Mosquito Transposable Elements and *piwi* Genes

Mónica A. Álvarez Morales

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Masters
in
Biochemistry

Dr. Zhijian (Jake) Tu
Dr. Erin Dolan
Dr. Michael Klemba

May 16, 2008
Blacksburg VA

Keywords: Transposable elements, RNAi, *Belly*, *piwi*

Abstract

Vector control is an essential and effective approach for controlling transmission of vector-borne diseases. However, increasing resistance to insecticide and drugs suggests that new strategies to control vector-borne diseases are needed. One possible strategy involves replacing mosquito populations with disease-resistant transgenic mosquitoes. Transposable elements (TEs) are an important component in this new strategy due to their ability to integrate exogenous DNA into chromosomes. They could potentially be useful tools in assisting the spread of disease-resistant genes in mosquito populations.

This research focuses on two related subjects, TEs and their regulation. The first subject is on a Long Terminal Repeat (LTR) retrotransposon in the African malaria mosquito, *Anopheles gambiae*, namely *Belly*. The second subject focuses on the characterization of *piwi* genes in the dengue and yellow fever mosquito, *Aedes aegypti*.

For the first subject we characterized *Belly* by identifying the two identical LTRs and one intact open reading frame. We also defined the target site duplications and boundaries of the full-length *Belly* element. This novel retrotransposon has nine full-length copies in the *An. gambiae* sequenced genome and their nucleotide similarity suggests that there has been fairly recent retrotransposon. We have shown that *Belly* is transcribed and translated in *An. gambiae*. Single LTR circles were recovered from *An. gambiae* cells, which is consistent with active transposition of *Belly*.

The second subject focuses on the *piwi* genes of *Ae. aegypti*. We found nine potential *piwi* genes in *Ae. aegypti* and two in *An. gambiae*. Phylogenetic analysis suggests that these *piwis* formed two subgroups and gene duplication within each group occurred after the divergence between the two mosquito species. RT-PCR and transcriptome analysis showed *Ago3* as well as all the seven tested *piwi* genes were expressed either in germline tissues or developing embryos. Differential expression patterns were observed. While most *piwis* were transcribed in the ovaries, testis, and embryos, two *piwis* appear to have a zygotic expression. Three *piwi* genes (*piwi* 3, *piwi* 4, and *Ago3*) were also detected in adult somatic tissues of *Ae. aegypti*. The expansion of the number of *piwi* genes in *Ae. aegypti* compared to *An. gambiae* and *D. melanogaster* may be correlated with a larger genome size and greater amount of TEs. The finding of *piwi* expression in adult somatic tissues is intriguing. It is possible that these *piwi* genes were expressed in the adult stem cells. It is also possible that they may be involved with anti-viral defense. Both of these hypotheses require further testing.

Para Ranj, Lola y Valentina

Acknowledgements

I would like to give a special thanks to my advisor, Dr. Zhijian (Jake) Tu for his support, advice and endless patience. Jake took me in to his lab before I had even been accepted to Graduate School. During this period he supported and encouraged me to further my studies and for this I will be forever grateful. He and his lab have been a positive experience and one that I won't soon forget. I am also very thankful to my lab mates: Dr. Monique Coy, Dr. Senay Sengul, Dr. Jim Biedler, Ray Miller, Randy Saunders, Dr. Yumin Qi, and Andy Mead. I have to thank Dr. Monique Coy for always helping me with my project and for her good sense of humor, time passed faster when she was around. I also have to thank Dr. Senay Sengul for her constant support and hugs. Not only did she teach me a lot of things in the lab but she also was a good confidant of mine. I have to thank Dr. Jim Biedler for our endless discussions on my project and providing a valuable source of ideas. He also has a great sense of humor, which made the lab environment enjoyable. I would also like to thank Ray Miller for his help regardless of his own workload. To Randy Saunder I have to give special thanks for always providing me with all the mosquitoes that I needed and for the endless patience he had with me. To Dr. Yumin Qi I have to thank for teaching me most of the techniques for the *Belly* project. And last but certainly not least, I would like to thank Andy Mead for providing me with good ideas on my project and for always picking the packages for the lab, which is quite a tedious task.

I would like to thank my committee Dr. Michael Klemba and Dr. Erin Dolan for being a valuable source of ideas and support. They have always been

accessible and willing to help me with advice and guidance. My sincere gratitude is expressed to both of them.

I would also like to give a big special thanks to Dr. Ed Smith, who is the reason I was able to get into Graduate School at Virginia Tech. He accepted me in the VT PREP program and gave me the tools and resources to get into the Biochemistry Department. Dr. Smith not only cared for me on an academic level but also on a personal level. I will also be forever grateful to you for all you help and understanding.

Last but not least I also would also like to thank my beloved husband Ranjit Randhawa, my parents Julio M. Alvarez and Eileen Morales, and my grandparents Carmen Coira, Nerys Alvarez and Julio Alvarez for giving me unconditional support and love always. I will always love you guys.

Table of Contents

List of Figures

List of Tables

Chapter One

Introduction

1.1 Transposable elements and their genomic impact

Transposable Elements (TEs) are mobile DNA segments that can change positions, or transpose, within the genome. These mobile sequences were first discovered in maize in the mid 1940´s by Barbara McClintock (McClintock 1951), for which she won a Nobel prize in 1983. TEs are present in all three domains of life (Craig 2002). They are abundant in eukaryotic genomes and are thought to play important roles in genome evolution.

TEs can have an influence on fitness and evolution of their host organism, by inserting themselves into coding sequences and causing chromosomal breaks, deletions, and translocations. TEs can also contribute to the variation of genome size within related species and thus contribute to the C-value paradox, which describes the lack of correlation between the biological complexity of an organism and the size of its genome. For example, 16% of the *Anopheles gambiae* genome (Holt et al. 2002), and 50% of the *Aedes aegypti* genome (Nene et al. 2007) are TE-derived sequences. It is evident that there are significant differences within mosquitoes since the *Ae. aegypti* genome is approximately 1300Mb and the *An. gambiae* genome is approximately ~5x smaller at 280Mb.

TEs were often called “junk” DNA as they were thought to have little or no function. Recent evidence has challenged this concept (Brosius 1999;

Makalowski 2000; Kidwell and Lisch 2001). Some TEs have been shown to have functions such as directing the assembly of chromatin structures and altering gene expression (Lerman et al. 2003; Fontanillas et al. 2007). However, most copies of TEs are inactive (unable to transpose) since they decay over time due to accumulated mutations and/or a slow amplification rate.

There are two main categories of TEs, Class I and Class II. Class I elements transpose via an RNA-mediated mechanism similar to retroviruses and are therefore sometimes referred to as retrovirus-like elements or retrotransposons. They synthesize cDNA molecules by reverse transcription and then integrate these cDNAs into the genome. The RNA-mediated TEs contain three subclasses, which include long terminal repeat (LTR) retrotransposons, non-LTR retrotransposons, and short interspersed elements (SINEs). These TEs contain similarities to retroviruses in their structures and protein sequences. They also contain transcriptional regulatory sites, which flank the open reading frame(s) that are important for their transposition. In Class II elements, or transposons, transposition occurs directly from DNA to DNA through a “cut and paste” or a “rolling circle” mechanism (Craig 2002). In these DNA-mediated TEs there are also subclasses which include “cut and paste” DNA transposons, miniature inverted-repeat TEs (MITEs), and *Helitrons* (Kapitonov and Jurka 2001; Kapitonov and Jurka 2003). They generally have inverted terminal repeats (TIRs) of 10-200 bp long that flank the coding region for transposase. Transposase is an enzyme that binds the TIRs of the transposon as well as the target DNA site mediating the transposition. Transposase is also responsible for joining the

transposon to its new insertion site.

In order for a TE to remain active in a population a functional copy needs to transpose in the germline, and therefore pass onto the next generation. However, an excess of transposition events can be detrimental to the host and therefore nature has evolved regulatory mechanisms to limit or suppress TE activity. There is evidence that transposition events by active TE copies are under tight control. There are two hypotheses on how TEs might be regulated, one of them is a TE autoregulatory mechanism, where the rate of transposition decreases when the copy number increases. This self-regulatory mechanism has been documented for P and *marine*r elements in *D. melanogaster* (Hartl et al. 1997; Kidwell and Lisch 2001). Autoregulation of TEs, includes insertion into heterochromatin and tissue-specific regulation such as alternative splicing (Hartl et al. 1997; Kidwell and Lisch 2001). The second and most recent hypothesis on TE regulation involves the repeat associated short interference RNA (rasiRNA) or piwi interacting RNA (piRNA) (Finnegan et al. 2001; Vastenhouw and Plasterk 2004), discussed in detail in the next section. Nevertheless, TEs have been able to thrive despite all the mechanisms that have evolved to silence them and thus understanding the dynamics of TE regulation is necessary.

1.2 RNA interference (RNAi)

The RNAi pathways are evolutionarily conserved control mechanisms that use small RNA molecules to inhibit gene expression through mRNA degradation, translational repression, or chromatin modification (Bernstein and Allis 2005;

Filipowicz et al. 2005; Zamore and Haley 2005). They are thought to be involved in protecting the genome from viruses and from random integration of TEs (Slotkin and Martienssen 2007). Such small RNA molecules include small interfering RNAs (siRNAs), microRNAs (miRNAs), and rasiRNAs or piRNAs.

siRNAs and miRNAs are generated through processing of double stranded RNA (dsRNA) and hairpin precursors, respectively by an RNase III enzyme. In flies, the long dsRNA is processed by Dicer-2 (Dcr-2) to yield siRNA (21 nt) duplexes. In the miRNA pathway the first step is different since there is a primary miRNA (pri-miRNA), which is cleaved by Drosha in the nucleus to produce precursor miRNA (pre-miRNA). The pre-miRNA is then exported to the cytoplasm where it is furthered processed by Dicer-1 (Dcr-1) to yield miRNA (22 nt) duplexes. The formation of small RNAs is sequentially catalyzed by dsRNA-specific endonucleases generating single stranded RNA (ssRNA) duplexes. These small sequences are subsequently loaded into a complex named, RNA induced silencing complex (RISC), which contains a member of the Argonaute (Ago) protein family, leading ultimately to gene expression inhibition (See Figure 1 and 2).

The presence of a member of the Argonaute gene family is crucial for RNAi-mediated silencing. These proteins provide the biochemical activity and the protein/protein interaction while the small RNAs bind to the nucleic acid targets. The Argonaute protein family is highly conserved, and contains two characteristic domains, a PAZ domain at the amino terminus and a PIWI domain at the carboxy terminus. The PAZ domain is approximately 110 amino acids and forms a nucleic

acid-binding pocket essential for small RNA binding. The PIWI domain is approximately 300 amino acids in length and shares structural similarities with ribonucleases. This family of proteins is subdivided into two subfamilies, the Ago subfamily and the Piwi subfamily. The Ago subfamily is involved in transcriptional and post-transcriptional silencing mechanisms while the Piwi subfamily has only been identified to interact with a subclass of small RNAs, called piRNAs in vertebrates and rasiRNA in invertebrates, which were reported to be present in the germline (Aravin et al. 2006; Girard et al. 2006; Lau et al. 2006). In *D. melanogaster* the Argonaute family is composed of 5 proteins: Argonaute 1 (Ago1), Argonaute 2 (Ago2), Argonaute 3 (Ago3), Aubergine (Aub), and Piwi, the last three of which belong to the Piwi subfamily. There are different numbers of Argonaute proteins among different organisms.

The rasiRNAs or piRNAs were discovered when efforts to characterize miRNAs were made by cloning total small RNAs from different developmental stages of *Drosophila* (Aravin et al. 2003). Their efforts not only yielded the 22 nt miRNA but also the 24 to 29 nt long RNAs, which corresponded to genomic repeats or transposable elements of *Drosophila*. Our limited knowledge thus far suggests that rasiRNA differ from siRNAs and miRNAs in several ways. First, the rasiRNAs are 24-31 nt long ssRNA instead of the approximately 21 nt of the miRNAs and siRNAs. Second, rasiRNAs mostly match repetitive sequences and TEs. Third, rasiRNA biogenesis appears to differ from those of siRNAs and miRNAs. In flies it was reported that the loss of Dcr-1 and Dcr-2 had no effect in the rasiRNAs levels and in TE repression (Vagin et al. 2006). The loss of these

two proteins suggests that rasiRNAs have an alternative biogenesis pathway. Lastly, unlike the siRNAs and the miRNAs, the rasiRNAs bind to members of the Piwi subfamily but not to the Ago subfamily.

This new type of small RNAs match repetitive sequences in both sense and antisense orientations (Djikeng et al. 2001; Llave et al. 2002; Reinhart and Bartel 2002) causing mRNA degradation, and/or chromatin modification (Lippman and Martienssen 2004; Matzke and Birchler 2005). However the majority of those rasiRNAs matching TEs have been shown to be in their antisense orientation (Vagin et al. 2006). It was also reported that sequences of rasiRNAs bound to piwi and aub were mostly antisense to TEs in the fly genome (Aravin et al. 2006). RasiRNAs were also found to bind to Ago3 but in the sense orientation. Furthermore, the rasiRNAs that were associated to piwi and aub had a tendency to contain a 5´uridine whereas the Ago3 associated rasiRNAs does not exhibit a 5´uridine preference. However, Ago3 rasiRNAs have a strong preference for adenine at position 10. Because of this observation Brennecke (Brennecke et al. 2007) and Gunawardane (Gunawardane et al. 2007) proposed an rasiRNA biogenesis pathway, called the “ping-pong” model (see Figure 3).

The proposed rasiRNA biogenesis pathway would require a collection of antisense rasiRNAs from an unknown initial step. These antisense rasiRNAs would then be loaded into piwi and aub, guiding them to active TE transcripts where a slicer mediated mechanism occurs between nucleotides 10 and 11 of the target RNA. This step is supposed to yield the 5´end of the sense rasiRNA. It is thought that the 3´end is processed by an unknown endonuclease, making it

accessible for loading into Ago3. The sense rasiRNA and Ago3 complex then targets complementary transcripts, where Ago3 slicing activity cleaves the 5´end of the antisense rasiRNA and an unknown endonuclease cleaves the 3´end completing the cycle. Further investigation would have to be done to support this model. These small RNAs are widely distributed among different species, such as plants (Hamilton et al. 2002; Llave et al. 2002; Mette et al. 2002), *Trypanosoma brucei* (Djikeng et al. 2001), *D. melanogaster* (Aravin et al. 2001; Pal-Bhadra et al. 2002; Aravin et al. 2003), and fission yeast (Volpe et al. 2002).

RasiRNAs have also been demonstrated to play a key role in embryonic development (Aravin et al. 2001) and to interact with only the Piwi subfamily (Aravin et al. 2006; Girard et al. 2006; Lau et al. 2006). The Piwi subfamily has been suggested in transcriptional silencing (Pal-Bhadra et al. 2002; Pal-Bhadra et al. 2004), translational repression (Grivna et al. 2006), and silencing of TEs (Aravin et al. 2001). *Piwi* genes are known to be involved in stem cell division, gametogenesis, and germline specification (Bohmert et al. 1998; Cox et al. 1998; Moussian et al. 1998; Tabara et al. 1999) in lower eukaryotes and plants. For example, in *Drosophila piwi* genes are required for the self-renewal of germline stem cells during gametogenesis and the loss of function of *piwi* results in a complete depletion of germ line stem cells in both males and females, leading to sterility of both sexes (Lin and Spradling 1997; Cox et al. 1998). On the other hand, the overexpression of *piwi* increases the number of germline stem cells and their rate of mitosis (Cox et al. 2000). In *Drosophila,* a *piwi* homologue called *aubergine,* is required for spermatogenesis (Schmidt et al. 1999) and both *piwi*

and *aub* have been reported to be involved in silencing of endogenous LTR retrotransposons testes and ovaries (Vagin et al. 2004; Vagin et al. 2006). Similarly in *C. elegans, piwi* homologs (prg-1 and prg-2) are important in germline maintenance during gametogenesis (Cox et al. 1998), while *piwi*-like genes in *Arabidopsis* (zwille and argonaute) are essential for meristem maintenance (Bohmert et al. 1998; Moussian et al. 1998).

1.3 Mosquitoes and application of TEs

Ae. aegypti and *An. gambiae* are the vectors responsible for the transmission of devastating diseases throughout the world. *Ae. aegypti* is the key vector for dengue and yellow fever. Dengue fever and dengue hemorrhagic fever are caused by one of four closely related virus serotypes (DEN-1, DEN-2, DEN-3, and DEN-4). These two diseases are primarily in tropical and sub tropical areas. Dengue occurs in over 100 countries. More than 2.5 billion people are at risk world wide and an estimated 50 million infections occur per year (WHO 2002). Yellow fever on the other hand affect an estimated 200,000 cases and causes 30,000 deaths per year (WHO 2002). The *An. gambiae* mosquito is the principal vector of malaria in Africa. This disease afflicts more than 500 million people and causes more than 1 million deaths each year in Africa (WHO 2002). Dengue, yellow fever and malaria are maintained in a cycle that involves humans and their vectors, the *Ae. aegypti* and *An. gambiae* mosquitoes. Mosquitoes are also important vectors for a number of other viruses and parasitic agents, including West Nile virus and filarial parasites. As such it is very important to study the

basic genetics of disease vectors. Understanding TE regulation will contribute to this basic genetic information and help improve transgenic tools for basic research on mosquito-pathogen interactions.

TEs are potentially a valuable resource for the development of molecular tools to deliver and drive refractory genes (genes that could make mosquito less effective vectors). They have been used for genetic manipulation, by taking advantage of their ability to integrate exogenous DNA into chromosomes. They can be used for mutagenesis, transgenesis, and population genetics. They can also be used as gene traps and as enhancer traps to find gene regulatory elements. They could also potentially be utilized as a novel strategy to control vector-borne diseases. Significant efforts are ongoing in order to use TEs to deliver and drive refractory genes as a novel strategy to control vector-borne diseases. Thus, TEs are potentially powerful tools for both basic research on mosquito genetics as well as for the control of vector-borne diseases. However, a number of difficulties need to be overcome in order for TEs to fulfill the promise.

Almost all of the TE-based transgenic vectors are DNA-mediated and only these types of elements have been used in mosquitoes (Adelman et al. 2002; Moreira et al. 2002; Robinson et al. 2004). For example, Class II TE derived from non-mosquito species have been used as vectors although their behavior in mosquitoes has not been well understood. Class II TEs *Hermes, Mos1, Minos* and *piggyBac* have been used to generate transgenic mosquitoes (O'Brochta et al. 1996; Tu and Coates 2004). In the case of the *Hermes* element, it has been shown to transpose into the germlines of several insect species, but its

transformation frequency is not nearly as efficient in *Ae. aegypti* as in *D. melanogaster* (O'Brochta et al. 1996). More importantly, currently available TE-based tools cannot be remobilized in *Ae. aegypti* once they are integrated into the germline (Tu and Coates 2004). In summary, understanding TE regulation and activity is vital in order to develop better TE tools to overcome these shortcomings.

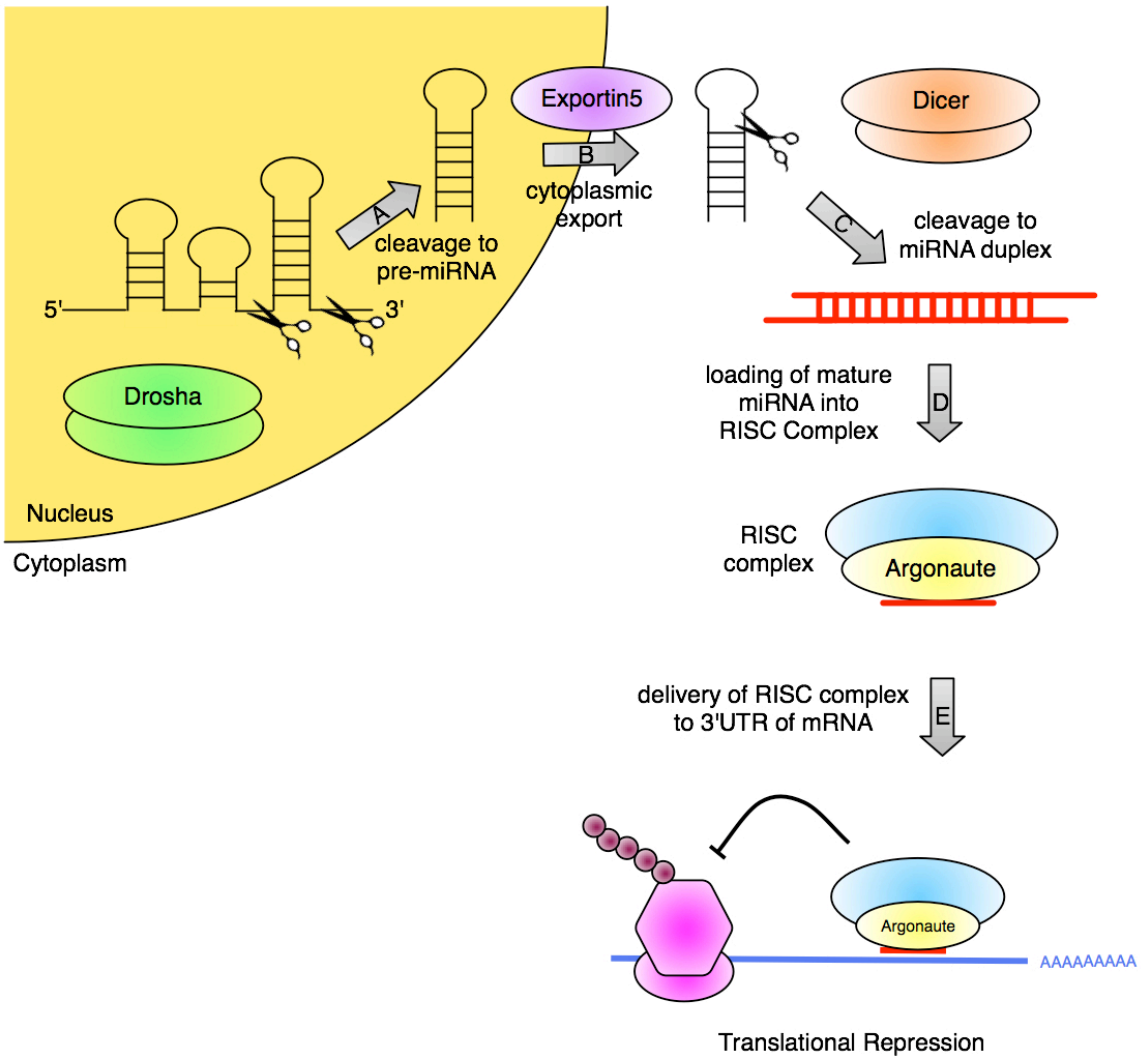


Figure 1: miRNA biogenesis pathway. The miRNA gene is first transcribed into a primary miRNA transcript (pri-miRNA). Subsequent steps include: A) Pri-miRNA is cleaved by Drosha to a pre-miRNA hairpin; B) Pre-miRNA is transported out of the nucleus by exportin-5; C) Pre-miRNA is cleaved by Dicer to form a short double-stranded miRNA duplex; D) miRNA duplex separates into single-stranded mature miRNAs and complexes with a RISC-like structure; E) miRNA/RISC complex binds the target mRNA and the mRNA is translationally repressed. The picture was adapted from de Fougerolles 2007.

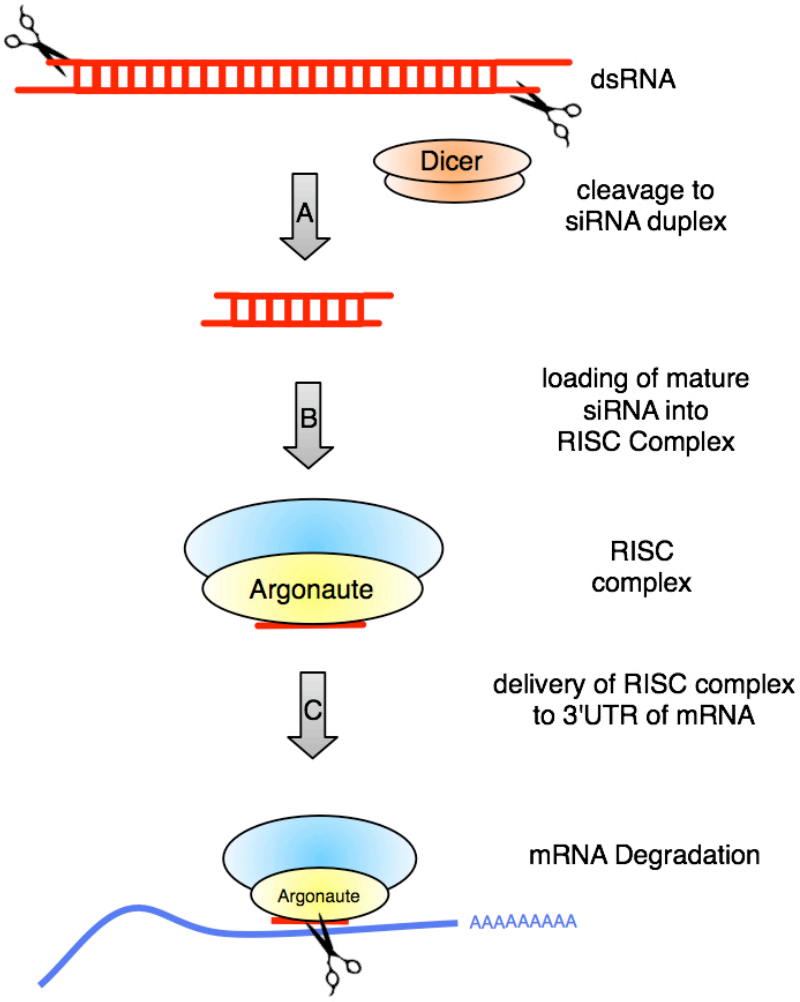


Figure 2: siRNA biogenesis pathway. The pathway starts with long dsRNA. Subsequent steps include: A) the dsRNA is cleaved by Dicer to form a short double-stranded siRNA duplex; B) siRNA duplex separates into single-stranded siRNAs and complexes with a RISC-like structure; C) mRNA binds with siRNA/RISC complex and mRNA is degraded. The picture was adapted from de Fougerolles 2007.

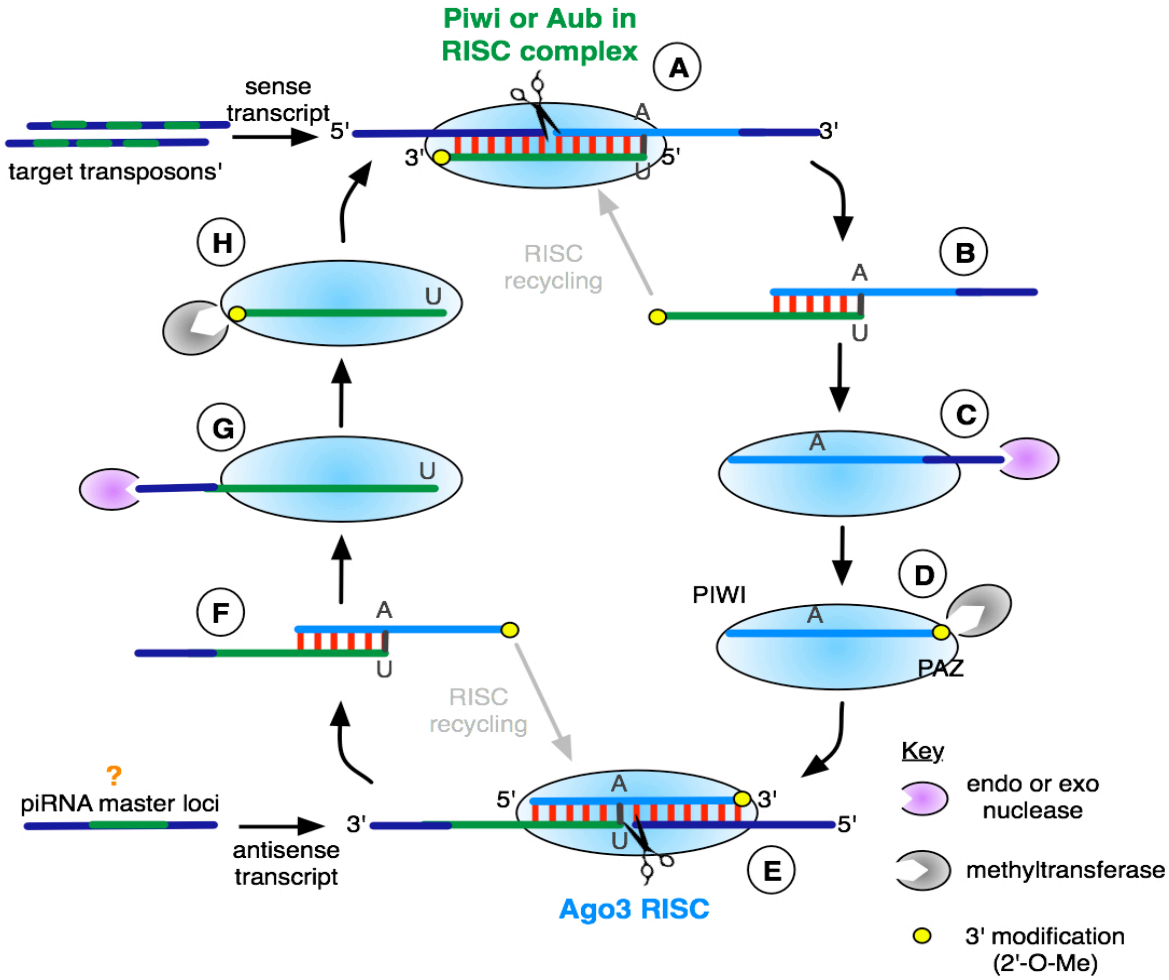


Figure 3: Proposed rasiRNA biogenesis pathway (Ping-Pong model). The proposed pathway is cyclical with an unknown initial step. A) The antisense rasiRNA in Aub or Piwi in RISC guides them to active TE transcripts (sense) where a slicer mediated mechanism occurs between nucleotides 10 and 11 of the target RNA. B) This step is proposed to yield the 5´end of the sense rasiRNA. C) The 3´end is thought to be processed by an unknown endo or exonuclease. D) The 3´end of the sense rasiRNA is further modified by an methyltransferase (2´-O-Me). E) The sense rasiRNA and Ago3 complex then targets complementary transcripts (from an unknown source), where Ago3 slicing activity cleaves the 5´end of the antisense rasiRNA. F) This step is proposed to yield the

5´end of the antisense rasiRNA. G) The antisense transcript is cleaved by an endo or exonuclease (unknown). H) The antisense transcript is further modified by an methyltransferase (2´-O-Me) and loaded into the Piwi or Aub in RISC completing the cycle. The picture was adapted from Hartig 2007.

Chapter Two

Belly, a potentially active LTR retrotransposon in the African malaria mosquito, *Anopheles gambiae*

Abstract

Retrotransposons are a group of mobile genetic elements that transpose through an RNA intermediate. LTR (Long Terminal Repeat) retrotransposons are similar to retroviruses in their structures and protein sequences. The LTR sequences contain transcriptional regulatory sites, which flank the open reading frame(s) that are important for their transposition. Here we report the characterization of a novel retrotransposon we named *Belly,* from the African malaria mosquito *Anopheles gambiae*. Comparison between empty and insertion alleles defined the boundaries of *Belly* and its 5bp target site duplication. This novel retrotransposon has all the properties necessary for retrotransposition including two identical direct LTRs flanking one open reading frame that encode a polypeptide with three zinc fingers, protease, reverse transcriptase, RNase H and Integrase motifs. *An. gambiae* has nine full-length copies of *Belly* throughout the sequenced genome and their nucleotide similarity suggests recent retrotransposition. The *Belly* transcript termination site was determined using 3' RACE, which also demonstrates the existence of processed transcript. The proteomic data, from the *An. gambiae* midgut, indicate that *Belly* is expressed.

Single LTR circles were recovered from *An. gambiae* cells (Sua1B) using nested inverse PCR, which is also consistent with active transposition of *Belly*.

2.1 Introduction

Retroelements comprise a common group of repetitive DNA elements, which are found dispersed in many organisms. They are a group of mobile genetic elements that transpose through an RNA intermediate. They can be a source of mutations that could have significant impact on evolution. The effect of retrotransposons in genome size variation depends on both their rate of retrotransposition and their frequency of elimination (Petrov 2001; Kidwell 2002). The efficiency of retrotransposition is determined by the number of active elements in the genome and by their overall activity.

Long Terminal Repeats (LTR) are a type of retrotransposons similar to retroviruses in their structures and protein sequences, like GAG and POL. The gag-like (GAG) proteins in retroviruses encodes for the structural proteins involved in creating the viral particle and POL is a polyprotein, processed to produce the enzymes necessary for retrotransposition. Retroviruses contain an additional gene called the *envelope* (*env*) gene, which is needed for their extracellular phase in their life cycle. However some retrotransposons do encode an envelope-like protein making them even more similar to retroviruses (Kim et al. 1994; Song et al. 1997).

There are three highly divergent clades of LTR retrotransposons: the *Gypsy/Ty3* group, the *Copia/Ty1* group and the *Pao/BEL* group. The three groups can be distinguished based on the organization of the POL proteins, which are protease (PR), reverse transcriptase (RT), RNase H and integrase (IN). These proteins are necessary for their transposition event. The PR conducts

proteolysis, hydrolyzing the peptide bonds that link amino acids together in the polypeptide chain. The RT is the enzyme responsible for creating double stranded DNA from an RNA template. The enzyme RNase H is a ribonuclease that catalyzes the cleavage of RNA via a hydrolytic mechanism, once the cDNA is made. Integrase is the enzyme responsible for enabling the genetic material (cDNA) to be integrated into the DNA (retroviruses like HIV contain this enzyme). The order of domains of the POL porteins in *Ty1/copia* elements is PR, IN, RT and RNase H, while in *Ty3/gypsy* elements, *Bel* elements, and the retroviruses; the order is PR, RT, RNase and IN (see Figure 4). When phylogenetic trees are constructed for LTRs the RT amino acid sequences are used to distinguish a number of clades. The *Bel* elements have the same overall organization as the *Gypsy* elements, but their RT is phylogenetically different thus resulting in two different clades. Active LTR sequences also contain transcriptional regulatory sites, which flank the internal open reading frame(s) that are important for their transposition. It has been demonstrated in retrotransposons, that transcription is initiated with the 5'LTR and to terminate in the 3'LTR, resulting in a polyadenylated transcript. Our initial genomic analysis that screen for likely active TEs, which is based on the identification of nearly identical copies, uncovered a novel retrotransposon we called *Belly*. Subsequently, this element was reported by another group as one of the 210 *Pao-Bel* retrotransposons in the *An. gambiae* genome (Massimiliano Marsano and Caizzi 2005). In the present study, we present detailed characterization of *Belly* and evidence for it being active in *An. gambiae* cells.

2.2 Materials and methods

Target site duplication

The PCR assay were conducted using DNA isolated from pooled of *An. gambiae* adults of the M and S molecular forms. The primers used are sequences flanking one *Belly* copy: 5'-CGTCATTCATACAAAGCGC-3' and 5'-GAAAGCAACA-AATGAGTTCAAG-3'. The PCR program uses was 3 mins. of initial denaturation at 94°C and 36 cycles consisting of 94°C for 45 sec, 52°C for 45 secs. and a final enlongation step at 72°C for 5 mins. A 1% agarose gel was run. A band that was at approximately 200bp was produced. The band was excised and purified using the GFX PCR DNA and Gel Band Purification Kit (Amersham Biosciences) following the manufacturer's protocol. The purified DNA was sent to sequence at the Virginia Bioinformatics Institute.

Rapid Amplification of cDNA Ends (3´ RACE)

Total RNA was isolated from Sua1B cells passage 34 following TRIzol protocol (Invitrogen). The cells were collected by detaching them from the flask a with spatula and washing them in 10 mL of PBS, this step was repeated 3 times. Following DNase treatment (Ambion), samples were reverse transcribed using Superscript II reverse transcriptase (Invitrogen) at 42°C for 1.5 h followed by heat inactivation at 70°C for 15 minutes. The BD SMART™ RACE cDNA Amplification Kit (Biosciences) was then used to prepare first-strand cDNA and to amplify 3' RACE products according to the manufacturer's instructions. A gene-specific primer used for the RT-PCR amplification, which was 5'-

TCAGAGTAAACCGGCAAAATC-3. The PCR program was an initial 3 mins. denaturation step at 94°C and 40 cycles consisting on 94°C for 30 secs, 60°C for 30 secs., 72°C for 1 mins., and a final elongation step at 72°C for 5 mins.

Nested inverse PCR

DNA from Sua1B passage 43was isolated following the Trizol (Invitrogen) protocol to run a PCR. The PCR reaction was prepared with 1µl of genomic DNA (from Sua 1B cells), 4µl of dNTPs, 4µl of 10X Buffer, 0.25µl of Taq (Takara), 36.75µl of autoclaved water and 2µl of each primer 5'-TAGTATCAGG-ACTAGGAACGGC-3' and 5'-GCTCGCGCTTTAGGATAAT-3'. After 3 mins of denaturation at 94°C and 36 cycles consisting of 94°C for 30 secs., 57°C for 30 secs., 72°C for 1 mins and a final enlongation step at 72°C for 10 mins. The PCR product was diluted with autoclaved water (1µl of the PCR sample and 9µl of autoclaved water). The dilution was used for a second round of PCR, the primers used were 5'-CTGTTGAAAGAGGATGGGAAG-3' and 5'-ACGAGGCG-TCAATTCTGG-3'. The same PCR program was used. The four bands of the first PCR and the three bands of the second PCR were excised and purified with GFX PCR DNA and Gel Band Purification Kit (Amersham Biosciences), following the manufacturers protocol. The purified DNA was sent for sequencing at Virginia Bioinformatics Institute.

2.3 Results and discussion

Characterization of *Belly* in *An. gambiae*

Belly is 6090bp in length and it is composed of 287bp identical terminal direct repeats, which flank the one 5 kb open reading frame. The *Belly* element is a low copy number retrotransposon, it occurs in 9 full-length copies throughout the *An. gambiae* genome. Eight of the nine copies of are highly conserved in the genome (see Table 1). The complete nucleotide sequence of the *Belly* element is shown in Figure 5. Motifs and putative encoded polypeptides that are characteristic among retroelements, were identified by comparison with other retroelements and are boxed and specified in Figure 5. The gag-like protein in *Belly* has three characteristic zinc finger nucleic acid binding motifs that were identified, C2C3H4C, C2C4H4C, and C2C3HH3H. The POL protein in *Belly* contains a protease, reverse transcriptase (with its seven characteristic motifs), RNase H and intergrase domains that were also identified and boxed (see Figure 5).

To define the boundary of *Belly* element and to determine the length of the target site duplication, a PCR was run using DNA from pooled *An. gambiae* mosquitos. The primers flank one copy of the *Belly* element. The 200bp product was excised, purified and sequenced (Figure 6). The sequence was aligned with the *Belly* locus. An empty site, devoid of *Belly*, was identified. The comparison between the empty site and the *Belly* insertion revealed the 5bp target site duplication flanking a complete *Belly* element.

Expression of *Belly* in *An. gambiae* cells and tissues

Proteomic analysis of the *An. gambiae* midgut samples revealed the presence of two peptides that matched perfectly the deduced amino acid sequences of *Belly* (shown as shaded peptides in Figure 5, data contributed by M. Davenport and M.

Jocobs-Lorena, Johns Hopkins University). We also performed 3´ RACE using RNA isolated from *An. gambiae* Sua1B cells. The amplified DNA was run on a gel (Figure 7A) and the 600bp band was excised and then sequenced. The sequence obtained was *Belly* with a polyA tail, evidence that that mRNA transcripts were processed (Figure 7B).

Recovery of a single-LTR circle

Nested inverse PCR was performed to identify possible intermediates of retrotransposition. Prior to the integration of a retroviruses or a retrotransposon into genomic DNA, circularized DNA intermediates have been identified. The unintegrated cDNA circles for LTR can either contain one or two LTR. To understand how nested inverse PCR is performed, see Figure 8A. The PCR products were ran on a gel, see Figure 8B. The 1Kb band on the second round of PCR was purified and sequenced. One LTR circle was identified for *Belly* in Sua1B cells, see Figure 8C.

In summary, we have characterized a *Bel*-like element that resides in the *An. gambiae* genome. *Belly* contains all the proteins required for transposition and sequence similarities among different *Belly* copies are very high, indicating at least recent transposition. We also gathered evidence suggesting that *Belly* is transcribed and translated in *An. gambiae*. We have also recovered a single LTR circle for *Belly* from an *An. gambiae* cell line, which may further indicate active retrotransposition by *Belly* because such circular molecule is regarded as one of the intermediates for retrotransposition. Future work will involve establishing retrotransposition assays, which may directly demonstrate transposition activity.

The identification of an active TE endogenous in the mosquito genome may provide tools for mutagenesis or transgenesis as well as the possibility to investigate TE regulation in mosquitoes.

Table 1: Nucleotide Percent Identity

Percent Identity (Paup)	1	2	3	4	5	6	7	8	9	10
1	--									
2	80.42	--								
3	74.92	96.19	--							
4	71.40	96.14	99.08	--						
5	71.68	96.45	99.32	99.56	--					
6	71.67	96.42	99.32	99.37	99.62	--				
7	71.64	96.35	99.13	99.29	99.41	99.39	--			
8	71.76	96.29	99.03	99.14	99.33	99.31	99.36	--		
9	69.39	96.13	98.90	99.08	99.27	99.10	99.11	99.11	--	
Consensus	71.93	96.76	99.53	99.59	99.75	99.75	99.69	99.59	99.49	--

This is the nucleotide similarity, in percentage, among the 9 full-length copies of *Belly* and the consensus sequence.

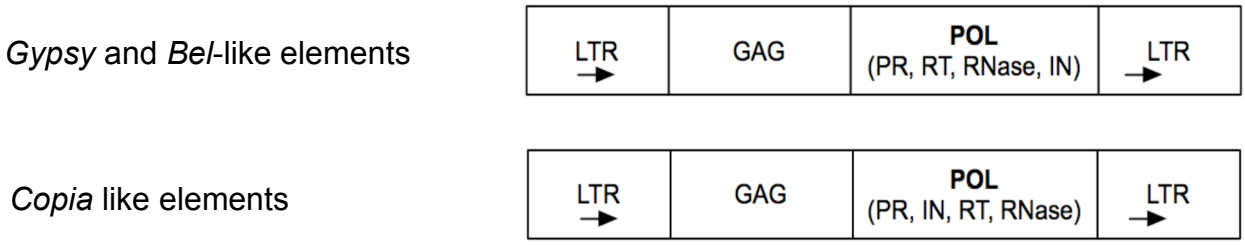


Figure 4: Schematic diagram of the organization of the polyprotein (POL) in the different types of LTR retrotransposons.

Tgttggaatgtaagggttatgaaacggtcattttgaattgtttgcggttgttttgtcagttgggaattaaaagttaaat
gtattttctggcagcactgccgatcgacaatttgtgattaagtatgtgtgcg**aataaa**gcggcactagcgcatgaaact 5´UTR
cgatacgagccggacgtgttctttactttgtctcctttggcgatcgaagacgacacaacaaaacacaacgtagggcgta
gaggcgtcaaggggaaaggaaccaacaaaccatgttccagaacgcaacattggtgccgtgaccaggatggtctaatat
ttgcggtttggtttgactggttaaagtgaagaactttgtgtgatttcttttggcaaaaatcgcgtaacgcgtgtgaaat
ctgaaagttccgtatcgagcgttggtggatttctgaataagaaaaaatcgcgaaagcgtgagaaatcattgtgctagtg
ggtgttctttgcgtgtgtatgcattttgtgtcgcgtattagaa

MSVSEDFAGFSDASGSNNTMQYTQAEDLDMIILKRERDRVVQSLTRIDTFLAQYKESDFPELTPRLDLLNE

RWREFQALARNIGAKDYSEDNDVLYGEIEDKVMLLKGKLLGKLRAGAEIPSVKQERVCDDYNSVRLPQLTL gag-like

PQFSGKYDEWLPYHDMFVVTVHENEKLSQVEKMLYLKGSLKGEALKVVDTLQACNSNYDVAWDALKKRYSN

EYILKKRHVNAMLQWPRMKVMNTVGIHGLIDCFERNLQILKQLGEVTEQWGCLIIQIIISKLDESTQQKWE

RHVEESEQKTVTDLLNFLRTQTRIMDAFAVDRPMAAGKSTSERRVASNVAAEAKCAKCDGSHMVENCDSFR

SLTLPRRREVVEAKKLCLNCLRQGHFQAKCWSRARCNICNRKHHSLLHGENVSETIGPSVREELPSTSGTQ 3 Zinc Fingers

NVVVNASSNRECTSVLLSTAIVSVRACGKKWLSARALIDSGSQVNLMTKGLAARLKLPQYESKTALSGVGH PR

SKVDITMSVTTVIRSKSCNYQERMQFLVLPRISSYRPVTGNQISRQNLPMNFVLADPNFDSDAEVDLLMGS

EFYATFLKPDNRGKIRLELPALPTFISTVFGWVATGKVPLASEGSNYVTCGTCTRLDDLIERFWIIEEIRE

PLQHSQEERDCEAHFVQTHQRDSEGRYIVKLPFKCDLQNQLGPSSAIARKRFLQLERRFNRDPWLKQKYTA RT

VINDYIDKGILVKVAANPDSEEAHGHYFLPHHPVIKASSTSTKVRPVFDGSASTDGGKSLNDLLMTGPVIQ Motif I & II

ENLLALLLKFRMRSVALVADIKQMYLQVKVHPDDTRFQRVLWRGSSVESIEVYELQRVTFGLAPSSFLAIR Motif III & IV

VLQQLAIDEGENFPLARQALLEDFYVDDYIGGASSEEEAVRLQAELTLLLRKGGFHLTKWNSNKPDVLSSV Motif V & VI

SAEDRATSNVKMFEVPEEPIKTLGIAWLPESDQLYIDSNIQMNNESWSRRKVYSLVARIYDPLGLVAPVTS Motif VII

WAKINMQSLWLATDDWDEEIPAVMQERWYAFQSQLGLLKEVKFSRHAVVHNPVAVQLHCFSDASEAAYGAC RNase H

VYVRTIGSSGEVVVELLAAKSRPAPLKRVSLARLELCGALLAARLQKVVRQALRIPDVETFMWTDATIVLH

WIRAPSHSWATYVANRVSEIQELTHGYKWMHVKGVDNPADIVSRGAMPNELLASKLWFHGPGWLQLSEEEWK

KNASGVLAIPEEELLERRKSSLVAAVSSESDDWCDRFSNYDKLLRITAYCMRFIRCCQRKLDPKHKGVLLVS

ELAEAKIRLVKREQRIYFAAEIKELSAGQTVRPKSSLKTLGAFLDGDGLLRVGGRLHRAKAMQVCSRFPLAL IN

PKKSRFTRLMAEYYHRLALHGGPTATLSALRREFWPIQGRSLVNSVCRGCLVCFRMNPALVQQPPGQLPVSR

AMPARPFSIVGVDFCGPIYLKPVHRRAAAEKAYISIFVCFSVKAVHIELVESLSTHAFLAAFRRFVARRGLP

SEVYSDNGLNFQGASKVIDDFYTLMNSDSAVEDISRYAVGAGVKWHFIPPHAPNFGGLWEAAVKAAKRVLLK

VVGDRQLAFGEMSTVLAQVEAQLNSRPLTPLSEDPEELDVLTPGHFLIGAPMNALPEPDVGDVPINRLKRYE

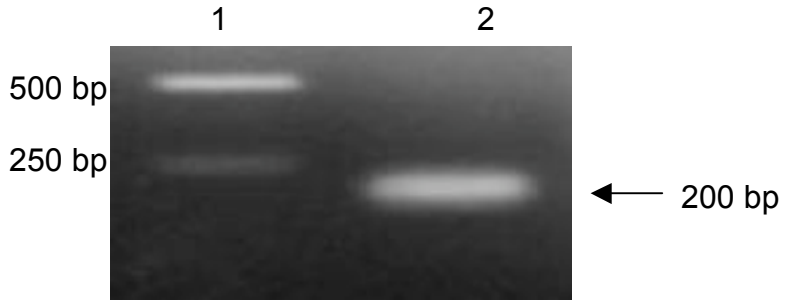
ELRRVVQNHWARWRREYFSELHNEHQRGKAVVELKVGQMVLLKEDGKTLHHWPMGRIAEVFPGPDGVVRVVS

IRTRNGLYKRPANRISLLPFERVN*

atatcataaaggcattttgtgaagtaacggcaagaagtaaatttggtaaatttaggtggccgctatgttggaatgtaagg
gttatgaaacggtcattttgaattgtttgcggttgttttgtcagttgggaattaaaagttaaatgtattttctggcagca
ctgccgatcgacaatttgtgattaagtatgtgtgcg**aataaa**gcggcactagcgcatgaaactcgatacgagccggacgt 3´UTR
gttctttactttgtctcctttggcgatcgaagacgacacaacaaaacacaacgtagggcgtagaggcgtcaaggggaaa
ggaaccaacaaaccatgttccagaacgcaaca

Figure 5: Characterization of *Belly*. The underlined sequence is the 287 bp LTRs and the boxed sequences correspond to the most conserved sequences of Zinc Fingers, Protease (PR), Reverse Transcriptase (RT) seven Motifs, RNase H, and Integrase. The shadowed protein sequences correspond to the sequences identified in a midgut proteomic study.

A

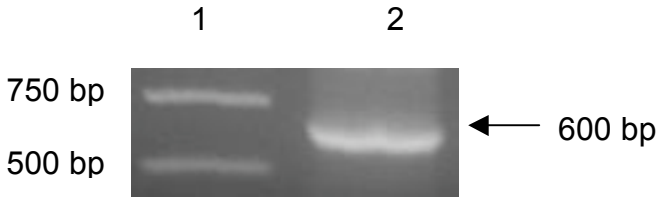


B

```
Belly 1 ttcacacacaaaattgagtaatagtggt  Belly  gtggtaatagttttgtcgccgaaatttt  343
        ||||||||||||||||||||||||||||                ||||||||||||||||||||||
PCR 1   ttcacacacaaaattgagtaatagtggt                aatagttttgtcgccgaaatttt  51
```

Figure 6: Target site duplication. A) Gel with primers flanking a *Belly* locus. Lane 1 is the 1Kb ladder and Lane 2 contains the PCR product amplified using primers flanking a *Belly* locus. **B**) The 200bp band was excised, sequenced, and aligned with the allele that contains *Belly* insertion. An empty site, devoid of *Belly*, was revealed. The underlined sequences are the 5bp target site duplication.

A

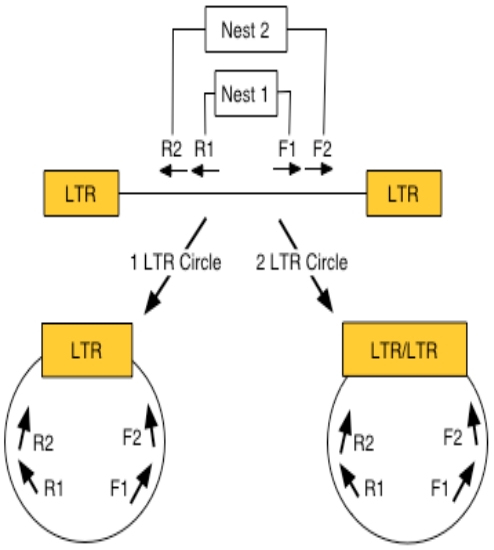


B

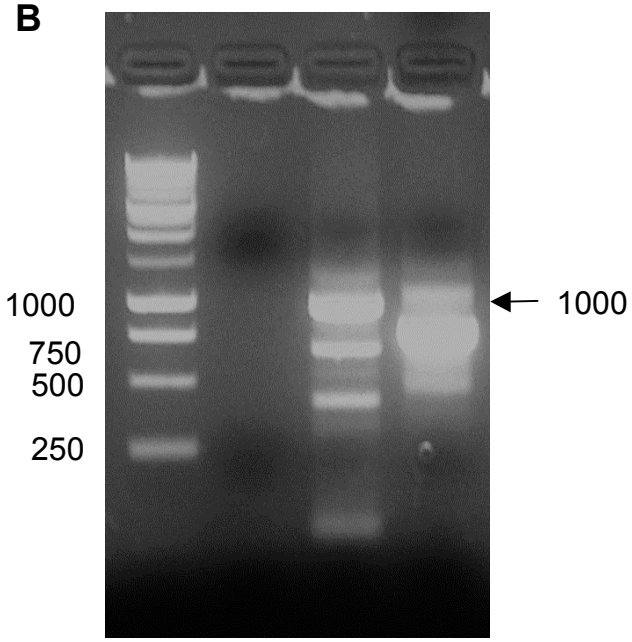
```
3′ RACE tgttggaatgtaagggttatgaaacggtcattttgaattgtttgcggttgttttgtcagt
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Belly   tgttggaatgtaagggttatgaaacggtcattttgaattgtttgcggttgttttgtcagt

3′ RACE tgggaattaaaagttaaatgtattttctggcagcactgccgatcgacaatttgtgattaa
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Belly   tgggaattaaaagttaaatgtattttctggcagcactgccgatcgacaatttgtgattaa

3′ RACE gtatgtgtgcgaataaagcggcactagcgcatgaaactcgaaaaaaaaaaaaaaaaaaaaaaaaaaaa
        ||||||||||||||||||||||||||||||||||||
Belly   gtatgtgtgcgaataaagcggcactagcgcatgaaactcgatacgagccggacgtgttctttactttgt
```

Figure 7: Evidence of transcription and determination of the 3' end of the transcript. A) Gel for 3'RACE-PCR. Lane 1 is the 1Kb ladder (Promega), Lane 2 contains the 3'RACE PCR product. **B**) The 600bp band was sequenced and aligned with *Belly*. Underlined is the putative poly(A) addition signal.

A



B



C

```
Belly 5635 gttagtatcaggactaggaacggcttgtataagaggccagcgaataggattagtcttctt
           |||||||||||| | ||||||||||||||||||||||||||||||||||||||||| |||
PCR   1    gttagtatcagaattaggaacggcttgtataagaggccagcgaataggattagtctcctt

Belly 5707 ccgtttgagagagtgaattagatatcataaaggcattttgtgaagtaacggcaagaagta
           |||||||||||||||||| |||||||||||| ||||||||||||||||||||||||||
PCR   61   ccgtttgagagagtgaactagatatcataaatgcattttgtgaagtaacggcaagaagta

Belly 5767 aatttggtaaatttaggtggccgcta [LTR] ttggtgccgtgaccaggatggtctaat
           |||||||||||||||||||||||||||     ||||||||||||||||||||||||||||
PCR   121  aatttggtaaatttaggtggccgcta [LTR] ttggtgccgtgaccaggatggtctaat


Belly 314  atttgcggtttggtttgactggttaaagtgaagaactttgtgtgatttcttttggcaaaa
           |||||||||||||||||| |||||||||||||||||||| ||||||||||||||||||
PCR   461  atttgcggtttggtttgagtggttaaagtgaagaactttgtttgatttcttttggcaaaa

Belly 374  atcgcgtaacgcgtgtgaaatctgaaagttccgtatcgagcgttggtggatttctgaata
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PCR        ttcgcgtaacgcgtgtgaaatctgaaagttccgtatcgagcgttggtggatttctgaata

Belly 434  agaaaaaatcgcgaaagcgtgagaaatcattgtgctagtgggtgttctttgcgtgtgtat
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

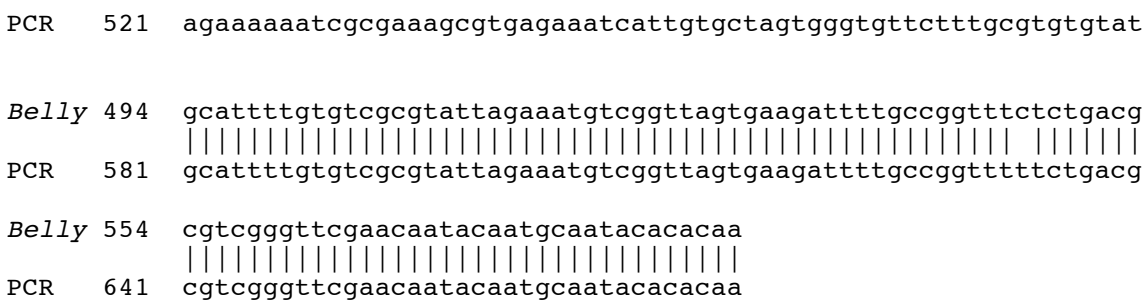


Figure 8: Recovery of a single LTR circle. A) Schematic diagram for nested inverse PCR, with possible outcomes of one LTR and two LTR circles. **B**) Gel for nested inverse PCR with DNA from Sua1B cells. Lane 1 is the 1Kb Ladder (Promega), Lane 2 is the negative control, Lane 3 is the first PCR with F1 and R1 primers, and Lane 4 is the second PCR with F2 and R2 primers. **C**) The 1000bp band, from the second PCR, was sequenced and aligned with *Belly*.

Chapter Three

Identification and Expression of *Piwi* genes in the Dengue and Yellow Fever mosquito, *Aedes aegypti*

Abstract

The Argonaute family of proteins in *D. melanogaster* was used to search for homologs in the sequenced genomes of *Ae. aegypti* and *An. gambiae.* We found nine potential *piwi* genes in *Ae. aegypti* and two in *An. gambiae*. Phylogenetic analysis suggests that these *piwis* formed two subgroups and gene duplication within each group occurred after the divergence between the two mosquito species. RT-PCR and transcriptome analysis showed that Ago3 and all seven tested piwi genes were expressed either in germline tissues or developing embryos. Differential expression patterns were observed. While most *piwis* were transcribed in the ovaries, testis, and embryos, two *piwis* appear to have a zygotic expression. Three genes from the Piwi subfamily (*piw*i 3, piwi 4, and *Ago3*) were also detected in adult somatic tissues of *Ae. aegypti*. The expansion of the number of *piwi* genes in *Ae. aegypti* compared to *An. gambiae* and *D. melanogaster* may be correlated with a larger genome size and a greater amount of TEs. The finding of *piwi* expression in adult somatic tissues is intriguing. It is possible that these *piwi* genes were expressed in the adult stem cells. It is also possible that they may be involved with anti-viral defense. Both of these hypotheses require further testing.

3.1 Introduction

TEs are major players of genetic change and constitute a large component of the genome of nearly every living organism. An RNAi related mechanism has been proposed as a way to limit or silence TE activity. This silencing mechanism is critical in the germline where TE activity can create a considerable mutational burden, which would accumulate with each passing generation. Such silencing may also be important in stem cells (Saito et al. 2006; Vagin et al. 2006; Brennecke et al. 2007), which are able to self-renew and differentiate into specialized cell types (Lin 1997; Lin and Spradling 1997).

The Argonaute proteins make up a highly conserved family whose members have been implicated in the RNAi related mechanisms of many organisms. They are thought to influence development, and at least a subset are involved in stem cell fate determination. Argonaute proteins are ~100-kD and contain two common domains, named the PAZ and PIWI domains (Cerutti et al. 2000). The Argonaute family of proteins is subdivided into two subfamilies, the Ago subfamily and the Piwi subfamily. The Piwi subfamily of proteins has been shown to strongly correlate with the emergence of germ cells. In *Drosophila*, the self-renewing ability of germline stem cells requires an evolutionarily conserved clade of genes called *piwi* (Cox et al. 1998). *Piwi* homologues have been identified in *C. elegans* (*prg-1* and *prg-2*) and human (*hiwi*) (Cox et al. 1998). In *C. elegans*, a drastic reduction of germline proliferation is seen when the activity of *prg-1* and *prg-2* is decreased (Cox et al. 1998). Most organisms separate germline and somatic cells early in development. *piwi* expression is mainly restricted to the germline. *Drosophila* has three genes of

the Piwi subfamily (*Piwi, Aubergine* and *Ago 3*), which are all expressed in germline cells. Piwi has also been reported to be present in the somatic cells of the *Drosophila* ovary (Cox et al. 2000).

In my research project I focused on *piwi* genes in the yellow fever mosquito, *Ae. aegypti.* There appears to be an expansion of the number of *piwi* genes in *Ae. aegypti* and we showed that *Ago3* and all seven *piwis* tested were expressed either in the germline tissues or the developing embyos. Two *piwis* (*piwi* 1 and *piwi* 2) appear to be expressed mainly zygotically during embryonic development. In adult somatic tissues two *piwis* (*piwi* 3 and *piwi* 4) and *Ago3* are also expressed. It is possible that these *piwis* may be expressed in adult stem cells. We also discuss the evolutionary implications of the expansion of *piwi* genes in *Ae. aegypti*.

3.2 Material and methods

DNA sequence retrieval

A systematic strategy which incorporates multi-query BLAST (Altschul et al. 1997) using the Ensemble (http://www.ensembl.org/) database (Birney et al. 2004), was used to search the Argonaute DNA sequences in *Ae. aegypti* and *An. gambiae.* The *D. melanogaster* Argonaute family was used as query, including Ago1 (FBgn0026611), Ago2 (FBgn0046812), Ago3 (FBgn0086780), piwi (FBgn0004872), and aub (FBgn0000146). The results of our findings are summarized in Table 2.

Phylogentic analysis

Phylogenetic analyses were performed using multiple sequence alignments of amino

acid of the Argonaute family. These alignments were obtained using ClustalX v1.83 for Mac OSX (Thompson et al. 1997). Parameters used for these alignments were: pairwise gap penalty (open=10, extension=0.2), protein gap separation parameters (separation distance=4). The phylogenetic analysis was performed PHYLIP v3.67. Neighbor-Joining and maximum parsimony trees were constructed with bootstrap support of 1000 replicates. We were able to compare the Argonoute family of *Ae. aegypti, An. gambiae,* and *D. melanogaster* (Figure 9).

The Amino Acid Divergences table was calculated by using MultiAlin (Corpet 1988), CLUSTALX, and PUAP (Swofford 2002) v4.0b10 for Macintosh. Amino acid sequences for all nine *piwi* sequences were obtained from ENSEMBLE and were entered into MultiAlin as a FASTA file. Parameters used were: Consensus levels (high = 70% and low = 40%). The results were saved in a FASTA file, which was then edited to remove the dashes. The newly created FASTA file was loaded into CLUSTALX for another round of alignment. Parameters used for this alignment were: pairwise gap penalty (open=10 and extension=3). The resulting alignment was saved as a Nexus file (.nxs), pairwise distances or pairwise sequence divergences were calculated using PAUP (see Table 3).

Transcriptome sequencing

Total RNA was isolated from eggs that were collected 0-2 hr, 2-4 hr, and 4-8 hr after deposition by following the Trizol protocol (Invitrogen). *Aedes aegypti* mosquitoes (Liverpool strain) that were used for egg-laying were maintained 16hrs in light and 8hrs in the dark cycle. The RNA samples were sent to illumina for sequencing

(http://www.illumina.com/). illumina sequencing produced several millions of small sequence fragments (33-35 bp) per sample. Sequences collected were compared to the seven *piwis* and *Ago3* using BLAST, to count the number of hits. A customized Perl script used the blast results to generate a summary table using an E-value cutoff of 1e-07. This summary table (see Table 4) contained the frequencies for each of the seven *piwi* genes and the *Ago 3* gene during each time point (0-2 hr, 2-4 hr, and 4-8hr). Since some of the *piwis* (*piwi* 5, *piwi* 6, and *piwi* 7) from *Ae. aegypti* share some degree of similarity, the frequency of unique occurrence these particular *piwis* sequences were calculated (see Appendix A). The number of times a single sequence hit multiple *piwis* (cross-match) was also determined using an E-value cutoff of 1e-07. Only *piwi* 5, *piwi* 6, and *piwi* 7 showed cross-match and the results are summarized in Appendix A.

RT-PCR analysis

Total RNA was isolated from eggs that were collected 0-1 hour, 0-2hr, 2-4hr, 4-8hr, and 34-36 hours, after they were laid. Also RNA was extracted from the dissected ovaries, midgut and thorax from non-blood fed (NBF) females and females 24 hours after blood feeding (BF). Testis from 4-6 day old males were also dissected. RNAs were isolated from all above samples using Trizol reagent (Invitrogen). The RNA was treated with DNase-free (Ambion) for the removal of contaminating DNA from the RNA samples. The first strand of cDNA was synthesized with 3µg/µl of total RNA using the SuperScript II reverse transcriptase (Invitrogen) and oligo(dT) primer were analyzed by PCR with 1µl of cDNA, 4µl of dNTPs, 4µl of 10X buffer, 0.25µl of Taq

polymerase (Takara), 36.75μl of autoclaved water and 2μl of each primer. After 2 minutes of denaturation at 94°C and 35 cycles consisting of 94°C for 30 seconds, 55°C for 30 seconds and a final elongation step at 72°C for 1:30 minutes PCR primers used and expected product sizes are shown in Table 5. Each *piwi* RT-PCR product was sequenced to confirm the amplification of inteneded cDNA. It is important to note that our primer design does not distinguish between *piwi* 5, *piwi* 6, and *piwi* 7.

3.3 Results

Previous investigations demonstrated that the Argonaute falmily in *Drosophila* consisted of Ago1, Ago2, Ago3, piwi, and aubergine (Williams and Rubin 2002). These protein sequences in *Drosophila* were utilized to acquire homologous sequences in *Ae. aegypti* and *An. gambiae.* We found nine *piwi* genes, two *Ago1* genes, one *Ago2* gene and one *Ago3* gene in *Ae. aegypti*. Of the nine *piwi* genes, seven were previously annotated (www.ensembl.org) and named *piwi*s 1-7. The remaining genes are highly similar to *piwi* 1 and *piwi* 7, respectively and are possibly smaller variants. Thus they were named *piwi* 1b and *piwi* 7b and the original *piwi* 1 and *piwi* 7 were named *piwi* 1a and *piwi* 7a, respectively. The structures of the Piwi subfamily genes that are under investigation are shown in Figure 10. The same naming system was used for *Ago1* since two were found. Two *piwi* genes, one *Ago1* gene, three *Ago2* genes, and one *Ago3* gene were found in *An. gambiae*. The two *piwi* genes in *An. gambiae* were named *piwi* 1 and *piwi* 2. The three *Ago2* genes are

very similar to each other and were named *Ago2a, Ago2b* and *Ago2c*. Shown in Table 2 are the gene ID (according to Ensemble), description, chromosomal location, strand direction, start, end, number of of exons, transcript length, and deduce protein length for the Argonaute families in *Ae. aegypti* and *An. gambiae.* The genes without ID were identified and characterized during this study by doing a Blast with each *piwi* gene against the *Ae. aegypti* genome using ENSEMBLE. The *Ago2* of *Ae. aegypti* was identified with the sequence reported in Campbell et al. 2008.

The Phylogenetic analysis delineates the Argonaute sub-families in the three organism evaluated, Ago1 and Ago2 comprising the Ago subfamily, while Ago3 and piwi proteins comprising the Piwi subfamily (Figure 9). Within the Piwi subfamily, *Ae. aegypti piwi* 1-3 are grouped together with *An. gambiae piwi* 1 while *Ae. aegypti piwi* 4-7 are grouped together with *An. gambiae piwi* 2. Within these two groups, *Ae. aegypti piwis* form monophyletic clades, indicating that the duplication of *piwi* genes in *Ae. aegypti* might have occurred after the divergence between *Ae. aegypti* and *An. gambiae.* Sequence divergence between nine *piwi* genes is consistent with the existence of two subgroups, *piwi* 1-3 and *piwi* 4-7 (see Table 3).

To determine whether these *piwis* and *Ago3* are expressed in *Ae. aegyti*, an expression pattern analysis was performed using gremline tissues, developing embryos, and adult somatic tissues. The expression pattern of seven of the *piwis* and Ago3 of *Ae. aegypti* was investigated using RT-PCR. *Actin* gene was used as the positive controls for all samples. In addition, *nanos* and *oskar* were used as positive controls in germline tissues and embryos and as negative controls in adult somatic

tissues since they have been reported to be expressed only in the germline and embyos (Calvo et al. 2005; Juhn and James 2006).

In all embryonic stages tested as well as dissected germline tissues such as the ovaries of blood-fed and sugar-fed females and male testis, transcripts of *piwis* 3-7 and *Ago3* were clearly detectable (see Figure 11). Thus *piwis* 3-7 and *Ago3* are maternally and paternally expressed and their transcripts were detectable through embryonic development. On the other hand, *piwi* 1 and 2 appear at 4-8hr, and 34-36hr embryos but not in germline tissues and early embryos, suggesting zygotic transcription. This pattern was confirmed with the transcriptome sequencing data of early embryos (Table 4) where the total number of hits corresponding to *piwi* 1 and *piwi* 2 dramatically increases after 4hr. The number of hits for *piwi* 5-7 decreased during development while *piwi* 3, *piwi* 4 and *Ago3* remain high (Table 4). It is important to point out that since *piwi* 5, *piwi* 6 and *piwi* 7 are so similar the total amount of hits was shown in Table 4 but it contain some hits that are overlapping. To see the amount of hits that match each of the three *piwis* see Appendix A. As for all the somatic tissues tested only, *piwi* 3*, piwi* 4 and *Ago3* were found to be expressed (see Figure 12). It should be understood that the expression profile for the *piwis* and *Ago3* does not reflect the relative expression level as it is not quantitative RT-PCR and detection of some of the genes under investigation could be due to leaky expression.

It is also important to note that the expression data from the testis showed the presence of *nanos* and *oskar*, contrary to what has been reported (Calvo et al. 2005; Juhn and James 2006). However, the reports of the absence of these two genes had

been for RNA extracted from the whole males and not for dissected testis. These two genes were always detectable at low levels during several replicates using new dissections and reagents. Although we cannot completely rule out the possibility of contamination, our –RT controls have been always clean. It is possible that low level of *nanos* and *oskar* expression may exist in testis.

3.4. Discussion

Recent advancements in small RNA research have elucidated the role of the Argonaute subfamily but a lot remains unclear, particularly in the Piwi subfamily. Previous investigation in *Drosophila,* predicted that *piwi* is a novel protein required for the self-renewing division of germline stem cells in both males and females (Boisson et al. 2006). However, very little is known of this subset of proteins in somatic tissues. To elucidate further the possible role of *piwi* in regulating germline stem cell division, we studied the expression of the Piwi subfamily in the embryos of *Ae. aegypti* at different time points. We also investigated their presence in somatic tissues in the *Ae. aegypti* mosquito.

The findings of the expansion of the number of *piwis* in *Ae. aegypti* compared to *An. gambiae* and *D. melanogaster* may suggest that there are more *piwis* in *Ae. aegypti* than in *An. gambiae* and *D. melanogaster* due to a larger genome size and a higher number of TEs. We have evidence showing all tested *piwi* genes are transcribed in the germline tissues and developing embryos. They are possibly involved in regulating the activity of TEs in these cells. Our laboratory recently

isolated a large number of rasiRNAs from *Ae. aegypti* embryos, which is consistent with the involvement of *piwi* in TE regulation (Li and Tu, unpublished data). *Piw*i 3, *piwi* 4, and *Ago3* are also transcribed in adult somatic tissues. It is possible that these *piwi* genes were expressed in the adult stem cells in the midgut or thoracic tissues. It is also possible that they may have to do with anti-viral defense. Both of these hypotheses require further testing.

Table 2: List of genes in the Argonaute family in *Ae. aegypti* and *An. gambiae*

Species	Gene	Gene ID	Description	Chromosomal (supercontig) Location	Strand	Start	End	Exons	Transcript Length (bp)	Translation Length (a.a.)
Ae. aegypti	Piwi 1a	AAEL006287	PIWI	SuperContig 1.199: contig_10239	-1	919681	922756	4	2885	945
	Piwi 1b	No ID	No Description	SuperContig 1.199: contig_10239	1	928353	929458	1	1106	368
	Piwi 2	AAEL013233	PIWI	SuperContig 1.809: contig_25028	1	254544	258715	5	3674	883
	Piwi 3	AAEL013227	PIWI	SuperContig 1.809: contig_25028	1	286693	293639	4	2754	888
	Piwi 4	AAEL007698	PIWI	SuperContig 1.275: contig_12989	1	893644	919878	6	3159	860
	Piwi 5	AAEL008076	PIWI	SuperContig 1.300: contig_13762	1	268758	277171	7	2300	674
	Piwi 6	AAEL013692	PIWI	SuperContig 1.896: contig_26310	-1	387881	414514	5	2649	882
	Piwi 7a	AAEL008098	PIWI	SuperContig 1.300: contig_13762	-1	247986	251578	5	2919	875
	Piwi 7b	No ID	No Description	SuperContig 1.896: contig_26311	1	427843	430949	4	2253	750
	Ago3	AAEL007823	PIWI	SuperContig 1.284: contig_13292	-1	1137093	1161534	7	3666	944
	Ago1a	AAEL015246	eukaryotic translation initiation factor 2c	SuperContig 1.165: contig_32598	-1	485	22803	10	2682	825
	Ago1b	AAEL012410	eukaryotic translation initiation factor 2c	SuperContig 1.693: contig_22990	1	414842	455831	12	2844	947
	Ago2	No ID	No Description	SuperContig 1.89: contig_5339	1	344396	388339	4	3255	992
An. gambiae	Piwi 1	AGAP011204	No Description	Chromosome 3L	1	19302937	19315980	4	2720	843
	Piwi 2	AGAP009509	No Description	Chromosome 3R	1	34973060	34976602	7	2930	889
	Ago3	AGAP008862	No Description	Chromosome 3R	-1	19875729	19881002	3	3336	930
	Ago1	AGAP011717	No Description	Chromosome 3L	1	32108213	32126103	13	3202	981
	Ago 2a	AGAP011537	No Description	Chromosome 3L	-1	26912149	26931547	4	2521	840
	Ago 2b	No ID	No Description	Chromosome 3L	-1	26916737	26926812	6	2523	840
	Ago 2c	No ID	No Description	UNKN	1	40892650	40892958	1	246	82

This table shows the *piwi*, *Ago1*, *Ago2,* and *Ago3* genes that are found in *Ae. aegypti* and *An. gambiae*. Each gene has an Ensemble Gene ID, description, chromosomal location, strand direction, start and end in the genome, amount of exons, transcript length, and translation length. The genes that do not have a Gene ID have not been annotated in Ensemble. To assemble this missing information (genes that had a, b, and c) we did a Blast (with each original gene) to identify the truncated versions. *Ago2* for *Ae. aegypti* was the same as reported in .

Table 3: Amino Acid Percent Similarity Between *Ae. aegypti* Piwi Protein Sequences

Piwi´s	Piwi 1a	Piwi 1b	Piwi 2	Piwi 3	Piwi 4	Piwi 5	Piwi 6	Piwi 7a	Piwi 7b
Piwi 1a	-								
Piwi 1b	100	-							
Piwi 2	51	52	-						
Piwi 3	52	52	65	-					
Piwi 4	36	36	53	54	-				
Piwi 5	53	53	56	55	80	-			
Piwi 6	38	39	52	52	80	93	-		
Piwi 7a	38	38	52	53	83	86	90	-	
Piwi 7b	49	49	58	58	86	86	93	100	-

The values represent the mean percent similarity of amino acid between each pairwise comparison of the *piwi* proteins found in *Ae. aegypti*. The amino acid percent similarity ranges between 36-100% as determined by PAUP v4.0b10.

Table 4: Transcriptome sequencing

Piwi	Eggs 0-2 hr	Eggs 2-4 hr	Eggs 4-8 hr
Piwi1	0	12	1639
Piwi2	1	3	502
Piwi3	323	1498	4658
Piwi4	1120	1191	2104
Piwi5	1074	286	132
Piwi6	1144	346	150
Piwi7	235	33	25
Ago3	474	1179	606
Total	2087707	4810435	3920821
Median	54	105	71

This table summarizes the number of hits from the sequencing data that match the *piwis* and *Ago3*. The samples that were sent for sequencing corresponds to *Ae. aegypti* embryos collected at the different time points (Eggs 0-2hr, Eggs 2-4hr and Eggs 4-8hr) after deposition. The e value cut off is 10-7. Total and median values refer to the entire transcriptome.

Table 5: List of primers and expected PCR product size

Genes	Forward Primer	Reverse Primer	Product Size of cDNA (bp)	Size with Genomic Contamination (bp)
Piwi 1	5´-CGTTTACAGGATTCCAAAGAAAG-3´	5´-ATTCGTGTGAAATACATGTGATATG-3	1385	1507
Piwi 2	5´- ATCATCGTGTACCGTGATGG-3´	5´-CACAAATCTATTCGTAAAACAAAAGC-3	1514	1571
Piwi 3	5´-CCTGATGCGAGCGATTG-3´	5-CGTTTCATTACCGTCAATTTATTAA-3´	1501	5627
Piwi 4	5´- AATGATGCGTGCCATGG-3´	5´- TTCATTTATTTGTCATCGAACGTTT-3´	1825	15416
Piwi 5	5´-AATGATGCGCGCCATGTC-3´	5´-TAATGTATCAACAAACTAGTCAATTCCAA-3	1646	1775
Piwi 6	5´- AATGATGCGCGCCATGTC-3´	5´-CTACAAGAAGTAGAGCTTCTTCTCGAGC-3´	1411	1540
Piwi 7	5´-AATGATGCGCGCCATGTC-3´	5´-GTTTATATAATTCATACGACGCGGTAGG-3´	1164	1778
Ago3	5´-TCAACGGATTGTGTGCATCG-3´	5´-TCAGCGGGAAAATCATGTCG-3´	743	938
Actin	5´-ATGTGTGACGATGATGTTGC-3´	5´-CTTCATGATCGAGTTGTAGACG-3´	855	3037
Nanos	5´-TTCGAATCCGATCCATCGTT-3´	5´-GCAGGTCGTCTGGTGTGATG-3´	583	10355
Oskar	5´-CGCCGAAATGATCAAGAAGC-3´	5´-GCTTCCGGTGACTTGTCTGG-3´	802	13082

This table summarizes the primers used to amplify the cDNA, the expected product size for each for all the genes tested and their controls. It also shows the product size that would be seen if there was genomic contamination.

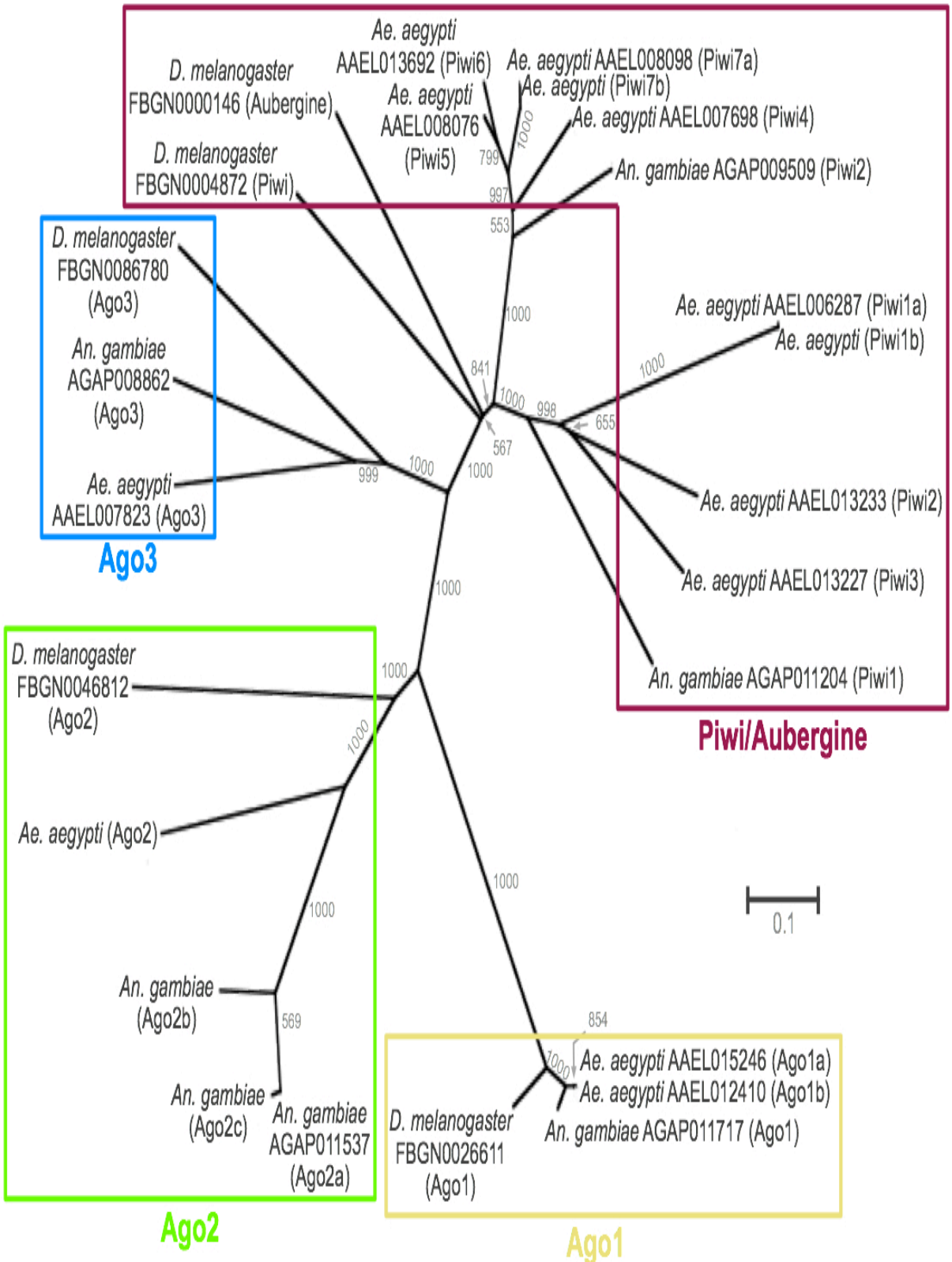

Ae. aegypti AAEL013692 (Piwi6)
Ae. aegypti AAEL008098 (Piwi7a)
Ae. aegypti (Piwi7b)
D. melanogaster FBGN0000146 (Aubergine)
Ae. aegypti AAEL008076 (Piwi5)
Ae. aegypti AAEL007698 (Piwi4)
D. melanogaster FBGN0004872 (Piwi)
An. gambiae AGAP009509 (Piwi2)
D. melanogaster FBGN0086780 (Ago3)
Ae. aegypti AAEL006287 (Piwi1a)
Ae. aegypti (Piwi1b)
An. gambiae AGAP008862 (Ago3)
Ae. aegypti AAEL007823 (Ago3)
Ago3
Ae. aegypti AAEL013233 (Piwi2)
Ae. aegypti AAEL013227 (Piwi3)
An. gambiae AGAP011204 (Piwi1)
Piwi/Aubergine
D. melanogaster FBGN0046812 (Ago2)
Ae. aegypti (Ago2)
An. gambiae (Ago2b)
An. gambiae (Ago2c)
An. gambiae AGAP011537 (Ago2a)
Ago2
0.1
Ae. aegypti AAEL015246 (Ago1a)
Ae. aegypti AAEL012410 (Ago1b)
D. melanogaster FBGN0026611 (Ago1)
An. gambiae AGAP011717 (Ago1)
Ago1

Figure 9: Phylogenetic tree of Argonaute proteins. A Phylogenetic tree was constructed for the Argonaute family using PHYLIP. This is an unrooted Neighbor-Joining tree with 1000 bootstrap replicates. Maximum parsimony analysis produced the same phylogenetic relationship. The alignment used for the phylogenetic reconstruction was obtained using CLUSTALX and included the deduced amino acid sequences of the members of the Argonaute family from *Ae. aegypti, An. gambiae* and *D. melanogaster*.

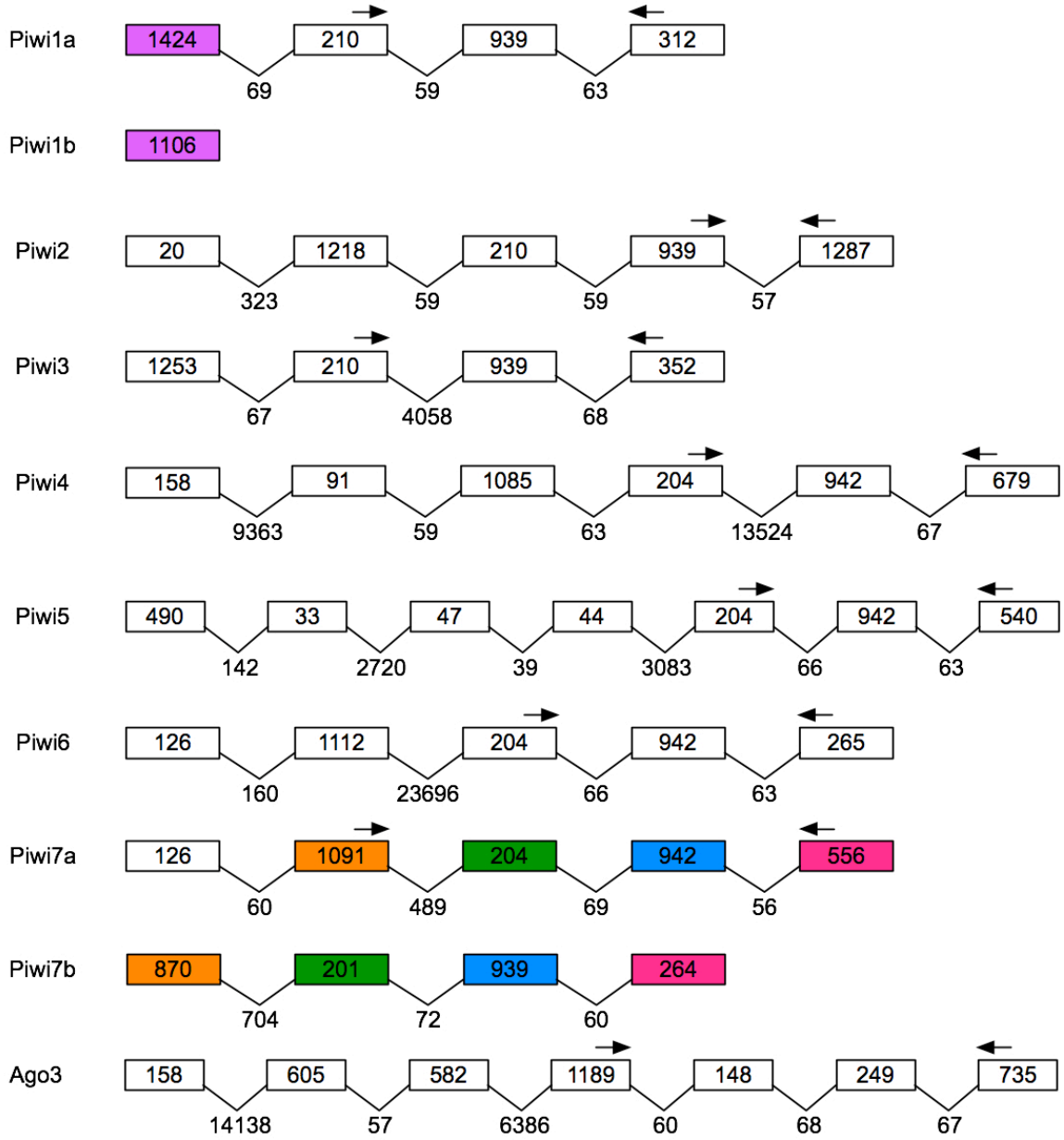


Figure 10: Gene structure of the Argonaute genes that are under investigation. The arrows correspond to the location of the primers used for RT-PCR. The *piwis* that have low percent of nucleotide divergence have an a and b versions and the exons that are similar among these piwis are colored.

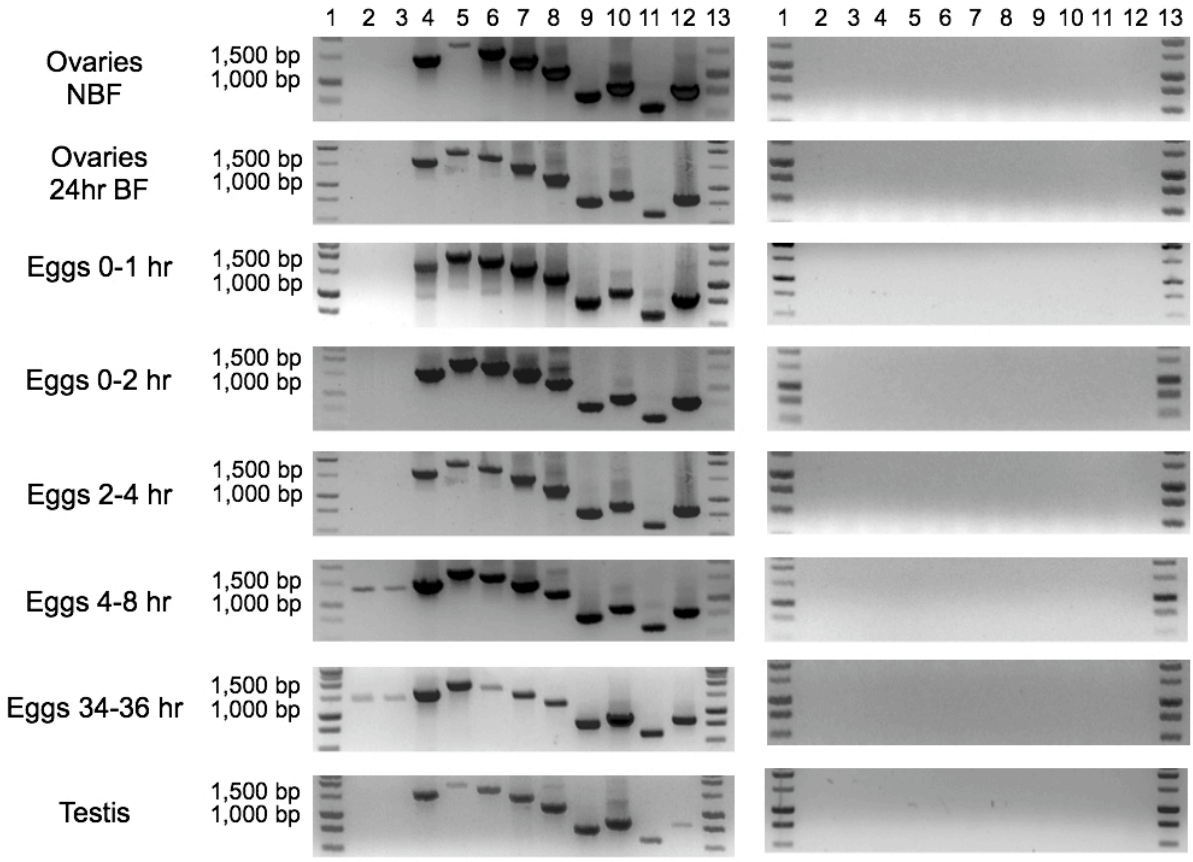


Figure 11: *piwi* expression pattern in the germline tissues and embryos. Lanes 1 and 13 are the 1Kb ladder (Promega), Lanes 2-8 are piwi 1-7, Lane 9 is Ago3, Lane 10 is actin, Lane 11 is nanos, and Lane 12 is oskar. The gels on the left side are with reverse transcriptase (RT) and the gels on the right side are the control gels without RT.

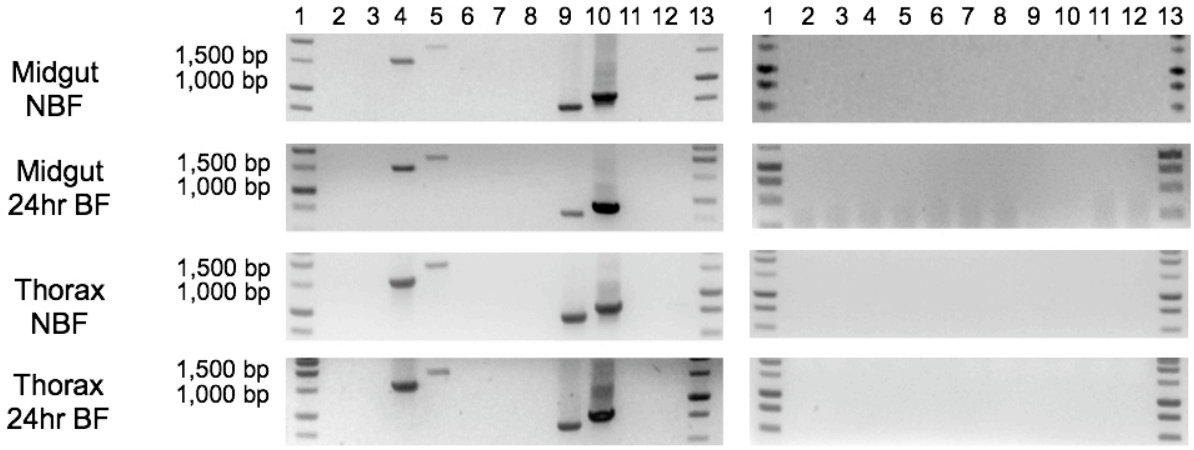


Figure 12: *piwi* expression pattern in adult somatic tissues. Lanes 1 and 13 are the 1Kb ladder (Promega), Lanes 2 to 8 are RT-PCR products for piwi 1 to 7, Lane 9 is Ago3, Lane 10 is actin, Lane 11 is nanos, and Lane 12 is oskar. The gels on the left side are with RT and the gels on the right side are the control gels without RT.

Chapter Four

Future perspectives

There are a number of possible future directions. For example, we could use RNAi to knock down some of the *piwi* genes in *Ae. aegypti*. RNAi is now a popular and common tool used for reverse genetics studies attempting to generate a loss of function through the introduction of dsRNA or shRNA complementary to the target mRNA transcript. The introduction of these complementary sequences can be accomplished through direct injection into newly laid eggs and adult mosquitoes by using a fine glass needle. The early embryo is a syncytium so injection at this stage is the same as direct cellular injection; therefore the foreign nucleic acids are in close proximity to the chromosomal DNA. In adult mosquitoes, the introduction of dsRNA may be accomplished by intra-throracic injection. There has been report of knockdown of gene expression in *An. gambiae* oocytes (Hoa et al. 2003) using relatively high concentrations of dsRNA. It is feasible to knockdown a component of RNAi with RNAi (Campbell et al. 2008).

As a negative control and to determine the mortality rate due to the injection trauma, buffer can be injected into the eggs and adults mosquitoes. This will help in determining the percent of adults that died due to the trauma as well as the amount of eggs that did not hatch for the very same reason. After the injections, the eggs and adults mosquitoes should be under observation to determine if there is any obvious phenotypic change, such as abnormal or slower development and ability to survive.

If the knockdowns are successful there are a number of interesting directions for this project. We could make transgenic lines that express the shRNAs to achieve the knockdown of *piwi*. Then we could determine, by a TE microarray, whether TEs activity was altered. It is hypothesized that release of suppression will increase transcription of potentially active TEs. Positive results could be further verified by northern blots. Another possibility would be to expose the mosquito to a virus. The rasiRNA pathway is thought to be involved in protecting the genome from viruses. When the mosquitoes are without this defense mechanism there is a probability that the virus might be able to replicate at a faster rate. Perhaps over expression of the members of the Agonaute family might not let the virus replicate at all. Determining the rate of replication could possibly be done by quatitative PCR of the virus several days post infection.

Bibliography

Adelman, Z. N., N. Jasinskiene, et al. (2002). "Development and applications of transgenesis in the yellow fever mosquito, Aedes aegypti." Mol Biochem Parasitol **121**(1): 1-10.

Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-402.

Aravin, A., D. Gaidatzis, et al. (2006). "A novel class of small RNAs bind to MILI protein in mouse testes." Nature **442**(7099): 203-7.

Aravin, A. A., M. Lagos-Quintana, et al. (2003). "The small RNA profile during Drosophila melanogaster development." Dev Cell **5**(2): 337-50.

Aravin, A. A., N. M. Naumova, et al. (2001). "Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline." Curr Biol **11**(13): 1017-27.

Bernstein, E. and C. D. Allis (2005). "RNA meets chromatin." Genes Dev **19**(14): 1635-55.

Birney, E., T. D. Andrews, et al. (2004). "An overview of Ensembl." Genome Res **14**(5): 925-8.

Bohmert, K., I. Camus, et al. (1998). "AGO1 defines a novel locus of Arabidopsis controlling leaf development." Embo J **17**(1): 170-80.

Boisson, B., J. C. Jacques, et al. (2006). "Gene silencing in mosquito salivary glands by RNAi." FEBS Lett **580**(8): 1988-92.

Brennecke, J., A. A. Aravin, et al. (2007). "Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila." Cell **128**(6): 1089-103.

Brosius, J. (1999). "RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements." Gene **238**(1): 115-34.

Calvo, E., M. Walter, et al. (2005). "Nanos (nos) genes of the vector mosquitoes, Anopheles gambiae, Anopheles stephensi and Aedes aegypti." Insect Biochem Mol Biol **35**(7): 789-98.

Campbell, C. L., K. M. Keene, et al. (2008). "Aedes aegypti uses RNA interference in defense against Sindbis virus infection." BMC Microbiol **8**: 47.

Cerutti, L., N. Mian, et al. (2000). "Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain." Trends Biochem Sci **25**(10): 481-2.

Corpet, F. (1988). "Multiple sequence alignment with hierarchical clustering." Nucleic Acids Res **16**(22): 10881-90.

Cox, D. N., A. Chao, et al. (1998). "A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal." Genes Dev **12**(23): 3715-27.

Cox, D. N., A. Chao, et al. (2000). "piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells." Development **127**(3): 503-14.

Craig, N. A. (2002). Mobile DNA: an Introduction. Mobile DNA II, ASM Press**:** 3-11.
Djikeng, A., H. Shi, et al. (2001). "RNA interference in Trypanosoma brucei: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs." Rna **7**(11): 1522-30.
Filipowicz, W., L. Jaskiewicz, et al. (2005). "Post-transcriptional gene silencing by siRNAs and miRNAs." Curr Opin Struct Biol **15**(3): 331-41.
Finnegan, E. J., M. Wang, et al. (2001). "Gene silencing: fleshing out the bones." Curr Biol **11**(3): R99-R102.
Fontanillas, P., D. L. Hartl, et al. (2007). "Genome organization and gene expression shape the transposable element distribution in the Drosophila melanogaster euchromatin." PLoS Genet **3**(11): e210.
Girard, A., R. Sachidanandam, et al. (2006). "A germline-specific class of small RNAs binds mammalian Piwi proteins." Nature **442**(7099): 199-202.
Grivna, S. T., B. Pyhtila, et al. (2006). "MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis." Proc Natl Acad Sci U S A **103**(36): 13415-20.
Gunawardane, L. S., K. Saito, et al. (2007). "A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila." Science **315**(5818): 1587-90.
Hamilton, A., O. Voinnet, et al. (2002). "Two classes of short interfering RNA in RNA silencing." Embo J **21**(17): 4671-9.
Hartl, D. L., A. R. Lohe, et al. (1997). "Modern thoughts on an ancyent marinere: function, evolution, regulation." Annu Rev Genet **31**: 337-58.
Hoa, N. T., K. M. Keene, et al. (2003). "Characterization of RNA interference in an Anopheles gambiae cell line." Insect Biochem Mol Biol **33**(9): 949-57.
Holt, R. A., G. M. Subramanian, et al. (2002). "The genome sequence of the malaria mosquito Anopheles gambiae." Science **298**(5591): 129-49.
Juhn, J. and A. A. James (2006). "oskar gene expression in the vector mosquitoes, Anopheles gambiae and Aedes aegypti." Insect Mol Biol **15**(3): 363-72.
Kapitonov, V. V. and J. Jurka (2001). "Rolling-circle transposons in eukaryotes." Proc Natl Acad Sci U S A **98**(15): 8714-9.
Kapitonov, V. V. and J. Jurka (2003). "Molecular paleontology of transposable elements in the Drosophila melanogaster genome." Proc Natl Acad Sci U S A **100**(11): 6569-74.
Kidwell, M. G. and D. R. Lisch (2001). "Perspective: transposable elements, parasitic DNA, and genome evolution." Evolution Int J Org Evolution **55**(1): 1-24.
Kim, A., C. Terzian, et al. (1994). "Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of Drosophila melanogaster." Proc Natl Acad Sci U S A **91**(4): 1285-9.
Lau, N. C., A. G. Seto, et al. (2006). "Characterization of the piRNA complex from rat testes." Science **313**(5785): 363-7.
Lerman, D. N., P. Michalak, et al. (2003). "Modification of heat-shock gene expression in Drosophila melanogaster populations via transposable elements." Mol Biol Evol **20**(1): 135-44.
Lin, H. (1997). "The tao of stem cells in the germline." Annu Rev Genet **31**: 455-91.

Lin, H. and A. C. Spradling (1997). "A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary." Development **124**(12): 2463-76.
Lippman, Z. and R. Martienssen (2004). "The role of RNA interference in heterochromatic silencing." Nature **431**(7006): 364-70.
Llave, C., K. D. Kasschau, et al. (2002). "Endogenous and silencing-associated small RNAs in plants." Plant Cell **14**(7): 1605-19.
Makalowski, W. (2000). "Genomic scrap yard: how genomes utilize all that junk." Gene **259**(1-2): 61-7.
Massimiliano Marsano, R. and R. Caizzi (2005). "A genome-wide screening of BEL-Pao like retrotransposons in Anopheles gambiae by the LTR_STRUC program." Gene **357**(2): 115-21.
Matzke, M. A. and J. A. Birchler (2005). "RNAi-mediated pathways in the nucleus." Nat Rev Genet **6**(1): 24-35.
McClintock, B. (1951). "Chromosome organization and genic expression." Cold Spring Harb Symp Quant Biol **16**: 13-47.
Mette, M. F., T. Kanno, et al. (2002). "Endogenous viral sequences and their potential contribution to heritable virus resistance in plants." Embo J **21**(3): 461-9.
Moreira, L. A., J. Ito, et al. (2002). "Bee venom phospholipase inhibits malaria parasite development in transgenic mosquitoes." J Biol Chem **277**(43): 40839-43.
Moussian, B., H. Schoof, et al. (1998). "Role of the ZWILLE gene in the regulation of central shoot meristem cell fate during Arabidopsis embryogenesis." Embo J **17**(6): 1799-809.
Nene, V., J. R. Wortman, et al. (2007). "Genome Sequence of Aedes aegypti, a Major Arbovirus Vector." Science.
O'Brochta, D. A., W. D. Warren, et al. (1996). "Hermes, a functional non-Drosophilid insect gene vector from Musca domestica." Genetics **142**(3): 907-14.
Pal-Bhadra, M., U. Bhadra, et al. (2002). "RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila." Mol Cell **9**(2): 315-27.
Pal-Bhadra, M., U. Bhadra, et al. (2004). "Interrelationship of RNA interference and transcriptional gene silencing in Drosophila." Cold Spring Harb Symp Quant Biol **69**: 433-8.
Reinhart, B. J. and D. P. Bartel (2002). "Small RNAs correspond to centromere heterochromatic repeats." Science **297**(5588): 1831.
Robinson, A. S., G. Franz, et al. (2004). "Insect transgenesis and its potential role in agriculture and human health." Insect Biochem Mol Biol **34**(2): 113-20.
Saito, K., K. M. Nishida, et al. (2006). "Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome." Genes Dev **20**(16): 2214-22.
Schmidt, A., G. Palumbo, et al. (1999). "Genetic and molecular characterization of sting, a gene involved in crystal formation and meiotic drive in the male germ line of Drosophila melanogaster." Genetics **151**(2): 749-60.

Slotkin, R. K. and R. Martienssen (2007). "Transposable elements and the epigenetic regulation of the genome." Nat Rev Genet **8**(4): 272-85.
Song, E. S., V. Lee, et al. (1997). "Antigen presentation in retroviral vector-mediated gene transfer in vivo." Proc Natl Acad Sci U S A **94**(5): 1943-8.
Swofford, D. L. (2002). "PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)."
Tabara, H., M. Sarkissian, et al. (1999). "The rde-1 gene, RNA interference, and transposon silencing in C. elegans." Cell **99**(2): 123-32.
Thompson, J. D., T. J. Gibson, et al. (1997). "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." Nucleic Acids Res **25**(24): 4876-82.
Tu, Z. and C. Coates (2004). "Mosquito transposable elements." Insect Biochem Mol Biol **34**(7): 631-44.
Vagin, V. V., M. S. Klenov, et al. (2004). "The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of Drosophila melanogaster." RNA Biol **1**: 54-58.
Vagin, V. V., A. Sigova, et al. (2006). "A distinct small RNA pathway silences selfish genetic elements in the germline." Science **313**(5785): 320-4.
Vastenhouw, N. L. and R. H. Plasterk (2004). "RNAi protects the Caenorhabditis elegans germline against transposition." Trends Genet **20**(7): 314-9.
Volpe, T. A., C. Kidner, et al. (2002). "Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi." Science **297**(5588): 1833-7.
WHO. (2002). "Strategic direction for research in Dengue." 2006, from http://www.who.int/tdr/diseases/dengue/direction.htm.
Williams, R. W. and G. M. Rubin (2002). "ARGONAUTE1 is required for efficient RNA interference in Drosophila embryos." Proc Natl Acad Sci U S A **99**(10): 6889-94.
Zamore, P. D. and B. Haley (2005). "Ribo-gnome: the big world of small RNAs." Science **309**(5740): 1519-24.

Appendix A

Table 6: Piwi summary occurence

Piwi	Piwi5	Piwi6	Piwi7
Piwi5 0-2 hr	728	353	49
Piwi5 2-4 hr	193	194	14
Piwi5 4-8 hr	88	86	9
Piwi6 0-2 hr	353	798	45
Piwi6 2-4 hr	194	253	12
Piwi6 4-8 hr	86	106	10
Piwi7 0-2 hr	49	45	194
Piwi7 2-4 hr	14	12	24
Piwi7 4-8 hr	9	10	16

The single occurrence column shows the number of hits that only pertains for that particular *piwi* and is not found in any other *piwi*. The rest of the columns represent the different *piwi´*s and they show the number of hits that are shared between them.