012). Default settings and options were used while running the scripts to prepare the libraries for the ProTRAC run, and then for the ProTRAC run itself.

3.2.11 piRNA cluster analysis with piClusterBuster

To compile data of interest on the piRNA cluster calls generated by ProTRAC, the genomic coordinates of called piRNA clusters in a BED file format were used as an input for a pipeline developed by another student in the lab, Patrick Schreiner, called piClusterBuster (Schreiner and Atkinson, 2017). Default settings were used for piClusterBuster to generate summaries for either the top 30 or top 50 piRNA cluster calls, as determined by density of normalized piRNA hits per kb of genomic loci sequence. Summaries generated include nucleotide occupancy of feature content (unannotated, transposon, or gene), transposon family overviews, as well as the plus/minus orientation of the transposon sequences in the genome.

3.3 Results

3.3.1 Nucleotide bias of *Ae. aegypti* piRNA populations

Examination of the nucleotide bias and ping-pong Z-scores of *Ae. aegypti* libraries led to a number of observations. The two germline libraries, 2 – 4 hour embryos and blood fed ovaries, had profiles similar to each other as well as to previously characterized *D. melanogaster* libraries. An *Ae. aegypti* embryonic library (32 million reads) contained a population of sRNA that were 74.7% U1 and 38% A10. An *Ae. aegypti* blood fed ovarian library (82 million reads) contained a population of sRNA that were 75.1% U1 and 33.1% A10. This is similar to an short-read archived *D. melanogaster*

ovarian library (SRR618933, 20.5 million reads) which had a population of sRNA that were 58.72% U1 and 30.67% A10 (Rosenkranz, 2016). Aligning these germline libraries to annotated transposon sequences and looking at the nucleotide composition of only the reads that mapped did not significantly alter these U1 or A10 biases – for instance, TE mapping embryonic sRNAs (13.6 million) were 76.23% U1 and 37% A10, while TE mapping ovarian sRNAs (31 million) were 74.6% U1 and 33.3% A10. (Table 3.1 and Table 3.2.)

In contrast, sRNA libraries made from more somatic tissue showed differences in nucleotide composition depending on whether the whole library was analyzed, or if only the subset of the library which mapped to transposable elements was analyzed. In addition, somatic libraries, compared to the germline libraries, had fewer reads mapping to transposable elements. In male 4th instar larvae sRNA (14 million reads), a composition of 58.6% U1 and 26.1% A10 was found. In larval gastric caecae and salivary glands sRNA (39 million reads), a composition of 63.9% U1 and 31.2% A10 was found. In the subset of reads (1.2 million) that mapped to transposons in whole 4th instar male larvae, a greater U1 composition of 85.7% U1 and 23% A10 was found. In the subset of reads (2.8 million) that mapped to transposons in larval gastric caecae and salivary glands, a greater U1 composition of 83.4% U1 and 20.6% A10 ocurred. (Table 3.1 and Table 3.2.)

The subset of reads in the somatic libraries that mapped to genes was higher than the subset of reads that mapped to transposons, and had weaker U1 signatures

compared to the library as a whole. In the subset of reads (2.6 million) that mapped to genes in whole 4th instar male larvae, 34.7% contained a U1, and 39.2% contained a A10. In the subset of reads (9.2 million) that mapped to genes in the larval gastric cacae and salivary gland libraries, 42.1% contained a U1, and 32.9% contained a A10 (Table 3.1 and Table 3.2).

Examining the sRNA libraries as a whole tended to mask the A10 bias because antisense, U1 sRNA are the dominant species. To better identify potential ping-pong signatures, I also looked at the nucleotide composition of the subset of sRNA that mapped to transposons in both the antisense as well as the sense orientation (Table 3.1). When transposon mapping reads were analyzed in a orientation specific manner, a clear trend could be seen. Antisense mapping reads had a stronger U1 bias, and sense mapping reads had a stronger A10 bias, for both germline and somatic libraries. For instance, in the male control larval libraries, 44.23% of reads that mapped to transposons in the sense orientation had a U1, compared to 89.13% of reads that mapped to transposons in the antisense orientation. 40.57% of the reads that mapped to transposons in the sense orientation had an A10, compared to only 20.97% of the reads that mapped to transposons in the antisense orientation. Similarly, in the ovarian libraries, antisense mapping reads had the stronger U1 bias (75.7% compared to 60.85%) and the weaker A10 bias (29.65% compared to 41.45%). (Table 3.1, Figure 3.1).

Examination of 5' 10 nt overlap Z-scores revealed that the germline libraries had strong Z-scores (Figure 3.2). *Ae. aegypti* ovarian libraries had an average Z-score of 65.93,

while *Ae. aegypti* embryonic libraries had an average Z-score of 34.06. For comparison, a *D. melanogaster* ovarian library had a Z-score of 52.5 (Rosenkranz, 2016). Based on these results, the amount of ping pong amplification that occurs in the *Ae. aegypti* germline and the *D. melanogaster* germline is similar. For non-germline libraries, the Z-scores were remain significant but were much lower, representing a lower number of precise 10 nt overlaps and less ping-pong amplification. Male 4th instar larvae had an average Z-score of 1.66, which corresponds to a p-value of less than 0.05, and larval gastric caecae and salivary glands had an average Z-score of 6.08, which corresponds to a p-value of less than 0.01. When mapping the distribution of overlaps, the *D. melanogaster* and *Ae. aegypti* ovarian libraries had very even distributions of complementary reads with overlaps of 1-9 and 11-20 base pairs – noise or overlaps that occur by chance – with one single large peak at the 10 base pairs of overlap due to the molecular mechanism of ping-pong amplification and piRNA biosynthesis. The somatic *Ae. aegypti* libraries showed a different profile – there was a much less even distribution of overlaps between complementary pairs, with peaks both at 5 nucleotides of overlap as well as 10 nucleotides of overlap. For the male 4th instar larvae library, the peak at 5 nucleotides was larger than the peak at 10 nucleotides. Looking at the nucleotide distributions of these libraries, base position five did not have a A10 bias stronger than other positions that might explain this peak (Figure 3.5).

3.3.2 proTRAC piRNA cluster calls

General trends for proTRAC calling of piRNA clusters focused more on the transposon content of the library, not on the total library size, due to the methods of normalization as described in section 3.2.10 (Table 3.3, Figure 3.2, Figure 3.3). Libraries with a lower relative transposon content (Ago3 knockdown 4th instar male larvae and whole body adults) had less of the *Ae. aegypti* genome called as a piRNA cluster – 1.44 to 1.55 million base pairs on average – and these clusters could be the origin of a smaller fraction of the total library – 12.99 to 36.11%, on average. Somatic libraries with a higher relative transposon content had more of the genome called as a piRNA cluster – 2.78 to 3.04 million base pairs on average – and these clusters could be the origin of a greater proportion of the library – 47.57 to 67.05%, on average. The germline libraries, which had a very high relative transposon content, had 2.26 to 2.56 million base pairs of the genome on average called as a piRNA cluster, which could be the origin of 23.38% to 28.06% of those tissue's sRNA populations, on average. Based on these results, the proTRAC method of calling may have under-called the number of piRNA clusters in germline tissues, but did a better job of calling somatic piRNA clusters. The results also suggest that there is a greater diversity of piRNA clusters expressed in the germline than in the soma.

3.3.3 Somatic piRNA clusters

piRNA clusters that were only expressed in somatic tissues were identified by running proTRAC on libraries generated from somatic tissues. piRNA cluster calls generated from somatic tissue sRNA libraries tended to have very high concordance with one another, both within biological replicates and also between tissue types (Table 3.11, Table 3.12, Figure 3.6). For instance, the base pairs of overlap of piRNA cluster loci within the larval somatic libraries (3 larval gastric caeca & salivary gland libraries and 3 whole 4th instar male larvae libraries) was 2.22 million ± 0.126 million base pairs; on average, clusters called from somatic libraries were 76.3% identical with each other. These somatic libraries shared some similarities with the ovarian libraries (2 libraries, one from the Liverpool strain and one from the Orlando strain of *Ae. aegypti),* with 0.952 ± .0358 million bp of overlap, on average 32.7%, and dissimilar to the early embryonic libraries (2 libraries, same strains as the ovarian libraries), with only 0.105 ± 0.077 million bp of overlap, or on average 3.6% identify. As another measure of the similarity in cluster calls within somatic tissues, in the top 25 piRNA clusters ranked by normalized piRNA hit count, there was only one called from the gastric caeca library, located on sc1.535, that was not found in the larval libraries (Table 3.3). In the top 25 piRNA clusters ranked by normalized piRNA hit count, there was only one called from whole larval libraries, located on sc1.33, that was not found in the gastric caeca libraries.

In the somatic libraries, there was a large difference in normalized hits between the top piRNA dense clusters. For instance, cluster 1 called from the gastric caeca and

salivary gland library had a normalized hit number of 214,943; cluster 25 had a normalized hit number of only 1,225. Cluster 1 was the origin of 130 times more sequenced piRNA than cluster 25. In male 4th instar larvae, cluster 1 was the origin of 68 times more sequenced piRNA than cluster 25. This suggests that, in the soma, the majority of piRNA come from a few highly expressed cluster, and helps explain close to a majority of library sRNA could be assigned to proTRAC called clusters in the soma. (Table 3.4 and Table 3.5).

In the somatic libraries, only 0.4% to 2.2% of the total called piRNA clusters were bidirectional (Table 3.10). The overwhelming majority of somatic clusters were unidirectional; that is, the majority of genomic features were arranged in the same orientation, either on the plus or minus strand, and piRNA were likewise derived from only one strand, implying precursor transcription from one end of the cluster to generate antisense piRNA. All of the piRNA mapping to these clusters had a strong U1 bias, but only 9.4% to 10.6% of the clusters had piRNAs mapping with a A10 bias over 33%, consistent with a predominant somatic primary piRNA pathway.

3.3.4 Example somatic piRNA clusters

An example of a simple unidirectional *Ae. aegypti* somatic piRNA cluster is the cluster on supercont1.478 (sc1.478), located from positions 495182 to 565395, a size of 70 kb. This cluster was called in libraries with somatic tissue – whole larva, ovaries, larval gastric caeca & salivary gland, and whole body adults, but not in the embryonic

libraries. The cluster is composed primarily of unannotated sequences, as well as transposable elements; the majority (greater than 90%; Figure 3.9) of the transposable elements are located on the minus strand of the genome. The predominant transposon family is the Gypsy family, followed by BEL and I elements.

Another somatic piRNA cluster is located on sc1.1, from 1151066 to 1486014, about 370 kb. Similar to the cluster on sc1.478, this cluster was only called in somatic tissue. The cluster is composed primarily of unannotated sequences, as well as transposable elements. Unlike previously described somatic clusters, this cluster has a unique feature – the first half of the cluster (about 130 kb) has the transposable elements arranged in the genome on the minus strand; the mapping piRNA are antisense to these elements and map to the plus strand. The latter half of the cluster has the transposable elements arranged in the genome on the plus strand; the mapping piRNA are antisense to these elements and map to the minus strand. Thus, the orientation of the mapping sRNA follow the split in cluster feature orientation (Figure 3.15). The predominant transposon family was again the Gypsy family, followed by BEL elements (Figure 3.10, 3.11, 3.12). Other somatic clusters found to have the split architecture of piRNA cluster sc1.1 were clusters on sc1.105, sc1.299, and sc1.489 (Table 3.13). For a full sRNA mapping and feature description of the 254 gastric caeca clusters and the 226 larval clusters, please see the supplemental data.

3.3.5 Germline piRNA clusters

By running proTRAC on libraries generated from germline tissue, clusters that were only expressed in germline tissues were identified. piRNA cluster calls from germline tissue enriched sRNA libraries tended to be more variable, in contrast with the uniformity of the calls in the soma. For instance, within the two early embryonic libraries there were 1.06 million bp of overlap (41.4%), and within the two ovarian libraries there was 0.946 million bp of overlap (41.9%). In comparison, the larval somatic libraries had 2.22 ± .126 million bp of overlap within samples (76.3% on average). Looking at the overlap between the two germline enriched tissues, ovaries and embryos, there was some piRNA cluster overlap of 0.725 million ± 0.232 million bp (28.3%). (Table 3.11, 3.12; Figure 3.6).

Unlike the somatic clusters, many germline clusters were called to be bidirectional; that is, genomic features were arranged in both orientations, sense and antisense, and piRNA were derived from both strands, implying precursor transcription from both ends of the cluster (Table 3.10). Unlike the piRNA clusters called from somatic clusters (0.4% to 2.2% bidirectional and 9.4% to 10.6% A10 fraction), ovarian and Ago3 pulldown libraries had a bidirectional cluster content of 18.5% and 22.4%, and 38% and 37.6% A10 fraction, respectively. Higher bidirectional cluster content and A10 enrichment are both hallmarks of the secondary piRNA pathway and ping-pong amplification. Bidirectional and A10 enrichment was strongest in the embryonic and Piwi 2 pulldown libraries – 28.8% and 35.8% bidirectional and 59.7% and 45.7% A10,

respectively. This could be due to the contribution of the somatic follicular cell piRNA pool to the ovarian libraries.

In addition, the ovarian library clusters, which contain both germline tissue as well as follicular somatic cells, shared called clusters with both the embyronic and the somatic library clusters. There were very few shared cluster calls between the embryonic and the somatic library clusters (Figure 3.6 and 3.7).

The spread of mapping piRNA between the top 25 clusters by piRNA hit density was not as great in the germline tissues as was found in the somatic tissues (130x and 68x). In the ovaries, the spread of mapping piRNA between cluster 1 and cluster 25 was only 17 fold; cluster 1 was the origin of 17 times more sequenced piRNA than cluster 25. In Ago 3 pulled down ovaries, the spread was nine fold. The spread was even lower for the more germline piRNA enriched libraries – three fold for the embryonic library, and 3.7 fold for the Piwi 2 pulled down ovaries (Tables 3.6 to 3.9). This suggests that, unlike in the soma, piRNA in the germline tissues aren't expressed predominately from a few major clusters but are expressed from a wider range of genomic loci.

3.3.6 Example germline piRNA cluster

An example of a *Ae. aegypti* germline cluster is the cluster located on supercont1.192, from 1716002 to 1760015, about 44kb. This cluster was called from sRNA libraries made with significant amounts of germline tissue – the ovarian and early embryonic libraries. It was not called in the more somatic libraries – the whole 4th instar

or gastric caecae libraries. This cluster had sRNA mapping to both strands of its genomic loci, and its transposon features were not orientated in a unidirectional fashion. Instead, features were almost equally split on the minus and positive strand. The predominant transposon family was not Gypsy; instead it was the R1 family, followed by DNA transposons and Gypsy elements (Figure 3.13, 3.14). For a full sRNA mapping and feature description of the 200 ovarian clusters, 243 embryonic clusters, 125 Ago 3 ovarian pull down clusters, and 162 Piwi 2 pulldown clusters, please see the supplemental data.

3.4 Discussion

3.4.1. Nucleotide bias and Z-score overlap of *Ae. aegypti* sRNA libraries

The nucleotide biases and the standard deviation Z-score of the 10 bp 5' U1/A10 ping-pong overlaps of the sRNA libraries generated from germline enriched *Ae. aegypti* tissues – 2 day blood fed ovaries and 2-4 hr early embryos – were broadly similar to the piRNA composition trends seen in *D. melanogaster.* There was a strong U1 bias (75%) and a weaker A10 bias (around 30%), reflecting an overall greater abundance of piRNA that map in the sense orientation of transposons, which in *D. melanogaster* are bound to Ago3 and tend to have an A at nucleotide position ten to complement the antisense piRNA with a U at position one (Malone et al., 2009). These libraries also had the greatest relative proportion of reads mapping to transposable elements, though not to the degree seen in *D. melanogaster.* Instead, the transposon mapping proportion in the *Ae. aegypti* germline libraries – around 20 - 30% - was more similar to that seen in *An. gambiae* (23.6

– 39.4%). In *D. melanogaster*, transposon mapping piRNAs are the majority (65.4 – 81.6%) (George et al., 2015). Restricting the reads analyzed from the germline libraries to reads that map to transposons did not effect these nucleotide biases determined from the whole library. Filtering the reads based on whether or not the reads mapped sense or antisense to a transposon did make a difference however; as expected from previous work in other species, sense mapping piRNA reads had a greater A10 bias and antisense mapping reads had a greater U1 bias.

The ping-pong Z-scores generated from the degree of 5' overlap of complementary pairs (from 1 bp of overlap to 20 bp of overlap) were likewise similar among the germline *Ae. aegypti* tissues and *D. melanogaster* ovaries (65.93, 34.06; 52.5). The numbers of overlaps followed an even distribution, with a peak at a 5' overlap of 10 base pairs.

The nucleotide biases and the Z-scores generated from the somatic libraries had some unexpected characteristics. The weakest U1 bias was seen in the reads that mapped to genes. Although reads that mapped to genes made up more of the somatic libraries than reads that mapped to transposons, they had a U1 bias that was weaker than that of the whole library, implying that if the piRNA pathway is silencing gene transcripts in the larva, it's doing so by a mechanism different from the canonical anti-sense complementarity, perhaps with one of the expanded somatic PIWI family members in *Ae. aegypti* such as Piwi 4, 5, or 6.

In addition, the U1 and A10 biases in the whole libraries were weaker in terms of a piRNA signature when compared to the U1 and A10 composition of only the reads that mapped to transposons. One explanation for this is the much lower transposon content in the somatic libraries; however, even in the germline libraries, where no difference was seen between the whole library of reads and the transposon set of reads, roughly 70-80% of the reads did not map to transposons.

A further finding was the lack of an A10 signature in the larval gastric caeca and salivary gland sRNA, where I have shown there is an enrichment of Ago3. Even when looking solely at sense-transposon mapping reads, the A10 bias is what could be expected by chance – 26.97% on average. It could be that in the somatic tissues, only the primary piRNA pathway as described in *D. melanogaster* is functioning, and therefore no ping-pong amplification is occurring. Instead, many primary anti-sense piRNA are being produced. *Ae. aegypti* Ago3, perhaps in concert with its expanded somatic role in the mosquito, may be functioning less like *D. melanogaster* Ago3 and more like *D. melanogaster* Piwi. In the whole 4th instar male larva however, an A10 bias is observed. This could be the influence of a germline compartment within the larva, not the gastric caeca, that expresses the full suite of *Ae. aegypti* Piwi proteins that may be necessary for ping-pong amplification.

This theory is partially contradicted by the finding that a Z-score of 6.08 is calculated for the gastric caeca library; that is, there are 6.08 standard deviations more 5' overlaps of exactly 10 base pairs of complementary pairs than is expected from

background, which is unlikely to happen by chance. The Z-score for the whole larva is lower, at 1.66. One explanation is a small subset of transposons, genes, or other genetic elements that are regulated through a secondary piRNA mechanism. The comparatively small amount of piRNA mapping to these elements is not enough to bias the nucleotide distributions of the whole library; however, there may be enough to register above the background noise of the few piRNA expected to have a perfect 10 nt overlap by chance.

One unexpected finding for the overlap plots for both somatic library sets was a major difference in distribution compared to the germline mosquito and fly libraries – whereas the germline libraries had an even distribution of overlaps from 1-20, with an exception of position 10, the somatic libraries had wildly ranging distributions. In addition, there was an unexplained large peak at 5 base pairs of overlap in both the somatic libraries. It could be that there is another mechanism of piRNA synthesis in the mosquito soma, perhaps related to the expanded Piwi family, both in terms of proteins and expression domain. However, unlike the strong germline Z-scores observed as a consequence of a strong U1 and A10 bias, there were no strong nucleotide biases at position 5 of the gastric caeca library that might explain this overlap peak.

3.4.2. Somatic piRNA clusters in *Ae. aegypti*

The clusters called from sRNA libraries made from larval somatic tissue – both the whole larva and the gastric caeca libraries – were very similar to each other. The degree of overlap ranged from 60-75%, and among the top 25 piRNA clusters,

agreement was almost universal. There was only 1 piRNA cluster among their tissue's top 25 that was unique to either gastric caeca or whole larva. In addition, the proportion of library reads that could be assigned to the piRNA cluster calls was high for the somatic library – 48 to 68% - and unlike in the previous methodology, the amount of the genome that was called as a piRNA cluster was much lower, only about 3 million base pairs instead of 20% of a 1.38 GB genome. This gives more confidence that there are no false positives among the piRNA calls, while still retaining sensitivity for the piRNA clusters that originate the majority of the piRNA. Supporting this conclusion is the finding that the top piRNA cluster in the soma originates as many as a hundred times more piRNA than the 25th cluster, and a thousand times more piRNA than the most weakly expressed called clusters.

Of the somatic clusters that were revealed from this analysis, characteristic-wise they are very similar to the *flamenco* locus in *D. melanogaster.* Somatic clusters in *Ae. aegypti* are likewise unidirectional, with transposon features all orientated in the same direction in the genome either on the plus or minus strand. The mapping sRNA map onto whichever strand of the genome is anti-sense to the transposons contained, implying that a precursor transcript is being transcribed from the somatic piRNA cluster antisense to the transposons. This antisense precursor transcript is then processed into many different antisense primary piRNA, which can go on to complement and silence transposon mRNA. The transposon families silenced is also similar to the fly – the predominant transposon species contained in these clusters and which give rise to

piRNA are elements in the Gypsy family, which possess the capability to invade the germline from the somatic follicular cells of the ovary. The utility of this is less clear in the wider soma of *Ae. aegypti*, such as in the gastric caeca or the whole 4th instar larva, but these Gypsy-heavy somatic clusters are also expressed in those tissues, judging by their sRNA populations. It could be that the mosquito benefits in fitness from controlling Gypsy transposition and mutation in the wider soma as well.

One unexpected feature was the dual nature of some of the somatic clusters. The first halves of these clusters have transposons on the minus strand and sRNA mapping to the plus strand – antisense to the transposable elements. The second half of the cluster however is reversed – transposons map to the plus strand and sRNA map to the minus strand, again antisense to the transposable elements. There is little to no crosstalk at the boundary between the two different halves of the cluster; that is, no sense piRNA were detected at the boundary. This cluster is unlike a somatic unidirectional cluster, where transcription only occurs in one direction – instead, transcription of the precursor occurs at both ends. However, it is also unlike a germline bidirectional cluster, where transcription occurs in both directions and goes through the whole cluster, generating antisense piRNA for silencing and sense piRNA to participate in the ping-pong amplification loop, generating even more antisense piRNA. Instead, transcription of the precursor appears to stop in the middle of the cluster, judging by the lack of sense piRNA being produced. One possible explanation is that transcription is leaky and some sense piRNA are being produced, just not to the level detected by the sequencing

coverage of the libraries. These leaky sense piRNA could then lead to a small amount of ping-pong amplification in the soma, if the right Piwi protein partners were available. Another potential explanation is that these are two unidirectional clusters that are located near one another by chance or a chromosomal rearrangement/inversion.

3.4.3 Germline piRNA clusters in *Ae. aegypti*

The clusters called from the germline enriched tissue in *Ae. aegypti* were less similar to each other than the clusters called from the somatic tissue – both within replicates and also across tissues. In addition, these clusters, though occupying roughly the same genomic size as the somatic clusters, could account for a lesser proportion of the sRNA population of the libraries they were called from. One explanation for this is that the piRNA diversity in these germline libraries is very high. Since each library can only take a snapshot of the piRNA population present in each sample, tissues with higher piRNA diversity are subject to more variability from one replicate to the next. Thus, the relative consistency of the somatic clusters may be due to a lower piRNA diversity and fewer overall somatic clusters being expressed in the genome, at least compared to the germline tissues.

A higher piRNA diversity in the germline tissues may also account for the relatively low proportion of cluster-mapping piRNA in these libraries. Due to its sensitive method of mapping and normalization, proTRAC may be missing bona-fide piRNA clusters due to a large subset of piRNA mapping to genomic loci at a moderate

level being mistaken as background noise. Instead, only the very high density piRNA clusters are called. Thus, in terms of germline clusters, there may need to be a less sensitive methodology and a common ground between calling 20% of the genome as a cluster (Arensburger et al., 2011), and only 0.015%. The clusters that proTRAC does call are still more likely to be the more significant and top piRNA generating loci. However, the difference between the top piRNA generating loci and the bottom piRNA generating loci in the germline is not as marked as in the soma. Instead of hundreds or thousands more piRNA being generated from the top cluster, in the germline the top cluster only generates three to ten fold more piRNA. This provides more support that there are bona-fide piRNA clusters in the germline that fail to meet proTRAC's density requirement to be called as a cluster. A possible explanation for this phenomenon is a more diffuse, spread out expression of many piRNA clusters in the germline. This may be an adaption to the large number of unique transposons and transposon families present in the *Ae. aegypti* genome, all of which may be activated in the germline as part of their transmission strategy and require repression to preserve genome integrity; in contrast, in the soma, there is less element variety with the main active species being Gypsy. Fewer elements would require fewer piRNA clusters/genome occupancy to control.

In terms of similarity to the somatic cluster calls, there is a degree of similarity between piRNA loci in ovarian tissue and the wider somatic tissue. This is most likely due to the influence of the somatic follicular cells in the ovary. The degree of similarity of piRNA loci in early embryonic tissue and the wider somatic tissue is roughly ten imes

smaller in comparison. *Ae. aegypti* develops slower than *D. melanogaster* and at the 2-4 hour stage these embryos were collected at, the maternal to zygotic transition is not yet complete. Therefore, these embryos likely only contain germline piRNA, which could explain their marked difference from the somatic clusters, which are called from somatic piRNA populations.

Looking at the characteristics of germline clusters in *Ae. aegypti*, they are similar to germline clusters in *D. melanogaster* such as 42AB. Clusters called from the embryonic library are often bidirectional, with features exhibiting no bias for either the plus or minus strand within clusters. In addition, sRNA map to both strands of the cluster loci and there is a greater A10 bias in mapping piRNA than seen in the piRNA that map to somatic clusters, all hallmarks of ping-pong amplification. There is also a greater diversity in transposon sequences contained within germline clusters. Unlike in somatic clusters, where Gypsy predominates, other elements such as DNA transposons predominate.

3.4.4 Conclusions

Based on data from new sRNA libraries, true somatic piRNA and piRNA clusters were discovered in *Ae. aegypti.* These somatic clusters are similar to the clusters described in the somatic support cells of the *D. melanogaster* ovary and their architecture and associated piRNA seem to primarily support primary piRNA production; transposon features were orientated in a single direction, allowing for transcription and

production of anti-sense piRNA. These piRNA had a strong U1 bias but a weaker A10 bias. However, there were precise ten-nucleotide overlaps that suggest some ping-pong amplification is present in these libraries.

One unique feature of somatic clusters found in *Ae. aegypti* is that a small fraction (around 1 percent) have a unique configuration where a single piRNA cluster produces only anti-sense piRNA, but from both the minus and plus strands of the genome. The minus strand orientated features and the plus strand orientated features are immediately adjacent to each other, but no or very few sense piRNA are detected.

Bidirectional germline clusters were also found. These were found to be similar to the germline clusters described in *D. melanogaster,* with mapping piRNA found on both strands of the cluster and transposon and gene features likewise oriented on both the minus and plus strands. These clusters produce both sense and anti-sense transcript, which allows for ping-pong amplification. Accordingly, the piRNA mapping to these clusters had both a U1 bias and an A10 bias.

A new method of cluster calling, proTRAC, was used to identify both germline and somatic clusters. In the soma, proTRAC was able to assign the majority or close to the majority of library sRNA reads to much fewer genomic loci than described previously. Based on the piRNA mapping statistics of these clusters, it appears as if the majority of somatic mapping piRNA – predominately against Gypsy elements – are produced from only a handful of clusters. For instance, the top 5 clusters in the gastric caeca libraries produce more piRNA than the 200 other called clusters. In contrast, the

proTRAC method of calling was not sensitive enough in calling clusters from germline tissue. This may be due to the fact that piRNA clusters in the *Ae. aegypti* germline appear to be numerous with high piRNA diversity; no one called germline cluster commanded as high a mapping percentage of piRNA reads as seen in the somatic clusters. A moderate-to-low amount of piRNA generated from numerous genomic loci in the germline may not produce sharp enough peaks of piRNA that allowed for the accurate calling of clusters in the soma.

3.5 References

Arensburger, P., Hice, R.H., Wright, J.A., Craig, N.L., and Atkinson, P.W. (2011). The mosquito Aedes aegypti has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs. BMC Genomics *12*, 606.

Assis, R., and Kondrashov, A.S. (2009). Rapid repetitive element-mediated expansion of piRNA clusters in mammalian evolution. Proc. Natl. Acad. Sci. *106*, 7079–7082.

Betel, D., Sheridan, R., Marks, D.S., and Sander, C. (2007). Computational Analysis of Mouse piRNA Sequence and Biogenesis. PLoS Comput. Biol. *3*, e222.

Biedler, J.K., Hu, W., Tae, H., and Tu, Z. (2012). Identification of Early Zygotic Genes in the Yellow Fever Mosquito Aedes aegypti and Discovery of a Motif Involved in Early Zygotic Genome Activation. PLoS ONE *7*, e33933.

Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A., and the Galaxy Team (2010). Manipulation of FASTQ data with Galaxy. Bioinformatics *26*, 1783–1785.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell *128*, 1089–1103.

Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. Science *322*, 1387–1392.

Chinwalla, A.T., Cook, L.L., Delehaunty, K.D., Fewell, G.A., Fulton, L.A., Fulton, R.S., Graves, T.A., Hillier, L.W., Mardis, E.R., McPherson, J.D., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.

Coline, G., Théron, E., Brasset, E., and Vaury, C. (2014). History of the discovery of a master locus producing piRNAs: the flamenco/COM locus in Drosophila melanogaster. Front. Genet. *5*.

Desset, S., Meignin, C., Dastugue, B., and Vaury, C. (2003). COM, a heterochromatic locus governing the control of independent endogenous retroviruses from Drosophila melanogaster. Genetics *164*, 501–509.

Gaunt, M.W., and Miles, M.A. (2002). An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. Mol. Biol. Evol. *19*, 748–761.

George, P., Jensen, S., Pogorelcnik, R., Lee, J., Xing, Y., Brasset, E., Vaury, C., and Sharakhov, I.V. (2015). Increased production of piRNAs from euchromatic clusters and genes in Anopheles gambiae compared with Drosophilamelanogaster. Epigenetics Chromatin *8*.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199–202.

Hirakata, S., and Siomi, M.C. (2016). piRNA biogenesis in the germline: From transcription of piRNA genomic sources to piRNA maturation. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1859*, 82–92.

Holt, R.A. (2002). The Genome Sequence of the Malaria Mosquito Anopheles gambiae. Science *298*, 129–149.

Krzywinski, J., Grushko, O.G., and Besansky, N.J. (2006). Analysis of the complete mitochondrial DNA from Anopheles funestus: An improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. Mol. Phylogenet. Evol. *39*, 417–423.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., et al. (2009). Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. Cell *137*, 509–521.

Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. Cell *137*, 522–535.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10–12.

Mével-Ninio, M., Pelisson, A., Kinder, J., Campos, A.R., and Bucheton, A. (2007). The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. Genetics *175*, 1615–1624.

Mohn, F., Sienski, G., Handler, D., and Brennecke, J. (2014). The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. Cell *157*, 1364–1379.

Nene, V., Wortman, J.R., Lawson, D., Haas, B., Kodira, C., Tu, Z., Loftus, B., Xi, Z., Megy, K., Grabherr, M., et al. (2007). Genome Sequence of Aedes aegypti, a Major Arbovirus Vector. Science *316*, 1718–1723.

Pélisson, A., Song, S.U., Prud'homme, N., Smith, P.A., Bucheton, A., and Corces, V.G. (1994). Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. EMBO J. *13*, 4401–4411.

Prud'homme, N., Gans, M., Masson, M., Terzian, C., and Bucheton, A. (1995). Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. Genetics *139*, 697–711.

Quesneville, H., Bergman, C.M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., and Anxolabehere, D. (2005). Combined evidence annotation of transposable elements in genome sequences. PLoS Comput. Biol. *1*, 166–175.

Rosenkranz, D. (2016). piRNA cluster database: a web resource for piRNA producing loci. Nucleic Acids Res. *44*, D223–D230.

Rosenkranz, D., and Zischler, H. (2012). proTRAC - a software for probabilistic piRNA cluster detection, visualization and analysis. BMC Bioinformatics *13*, 5.

Rosenkranz, D., Han, C.-T., Roovers, E.F., Zischler, H., and Ketting, R.F. (2015). Piwi proteins and piRNAs in mammalian oocytes and early embryos: From sample to sequence. Genomics Data *5*, 309–313.

Salvemini, M., Mauro, U., Lombardo, F., Milano, A., Zazzaro, V., Arcà, B., Polito, L.C., and Saccone, G. (2011). Genomic organization and splicing evolution of the doublesex gene, a Drosophila regulator of sexual differentiation, in the dengue and yellow fever mosquito Aedes aegypti. BMC Evol. Biol. *11*, 41.

Schreiner, P.A., and Atkinson, P. (2017). piClusterBusteR: Software For Automated Classification And Characterization Of piRNA Cluster Loci. BioRxiv 133009.

Smith, C.D., Shu, S., Mungall, C.J., and Karpen, G.H. (2007). The Release 5.1 annotation of Drosophila melanogaster heterochromatin. Science *316*, 1586–1591.

Wright, J.A. (2011). Investigating Transposable Elements for Use in Dipteran Systems. Ph.D. University of California, Riverside.

Yamanaka, S., Siomi, M.C., and Siomi, H. (2014). piRNA clusters and open chromatin structure. Mob. DNA *5*, 22.

Yu, Y., Gu, J., Jin, Y., Luo, Y., Preall, J.B., Ma, J., Czech, B., and Hannon, G.J. (2015). Panoramix enforces piRNA-dependent cotranscriptional silencing. Science *350*, 339–342.

Zanni, V., Eymery, A., Coiffet, M., Zytnicki, M., Luyten, I., Quesneville, H., Vaury, C., and Jensen, S. (2013). Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. Proc. Natl. Acad. Sci. *110*, 19842–19847.

3.6 Figures and tables

	% TE Reads with U1 Sense/Antisense	% TE Reads with A10 Sense/Antisense	Total # of TE Mapping Reads	Total % of TE Mapping Reads	5' Overlap Z-Score (Std. Dev.)
Gastric Caecae & Salivary Gland (3)	44.02 ± 2.35 %	26.97 ± 1.33 %	2.85 million	6.11 ± 0.93%	6.08 ± 2.12
	87.82 ± 3.23 %	20.34 ± 0.76%			
Larva, Male Control (3)	44.23 ± 1.78 %	40.57 ± 2.41 %	1.10 million	7.69 ± 0.64 %	1.66 ± 0.49
	89.13 ± 0.49 %	20.97 ± 2.01 %			
Larva, Male Induced (3)	36.77 ± 3.15 %	35.90 ± 2.61 %	0.47 million	4.91 ± 1.87 %	0.43 ± 0.52
	82.63 ± 3.51 %	22.33 ± 1.80 %			
Whole Body Adults Blood Fed (3)	18.20 ± 1.31 %	75.63 ± 4.02 %	1.11 million	3.52 ± 2.64 %	41.41 ± 12.48
	58.83 ± 5.82 %	36.23 ± 0.75 %			
Ovaries Blood Fed (2)	60.85 ± 2.48 %	41.45 ± 0.64 %	48.50 million	26.97 ± 5.11 %	65.93 ± 32.65
	75.7 ± 4.81 %	29.65 ± 0.49 %			
Embryos 2-4 HR (2)	62.05 ± 2.76 %	52.15 ± 1.48 %	29.73 million	30.93 ± 4.76 %	34.06 ± 10.38
	75.90 ± 4.95 %	34.45 ± 0.92 %			
A3 Pulldown Ovaries BF (1)	58.8%	45.9%	2.09 million	13.45%	36.41
	79.0%	28.8%			
P2 Pulldown Ovaries BF (1)	72.4%	41.2%	13.29 million	25.52%	33.12
	81.6%	31.1%			

Table 3.1: U1 and A10 biases of transposon mapping reads from various *Ae. aegypti* tissues, split by reads that map to the sense strand and reads that map to the antisense strand. Number in parenthesis indicates the number of biological replicates. Further columns indicate the total number of transposon mapping reads, the transposon content of the origin library, and a ping-pong Z-score calculated from the whole library as a measure of secondary piRNA signature/ping-pong amplification.

	% Library Reads with U1	% Library Reads with A10	Total # of Library Reads
Gastric Caecae & Salivary Gland	63.90%	31.20%	39.07 million
Larva, Male Control	58.60%	26.10%	14.06 million
Larva, Male Induced	43.60%	25.80%	9.23 million
Ovaries Blood Fed	75.10%	33.10%	31.74 million
Embryos 2-4 HR	74.70%	38.00%	82.20 million
A3 Pulldown Ovaries BF	49.90%	28.30%	13.72 million
P2 Pulldown Ovaries BF	76.40%	33.30%	47.84 million

Table 3.2: U1 and A10 biases of total sRNA library reads from *Ae. aegypti* tissues. Library read sample size is indicated in the last column.

	Genome Coverage (basepairs)	Cluster Mapping % of Only Unique Sequences	Cluster Mapping % of All Sequences	TE Mapping % of Reads in Origin Library
Gastric Caecae & Salivary Gland (3)	3.04 ± 0.33 million	24.13 ± 4.21 %	67.04 ± 13.84 %	6.11 ± 0.93 %
Larva, Male Control (3)	2.78 ± 0.17 million	35.05 ± 10.82 %	47.57 ± 5.56 %	7.69 ± 0.64 %
Larva, Male Induced (3)	1.55 ± 0.56 million	20.72 ± 4.82 %	36.11 ± 12.33 %	4.91 ± 1.87 %
Whole Body Adults Blood Fed (3)	1.44 ± 0.75 million	17.15 ± 7.46 %	12.99 ± 12.93 %	3.52 ± 2.64 %
Ovaries Blood Fed (2)	2.26 ± 0.29 million	12.91 ± 4.00 %	28.06 ± 11.62 %	26.97 ± 5.11 %
Embryos 2-4 HR (2)	2.56 ± 0.27 million	12.01 ± 1.61 %	23.38 ± 5.20 %	30.93 ± 4.76 %

Table 3.3: proTRAC-called piRNA clusters from various *Ae. aegypti* tissue sRNA libraries. Number in parenthesis indicates the number of libraries used in the analysis. The percent of reads that can be mapped to the called clusters are broken down into the set of unique reads (reads that only occur once in the library) and total library reads (includes all library reads, even if they are not unique).

Location	Start	Stop	Size	Hits (Normalized)	U1	A10	Main Strand Hits	Directionality	Strand
supercont1.124	973080	985016	11937	214943.564915613	96.80%	75.00%	100.00%	mono	minus
supercont1.478	495182	565395	70214	141010.964463876	92.40%	18.10%	100.00%	mono	plus
supercont1.1	1151066	1282898	131833	56475.466444595	94.50%	45.00%	100.00%	mono	plus
supercont1.286	1274056	1350988	76933	28309.1514716949	92.60%	20.20%	100.00%	mono	minus
supercont1.555	364010	373052	9043	13594.7587919728	87.90%	32.80%	100.00%	mono	minus
supercont1.286	1354000	1420985	66986	12958.2586182583	91.70%	21.90%	100.00%	mono	minus
supercont1.209	241007	283960	42954	10017.0689611057	93.40%	17.10%	100.00%	mono	minus
supercont1.1145	48675	123001	74327	8502.8260996499	94.10%	19.10%	100.00%	mono	minus
supercont1.209	295871	361010	65140	5953.110082	93.10%	14.40%	100.00%	mono	minus
supercont1.589	113784	133970	20187	5573.3350369956	96.00%	26.80%	100.00%	mono	plus
supercont1.943	219023	256028	37006	3283.7268033371	92.20%	12.00%	100.00%	mono	minus
supercont1.435	933103	972184	39082	3267.4931685047	89.60%	27.10%	100.00%	mono	minus
supercont1.379	30027	77026	47000	3094.5745363006	95.30%	26.00%	99.70%	mono	plus
supercont1.20	3155054	3204013	48960	2937.5182127708	92.30%	16.90%	100.00%	mono	plus
supercont1.302	768012	786013	18002	2743.8015163115	95.30%	30.30%	100.00%	mono	plus
supercont1.402	887223	928010	40788	2736.8296762179	93.00%	22.80%	100.00%	mono	minus
supercont1.1	1391001	1429009	38009	2736.6996344603	92.40%	14.90%	99.70%	mono	minus
supercont1.36	2883351	2891776	8426	2433.3745508167	98.40%	1.50%	100.00%	mono	plus
supercont1.215	673039	704015	30977	2164.6332797733	91.00%	19.20%	99.50%	mono	minus
supercont1.125	1644041	1652000	7960	2145.9707959807	90.90%	5.10%	100.00%	mono	plus
supercont1.535	*577339*	*583923*	*6585*	*2047.6647935705*	*57.50%*	*52.20%*	*100.00%*	*mono*	*plus*
supercont1.488	341683	361017	19335	1882.360720624	93.30%	23.10%	99.90%	mono	plus
supercont1.790	353176	369966	16791	1832.4512149247	93.30%	23.20%	99.90%	mono	plus
supercont1.943	201756	218022	16267	1648.9534654545	94.90%	15.90%	100.00%	mono	minus

Table 3.4: Example of proTRAC output. Top 25 piRNA clusters by mapping sRNA read density from a larval gastric caeca & salivary gland sRNA library. The cluster in italics is the only cluster that was not also called from whole larval libraries. Statistics on cluster size, U1/A10 content, piRNA density, and directionality of the cluster are provided.

Location	Start	End	Size	Hits (normalized)	U1	A10	Main Strand Hits	Directionality	Strand
supercont1.478	499320	565381	66062	107154.21803417	93.80%	26.30%	100.00%	mono	plus
supercont1.286	1273196	1350976	77781	31029.5656635469	93.80%	20.70%	100.00%	mono	minus
supercont1.555	364519	373135	8617	22399.6151210888	78.00%	70.40%	100.00%	mono	minus
supercont1.1	1184026	1282867	98842	20389.9547162762	95.20%	27.60%	100.00%	mono	plus
supercont1.124	976017	984379	8363	16355.3060790399	97.80%	57.30%	100.00%	mono	minus
supercont1.286	1353058	1420922	67865	15769.0792607352	93.20%	22.70%	100.00%	mono	minus
supercont1.1	1151072	1182929	31858	15473.3616400586	97.60%	56.30%	100.00%	mono	plus
supercont1.3	3031941	3037977	6037	7409.2005123836	64.00%	61.00%	100.00%	mono	plus
supercont1.209	242781	283592	40812	7373.7547603946	94.20%	18.10%	100.00%	mono	minus
supercont1.69	2239014	2245483	6470	6739.5399218984	77.70%	77.60%	100.00%	mono	minus
supercont1.1145	49805	123888	74084	6636.600098696	94.80%	22.30%	99.90%	mono	minus
supercont1.9	4141014	4147753	6740	5915.0413394979	56.20%	52.90%	100.00%	mono	minus
supercont1.209	306044	362014	55971	5231.8123616646	94.80%	15.70%	100.00%	mono	minus
supercont1.20	3155085	3222021	66937	3564.1322248737	94.60%	19.50%	99.80%	mono	plus
supercont1.435	935369	972154	36786	3484.8468359093	92.10%	28.40%	99.80%	mono	minus
supercont1.589	113837	132672	18836	2962.1687416275	95.30%	26.50%	100.00%	mono	plus
supercont1.302	769997	783787	13791	2746.9053361309	95.90%	29.70%	100.00%	mono	plus
supercont1.790	353628	370003	16376	2458.0031889565	94.90%	23.00%	99.90%	mono	plus
supercont1.488	344420	357523	13104	2419.8679933007	94.90%	23.40%	99.90%	mono	plus
supercont1.402	889240	929018	39779	2334.7490459885	93.90%	23.70%	99.80%	mono	minus
supercont1.140	1055804	1080548	24745	2173.2936218654	86.80%	16.30%	99.90%	mono	plus
supercont1.943	202282	217967	15686	2093.9580729017	95.70%	22.70%	100.00%	mono	minus
supercont1.379	23598	76964	53367	1836.1615038288	95.10%	26.70%	99.70%	mono	plus
supercont1.555	289745	296907	7163	1585.7268450278	83.20%	86.90%	100.00%	mono	plus

Table 3.5: Top 25 piRNA clusters by mapping sRNA read density from a male 4th instar larval sRNA library. Statistics on cluster size, U1/A10 content, piRNA density, and directionality of the cluster are provided.

Location	Start	End	Size	Hits (normalized)	U1	A10	Main Strand Hits	Directionality	Strand
supercont1.478	501058	565259	64202	18457.0021906686	91.70%	14.60%	99.90%	mono	plus
supercont1.320	613545	623568	10024	7813.2963772851	82.60%	17.10%	99.90%	mono	plus
supercont1.286	1276594	1334986	58393	4743.7062950917	88.90%	21.30%	99.70%	mono	minus
supercont1.98	1091021	1101028	10008	3954.567760551	81.10%	7.30%	97.10%	mono	plus
supercont1.1	1182038	1229002	46965	3652.2832808622	88.40%	25.70%	99.80%	mono	plus
supercont1.286	1357014	1395020	38007	3198.0231447532	87.90%	24.90%	99.70%	mono	minus
supercont1.83	1413057	1435013	21957	2712.3423783303	82.90%	31.70%	94.30%	mono	minus
supercont1.555	365005	375591	10587	1923.0611399418	23.20%	81.10%	98.70%	mono	minus
supercont1.1	1261017	1281989	20973	1921.2828334427	91.60%	17.00%	100.00%	mono	plus
supercont1.406	373002	384028	11027	1905.0256591392	87.70%	49.70%	94.70%	mono	minus
supercont1.83	1342384	1358920	16537	1901.0993326793	81.00%	43.50%	96.90%	mono	minus
supercont1.402	913022	926025	13004	1882.716553523	87.00%	14.00%	100.00%	mono	minus
supercont1.556	459023	468022	9000	1811.2842300634	85.30%	7.70%	96.00%	mono	plus
supercont1.234	892006	903959	11954	1795.4568088034	91.30%	16.10%	95.70%	mono	plus
supercont1.58	31083	43993	12911	1523.7016377979	84.60%	37.50%	94.40%	mono	minus
supercont1.2	2674015	2696018	22004	1520.4265537191	75.30%	48.30%	85.90%	bi	plus
supercont1.226	737037	749880	12844	1489.4639081187	86.60%	34.20%	92.30%	mono	minus
supercont1.9	4559091	4567995	8905	1361.234988796	11.90%	90.40%	99.00%	mono	minus
supercont1.2	337001	351879	14879	1292.3495954306	90.40%	7.20%	95.80%	mono	minus
supercont1.39	1012008	1028017	16010	1252.6887116601	86.50%	53.40%	87.10%	bi	minus
supercont1.403	782012	792016	10005	1239.663205877	90.50%	54.60%	94.10%	mono	minus
supercont1.173	1105000	1120024	15025	1224.0060176417	59.30%	52.40%	92.00%	mono	minus
supercont1.2	304007	318968	14962	1142.3751870221	80.00%	35.30%	75.30%	mono	plus
supercont1.209	264001	282020	18020	1095.7071810695	90.30%	23.80%	99.70%	mono	minus

Table 3.6: Top 25 piRNA clusters by mapping sRNA read density from a 2 day blood fed ovarian sRNA library. Statistics on cluster size, U1/A10 content, piRNA density, and directionality of the cluster are provided.

Location	Start	End	Size	Hits (normalized)	U1	A10	Main Strand Hits	Directionality	Strand
supercont1.226	730010	749486	19477	4212.6733093814	84.70%	41.60%	89.00%	mono	minus
supercont1.2	304003	320023	16021	3850.8093063452	84.90%	37.80%	77.00%	mono	plus
supercont1.83	1413470	1435980	22511	3706.2432840813	79.30%	30.80%	92.20%	mono	minus
supercont1.281	396014	407017	11004	2879.1579034595	78.30%	54.00%	89.60%	mono	plus
supercont1.98	1091019	1106023	15005	2831.3023067034	81.60%	14.40%	93.00%	mono	plus
supercont1.479	234002	251028	17027	2666.8110294679	80.60%	37.90%	88.00%	mono	plus
supercont1.234	891295	904004	12710	2317.7609076731	85.30%	31.00%	91.70%	mono	plus
supercont1.2	696003	708025	12023	2149.3910024691	79.10%	42.00%	92.60%	bi	plus
supercont1.258	331001	339976	8976	2087.4232114206	93.90%	37.70%	95.80%	mono	plus
supercont1.173	1082008	1094899	12892	2065.2726473595	78.80%	41.90%	80.40%	mono	plus
supercont1.2342	1098	8208	7111	1998.8091340782	77.50%	7.00%	100.00%	mono	minus
supercont1.826	352030	366901	14872	1973.2743298658	80.90%	35.90%	87.60%	bi	plus
supercont1.927	138016	149018	11003	1700.8439465419	87.70%	28.60%	92.30%	mono	plus
supercont1.38	3015007	3032009	17003	1675.1462796185	77.10%	27.80%	87.80%	mono	plus
supercont1.39	1011007	1028020	17014	1633.2005380814	87.00%	56.00%	89.70%	bi	minus
supercont1.2	2673034	2686009	12976	1616.505114849	71.10%	59.20%	87.90%	mono	plus
supercont1.58	31071	44020	12950	1600.6099382332	76.30%	36.20%	91.10%	mono	minus
supercont1.40	225007	237028	12022	1503.9773844174	78.90%	33.50%	90.10%	bi	plus
supercont1.376	84010	96013	12004	1495.29128955	69.00%	57.10%	82.20%	bi	minus
supercont1.35	1869345	1882982	13638	1348.7034314343	77.70%	34.80%	76.70%	bi	plus
supercont1.58	1935162	1950017	14856	1329.7386792493	79.60%	30.80%	95.00%	mono	plus
supercont1.405	480034	491016	10983	1323.6075244544	84.80%	29.00%	94.10%	bi	minus
supercont1.1017	52000	67004	15005	1298.4900913222	82.30%	33.90%	90.00%	mono	minus
supercont1.62	2767022	2777020	9999	1278.6659709498	87.40%	25.70%	94.80%	mono	plus

Table 3.7: Top 25 piRNA clusters by mapping sRNA read density from a 2-4 hour embryonic sRNA library. Statistics on cluster size, U1/A10 content, piRNA density, and directionality of the cluster are provided.

Location	Start	End	Size	Hits (normalized)	U1	A10	Main Strand Hits	Directionality	Strand
supercont1.478	514071	564952	50882	7626.9595527593	95.1%	62%	96.9%	mono	minus
supercont1.401	477091	482270	5180	7012.5214924041	84.5%	29.8%	89.6%	mono	plus
supercont1.376	86051	95870	9820	2318.719539767	92.4%	11%	79.6%	mono	minus
supercont1.286	1356014	1389985	33972	2149.9139848185	84.8%	34.5%	84.5%	mono	plus
supercont1.320	615131	622729	7599	1792.5549501909	93%	21.4%	99.8%	mono	minus
supercont1.518	148061	160000	11940	1755.1719851308	84.9%	11.9%	95.9%	mono	minus
supercont1.83	1415051	1435967	20917	1527.7925292988	85.7%	22.5%	91.7%	mono	minus
supercont1.403	770030	779952	9923	1409.9634650717	96.4%	6.4%	99.9%	mono	plus
supercont1.98	1091054	1101021	9968	1254.9222003417	85%	12.7%	94.8%	mono	plus
supercont1.124	973247	983745	10499	1253.3525521446	86.5%	49.1%	93.1%	mono	minus
supercont1.226	739001	748166	9166	1189.2136875011	83.7%	20.8%	94.7%	bi	plus
supercont1.286	1318000	1331996	13997	1116.3458734382	86.4%	25.4%	92.1%	bi	minus
supercont1.35	1872040	1882989	10950	1109.5634490457	89.2%	14.4%	94.5%	mono	plus
supercont1.286	1295007	1316022	21016	1070.724552154	89.5%	28.6%	96.7%	mono	minus
supercont1.479	235004	249997	14994	1056.9125920476	78.4%	47.9%	79.6%	mono	plus
supercont1.90	922001	934991	12991	1008.8740854454	93.2%	21.8%	100%	mono	plus
supercont1.405	480035	490920	10886	1007.0165075205	89%	30.8%	90.5%	mono	minus
supercont1.1	1252051	1273020	20970	1001.440808429	89.2%	14%	93.7%	mono	plus
supercont1.1	1189021	1208024	19004	993.7866762581	86%	35%	97.4%	mono	minus
supercont1.234	892512	902972	10461	873.5414531051	81.7%	75.3%	95.2%	mono	minus
supercont1.479	10114	18978	8865	855.129914252	75.2%	39.5%	95.3%	bi	minus
supercont1.17	1782004	1793020	11017	851.0036469114	85.6%	26.6%	94.9%	bi	plus
supercont1.2	304042	313882	9841	847.8036827618	93.6%	72.8%	95.3%	bi	minus
supercont1.843	142976	150951	7976	844.1595669306	87.8%	16.6%	91.1%	mono	plus

Table 3.8: Top 25 piRNA clusters by mapping sRNA read density from a 2 day blood fed ovarian sRNA library, pulled down with an Ago3 antibody. Statistics on cluster size, U1/A10 content, piRNA density, and directionality of the cluster are provided.

Location	Start	End	Size	Hits (normalized)	U1	A10	Main Strand Hits	Directionality	Strand
supercont1.478	529014	558029	29016	2931.888156259	95.5%	9.9%	99.6%	mono	plus
supercont1.376	86026	95954	9929	2257.5121394743	71.3%	52.9%	90.9%	mono	plus
supercont1.518	148061	160013	11953	1839.4804806795	77.1%	38%	95.7%	bi	minus
supercont1.403	770003	779934	9932	1656.8697986818	86.6%	46.9%	88.1%	mono	minus
supercont1.226	732011	749902	17892	1603.0270232888	91.2%	20.5%	93.6%	mono	minus
supercont1.401	477092	485815	8724	1522.7326286352	89.9%	95.3%	99.8%	mono	minus
supercont1.98	1091034	1105022	13989	1513.8428069397	84%	18.4%	76.1%	mono	plus
supercont1.90	921002	934903	13902	1512.2249023147	90.1%	32.9%	89%	bi	plus
supercont1.2	304001	319029	15029	1310.2846859722	88.3%	24.6%	83.6%	mono	plus
supercont1.9	4595012	4609026	14015	1218.9914063267	87.7%	12.2%	89.5%	bi	minus
supercont1.35	1870001	1883021	13021	1210.2035888823	90.1%	20.4%	93.5%	bi	plus
supercont1.479	235008	256999	21992	1174.5728773709	85.2%	22.5%	93.6%	mono	plus
supercont1.83	1425000	1435976	10977	1146.0341784796	90.9%	69.8%	98.3%	mono	minus
supercont1.234	892088	902984	10897	1121.4397931298	93.5%	17.3%	97.2%	mono	plus
supercont1.1041	248030	265009	16980	1119.5085312378	86.5%	14.6%	95.9%	mono	minus
supercont1.826	352030	366023	13994	1067.7358767242	86%	21.6%	95.4%	bi	plus
supercont1.83	1344001	1358922	14922	1035.9500103728	83.9%	27%	96.5%	mono	minus
supercont1.33	2190007	2199023	9017	916.9376660556	94.2%	64.1%	97.3%	bi	plus
supercont1.403	782053	792015	9963	878.10447379	91.8%	57.8%	94.6%	mono	minus
supercont1.927	139028	153177	14150	857.9480182212	80.1%	26.2%	93.2%	bi	plus
supercont1.192	1716001	1726026	10026	846.4792532791	85.3%	35.2%	86.1%	bi	minus
supercont1.226	834002	846018	12017	833.5732361167	87.1%	24.8%	89.5%	mono	minus
supercont1.39	1012010	1027926	15917	805.9200239421	87.7%	31.7%	91.3%	bi	minus
supercont1.17	1782031	1793033	11003	787.7498427719	85.1%	39.2%	94.3%	mono	plus

Table 3.9: Top 25 piRNA clusters by mapping sRNA read density from a 2 day blood fed ovarian sRNA library, pulled down with a Piwi 2 antibody. Statistics on cluster size, U1/A10 content, piRNA density, and directionality of the cluster are provided.

	Number of Called piRNA Clusters	Bidirectional Clusters	U1 fraction above 33%	A10 fraction above 33%
Gastric Caecae & Salivary Gland	254	1 (0.4%)	254 (99.6%)	24 (9.4%)
Larvae, Male Control	226	5 (2.2%)	226 (100%)	24 (10.6%)
Ovaries Blood Fed	200	37 (18.5%)	198 (99%)	76 (38%)
Embryos 2-4 HR	243	70 (28.8%)	243 (100%)	145 (59.7%)
A3 Pulldown Ovaries BF	125	28 (22.4%)	124 (99.2%)	47 (37.6%)
P2 Pulldown Ovaries BF	162	58 (35.8%)	161 (99.4%)	74 (45.7%)

Table 3.10: Summary of proTRAC cluster calls. Shown are the number of bidirectional clusters for each tissue type and the U1 and A10 biases of the mapping piRNAs.

	Self Overlap (Within Samples)	Overlap with Ovary	Overlap with Embryo
Larval Soma (6)	2.22 ± 0.126 million bp	0.952 ± 0.358 million bp	0.105 ± 0.077 million bp
2-4 HR Embryos (2)	1.06 million bp	0.725 ± 0.232 million bp	
2D BF Ovary (2)	0.946 million bp		

Table 3.11: Overlap, in total base pairs, between tissue replicates and between tissues samples. Number in parenthesis indicates number of biological replicates. Some cells were grayed out due to redundancy.

	Self Overlap (Within Samples)	Overlap with Ovary	Overlap with Embryo
Larval Soma (6)	76.3%	32.7%	3.6%
2-4 HR Embryos (2)	41.4%	28.3%	
2D BF Ovary (2)	41.9%		

Table 3.12: Average overlap, by percent of cluster base pairs, between tissue replicate and between tissue samples. Number in parenthesis indicates number of biological replicates. Some cells were grayed out due to redundancy.

Location	Start	Stop	Size	Hits (Normalized)	U1	A10	Main Strand Hits	Directionality	Strand
supercont1.1	**1151066**	**1282898**	**131833**	**56475.466444595**	**94.50%**	**45.00%**	**100.00%**	**mono**	**plus**
supercont1.1	1311009	1330019	19011	516.9697693575	91.00%	18.80%	99.90%	mono	minus
supercont1.1	1331005	1337024	6020	112.4726804526	87.90%	22.40%	100.00%	mono	minus
supercont1.1	1346032	1361020	14989	475.9444253877	93.80%	20.60%	99.90%	mono	minus
supercont1.1	1365015	1382022	17008	462.7408611256	92.40%	17.20%	99.90%	mono	minus
supercont1.1	1391001	1429009	38009	2736.6996344603	92.40%	14.90%	99.70%	mono	minus
supercont1.1	1433027	1439024	5998	104.7048720625	91.10%	14.10%	99.30%	mono	minus
supercont1.1	1440000	1450014	10015	256.0867628587	92.30%	21.80%	99.90%	mono	minus
supercont1.1	1468011	1486014	18004	1365.5057558138	88.40%	27.70%	87.80%	mono	minus
supercont1.105	**2235012**	**2240464**	**5453**	**312.6815881839**	**91.50%**	**80.40%**	**99.90%**	**mono**	**plus**
supercont1.105	2377524	2384899	7376	344.0404427294	94.90%	17.10%	100.00%	mono	minus
supercont1.299	**1310000**	**1319658**	**9659**	**277.2484588281**	**93.80%**	**26.10%**	**99.40%**	**mono**	**plus**
supercont1.299	1349100	1355328	6229	107.2596851091	93.20%	29.10%	99.50%	mono	minus
supercont1.489	**323007**	**337986**	**14980**	**451.7551404818**	**91.40%**	**6.70%**	**99.60%**	**mono**	**plus**
supercont1.489	358001	363023	5023	115.9231341126	93.20%	21.60%	97.60%	mono	minus

Table 3.13: proTRAC called somatic unidirectional clusters composed of halves that are transcribed from opposite strands (plus/minus) that are directly (within 30kb) adjacent to each other.

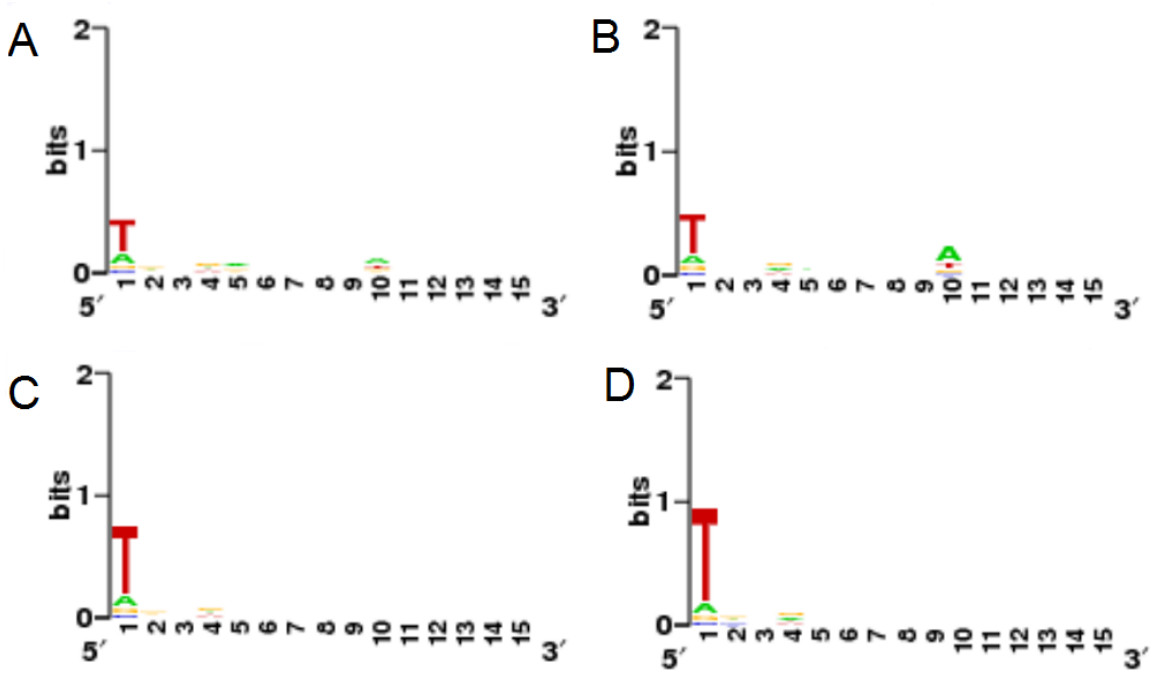

Figure 3.1: SeqLogo graphs showing over-represented nucleotides from positions 1 to 15 of a transposon mapping sRNA read. (A) Sense strand transposon mapping reads from a blood fed ovary sRNA library. N = 6.03 million. (B) Sense strand transposon mapping reads from a 2-4 HR embryo library. N = 2.75 million. (C) Antisense strand transposon mapping reads from a blood fed ovary sNRA library. N = 10.29 million. (D) Antisense strand transposon mapping reads from a 2-4 HR embryo library. N = 10.81 million.

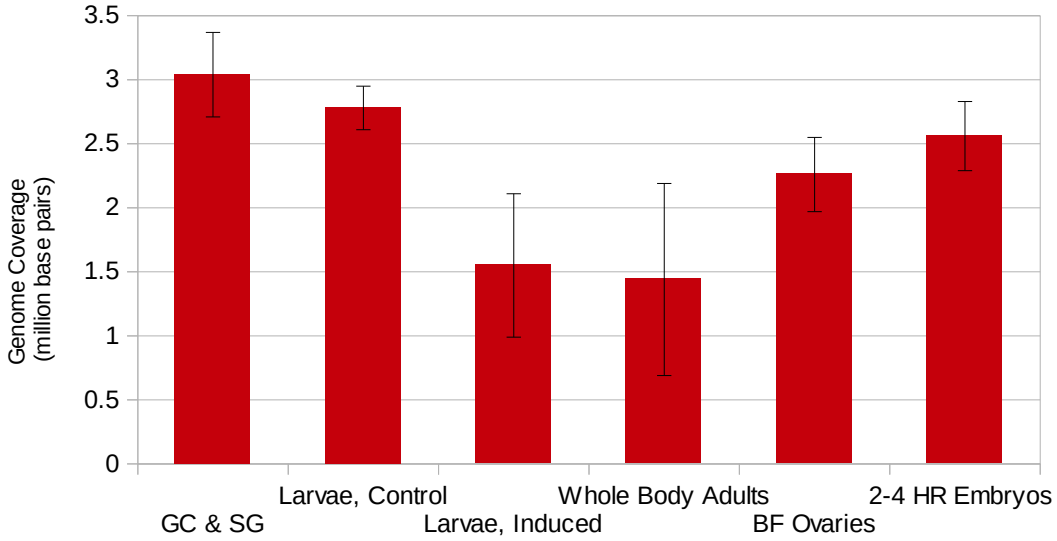

Figure 3.2: The genome occupancy of proTRAC called piRNA clusters in *Ae. aegypti* tissues.

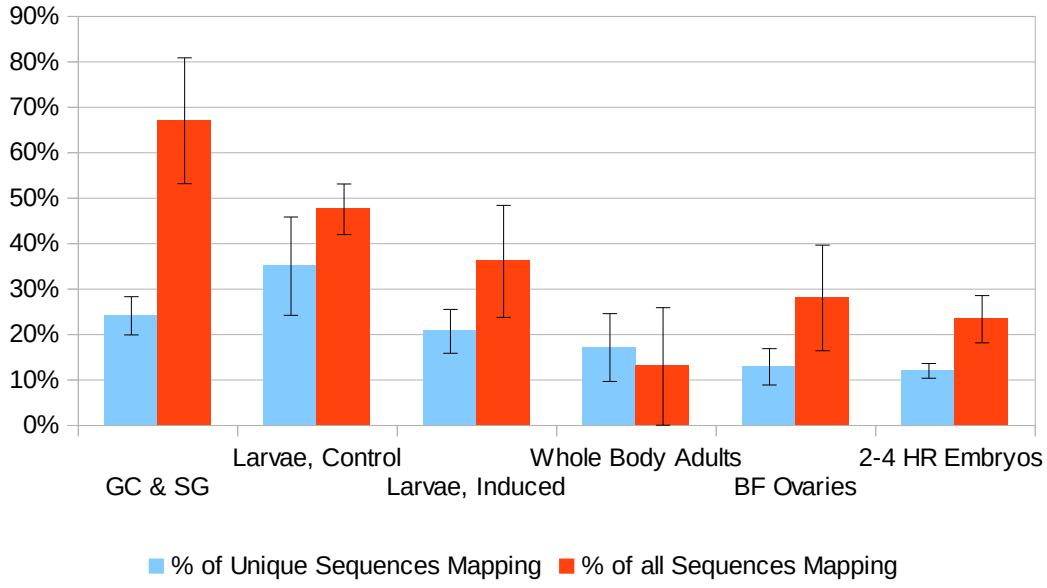

Figure 3.3: Percent of unique sRNA reads that can be assigned to proTRAC called clusters (blue) and percent of all sRNA reads that can be assigned (red).

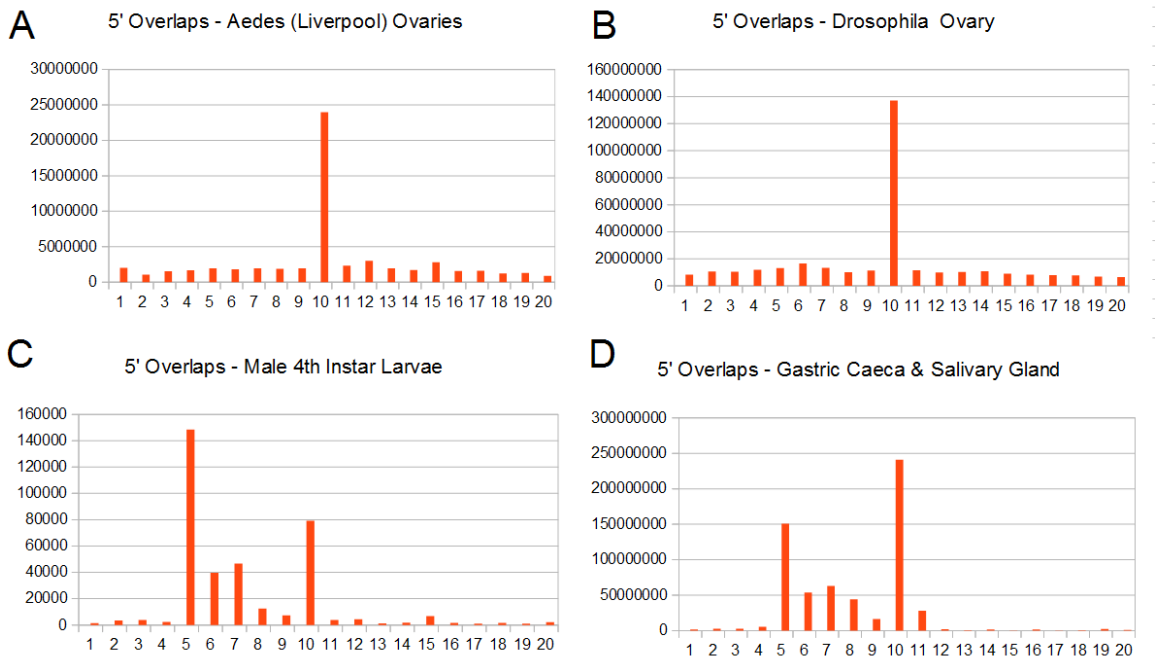

Figure 3.4: The number of 5' overlaps of complementary read pairs of a certain size, from 1 bp to 20 bp, shown on the x-axis. (A) *Ae. aegypti* Liverpool strain blood fed ovaries. (B) *D. melanogaster* ovaries. (C) *Ae. aegypti* male 4th instar larvae. (D) *Ae. aegypti* gastric caeca and salivary glands.

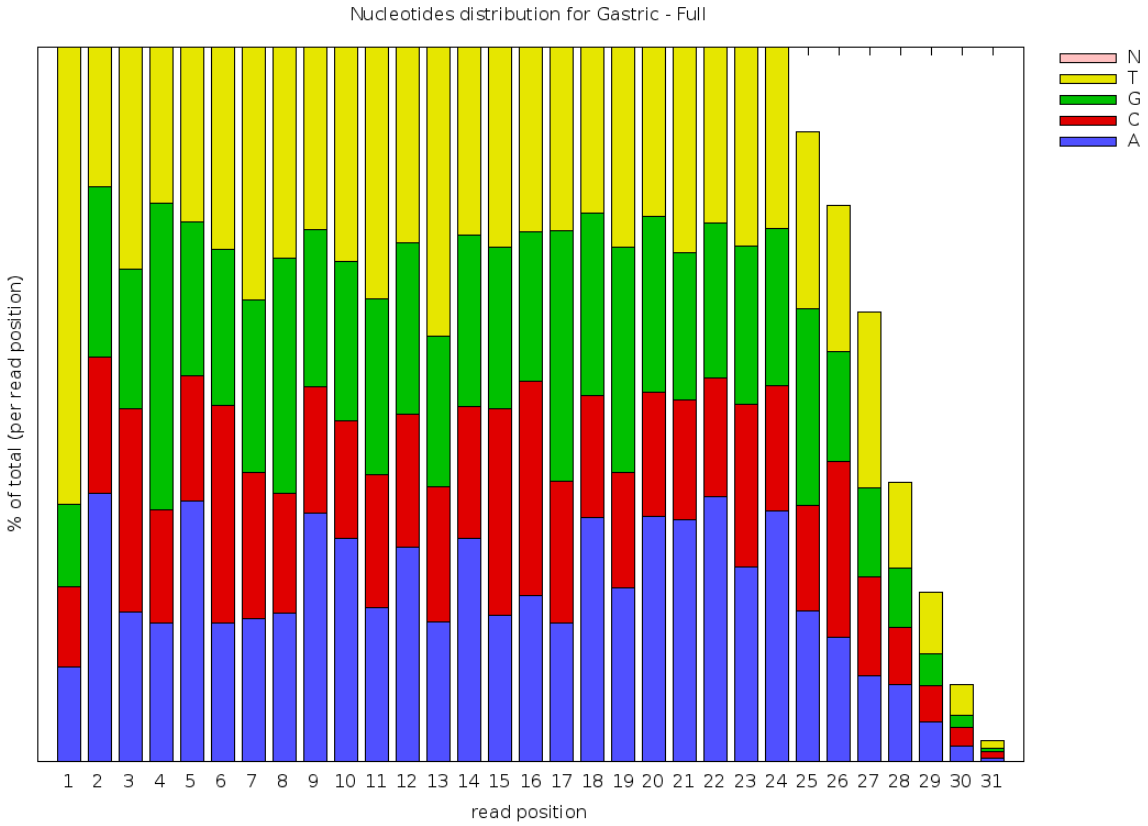

Figure 3.5: Nucleotide distrubtion of size selected (23-31 nt) reads from a gastric caeca and salivary gland sRNA library.

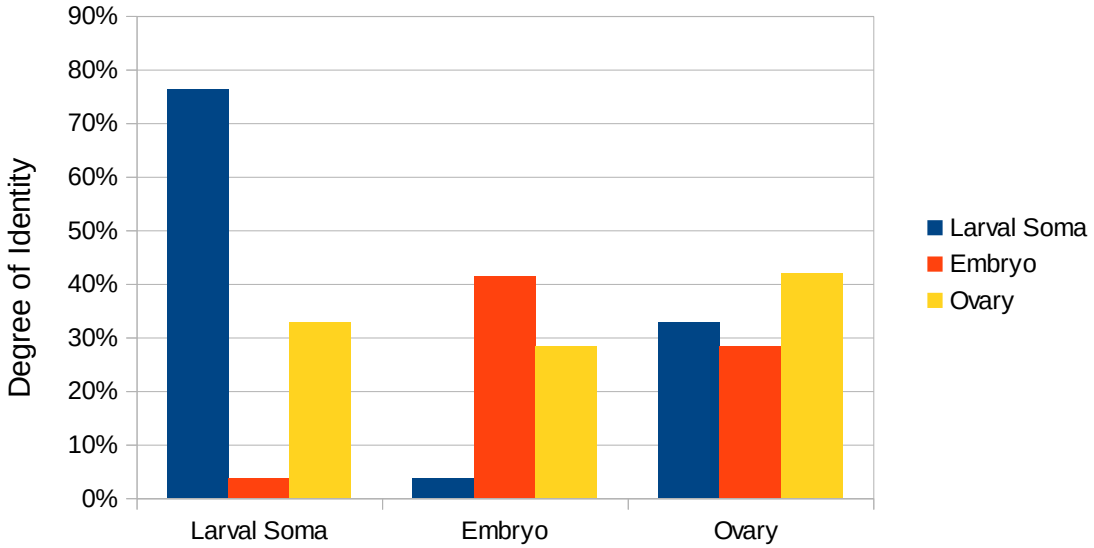

Figure 3.6: Average percent identify, by base pairs, of proTRAC called piRNA clusters, both within libraries within a sample set (larval soma, embryo, and ovary) and between sample sets.

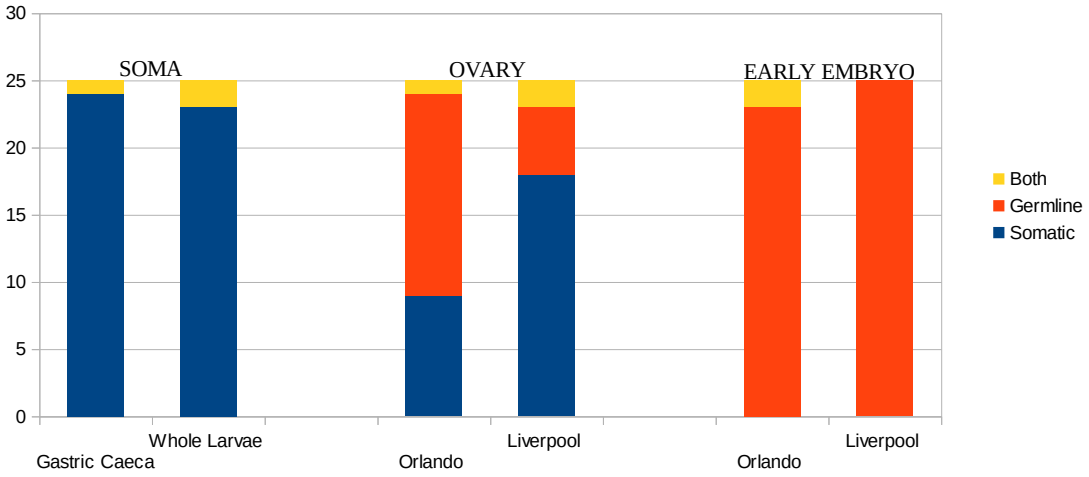

Figure 3.7: Commonality of the top 25 piRNA clusters from *Ae. aegypti* tissues. For the soma, two tissue types are used, gastric caeca and whole larvae. For ovary (mixed soma and germline) and early embryo (germline) graphs, Orlando and Liverpool indicate the strain used to generate the library.

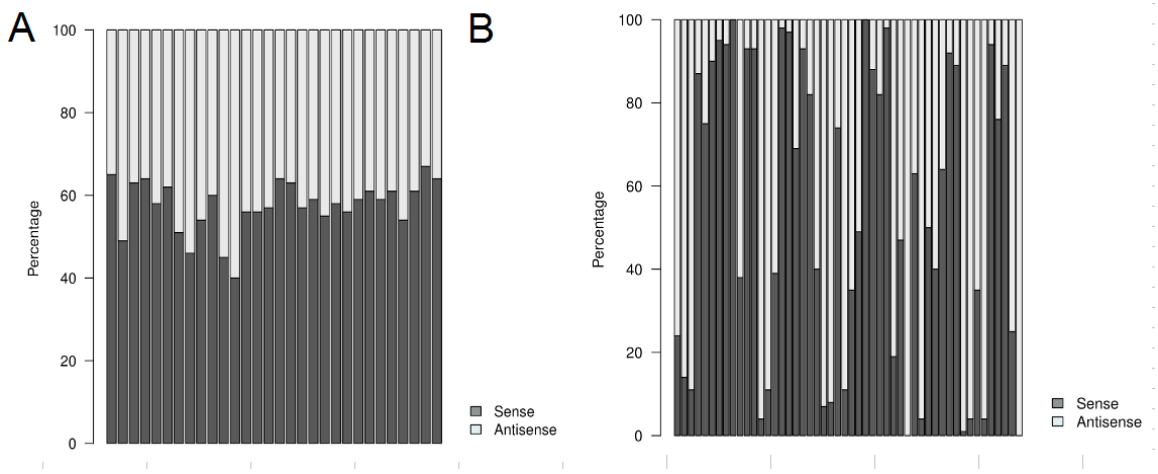

Figure 3.8: The sense/antisense orientations of features within individual piRNA clusters. (A) Top 50 piRNA clusters called from an *Ae. aegypti* Orlando embryonic library. (B) Top 50 piRNA clusters called from an *Ae. aegypti* Orlando larval gastric caeca library.

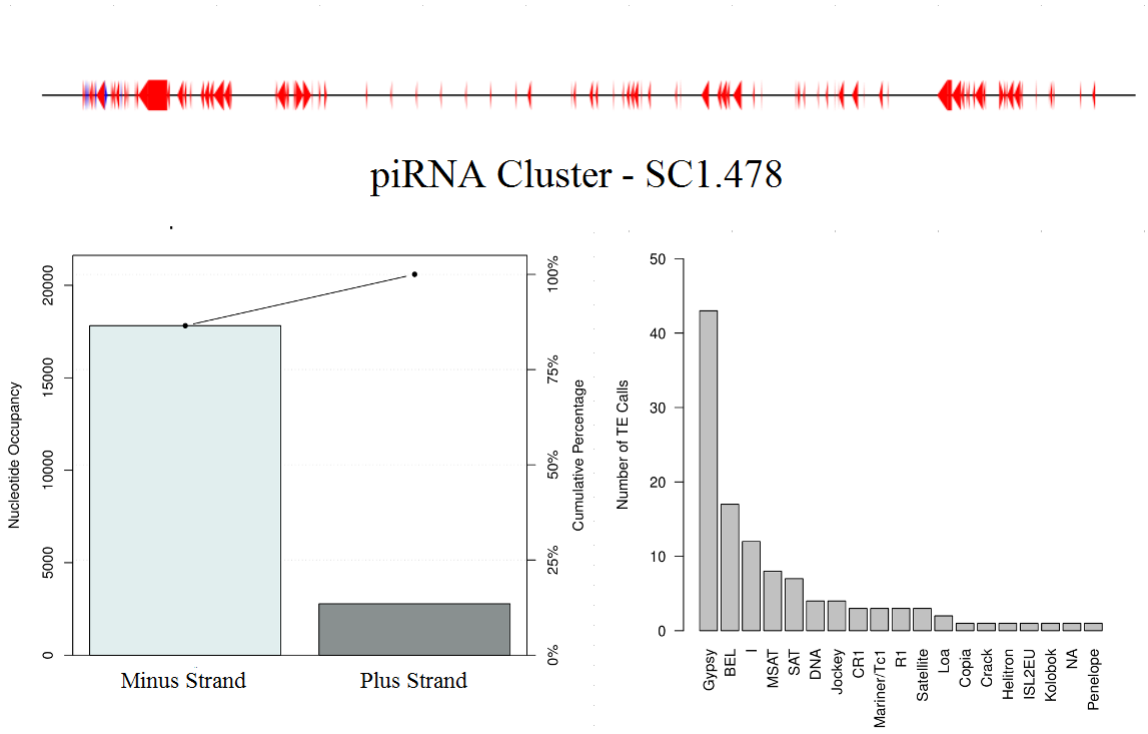

Figure 3.9: The somatic cluster SC1.478, showing feature orientation along the genomic locus and with a summary bar graph, and a chart of transposon content. Red features are transposons, blue features are genes.

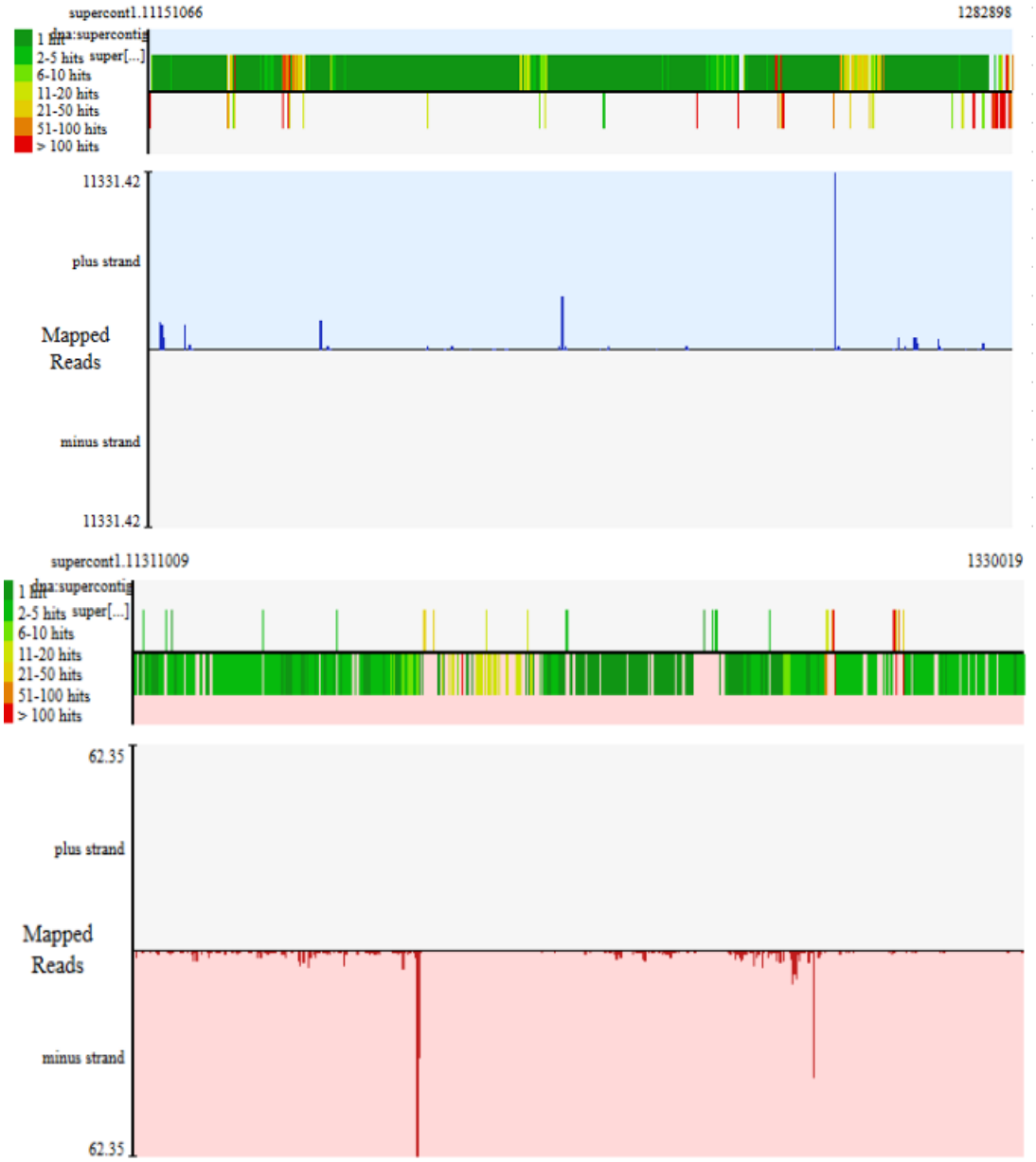

Figure 3.10: sRNA mapping plot of the two adjacent halves of somatic cluster supercont1.1. Peaks in blue are on the plus strand and are confined to the first half of the cluster; peaks in red are on the minus strand and are confined to the second half of the cluster.

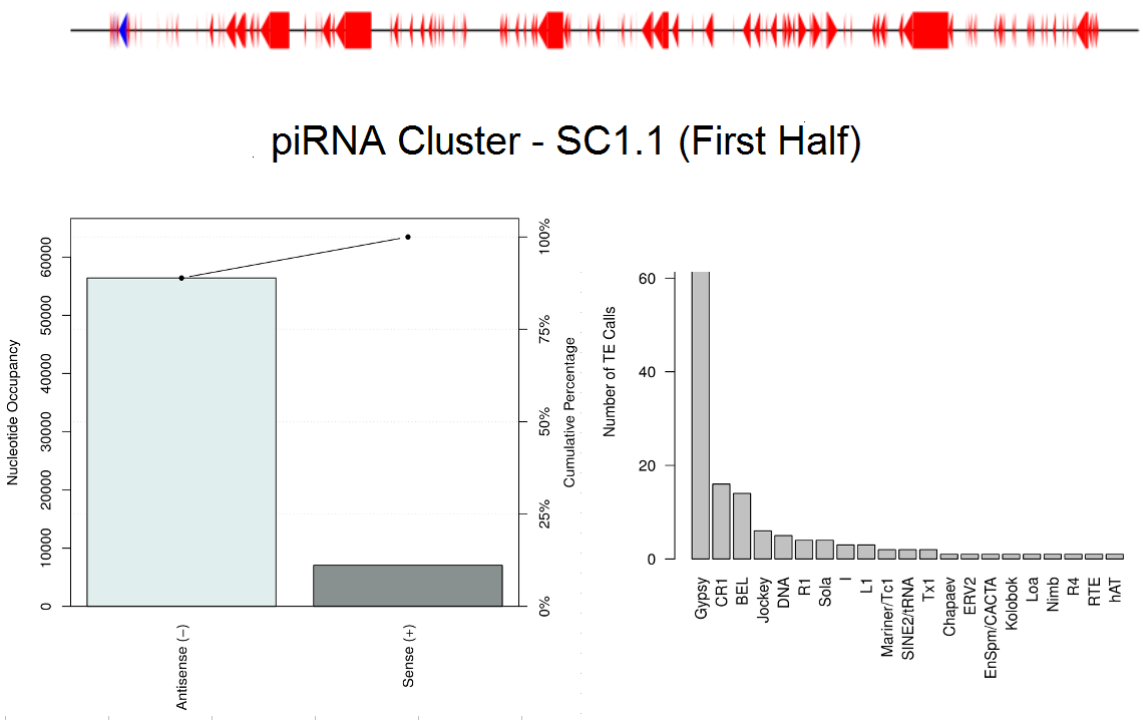

Figure 3.11: The first half of the somatic cluster on supercont1.1, showing feature orientation along the genomic locus and in a summary bar graph, and a chart of transposon content. Red features are transposons, blue features are genes.

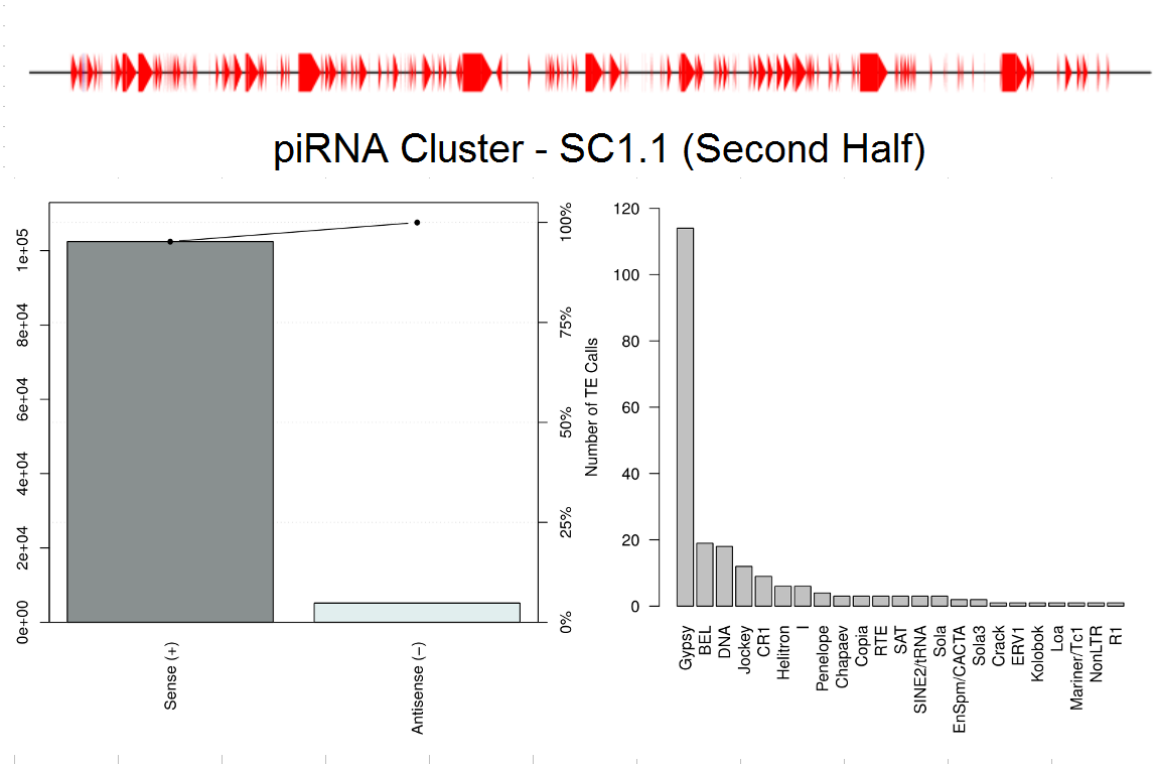

Figure 3.12: The second half of somatic cluster supercont1.1, showing feature orientation along the genomic locus and in a summary bar graph, and a chart of transposon content. Red features are transposons, blue features are genes.

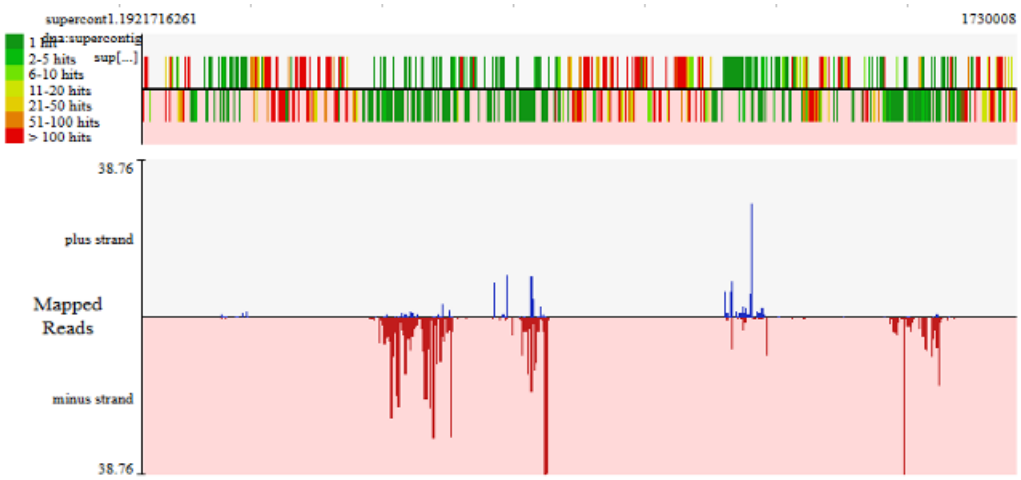

Figure 3.13: sRNA mapping plot of germline cluster supercont1.192. Peaks in blue are on the plus strand; peaks in red are on the minus strand.

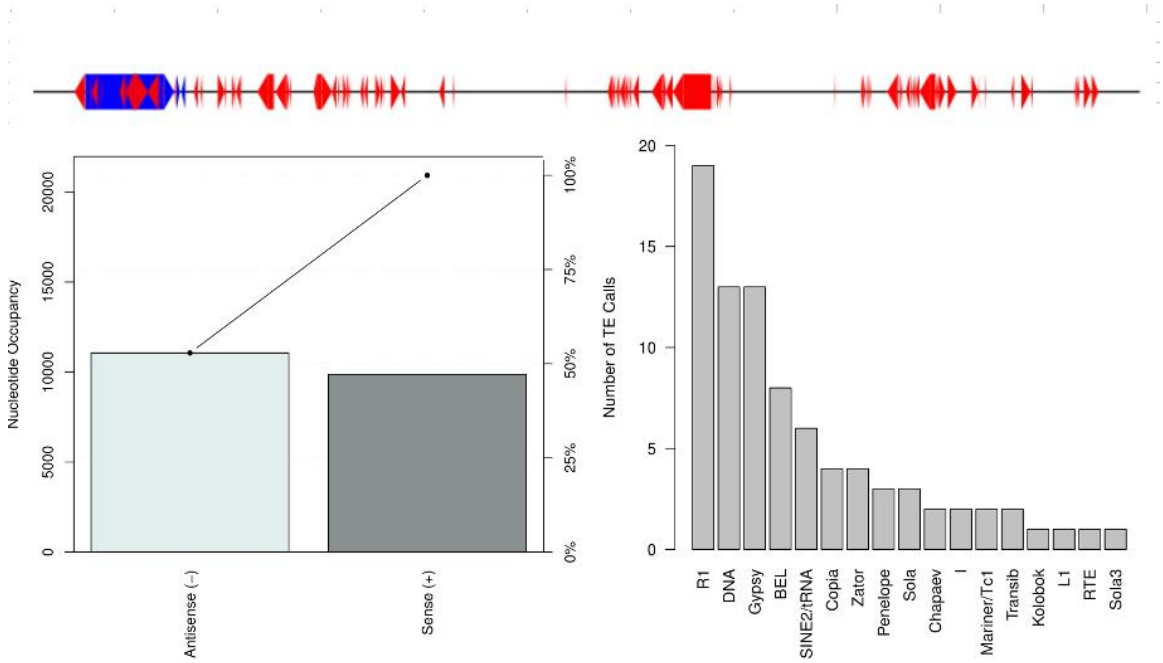

Figure 3.14: The bidirectional germline cluster supercont1.192, showing feature orientation along the genomic locus and in a summary bar graph, and a chart of transposon content. Red features are transposons, blue features are genes.

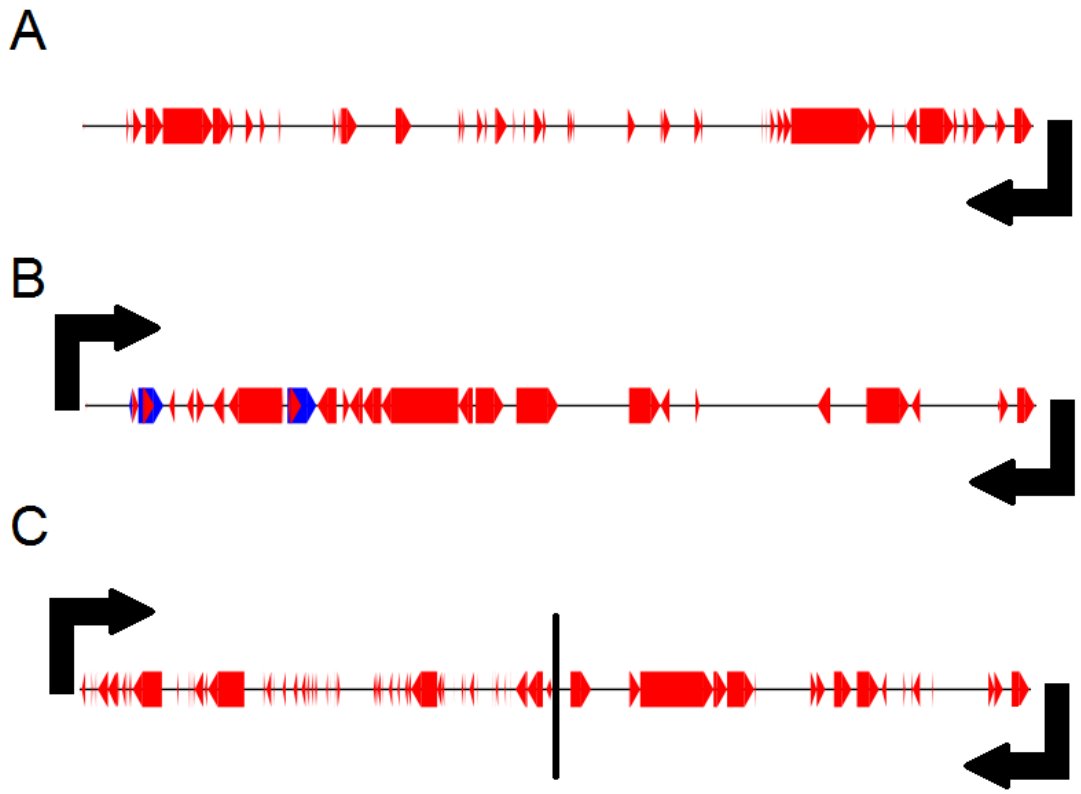

Figure 3.15: The different types of piRNA clusters observed in *Ae. aegypti.* Red features are transposons; blue features are genes. Arrows indicate feature orientation. (A) Unidirectional cluster producing piRNA anti-sense to cluster features. (B) Bidirectional cluster producing piRNA both sense and anti-sense to cluster features. (C) Adjacent undirectional clusters which produce piRNA anti-sense to cluster features; one cluster from the plus stand of the genome and the other from the minus strand.

Chapter 4 – A somatic role for Ago 3 in *Aedes aegypti*

4.1 Introduction

4.1.1 Ago 3 mutants and piRNA in *D. melanogaster*

In *D. melanogaster,* Ago3 mutants in both sexes have reproductive defects. a*go3* females are sterile; they produce few embryos, none of which hatch, and a subset of these embryos have patterning defects in the dorsal appendages (Li et al., 2009). Although they are not completely sterile, unlike mutant females, *ago3* male flies have reduced fertility. When the testis of *ago3* flies were tested for the presence of the germline factor Vasa, no Vasa expression was detected in the testis germline stem cell niche, which suggests that without Ago3, male flies cannot maintain their germline and their fertility (Li et al., 2009).

piRNAs in *D. melanogaster* are closely tied to one of the three PIWI proteins – sequencing of ribonucleoprotein (RNP) complexes from ovaries revealed that 73-87% of bound piRNA are unique to a specific protein, suggesting that there is no major overlap in the piRNA populations that different PIWI proteins bind (Brennecke et al., 2007). Ago 3 shows a binding preference for piRNAs that are sense to transposable elements (75%), and have an A bias at position 10 (73%). In contrast, piRNAs bound to Piwi and Aub preferentially bind antisense to transposable elements (76%, 83%) and have a U bias at position 1 (83%, 72%). (Brennecke et al., 2007). Although they share some characteristics, the piRNAs bound to Piwi and Aub are distinct in terms of their genomic origin. Piwi bound piRNAs are produced from somatic piRNA clusters such as *flamenco,* which is

made up of many transposon-derived sequences transcribed in an anti-sense orientation. Aub bound piRNA are produced from germline piRNA clusters such as 42AB, which are also made up of transposon-derived sequences. However, an antisense bias is not enforced for these sequences (Brennecke et al., 2007).

Taken as a whole, roughly 80% of *D. melanogaster* ovarian piRNAs map to transposons or simple local repeats, and 5-10% map to unannotated heterochromatic genomic regions, which may contain uncharacterized transposons (Brennecke et al., 2007). Although Ago 3 binds sense strand piRNAs, the majority of piRNAs are antisense to their targets, suggesting that Ago 3 is present at a low abundance compared to the other two Piwi proteins in the ovary. Based on sequencing, the pool of Piwi piRNA is 15 times higher than the pool of Ago3 piRNA; similarly, the pool of Aub piRNA is 6.4 times higher than Ago3 (Li et al., 2009).

Loss of function mutations in *ago3* in *D. melanogaster* had effects on piRNA populations in the ovary, as well as transposable element expression (Li et al., 2009). Broadly, transposable elements could be divided into two groups, with one transposon family that did not fit neatly into either category. One group of piRNA, which mapped against 68 of the 95 *D. melanogaster* transposon families, had their abundance greatly diminished by the loss of Ago3; both sense and anti-sense transposon-mapping piRNA were effected, but anti-sense piRNA saw the heaviest decrease (Li et al., 2009). Thus, although Ago3 binds sense piRNA, the loss of Ago3 protein had a larger effect on the anti-sense piRNA populations mapping to transposons in this group, probably due to

the catalytic amplification effect of the ping-pong cycle. The other group, representing 26 of the 95 *D. melanogaster* transposon families, were less heavily effected by the loss of Ago3. piRNAs mapping against these elements were more often found bound to Piwi, and to have sequences present in the *flamenco* locus, and thus an Ago3 independent mode of biogenesis (Brennecke et al., 2007).

4.1.2 Ago 3 mutants and piRNA in *Ae. aegypti*

Less research has been performed on the the effects of loss of function of Ago3 in *Ae. aegypti.* Investigation in an *Ae. aegypti* cell line, Aag2, which is immune-competent and of embyronic tissue culture origin (Barletta et al., 2012; Peleg, 1968) revealed that Aag2 cells have a PIWI family expression profile very similar to somatic tissue in the adult mosquito; expression of Piwi 4, 5, and 6 and Ago 3 are high, but expressions of Piwi 1, 2, and 3, which are more germline specific, and Piwi 7, which is only highly expressed in the early embryo, are all low (Miesen et al., 2015). The vast majority of TE-derived piRNAs sequenced from Aag2 cells were antisense (316,877) to annotated *Ae. aegypti* transposon sequences and had a strong U1 bias, as opposed to the sense fraction (24,610), which were less abundant. Sense piRNAs showed an A10 bias, similar to the ping-pong signature in *D. melanogaster* (Vodovar et al., 2012). Interestingly, Ago 3 dsRNA-mediated knockdown in these cells showed a mild effect on anti-sense TE derived piRNAs, but a stronger reduction in sense strand piRNAs, opposite of the anti-sense collapse seen in *D. melanogaster* Ago 3 mutants. In addition to Ago 3, further

dsRNA-mediated knockdowns showed that Piwi 4 and 5 were also important in the production of wild-type levels of TE-derived piRNAs. Although Piwi 4 knockdowns impacted TE-derived piRNAs, somewhat contradictory, Piwi 4 IPs were depleted of TE-derived piRNAs. (Li et al., 2009; Miesen et al., 2015). In addition to TE-derived piRNAs, Aag2 cells also produced viral derived piRNA (vpiRNA) in response to arbovirus infection; similar to TE-derived piRNA, these vpiRNA also had a ping-pong signature – U1 and A10 – and were dependent on Ago 3 and Piwi 5; there was, however, no dependence on Piwi 4 (Miesen et al., 2015, 2016).

4.1.3 Chapter aims

With the aim of examining the function of *Ae. aegypti* Ago3 in the whole animal, Ago3 RNAi lines were generated. As described in chapter 2, section 2.4.4 and Table 2.5, after doxycycline induction both mortality/developmental delay as well as a reduction in Ago3 transcript were observed in male fourth instar larvae. In addition, somatic expression of Ago3 in larvae was verified by qPCR and immunohistochemistry, with an enrichment of expression in the gastric caecum (Chapter 2, section 2.4.5). The mechanism by which doxycycline induction was causing mortality in the male larvae was unclear.

To obtain a molecular understanding of the effects of somatic Ago3 knockdown, mRNA-seq libraries were generated from wild type and Ago3 RNAi male and female larvae, with the aim of testing for any differences in gene or transposon expression in the

mRNA-seq libraries that could lead to the observed phenotype. In addition, the marked whole-body effect of Ago3 RNAi induction and the fact that the larval animal is mainly comprised of somatic tissue raised questions about the somatic role of Ago3 in *Ae. aegypti*. In the model organism *D. melanogaster,* no effect of Ago3 beyond the ovary or testes was seen, and mutants are typically sterile. Alongside the mRNA-seq libraries, sRNA-seq libraries were made from larval somatic tissue to examine the effects on sRNA populations following Ago3 knockdown.

4.2 Materials and methods

4.2.1 Mosquito rearing and RNAi induction

Strains used in these experiments were reared as described in chapter 2, section 2.2.1. Ago 3 RNAi lines were transformed at the University of Maryland Insect Transformation Facility with plasmids made by a Robert Hice, a research associate in the Atkinson laboratory. *Ae. aegypti* Orlando mosquitoes were transformed by microinjection of transformation donor (pBacTet-ago3SB) and a helper plasmid expressing piggyBac (pBac) transposase. These lines express a 372 base pair sequence of Ago 3 in a snapback configuration to form a double-stranded RNA molecule under the control of a tetracycline-induced promoter. The tetracycline reversible trans-activator is under the control of a heat shock promoter, hsp70 (Figure 4.1). To induce RNAi targeted against Ago 3, eggs were hatched under vacuum conditions in one liter of distilled water

sterilized by UV light. A 30mg/ml aqueous stock solution of doxycycline hyclate was then added, to create final rearing conditions of 30ug/ml of doxycycline.

4.2.2 Genome walking

To search for the integration sites of our Ago 3 RNAi snapback transformation construct in *Ae. aegypti* Ago 3 line M14, genomic DNA was extracted from transgenic mosquitoes using a Wizard Genomic DNA Kit from Promega. 500 ng of gDNA was used in a double digest with Fermentas FastDigest enzymes Afe1 and SnaB1. Digested DNA was then cleaned through a column with a Qiagen PCR cleanup kit. Eluted DNA was then ligated to genome walking adapters with T4 DNA ligase. Genome walking adapters were two sequences annealed together at 95°C , a top adapter, 5′-GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGT-3′ and a bottom adapter 5′-ACTATAGGGCACGCGTGGT-3. The ligase in the ligation reaction was then heat killed at 65°C for ten minutes.

1 ul of the heatkilled ligation reaction was used as template for a 25ul LongAmp (New England Biolabs) PCR reaction. Cycling conditions were 94°C for 30″, 6 × (94°C for 20″, 65°C for 5'), 31 × (94°C for 20″, 59°C for 30″, 65°C for 5'), 65°C for 10', 4°C hold. Primers were GWAP1, 5'-GTAATACGACTCACTATAGGGG and pBac RE 1F, 5'-GTTACTTTATAGAAGAAATTTTGAG. This reaction was diluted 1:50 and used as a template for a second round of amplification, using GWAP2, 5'-ACTATAGGGCACGGCGTGGT and pBac RE 2F, 5'- CGTACGTCACAATATGATTA.

Reactions were run out on a 1% agarose gel. If a distinct band was present, PCR products were column purified (Qiagen PCR Cleanup Kit) and sequenced at the IIGB Genomics Core. The sequences were then blasted against the *Ae. aegypti* genome using the NCBI blast tool to determine the integration site.

Final integration site verification was performed with primers in the genomic region to the left and right of the insertion site as determined via genome walking. PCR amplification was performed with pBac RE 1F and AAEL017329 F (5'-AATAACCATCACGTCCTCGTACTG) and with pBac LE 1F and AAEL017329 R (5'-GGTTGTTTCTCACACTTATCCGGATGAA) to verify the genomic location directly to both the left and right ends of the pBac transposon used to mediate transformation.

4.2.3 Genomic sequencing

Further integration sites were determined through genomic sequencing of the M14 line. Two micrograms of genomic DNA were fragmented on a Diagenode Biorupter at the IIGB Genomics Core, for six minutes at 50% power. DNA fragments of approximately 300 bp were saved and processed with the NEBnext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs), omitting the reverse transcriptase steps. After adapter ligation and purification of library products, fragments of around 400bp were sent to the IIGB Genomics Core for 100 bp paired-end sequencing reads.

To check libraries for sequenced genomic fragments that spanned the integration site, library reads were aligned to the transformation donor plasmid with Bowtie2 (Langmead and Salzberg, 2012). If both ends of a paired read mapped to the plasmid, they were discarded. If only one end of a paired read mapped to the plasmid, its partner was pulled out of the library file with a grep function, which pulls out a line from a file that contains the specified characters, in this case the flowcell location ID of the read. Both paired reads would have been sequenced from the same read present at one location in the flowcell. This read was then blasted against the *Ae. aegypti* genome to determine the genomic location of integration.

4.2.2 Tissue dissections

Tissue dissections were performed as described in chapter 2, section 2.2.4.

4.2.3 Quantitative PCR

qPCR was performed as described in chapter 2.2.6. Primers for qPCR analysis of transposon expression are located in Table 4.1.

4.2.4 Doublesex PCR sexing of larvae

Larvae were sexed as described in chapter 3.2.7.

4.2.5 Library preparation and sequencing

Libraries were prepared as described in chapter 3.2.3, 3.2.4, and 3.2.8.

4.2.6 Determination of differential expression of *Ae. aegypti* genes

To determine the differential expression of *Ae. aegypti* genes between different library sets, ribosomal RNA (rRNA) reads were first removed from all libraries using Bowtie2 to map and then discard all reads that aligned to a file containing arthropod rRNA sequences, downloaded from the Silva rRNA database (Quast et al., 2013). Remaining non-ribosomal reads were then aligned to the *Ae. aegypti* genome assembly downloaded from Vectorbase using STAR (spliced transcripts alignment to a reference) on the high memory node of the IIGB genomics biocluster (Dobin et al., 2013).

From the SAM (sequence alignment/map) file that was the output of STAR, a genome annotation file (GTF file) was used to inform Cuffdiff2 to determine which reads (aligned to the genome) corresponded to genomic regions that were annotated as genes (Trapnell et al., 2013) The GTF file for *Ae. aegypti* was downloaded from Vectorbase. Cuffdiff2 then assigned normalized read values for every gene present in the GTF file, and used library replicates to determine if significantly differential gene expression existed between library conditions. Both a p-value and a q-value, to control for false discovery rates inherent in testing thousands of genes (FDR), were generated for each gene. Genes with a q-value of less than 0.05 were considered to have significant differences in expression between the two library sets being compared.

4.2.7 Determination of differential expression of *Ae. aegypti* transposons

To determine the differential expression of transposons from mRNA-seq libraries, a different approach was used than the approach used for genes, following the workflow used by piPipes (Han et al., 2015). Cuffdiff2 requires features to be annotated in the genome and compiled in a GTF file format, which commonly does not include repetitive sequences like transposons, which can be present in the genome in hundreds or thousands of different locations. To address the challenge of accurately quantifying library reads to transposons, reads were first mapped not to the whole genome, but to a FASTA file containing *Ae. aegypti* transposons compiled from Repbase, as well as including transposons discovered in the lab, such as Muta1. Bowtie2 was used to map reads.

The total number of reads to different genomic features – transposable elements and viruses – were counted using a script called eXpress (Roberts and Pachter, 2012) from STAR mapping output of the processed sRNA libraries. eXpress uses an algorithm based on nucleotide biases and feature length to assign reads that don't map to features uniquely – that is, they map to multiple locations in the genome. This created more accurate read counts for features, especially repetitive features such as transposons. The output from the eXpress counting of reads, from the estimated_counts column, were then tabulated into a matrix that could be imported into the R environment and analyzed through the use of the Bioconducter package, DESeq2. Significantly

differentially sequenced counts to individual features were then called using DESeq2, using a false discovery rate of 10%.

Count numbers from eXpress were rounded to whole integers in R for DESeq2 input. DESeq2 calculated dispersion factors for the analysis and then performed hypothesis testing and FDR calculations (Love et al., 2014). To calculate size factors for DESeq2 analysis of transposon expression, size normalization factors for the set of reads that correspond to genes from library samples were used instead of calculating size factors from transposon counts only, which were a smaller proportion of the library, similar to methods used in Li et al, 2009. This also avoided masking of global chances in transposon expression in different sample conditions (for instance, in the Ago3 knockdown libraries) by normalization – in typical mRNA-seq experiments, the experimental condition will not cause a change in global transcription, but only in a subset of genes (Love et al., 2014).

4.2.8 Determination of differential populations of *Ae. aegypti* sRNA

sRNA libraries were trimmed of Illumina adapters as described in Chapter 3, section 3.2.9. To call differential populations of sRNA mapping to genes, the same strategy outlined in 4.2.6 was used. To call differential populations of sRNA mapping to transposons, the same strategy outlined in 4.2.7 was used, with one modification. To cut down on processing time, Bowtie2 mapping of reads was restricted from reporting all possible alignments down to reporting a maximum of 10 alignments per read.

4.3 Results

4.3.1 Characterization of Ago3 – M14 Line phenotypes

As described in chapter 2, section 2.4.4 and Table 2.5, Ago3 M14 mosquitoes induced with doxycycline hyclate showed developmental delay and male mortality. After homozygosity of the M14 line was established for the transgene, the line was maintained with relaxed screening (less frequently than every generation), after which new sex-specific phenotypes began to emerge beyond the canonical 3xP3-GFP expression.

The 3xP3-GFP transgene expresses green fluorescent protein (GFP) under the control of the 3xP3 promoter, which is most transcriptionally active in the eye (A. Wimmer et al., 1999; Berghammer et al., 1999). Male M14 mosquitoes expressed GFP in the eye, and also expressed GFP in other tissues (Figure 4.1). For instance, male M14 pupae exhibited fluorescence in the abdomen (GE/AP) or the developing legs (GE/LP). This may be due to multiple integrations near endogenous promoters. However, female M14 mosquitoes did not express GFP in the eye, despite the presence of the transformation construct and GFP coding sequence in the genome (Figure 4.2), which may be due to a sex-specific transcriptional silencing of the construct and transgenes.

4.3.2 Characterization of Ago3 – M14 Line integration sites

To determine the integration sites of the transgene as a possible explanation for the male lethality and/or GFP expression expansion phenotypes, I first used a genome

walking approach. Through genome walking, one integration site in the *Ae. aegypti* gene AAEL017329 was discovered (Fig 4.2). AAEL017329 is homologous to the *D. melanogaster mei-P26* gene, and the insertion was into the 3' untranslated region (UTR) of the *Ae. aegypti* gene. In *D. melanogaster, mei-P26* mutants of both sexes have meiosis exchange defects, leading to impaired germline differentiation and fertility; however, lethality is not seen (Page et al., 2000). No significant differences in expression, as called by DESeq2 from mRNA-seq data, were seen in levels of *Ae. aegypti mei-P26* transcript when comparing Ago3 M14 transgenic larva and wild type Orlando larva. Without a disruption in the coding sequence or change in expression, wild type function of AAEL017329 appeared to be preserved.

Two further integration sites were sequenced as junction DNA fragments in a genomic DNA library. These insertions were not in protein-coding or miRNA genes; one was in supercontig1.113, position 1,837,859, and the other was in supercontig1.384, position 310,244. The site in sc1.113 is 70kb away from the nearest gene (AAEL004277) and the site in sc1.384 is 100kb away from the nearest gene (AAEL0092677). No significant differences in expression, as called by DESeq2 from mRNA-seq data, were seen in the transcript levels of nearby genes when comparing Ago3 M14 transgenic larva and wild type Orlando larva. Based on these results, the integration sites of the transformation plasmid did not appear to effect wild type gene function.

4.3.3 M14 Ago3 and transposon expression via QPCR

Upon induction, Ago3 transcript in doxycycline induced male M14 larvae was on average 19.6% the level in uninduced male M14 larvae, with a P(H1), or the probability that the difference between the sample and control groups were due only to random choice, value of 0%. In addition to *Ago3*, the expression levels of a panel of six different transposons was tested with qPCR. These included class II cut-and-paste DNA elements including *AedesBuster 1* (AeB1), *Mutator 1* (Muta1), and *pogo12*, a SINE element, *Gecko*, a long terminal repeat (LTR) retrotransposon, *LOA_Lian* (Lian), and a non-LTR retrotransposon, *Ty1Copia 174* (Copia). Of these elements, two had significantly higher levels of expression (P(H1) scores of less than 5%). *pogo12* expression in induced male M14 larvae was on average 289% the level in uninduced male M14 larvae, with a P(H1) value of 3.4%. *Lian* expression in induced male M14 larvae was on average 479% the level in uninduced male M14 larvae, with a P(H1) value of 0%. The other four elements also had higher levels of expression, but with P(H1) values ranging from 9.5% to 17.1%, above the threshold for significance (Figure 4.3).

Similar values were seen when comparing induced M14 male larvae with uninduced wild type Orlando male larvae. Ago3 transcript in induced M14 male larvae was on average 13.8% the level in control M14 larvae, with a P(H1), or the probability of an alternate hypothesis, value of 0%. *pogo12* expression in induced male M14 larvae was on average 469% the level in uninduced male Orlando larvae, with a P(H1) value of 3.3%. *Lian* expression in induced male M14 larvae was on average 681% the level in

uninduced male Orlando larvae, with a P(H1) value of 0%. *Gecko* expression in induced male M14 larvae was on average 214% the level in uninduced male Orlando larvae, with a P(H1) value of 0% *Copia* and *AeB1* also had higher levels of expression, but with P(H1) values of 14.8% and 10.5%, respectively, above the threshold for significance. *Muta1* had no change in relative expression (Figure 4.4).

No significant differences in expression were seen in *Ago3* or any of the six transposons in induced (dox+) Orlando wild type male larvae compared to uninduced (dox-) Orlando wild type male larvae (Figure 4.5). These results suggest that the transposon derepression and Ago3 knockdown measured by qPCR in induced M14 males was not due to an effect of the tetracycline by itself, but due to the RNAi induced by the antibiotic. In addition, no significant differences in expression were seen in *Ago3* upon induction in other tissue types, including two day blood fed ovaries and two-to-four hour embryos, suggesting that the tetracycline inducible system was ineffective in driving RNAi in those tissues.

4.3.3 M14 Ago3 expression via mRNA-seq

mRNA-seq was next carried out in several different 4th instar larval sets. In the Ago3 M14 line, the sets included: male induced, male uninduced, and female induced, female uninduced. In Orlando wild type mosquitoes, the sets included: male induced (dox+), male uninduced (dox-) and female induced (dox+) and female uninduced (dox-).

The female larvae, consistent with the phenotype of EGFP silencing, did not show a reduction in Ago 3 transcript levels. From three induced M14 female libraries, Cuffdiff calculated an average of 13.7 FPKM (fragments per kilobase of transcript per million mapped reads). From three uninduced M14 female libraries, Cuffdiff calculated an average of 14.3 FPKM. In addition, the female libraries did not have many reads mapping to sequences within the transformation plasmid that would otherwise be transcribed, indicating that the plasmid was silenced. Induced M14 males had 21.7 RPM to the reverse tetracycline-controlled trans activator (rtTA) and uninduced M14 males had 0.54 RPM; induced M14 females had 0.01 RPM and uninduced M14 females had 0 RPM. Looking at sequencing of EGFP transcript fragments in these libraries, induced M14 males had 247 RPM and uninduced M14 males had 33.41 RPM; induced M14 females had 0.19 RPM and uninduced M14 females had 0.14 RPM. Orlando wild type males had no reads to pBacTet-ago3SB. Orlando wild type females had 0.29 RPM mapping to EGFP; this may be due to rearing contamination (Table 4.3).

Induced male M14 larvae did show a reduction in Ago3 transcript levels compared to uninduced M14 larvae and the Orlando wild type controls. Induced male M14 larva had an average RPKM to Ago3 (minus the foldback sequence) of 2.15 – 23.4% of the RPKM value of Ago3 in uninduced male M14 larvae and 28% of the RPKM value of Ago3 in the Orlando wild type controls, in line with previous qPCR results (Figure 4.6). The libraries supported the qPCR data in showing an Ago3 knockdown for male M14 mosquitoes only; in female mosquitoes Ago3 RNAi knockdown induction failed

and there was low expression of genes carried by the transformation plasmid, indicating a failure of the RNAi system in females.

4.3.3 M14 Line gene expression via mRNA-seq

Cuffdiff called fifteen genes as significantly differentially expressed between the M14 female induced and uninduced samples, but of these genes, the difference between the two sample sets was not large (Table 4.4). All fifteen had no expression, a FPKM value of 0, in the female induced libraries, and were expressed at a very low level (between 0.56 and 2.94 FPKM) in the female control libraries, corresponding to only a minor effect on global mRNA expression in female mosquitoes that could be explained by noise.

Examination of genes that were significantly differentially expressed between sexes in 4th instar larvae revealed that all eleven genes were expressed at higher levels in the female larvae, from 32 to 1024 times higher (Table 4.5). Of these genes, eight had an unknown function. Three of these sex-biased expression genes belong to a large Dipteran syntenic gene family called Osiris, which is lethal in *D. melanogaster* when the whole locus is deleted (Shah et al., 2012).

Looking at genes that were significantly differentially expressed between the M14 male induced and uninduced samples, four genes were expressed strongly in the induced sample libraries, and three genes were expressed only in the control sample libraries (Table 4.6). Of the three genes expressed only in the control libraries, their expression ranged from 2.98 to 9.52 FPKM. Two of these genes were of unknown function, and one was homologous

to a glucose dehydrogenase, AAEL011808. Mutants of the homologous dehydrogenase gene in *D. melanogaster* are viable (Dietzl et al., 2007). Of the four genes expressed more highly in the induced sample libraries, the increase in expression ranged from 8 to 16 times higher. Two of these genes, AAEL004280 and AAEL01433, are *Osiris* genes that were also seen to be differentially expressed between male and female larvae. One gene remained unnanotated, and the other was homologous to an alkaline phosphotase, AAEL003289. Mutants of the homologous alkaline phosphotase gene in *D. melanogaster* are viable (Dietzl et al., 2007). From these genes, there is no clear candidate for a cause of the male lethality phenotype as the genes are either unannotated or not known to cause lethal phenotypes in *D. melanogaster.*

4.3.3 M14 transposon expression via mRNA-seq

Transposon expression in M14 female larvae was largely unaffected by doxycycline induction. The majority of the 2086 transposons tested did not differ in expression by more than four fold, or a log two fold change of two (Fig 4.7). 221 transposons had no expression, or too low expression to test, or were count outliers as determined by Cook's test. 12 transposons were repressed by a log fold change of more than two; ten of these were *Gypsy* elements, one was a *Copia* element, and one was a *BEL* element. However, none of these changes in expression had a p-adjusted value of less than 10% and thus are likely not significant as determined by DESeq2. 12 transposons were upregulated by a log fold change of more than two; four of these were BEL elements, three were *Gypsy* elements, two were *SOLA* elements, and there was one each of *SAT*, *Mariner*, and *Kiri* elements. Again, none of these log fold expression changes had a p-adjusted value of less than 10%.

Transposon expression of certain elements in the M14 male larvae were affected by doxycycline induction of Ago 3 knockdown. 14 elements scored a p-adjusted value of less than 10%, and 23 elements scored a p-adjusted value of less than 20% (Table 4.7, Figure 4.8). Of the 23 elements with a p-adjusted value of less than 20%, 22 saw an increase in expression (ranging from a log fold change of 2.8 to 5.1). One element saw a significant decrease in expression (log fold change of 5.2). 12 of the 23 elements with a p-adjusted value of less than 20% were long terminal repeat (LTR) retrotransposons, including *Gypsy*, *Copia*, and *BEL* superfamilies. Nine elements were non-LTR retrotransposons, including *LOA*, *Crack*, *I*, and *Kiri* superfamililes. Two elements were class II DNA transposons: *piggyBac-1_AAE* and *Sola1-N9C_AAe*.

Based on these results. Ago3 knockdown had a modest effect on de-repression of mainly retro-elements.

4.3.3 sRNA populations following Ago3 RNAi knockdown

Upon Ago3 RNAi knockdown in M14 male 4th instar larvae, an overall reduction in transposon-mapping sRNAs was sequenced from larval sRNA libraries (Figure 4.9). Although 2064 *Ae. aegypti* transposons had non-zero mapping sRNA reads, 1753 (85%) of these were not tested by DESeq2 due to having total read-counts that were too low for an accurate analysis. Of the 311 remaining tested elements, significant differences in sRNA populations were seen in 57 elements with a p-adjusted value of less than 10%. 40 elements (12.9%) saw less sRNA mapping in the test Ago3 knockdown condition

compared to the control uninduced larvae; 17 elements (5.5%) saw greater numbers of sRNA mapping.

Of the 24 elements with a p-adjusted value of less than 5%, the majority of elements (16, 66.67%) were abundant *Gecko* family retroelements. The remainders were two *Sola* DNA transposons, three *Bel* retroelements, and one *Crypton* DNA transposon (Table 4.8). In terms of male 4th instar larvae sRNA populations, the greatest effect of the Ago3 RNAi knockdown was a reduction of sRNA mapping to *Gypsy* elements. In *D. melanogaster*, these elements are most active in the somatic follicular cells of the ovary (Mével-Ninio et al., 2007).

4.4 Discussion

4.4.1 M14 Line female mosquitoes

The loss of GFP expression in female M14 mosquitoes and the continued maintenance of GFP expression in the males suggests that there may be silencing of the transformation donor plasmid, and that this silencing is occurring in female larvae only. Based on PCR data, the donor plasmid is present in female larvae. However, GFP is not expressed, both in terms of observable GFP under a fluorescence microscope as well as detectable GFP transcript in the mRNA-seq libraries. Along with a lack of GFP transcript, females also lacked expression of other genes present in the transformation plasmid, such as the rtTA gene.

Since the rtTA gene is being expressed at very low levels (0.01 RPM), induction with doxycyline is unlikely to drive much transcription of the foldback Ago3 RNAi sequence or lead to much downstream silencing of Ago3 transcript. When Ago3 levels were tested with qPCR and mRNA-seq, no significant changes in expression of Ago3 were detected in the induced (dox+) M14 female samples. These samples included 4th instar larvae, blood fed ovaries, and embryos laid by gravid females. A further explanation for the lack of RNAi knockdown of Ago3 in ovaries and embryos (beyond female silencing) is that the induction method of doxycycline in rearing tubs and sugar water may not extend into the adult stage or permeate the germline tissues.

Along with lack of Ago3 knockdown in M14 females, analysis of the mRNA-seq libraries showed that there were no significant differences in expression in any transposable elements upon induction. Although there were fifteen protein-coding genes called as significantly differentially expressed, all fifteen had similar expression patterns, with no expression in the induced samples and very low expression in the control samples (between 0.56 and 2.94 FPKM), which corresponds to a very small difference in overall transcript abundance and may be due to technical factors such as differences in library coverage impacting very low abundance transcripts – low coverage libraries won't detect these transcripts. These results are in line with the wild-type levels of Ago3 in female mosquitoes. With wild type levels of Ago3 in induced female M14 mosquitoes, wild type levels of transposon and gene expression are also expected.

4.4.2 M14 Line phenotypes and integration sites

Upon induction of M14 mosquitoes, a marked sex bias was revealed – only adult females were seen. By tracking progress through mosquito developmental stages of a set number of embryos, it was determined that females weren't being converted to males. Instead, in the male larvae, developmental delay and mortality upon induction was seen at the 4th instar stage of larval development. Males arrested at the last larval stage and never pupated (Table 2.5). Since only females pupated, only female adults ever emerged from the induced M14 tubs. Doxycycline induction by itself had no adverse effects on the development of wild type mosquitoes, which do not contain the doxycycline sensitive Ago3 RNAi construct.

In M14 males, GFP was seen beyond the expression domain expected from 3xP3, a promoter that typically activates expression most strongly in the Dipteran eye. Male larvae were observed with GFP expression in the larval body, male pupae were seen with expression in the developing legs and abdomen, and adult males were seen with GFP expression in the antenna and abdomen, suggesting a pattern of multiple transformation donor integrations into the genome, next to endogenous promoters (Figure 4.1).

Three integration sites were identified in the M14 line, one by genome walking and two by whole genome sequencing, but more may exist. One site is within the 3' UTR of AAEL017329, *Ae. ageypti mei-P26*, but based on RNA-seq, transcription of the gene is unaffected. There may be effects on translation, but if there are effects solely from this

integration, they are unlikely to be the cause of the male morality phenotype, as the phenotype only appears upon induction. Based on the location of the site in the 3' UTR, activation and read-through transcription from the tet-operon would not result in a sense transcript or overexpression of *mei-P26*; the mRNA-seq libraries also do not detect a change in expression of this gene.

Two other integration sites, one in supercontig 1.113 and one in supercontig 1.384, are present in the M14 line. These sites are not located in genes, and the nearest genes are 70 to 100kb away. These nearby genes were not significantly differentially expressed in control (uninduced) M14 larvae, and induction with doxycycline did not cause the misexpression of these genes. Based on integration site distance and mRNA-seq data, it seems unlikely that integration-mediated disruption of the normal functioning of these genes caused the phenotype of male mortality. Since the phenotype of male mortality is only seen upon doxycycline induction, a more plausible explanation for the male mortality phenotype is the activation of the tet-operon and transcription of the Ago3 foldback sequence.

4.4.2 M14 line gene expression upon induction

Upon induction, a knockdown of Ago3 was seen in male M14 larvae. By qPCR and mRNA-seq fragment analysis, the level of Ago3 in induced larvae was 19.6% to 28% the level of control uninduced larvae. Along with the knockdown of Ago3, significant changes of expression in eight genes was seen. The analysis took into account replicate

libraries of wild type mosquitoes reared with doxycycline, so these changes in gene expression were not simply a result of the presence of the antibiotic in the environment. Three of these genes have an as of yet unknown function, and only one has a homologue in *D. melanogaster*.

Of the other genes, one, AAEL003289 has sequence similiarity to an alkaline phosphatase in *D. melanogaster,* CG1809. RNAi and P-element mediated disruptions of this gene result in flies that are fertile and viable (Dietzl et al., 2007). AAEL011808 has sequence similarity to a glucose dehydrogenase in *D. melanogaster,* CG6142. RNAi mediated disruptions of these gene result in flies that are fertile and viable. No obvious phenotypic differences were seen in these mutants; however, a comprehensive study has not been done (Dietzl et al., 2007). Although changes in expression were seen in these genes upon induction, in terms of understanding the mechanism of male mortality and developmental delay in the M14 line, the lack of evidence for severe or lethal mutant alleles of these two genes genes makes them weaker candidates for causing the phenotype.

The remaining two genes were identified as being upregulated (log two fold change of three) and are members of a large, syntentic gene family in Dipterans: Osiris. AAEL004280 has the strongest similarity to Osi9 in *D. melanogaster* – mutant alleles for this gene are viable in the fruit fly, but one developmental phenotype has been observed (Dietzl et al., 2007). Female fruit flies with a *p[GT1], P* element, insertion within the gene had significant defects in wing shape size and morphology. The mutation had no

significant effect on the wing shape of male flies (Carreira et al., 2011). AAEL0014433 has the strongest similarity to Osi6 in *D. melanogaster.* Mutant alleles for this gene in the fruit fly can either be viable or lethal, depending on the genetic source of GAL4, which may be related to the strength of the RNAi induction (Dietzl et al., 2007). These two *Ae. aegypti Osiris* genes were also seen in another analysis, looking at significant differences in gene expression between male and female 4th instar larvae (Table 4.5). The upregulation of these two *Osiris* genes in induced male larvae brings their expression closer to the levels seen in female larvae. If the mortality and developmental delay seen in induced male larvae is related to the upregulation of these two genes, it may be that female levels of expression of these genes is harmful in males. Their expression through developmental time may also play a role in the development phenotypes observed.

Studies on the *Osiris* gene family in *D. melanogaster* have focused on the initial finding of tight dosage control – the *Osiris* family syntenic gene locus was first identified as Triplo-lethal (Tp1) which is lethal when present in either one or three copies in an otherwise diploid organism, the fruit fly. Death occurs in late embryos or in the first instar of larval development (Dorer et al., 2003). The sensitivity to dosage in this gene family could provide a mechanism for male lethality in the M14 Ago3 knockdowns, due to misregulation of *Ae. aegypti* homologues for Osiris 6 and 9. Lethality in a later life stage, 4th instar larvae, could be due to the fact that only a subset of *Ae. aegypti* Osiris genes have been knocked down. In *D. melanogaster,* 23 Osiris genes are present; *Ae. aegypti* currently has 24 annotated (Dorer et al., 2003). These genes have a secretion

signal peptide and four domains in common, one of which is a putative transmembrane domain, and are thought to arise from a large gene duplication. They are expressed in various developmental stages, but their function remains unknown (Shah et al., 2012) Synteny of the gene family has been conserved in many insects, with *An. gambiae* and possibly *Ae. aegypti* being exceptions – the assembly of the *Ae. aegypti* genome is split into many supercontigs which makes identifying large syntenic regions more difficult (Shah et al., 2012).

4.4.3 M14 line transposon expression upon induction

Since *Ago3* in *D. melanogaster* has a well described function in transposon regulation, I performed a qPCR based assay testing the expression of six transposons upon Ago3 knockdown induction. An overall upregulation of transposon expression in the tested transposons was seen, compared to both wild type male larvae and uninduced M14 male larvae, though not always to a level exceeding significance. Two transposons that did consistently show significant increased expression in the induced male larvae were *pogo12* and *LOA_Lian*, 33% of the transposons tested.

A much smaller fraction of global transposon expression was identified as significantly differentially regulated based on the mRNA-seq triplicate libraries and the DESeq2 log fold change and FDR analysis. Of 1100 tested transposons, only 14 elements had a p-adjusted value of less than 10%. This represents 1.3% of tested transposons and 0.7% of total transposons. The small global effect seen could be due to a variety of

factors. One is redundancy of function – in *D. melanogaster,* there are only three *Piwi* genes, while *Ae. aegypti* has an expansion to seven. Knocking down only Ago3 could thus have a less severe phenotype of transposon deregulation. Another factor could be the incomplete knockdown of the RNAi system, instead of a mutant allele knockout. Since the M14 induced larvae retain some Ago3 function, the phenotype of transposon derepression may be less severe. A third factor could be an expanded role of Ago3 in other regulatory pathways – for instance, as shown in Chapter 3, a larger fraction of sequenced piRNA length sRNA in the larval libraries mapped to genes than transposons. If the focus of *Ae. aegypti* Ago3 is more towards binding gene derived piRNA instead of transposon derived piRNA, knocking down Ago3 would have less of an effect on transposon regulation.

In addition, the two transposons identified by the qPCR analysis, *pogo12* and *LOA_Lian* as consistently upregulated did not show up in the parallel mRNA-seq analysis. However, the majority of the six qPCR tested transposons were not tested for significance by DESeq2 from the mRNA-seq data because of very low expression values under both conditions. The exception was *Gecko*, a very abundant retroelement with high levels of expression. It is possible that in the absence of element-specific primers and PCR amplification as applied in the qPCR assay, the mRNA-seq libraries failed to achieve a sequencing depth or coverage that could confidently test the significance of differential read counts to these particular *pogo12* and *LOA_Lian* elements identified through qPCR that are nonetheless expressed at a low level in the organism.

4.4.4 M14 sRNA populations upon induction

Following Ago3 RNAi knockdown in male M14 4th instar larvae, a modest reduction in transposon-mapping sRNA populations was observed in larval sRNA libraries. 40 elements (12.9%) saw less sRNA mapping in the test Ago3 knockdown condition compared to the control uninduced larvae; 17 elements (5.5%) saw greater numbers of sRNA mapping. The majority of *Ae. aegypti* transposons (85%) did not have a significant number of mapping sRNA, which corresponds in part with the mRNA-seq data where 45% of the elements were not tested due to low expression. One explanation is the source of the libraries being predominately (apart from the germline rudiments) somatic larval tissue, where many transposons may be dormant as part of their transmission strategy and germline piRNA clusters are not heavily transcribed. In the absence of germline piRNA cluster transcription and active transposon mRNA triggers, sRNA populations against many transposons would be low or non-existent.

Supporting this theory is the finding that many (66.67%) of the elements which saw the most significant (p-adjusted value of less than 5%) changes in mapping sRNA populations between the test induced and control uninduced condition were *Gypsy* elements, elements known in *D. melanogaster* to be expressed heavily in the somatic follicular cells of the ovary from where they attempt to invade the germline in a retro-viral like fashion (Mével-Ninio et al., 2007). Although in *D. melanogaster Gypsy* elements are controlled by *Piwi* only in a subset of somatic tissue, the ovarian follicular cells, the situation may be different in *Ae. aegypti.* The expansion of the PIWI family in *Ae. aegypti,*

both in number of genes as well as the expression domain of these genes, may in part be due to a need to control the activity of certain highly active somatic transposable elements such as *Gypsy.*

4.4.4 Conclusions

A lethality and developmental delay phenotype was seen in the 4th instar larval stages of M14 males with an Ago3 knockdown of 19.6 to 26.8%. In females, a silencing of the donor plasmid and no Ago3 knockdown or EGFP expression was observed. Consistent with the lack of Ago3 knockdown, female larvae do not exhibit phenotypes or transposon depression upon induction, unlike the male larvae.

Three integration sites were discovered in the M14 line, but based on mRNA-seq data, it is unlikely that they contribute to the lethality phenotype. Two Osiris genes, homologues of *D. melanogaster Osi6* and *Osi9,* were found to be misregulated in induced male larvae. A mutant allele of *Osi6* is lethal in *D. melanogaster,* and differences in dosage, 1x or 3x, of the whole *Osiris* syntenic cluster is also lethal in *D. melanogaster.* In addition, *Osiris* genes are expressed variably throughout development. *Ago3* mediated regulation of *Osiris* in *Ae. aegypti* could thus be the mechanism for the phenotype of lethality and developmental delay seen upon induction.

Another potential explanation for the phenotype of male lethality is the misregulation of transposons. A subset of *Ae. aegypti* transposons (0.7%) were significantly upregulated upon induction and Ago3 knockdown in male larvae. These

transposons had higher levels of expression and may be more active than other elements in the whole-genome *Ae. aegypti* transposon list, some of which may now be immobile and inactive.

A major transposon family affected in larvae where Ago3 is knocked down was the *Gypsy* family. sRNA populations mapping against a subset of *Gypsy* elements (16) declined in a statistically significant fashion; in addition, a subset of *Gypsy* elements (8) saw their expression increase in a statistically significant fashion. Although the global effect on sRNA populations was modest, this may again be due to incomplete knockdown of Ago3 and a redundancy of PIWI family function in *Ae. aegypti.* An accumulation of somatic genetic damage throughout development could cause the lethality and developmental delay seen in the late instar male larvae, in a mechanism similar to the *piwi* mutant mediated stunting of fertility and both male and female gonads seen in *D. melanogaster.*

4.5 References

A. Wimmer, E., Berghammer, A.J., and Klingler, M. (1999). [No Title]. Nature *402*, 370–371.

Barletta, A.B.F., Silva, M.C.L.N., and Sorgine, M.H.F. (2012). Validation of Aedes aegypti Aag-2 cells as a model for insect immune studies. Parasit. Vectors *5*, 148.

Berghammer, A.J., Klingler, M., and A. Wimmer, E. (1999). Genetic techniques: A universal marker for transgenic insects. Nature *402*, 370–371.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell *128*, 1089–1103.

Carreira, V.P., Soto, I.M., Mensch, J., and Fanara, J.J. (2011). Genetic basis of wing morphogenesis in Drosophila: sexual dimorphism and non-allometric effects of shape variation. BMC Dev. Biol. *11*, 32.

Dietzl, G., Chen, D., Schnorrer, F., Su, K.-C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., et al. (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature *448*, 151–156.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dorer, D.R., Rudnick, J.A., Moriyama, E.N., and Christensen, A.C. (2003). A Family of Genes Clustered at the Triplo-lethal Locus of Drosophila melanogaster Has an Unusual Evolutionary History and Significant Synteny With Anopheles gambiae. Genetics *165*, 613–621.

Han, B.W., Wang, W., Zamore, P.D., and Weng, Z. (2015). piPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. Bioinforma. Oxf. Engl. *31*, 593–595.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., et al. (2009). Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. Cell *137*, 509–521.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*.

Mével-Ninio, M., Pelisson, A., Kinder, J., Campos, A.R., and Bucheton, A. (2007). The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. Genetics *175*, 1615–1624.

Miesen, P., Girardi, E., and van Rij, R.P. (2015). Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. Nucleic Acids Res. gkv590.

Miesen, P., Ivens, A., Buck, A.H., and van Rij, R.P. (2016). Small RNA Profiling in Dengue Virus 2-Infected Aedes Mosquito Cells Reveals Viral piRNAs and Novel Host miRNAs. PLoS Negl. Trop. Dis. *10*.

Page, S.L., McKim, K.S., Deneen, B., Van Hook, T.L., and Hawley, R.S. (2000). Genetic studies of mei-P26 reveal a link between the processes that control germ cell proliferation in both sexes and those that control meiotic exchange in Drosophila. Genetics *155*, 1757–1772.

Peleg, J. (1968). Growth of arboviruses in primary tissue culture of Aedes aegypti embryos. Am. J. Trop. Med. Hyg. *17*, 219–223.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. *41*, D590–D596.

Roberts, A., and Pachter, L. (2012). Streaming fragment assignment for real-time analysis of sequencing experiments. Nat. Methods *10*, 71–73.

Shah, N., Dorer, D.R., Moriyama, E.N., and Christensen, A.C. (2012). Evolution of a Large, Conserved, and Syntenic Gene Family in Insects. G3 GenesGenomesGenetics *2*, 313–319.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. *31*, 46–53.

Vodovar, N., Bronkhorst, A.W., van Cleef, K.W.R., Miesen, P., Blanc, H., van Rij, R.P., and Saleh, M.-C. (2012). Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. PloS One *7*, e30861.

4.6 Figures and tables

Primer	Sequence
Real RPS7 For	TCAGTGTACAAGAAGCTGACCGGA
Real RPS7 Rev	TTCCGCGCGCGCTCACTTATTAGATT
A3 qPCR r4 F	CGAAGCAGAAGAGCAACTCC
A3 qPCR r4 R	TTCGTACTCGGAGCACATTC
AeB1 Primer F	AGGAAGAGTTGCCCAGTACC
AeB1 Primer R	CGCATGCTTTCATTTGCCAG
Lian Primer F	GAAGAAGCATGTCCTCTGCG
Lian Primer R	AGCCGACTGACTTCACTCAA
Muta Primer F	GAAGGATTACGACGTTTAGCCC
Muta Primer R	CTAGATCGATCCCCAGCTCATT
Gecko F	GACGGACCTGGTGTAGTGG
Gecko R	GCTAGTTCATCTCGGGACCA
Pogo 12 F	TTTGTCATACACTCGCCCGA
Pogo 12 R	TCCAACCGCCTTTCCATAGT
Copia 174 F	TTTGTTTGCTTCGCGGATGA
Copia 174 R	CAGCCGTTAACAAGGATCCG

Table 4.1: Primers used for TE qPCR

	Total Reads	Total Reads (RPM): rtTA	Total Reads (RPM): EGFP
M14 Male Induced	38.6 million	836 (21.7)	9534 (247)
M14 Male Uninduced	53.3 million	29 (0.54)	1781 (33.41)
M14 Female Induced	94.3 million	1 (0.01)	18 (0.19)
M14 Female Uninduced	78.4 million	0 (0)	10 (0.13)
Orlando Male	142.3 million	0 (0)	0 (0)
Orlando Female	185.0 million	0 (0)	54 (0.29)

Table 4.2: Reads mapping to transformation donor plasmid genes in M14 male and female animals, as well as wild type Orlando animals.

Gene ID	Matching Method	Drosophila Match	Function/Domain	Female Induced FPKM	Female Control FPKM	q value
AAEL000694	Blast	CG13972	IQ Domain	0	0.827246	0.0375533
AAEL000725	VB Orthologue	CG17717	Zinc Finger Protein	0	0.668659	0.0375533
AAEL001911	Blast	CG31816	Unknown	0	1.77241	0.0375533
AAEL002052	VB Orthologue	CG42360	Ribonuclease like	0	0.581085	0.0375533
AAEL002096	VB Orthologue	CG18675	Unknown	0	1.44628	0.0375533
AAEL002275	VB Orthologue	CG3565	Kv channel interacting /Ca ion binding	0	1.11818	0.0375533
AAEL003217	VB Orthologue	CG10839	Dynein Light Chain-like	0	1.0459	0.0375533
AAEL005222	VB Orthologue	CG13526	Calmodulin-like	0	2.7033	0.0375533
AAEL005891	Blast	CG8472	Calmodulin	0	2.71795	0.0375533
AAEL009963	VB Orthologue	CG12681	Unknown	0	1.07961	0.0375533
AAEL010741	Blast	CG34181	Unknown	0	0.678091	0.0375533
AAEL012937	n/a	None	Unknown	0	2.93991	0.0375533
AAEL013918	VB Orthologue	CG18817	Tetraspanin	0	0.74093	0.0375533
AAEL013951	VB Orthologue	CG5107	Nucleosome assembly	0	0.556151	0.0375533
AAEL017390	Blast	None	Protein disulfide isomerase	0	2.41491	0.0375533

Table 4.3: Significantly differentially (q-value less than 0.05) expressed genes between M14 female induced larvae and control female larvae, as called by Cuffdiff.

Gene ID	Matching Method	Drosophila Match	Function/Domain	Male FPKM	Female FPKM	log2(fold_change)	q_value
AAEL014433	VB Orthologue	CG1151	Osiris 6	0.460845	629.311	10.4153	0.00624531
AAEL004915	n/a	None	Unknown	2.24151	2516.13	10.1325	0.00624531
AAEL004280	VB Orthologue	CG15592	Putative Osiris Imaginal wing disk	0.632378	668.641	10.0462	0.00624531
AAEL002962	VB Orthologue	CG1155	Osiris 14	1.07697	215.167	7.64233	0.00624531
AAEL012148	VB Orthologue	CG4702	Unknown	1.49274	167.012	6.80584	0.00624531
AAEL005416	VB Orthologue	CG42331	oxidase/peroxidase	1.076	64.7242	5.91056	0.00624531
AAEL014369	VB Orthologue	CG8927	Cuticle	0.481026	26.9927	5.81031	0.00624531
AAEL006259	VB Orthologue	CG5192	Opsin	0.742494	39.2414	5.72386	0.00624531
AAEL004564	VB Orthologue	CG8420	Unknown	1.57938	80.2304	5.66672	0.00624531
AAEL009186	VB Orthologue	CG12107	Unknown	0.438633	21.9421	5.64454	0.00624531
AAEL005960	VB Orthologue	CG2150	Unknown	0.523889	27.9294	5.73638	0.0444111

Table 4.4: Significantly differentially (q-value less than 0.05) expressed genes between male and female 4th instar larvae, *Ae. aegypti* Orlando strain, as called by Cuffdiff.

Gene ID	Matching Method	Drosophila Match	Function/Domain	Induced Male FPKM	Control Male FPKM	log2(fold_change)	q_value
AAEL000166	VB Orthologue	CG17162	Unknown	0	5.17849	inf	0.0443417
AAEL001692	n/a	None	Unknown	4026.47	219.693	-4.19595	0.0443417
AAEL003289	Blast	CG1809/10592	Alkaline phosphotase Fertile, non-lethal	59.029	4.09551	-3.84931	0.0443417
AAEL004280	VB Orthologue	CG15592	Putative Osiris Imaginal wing disk	131.488	3.34688	-3.24908	0.0443417
AAEL011808	VB Orthologue	CG6142	Glucose dehdyrogenase Fertile, non-lethal	0	2.98285	inf	0.0443417
AAEL014433	VB Orthologue	CG1151	Osiris 6	93.3085	1.80968	-3.14734	0.0443417
AAEL017517	n/a	None	Unknown	0	9.52471	inf	0.0443417

Table 4.5: Significantly differentially (q-value less than 0.05) expressed genes between M14 male induced larvae and control male larvae, as called by Cuffdiff.

	baseMean	log2FoldChange	lfcSE	Wald stat	pvalue	padj
Gypsy-6_AA-I	82.12715	4.70406	1.19006	3.95281	0.00008	0.03772
Gypsy-269_AA-LTR	25.75137	5.76254	1.50037	3.84075	0.00012	0.03772
L1-46_AAe	194148.09606	4.25881	1.14486	3.71992	0.00020	0.04085
Gypsy-6_AA-LTR	36.11605	4.41253	1.22749	3.59476	0.00032	0.04845
Copia-133_AA-LTR	10.17287	-5.24491	1.47988	-3.54416	0.00039	0.04845
Sola1-N9C_AAe	5.93172	5.08265	1.47843	3.43787	0.00059	0.05151
LOA_Ele6B_AAe	158.58868	4.05542	1.16682	3.47561	0.00051	0.05151
Gypsy-210_AA-LTR	6.14098	4.99999	1.50868	3.31416	0.00092	0.07066
Crack-21_AAe	128.78597	3.52139	1.07375	3.27953	0.00104	0.07105
I-12_AAe	47.63334	3.51741	1.08702	3.23582	0.00121	0.07459
L1-37_AAe	1787.10624	3.20288	1.01309	3.16151	0.00157	0.08775
piggyBac-1_AAe	10.09304	4.21299	1.36700	3.08192	0.00206	0.09732
I_Ele39	23.83370	3.30883	1.07365	3.08186	0.00206	0.09732
BEL-90_AA-LTR	10.25840	4.72868	1.55849	3.03415	0.00241	0.10596
Gypsy-85_AA-I	8.30818	4.08044	1.36811	2.98253	0.00286	0.10988
BEL-648_AA-I	10.23314	4.18393	1.39510	2.99901	0.00271	0.10988
Gypsy-595_AA-I	12.59889	4.02390	1.36823	2.94095	0.00327	0.11837
Kiri-10_AAe	11.12711	3.76710	1.32974	2.83296	0.00461	0.14182
Gypsy-55_AA-I	14.93287	3.93783	1.38089	2.85167	0.00435	0.14182
Gypsy-259_AA-I	474.38395	2.77060	0.97488	2.84198	0.00448	0.14182
Crack-1_AAe	10.95738	3.54361	1.28893	2.74927	0.00597	0.17492
BEL-651_AA-I	12.21193	3.91478	1.43923	2.72006	0.00653	0.18246

Table 4.6: DESeq2 called significantly differentially expressed transposons with a p-adjusted value of less than 20% from a mRNA-seq library. Test condition was doxycycline induction of Ago3 knockdown in 4th instar male larvae.

	Base Mean	log2FoldChange	lfcSE	stat	pvalue	padj
Gypsy-278_AA-LTR	1289.1637056997	-1.1362886201	0.238656988	-4.7611789186	1.92E-006	0.0005985671
Gypsy-135_AA-I	49.565030251	2.036190651	0.4459690245	4.5657669912	4.98E-006	0.0007738789
Gypsy-46_AA-LTR	1316.7123764942	-0.9611526871	0.24251218	-3.9633171711	7.39E-005	0.0057469282
Gypsy-259_AA-I	105.7176195899	1.2588027956	0.3171378609	3.9692605355	7.21E-005	0.0057469282
Sola1-N10_AAe	49.5533140892	1.3102813345	0.3578280849	3.6617621414	0.0002504864	0.0146573754
Gypsy-127_AA-I	50.6162562276	-1.2862744122	0.3574089535	-3.5988869324	0.0003195821	0.0146573754
BEL-1N_AA-LTR	65.3858765163	-1.3931474832	0.3879977637	-3.5906069918	0.0003299088	0.0146573754
Gypsy-232_AA-LTR	521.8997136603	-0.9193225847	0.2612301205	-3.5192059135	0.0004328406	0.014957048
BEL-10_AA-LTR	131.1738264215	-1.5231758853	0.4316139556	-3.5290237154	0.0004170958	0.014957048
Mariner-N1_AAe	167.6944893521	1.2295368197	0.3579267191	3.435163552	0.0005921963	0.0184173051
Gypsy-139_AA-I	1595.968282291	-0.9257485048	0.275167839	-3.3643048849	0.0007673672	0.021695564
BEL-616_AA-I	139.538809797	-1.3470392979	0.4179879288	-3.2226751181	0.0012699951	0.0329140393
Gypsy-232_AA-I	2836.8633098495	-0.7950724246	0.2495730716	-3.1857300129	0.0014438928	0.0345423573
Gypsy-607_AA-I	1342.0793530981	-0.774746458	0.2454697887	-3.1561784537	0.0015985103	0.035509764
CryptonI-1_AA	234.1043994461	1.15168238	0.369265471	3.1188466578	0.0018156042	0.037643528
BEL-155_AA-I	1182.713268428	-0.7841314397	0.2531603235	-3.0973709819	0.0019524536	0.0379508177
Sola2-N2_AAe	237.320475023	1.1147633696	0.3839827053	2.903160362	0.0036941742	0.041031721
Gypsy-614_AA-I	477.4768568998	-0.7476826456	0.2496539727	-2.9948758177	0.002745566	0.041031721
Gypsy-607_AA-LTR	220.4747027356	-1.0776705127	0.3640710199	-2.9600557411	0.0030758339	0.041031721
Gypsy-588_AA-I	1457.4119090242	-0.8969670357	0.3067140962	-2.9244402096	0.0034507636	0.041031721
Gypsy-46_AA-I	2447.9430523484	-0.8569246091	0.293971599	-2.9149911483	0.0035569863	0.041031721
Gypsy-285_AA-I	79.7997589995	0.9892947225	0.3339722583	2.9622062851	0.0030544307	0.041031721
Gypsy-274_AA-LTR	375.6504442699	-0.9909919371	0.326995687	-3.0305963549	0.002440713	0.041031721
Gypsy-233_AA-I	846.8423825651	-0.8556506651	0.2899298936	-2.9512329845	0.0031650807	0.041031721

Table 4.7: DESeq2 called significantly differentially expressed transposon-mapping sRNAs with a p-adjusted value of less than 5% from a sRNA-seq library. Test condition was doxycycline induction of Ago3 knockdown in 4th instar male larvae.

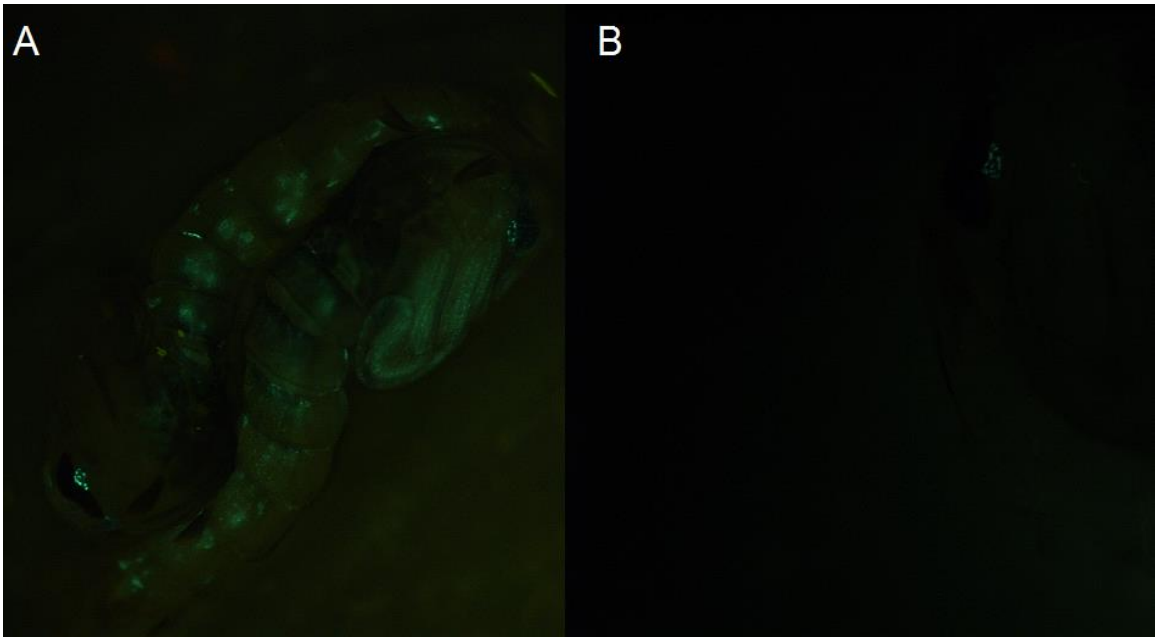

Figure 4.1: M14 male GFP phenotypes. (A) GFP in the eye, abdomen (GE/AP), and developing legs (GE/LP). (B) GFP in the eye only.

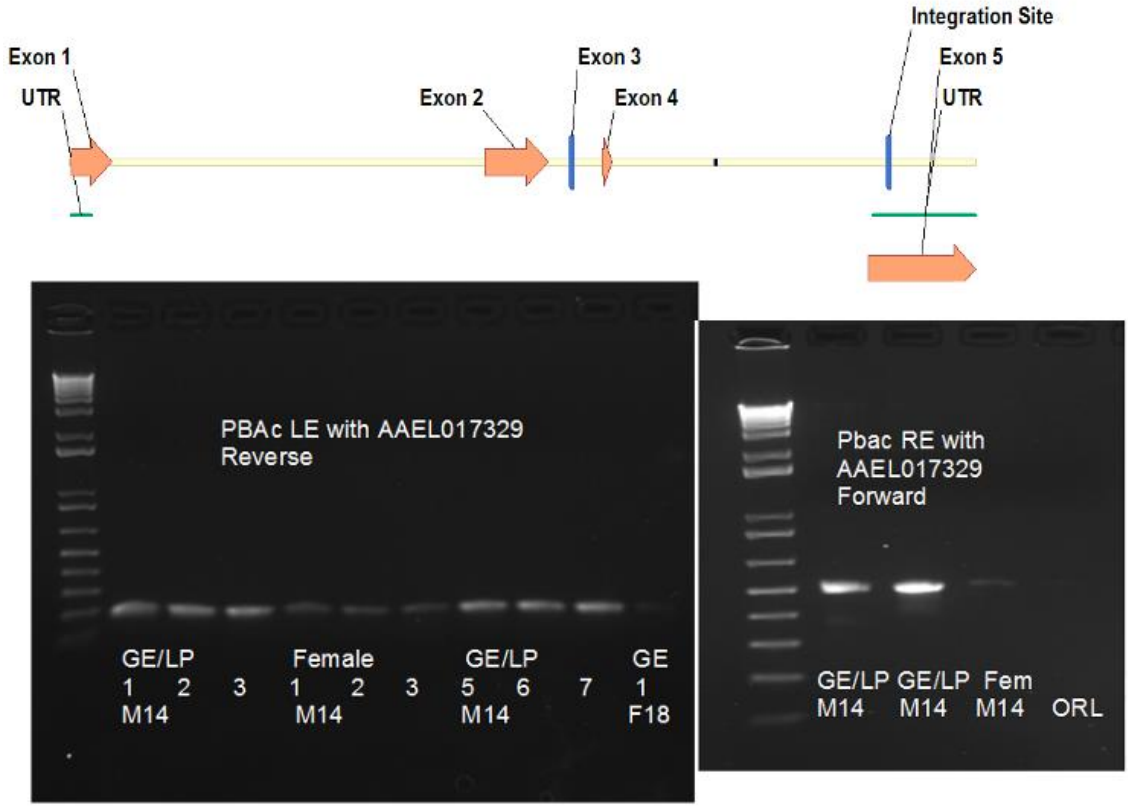

Figure 4.2A: Integration into AAEL0017329 3' UTR and PCR verification of integration from both transposon ends. Both male M14 mosquitoes with transgenic eye marker phenotypes (GE/LP) and female mosquitoes without eye marker phenotypes contain a pBac integration in exon 5 of AAEL017329. Wild type and F18 mosquitoes do not.

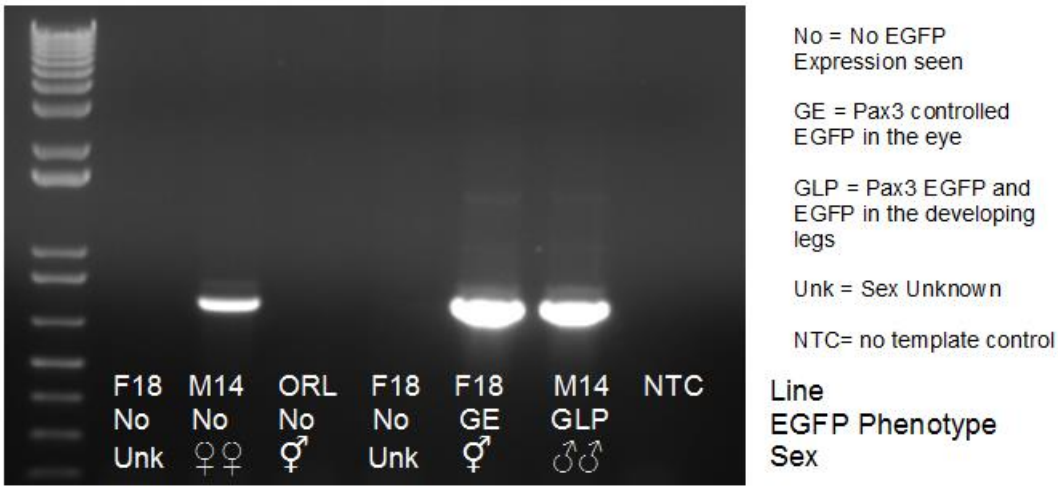

Figure 4.2B: EGFP PCR to test for the presence of the EGFP gene. M14 males contain the EGFP gene. M14 females who have lost observable EGFP phenotype also contain the EGFP gene. Orlando and F18 females who do not express observable EGFP do not contain the EGFP gene.

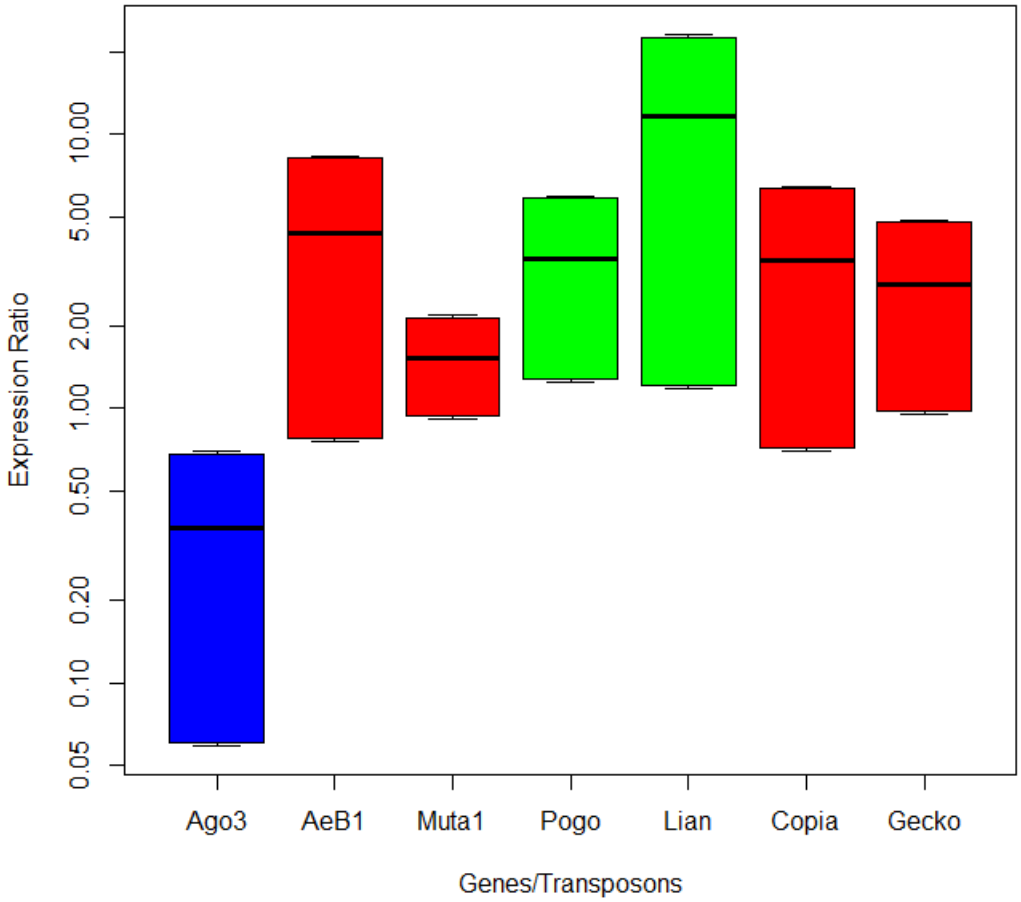

Figure 4.3: Expression levels of Ago3 and six transposons in induced M14 male larvae relative to expression levels in uninduced M14 male larvae. Y-axis is on a log-scale. Ago3 expression is down; Pogo and Lian expression is up. AeB1, Muta1, Lian, and Copia were not significantly upregulated in induced M14 males compared to uninduced males.

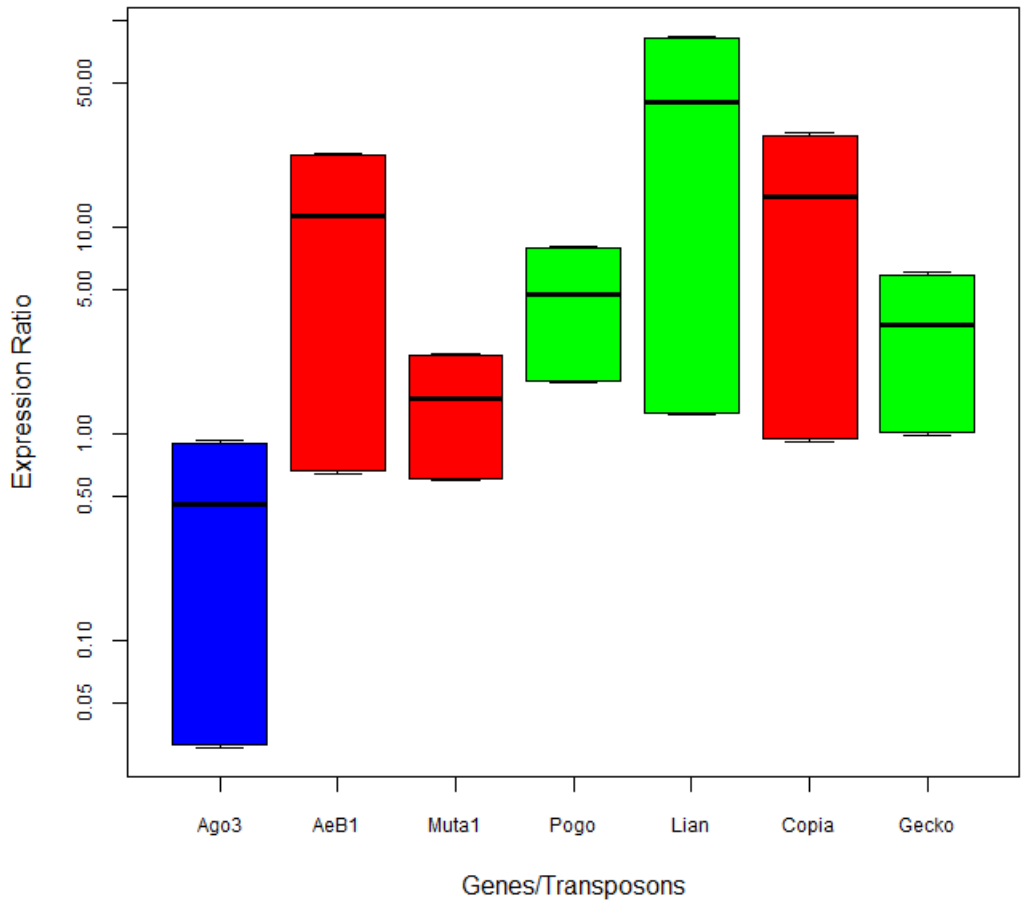

Figure 4.4: Expression levels of Ago3 and six transposons in induced M14 male larvae relative to expression levels in uninduced Orlando wild type male larvae. Y-axis is on a log-scale. Ago3 expression is down; Pogo, Lian, and Gecko expression is up. AeB1, Muta1, and Copia were not significantly upregulated in induced M14 males compared to Orlando wild type males.

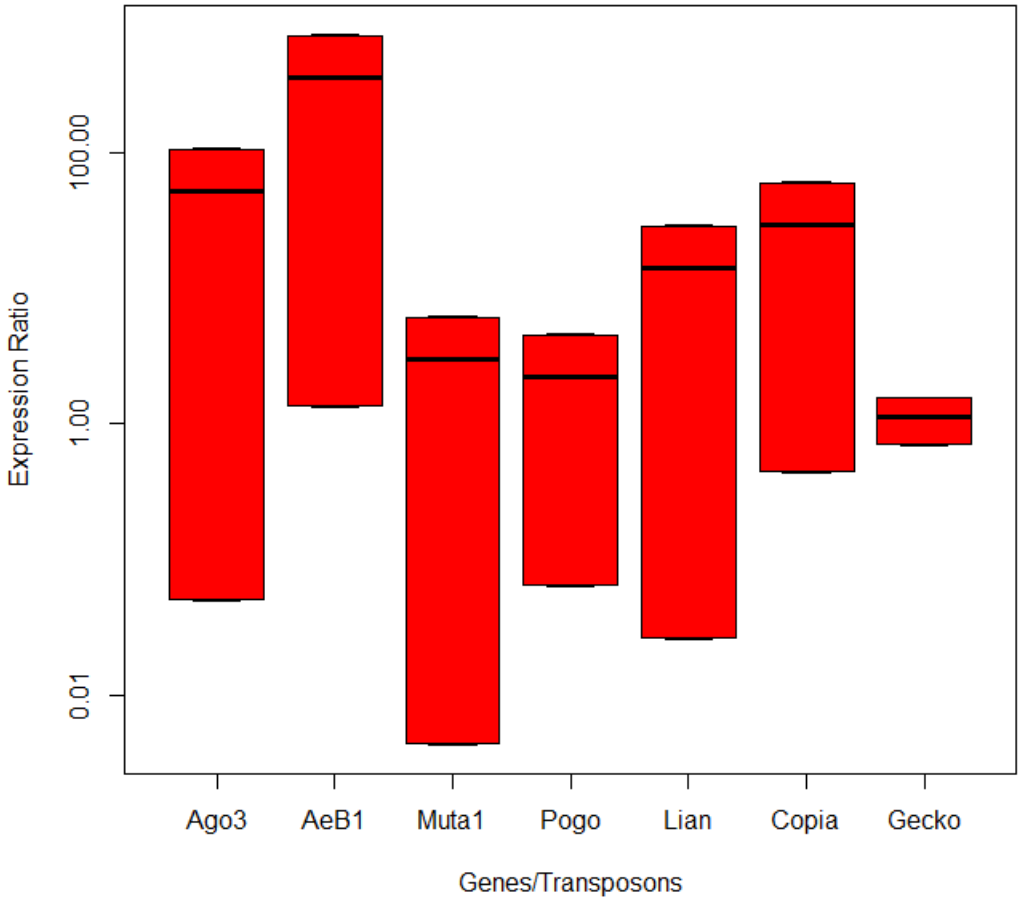

Figure 4.5: Expression levels of Ago3 and six transposons in induced Orlando wild type male larvae relative to expression levels in uninduced Orlando wild type male larvae. Y-axis is on a log-scale. No genes or transposons were significantly differentially expressed.

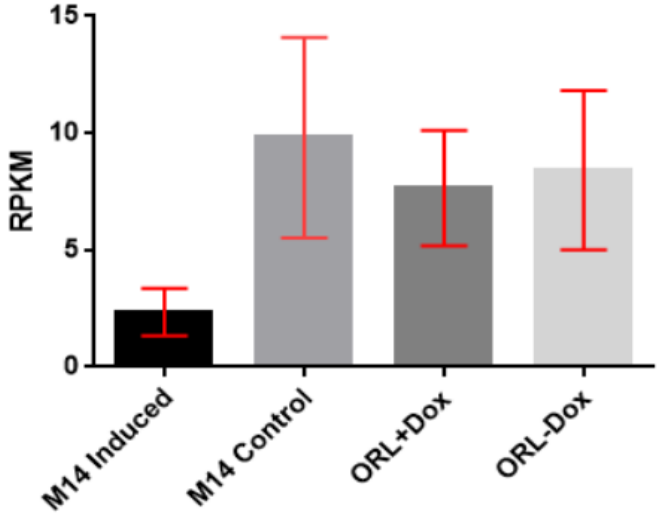

Figure 4.6: Ago 3 knockdown in male mosquitoes calculated from mRNA-seq RPKM data.

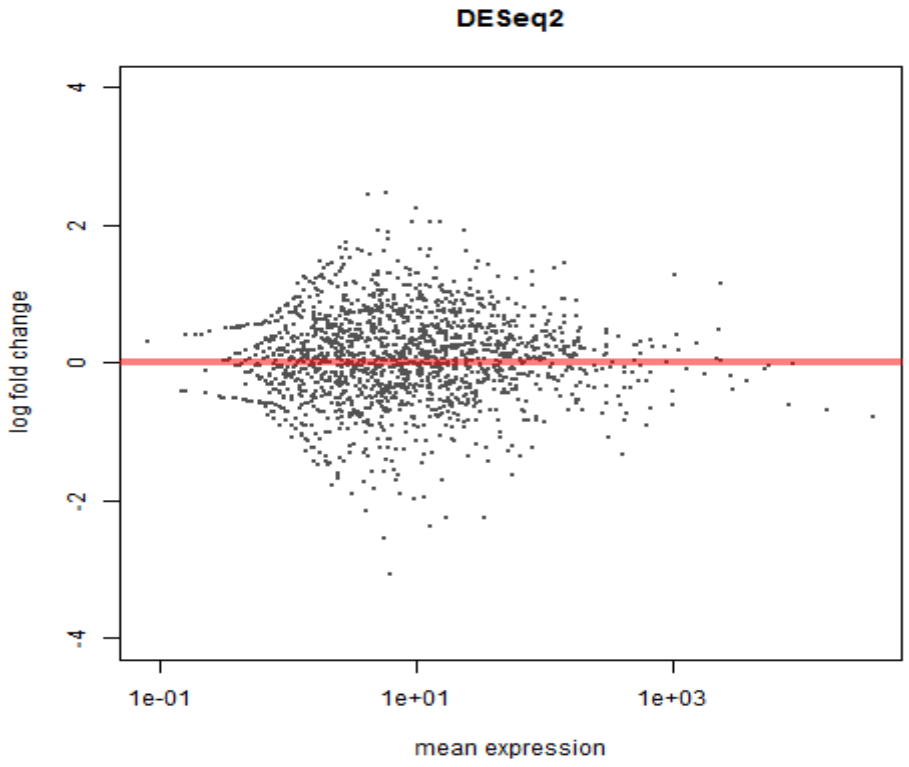

Figure 4.7: Log fold change and mean expression of *Ae. aegypti* transposons upon induction in female M14 larvae. Test condition was doxycyline induction in 4th instar female larvae. Any significant changes in transposon expression (p-adjusted less than 10%) are represented in red.

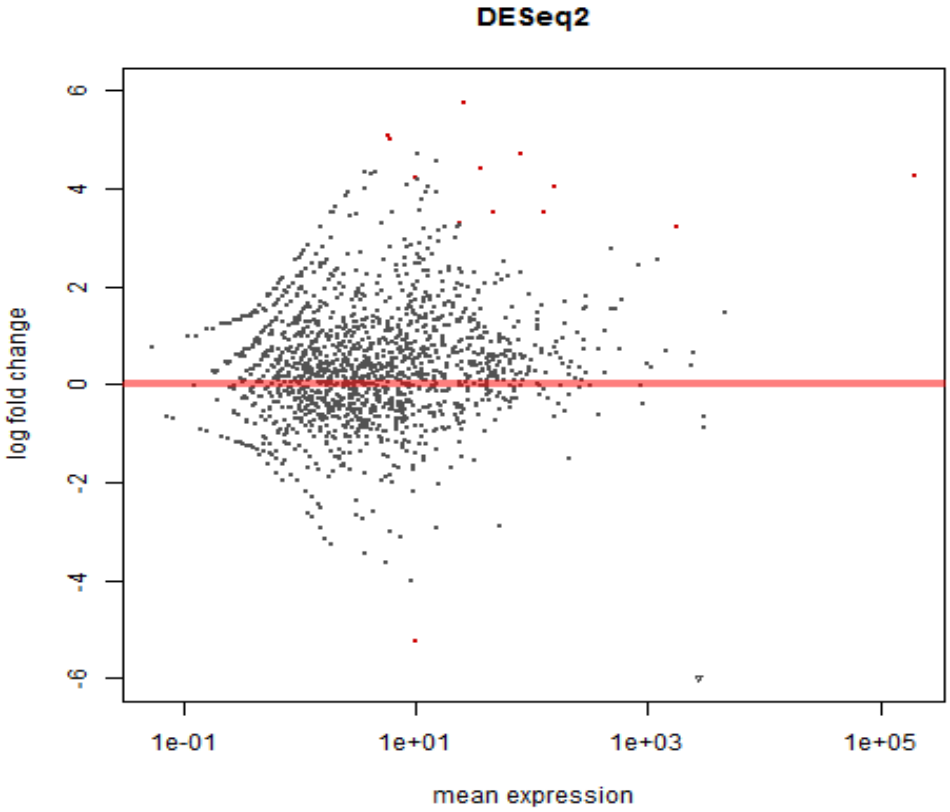

Figure 4.8: Log fold change and mean expression of *Ae. aegypti* transposons upon induction in male M14 larvae. Test condition was doxycyline induction of Ago3 knockdown in 4th instar male larvae. Any significant changes in transposon expression (p-adjusted less than 10%) are represented in red.

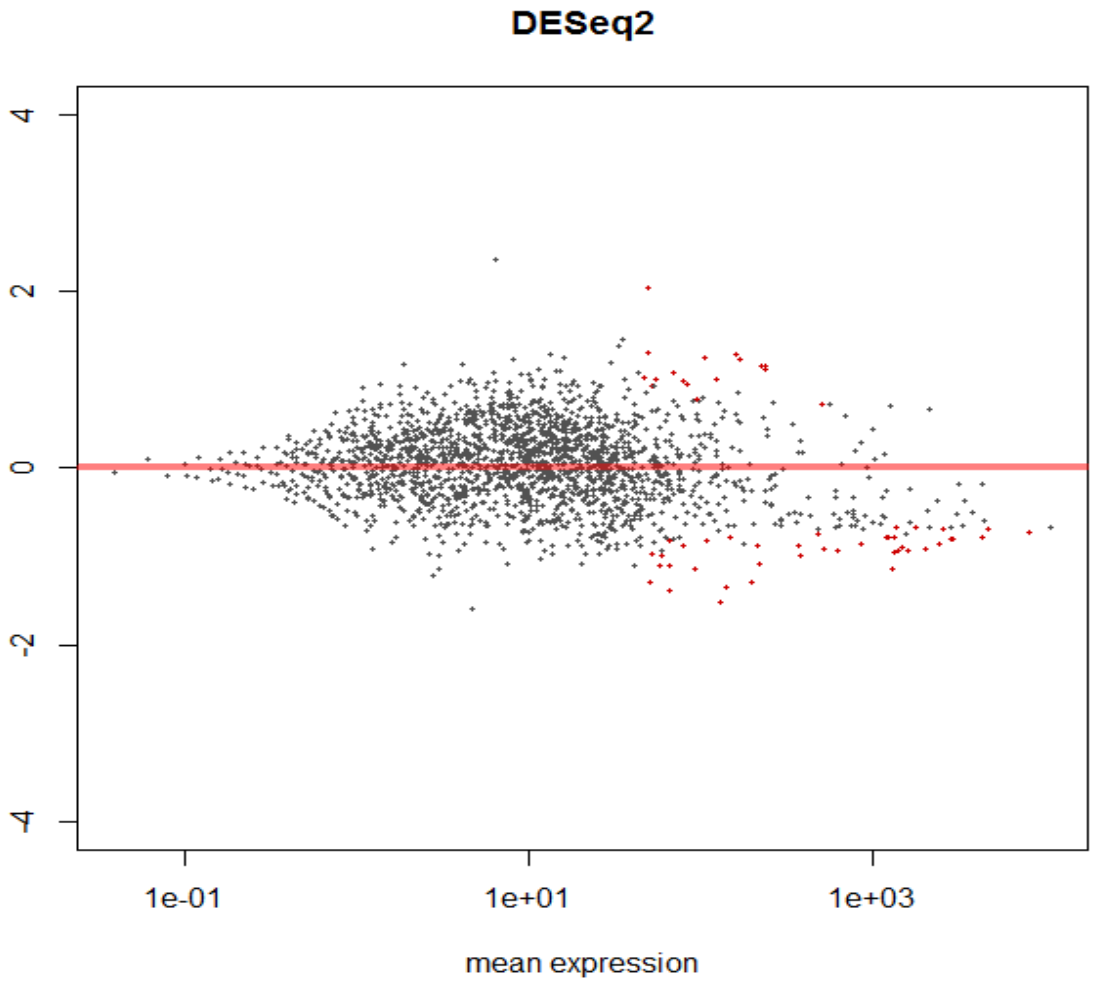

Figure 4.9: Log fold change and mean expression of sRNA populations mapping to *Ae. aegypti* transposons upon RNAi induction in male M14 larvae. Test condition was doxycyline induction of Ago3 knockdown in 4th instar male larvae; control was uninduced larvae. Any significant changes in mapping sRNA popluations (p-adjusted less than 10%) are represented in red.