A mosquito-specific *bZIP* transcription factor and the influence of a Y-specific gene on sex determination in *Anopheles stephensi*

Frank Criscione

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

Doctor of Philosophy
In
Biochemistry

Zhijian (Jake) Tu, Chair
Jianyong Li
Igor Sharakhov
Brenda Winkel
Jinsong Zhu

(7^{th} of June 2013)
Blacksburg, Virginia

Keywords: bZIP1, mosquito, miRNA, sex determination, Y chromosome, vector control, GUY1, female lethal

Abstract

Aside from few model organisms, little is known about early embryonic development or sex determination in insects, in particular mosquitoes which are major vectors of worldwide disease. The goals of this work were to investigate a mosquito-specific transcription factor and its intronic miRNA cluster and characterize a novel Y chromosome gene in *An. stephensi*. The aims of these experiments were to expand on the knowledge of genes involved in embryonic development and sex determination with potential application in vector control strategies.

In *Ae. aegypti* a mosquito-specific bZIP1 transcription factor was demonstrated to be conserved among divergent mosquito species. It was maternally and zygotically-expressed and knock-down of bZIP1 mRNA via siRNA microinjection in the embryo resulted in embryonic death. The expression profile of this gene was determined through the use of RT-PCR and qRT-PCR. Additionally, this gene contains a miRNA cluster that is also relatively conserved amongst members of the Culicidae family suggesting its evolutionary importance. The miRNAs are also maternally and zygotically expressed and are the most abundant embryonic miRNAs as determined by small RNA sequencing and Northern analysis. Promoter activity for bZIP1 was characterized and the promoter was used to direct maternal and zygotic transgene expression in *An. stephensi*.

Y chromosome genes were successfully identified in *An. stephensi* from Illumina sequencing data. This work focused on a gene unique to the Y 1 (GUY1). It was shown that GUY1 was male specific and linked to the Y chromosome. RT-PCR and single embryo analysis suggested that GUY1 was expressed during the maternal to zygotic transition and was only expressed in male embryos. It was shown in multiple transient and transgenic assays that the

ectopic expression of GUY1 can influence the sex of subjected individuals and skew sex distribution to a male bias.

There is still much to be investigated before a GUY1-based transgenic line can be tested and implemented for use in vector population control. However, the work in this dissertation represents a major step towards novel mosquito control strategies based on the manipulation of Y chromosome genes.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

miRNA	micro RNA
bZIP	basic leucine zipper
siRNA	small interfering RNA
RT-PCR	reverse transcriptase polymerase chain reaction
qRT-PCR	quantitative RT-PCR
ddPCR	digital droplet PCR
DENV	dengue virus
Mb	Mega base
RNAi	RNA interference
dsRNA	double stranded RNA
MEDEA	maternal-effect dominant embryonic arrest
UTR	untranslated region
tra	transformer
dsx	doublesex
fru	fruitless
sxl	sex-lethal
SIT	sterile insect technique
RIDL	release of insects carrying a dominant lethal
GMI	genetically modified insect
EtOH	ethanol
dNTP	deoxynucleotide
$MgCl_2$	magnesium chloride
DTT	dithiothreitol
cDNA	complementary DNA

DEPC	diethylpyrocarbonate
RACE	rapid amplification of cDNA ends
RPS7	ribosomal protein subunit 7
LC-MS/MS	liquid chromatography tandem mass spectrometry
TBE	tris/borate/EDTA
SSC	saline sodium citrate
SDS	sodium dodecyl sulfate
BLAST	basic local alignment search tool
RNAseq	RNA sequence
EST	expressed sequence tag
TSA	transcriptome shotgun assembly
GUY1	gene unique to the Y 1
DAPI	4', 6-diamidino-2-phenylindole
FISH	florescence *in situ* hybridization
gDNA	genomic DNA
PBS	phosphate buffer saline
EDTA	Ethylenediaminetetraacetic acid
RLU	relative light unit
bHLH/HTH	basic helix-loop-helix/helix-turn-helix
RPS4	ribosomal protein subunit 4
CNV	copy number variation
G0	generation zero

Chapter 1

Introduction

1.1 Mosquito development and disease

Vector borne disease is a global problem faced on nearly every continent of the planet, having detrimental effects on economies and the health of humans in both developed and underdeveloped countries. Mosquitoes contribute to the vast majority of vector borne disease and inspire the work described in this dissertation. Specifically, this work focuses on *Aedes aegypti*, the dengue vector, and *Anopheles stephensi*, the Asian malaria vector. Together, malaria and dengue account for more than 600 million cases of disease each year (www.who.int, www.cdc.gov).

Ae. aegypti has been widely studied due to its role in the transmission of dengue virus, yellow fever and chikungunya (Miller *et al.*, 1996). *Ae. aegypti* is a Culicinae mosquito with a 1.38 gigabase genome (47% of which is comprised of transposable elements), approximately 250 million years diverged from *Drosophila* (Nene *et al*. 2007) and approximately 145-200 million years divergent from *Anopheles gambiae*, the major African malaria vector (Krzywinski *et al*., 2006). *Ae. aegypti* mainly inhabits the tropical and subtropical regions of the planet (Patz *et al.*, 1996). Upon the laying of eggs there is a 96 hr developmental process before the egg is capable of hatching into the aquatic larval stage (Ramini and Cupp 1975). The eggs are able to survive long periods of desiccation which facilitates the problematic spread of the vector to new locations as climates tend to be warming. The larval stage consists of three molts before reaching the pupae stage. The time spent in the larval stage can vary from five days to two weeks depending on the water temperature and availability of food (Southwood *et. al*., 1972). During the two day pupae stage a size and developmental difference is observed between males and females, with females being larger and developing at a slower rate (Southwood *et al*., 1972). In

the adult stage there are distinct morphological differences between males and females (e.g. antennae structure) and sexual maturity, defined as the capability for females to be inseminated, occurs approximately 36 hours post emergence (Gwadz, 1972).

The viruses taken from an infectious blood meal spread to the salivary glands where it will then be released during the subsequent blood meal. In the case of the RNA virus dengue, there are four serotypes (DENV I-IV) differing in their surface antigens (Kurane *et al.,* 1992). Infection with one type of dengue does not afford protection against the other three types, rather the reaction to a second form of dengue usually results in severe symptoms (Gubler 1997). After the 4-7 day incubation period a high fever and vomiting ensues, for which the main treatment is fluid replenishment. Severe cases result in dengue hemorrhagic fever, the leading cause of death by dengue virus in humans. There is currently no effective vaccine against dengue virus (Simmons *et al*., 2012).

An. stephensi is a dipteran insect with a genome size of approximately 235 Mb (Jost and Mameli, 1972) and there are approximately 10 million years of divergence separating *An. stephensi*, the main malaria vector in Asia, from *An. gambiae* (Gaunt and Miles 2002). This species contains two diploid chromosomes and a set of sex chromosomes wherein XX and XY are the female and male karyotypes, respectively (Robinson *et al.* 1987, Clements 1992). The Y chromosome contributes an estimated 10% (Krzywinski *et al.,* 2004) of the total genome size. This insect has four life stages with the first three being obligate aquatic stages: embryo, larvae, pupae, and adult. The developmental process takes approximately two weeks from egg to adult (Clements 1992).

An. stephensi are susceptible to infection by a variety of *Plasmodium* species, the parasite that causes malaria. First, female mosquitoes ingest a blood-meal infected with *Plasmodium*

gametes. The 14-day sexual reproduction stage begins in the midgut with the combining of male and female gametes to form a fertilized oocyst. This stage ends with the rupturing of the oocyst and release of sporozites into the mosquito's body. The sporozites then migrate to the salivary glands where they are released with the mosquito's saliva into the blood stream of the victim animal during the next blood meal (James *et al*., 1932, Boyd 1938). The parasite has evolved to have minimal adverse effects on its mosquito host while maintaining infection for the duration of the mosquito's life (Collins *et al*., 2002).

Approximately one-third of the world's population is at risk of infection with members of the genus *Plasmodium*, the parasites that cause malaria. Symptoms of this disease include fever, headache, coma and death. The mortality malaria is highest in the elderly and young children (James *et al*., 1932). In the Middle East and Asia a major malaria vector is *An. stephensi*. Transmission of the disease to humans requires contact between the infected mosquito and a human host. Over 14 days after initial mosquito infection, the mosquito is capable of transmitting disease. Despite the odds of an infected mosquito surviving this period of time, as it must overcome obstacles in the wild, malaria remains a major vector borne disease.

1.2 miRNA biogenesis and application

MicroRNAs (miRNAs) are small RNA molecules that are part of a larger class of small RNAs involved in the process of RNA interference (RNAi). These small RNA molecules approximately 22 nucleotides in length (Bartel 2004, Ambros 2004), are present in plants and animals and modulate post-transcriptional regulation via mRNA degradation, translational repression or destabilization of mRNA by complementary targeting of the 3'UTR of mRNA (Prillai *et al*., 2007). In plants this complementation is usually a perfect base pair match;

however, in animals there may be some mis-matches (Rajewsky 2006, Bentwich 2005). The seed region, the first 2-7 nucleotides of the miRNA, is most often seen as having a perfect match to its target.

MicroRNAs are encoded by nuclear DNA and usually transcribed by RNA polymerase II (Lee *et al*., 2004) when derived from their own genes (Biemar *et al*., 2005). After transcription, the 5' end of the transcript is capped and the 3' end is polyadenylated. Characteristic of miRNA, they form a hairpin structure via self complementation known as the primary miRNA (pri-miRNA). This secondary structure is recognized by the nuclear protein Drosha, and then cleaved by Posha, releasing the pre-miRNA. Alternatively, miRNA can be found in the introns of genes. Such is the case of the miRNA described in these studies, located in the intron of a bZIP gene. In this case, after the intron is spliced from pre-mRNA by the spliceosome it is further processed by nucleotide excision repair proteins in conjunction with Dicer (Ying and Lin, 2004). Intronic miRNAs are usually under the governance of their host gene promoters; however, there are cases of incongruence between host gene and miRNA expression suggesting independent control of processing of these miRNAs (Isik *et al.*, 2010).

The pre-miRNA are then exported to the cytosol via Ran-GTP/Exportin mediated transfer where they will be further processed and incorporated into the RNA interference silencing complex (RISC). Upon exit from the nucleus, the loop of the hairpin is cleaved by Dicer, leaving a 3' overhang and generating a miRNA/miRNA* dsRNA molecule. The miRNA is loaded into the protein, Argonaut, and the miRNA* is usually degraded (Lee *et al*., 2004). In plants, cleavage of the hairpin is performed in the nucleus by a Dicer-like protein (Kutter and Svoboda, 2008). Once loaded into argonaut the miRNA acts as a guide strand for complementation-based

mRNA silencing or degradation. Occasionally, the miRNA* can also be loaded into argonaut and be functionally active (Yang *et al*., 2010).

From an evolutionary standpoint, miRNA evolve relatively slowly and often posses a range of targets. Highly divergent organisms maintain very similar miRNA and target couplings. Families of miRNA, or miRNA with similar target recognition motifs, tend to be conserved within the kingdoms of plants and animals (Maher *et al*. 2006, Tanzer and Stadler, 2004).

miRNA targeting in plants is usually based on perfect complementation, making target prediction relatively easy. However, in animals the allowing of mismatches makes target prediction a more difficult task. Algorithms have been designed to both use bioinformatics to identify potential miRNA based on secondary structure and thermodynamic stability, and predict miRNA targets based on complementation (Rajewsky 2006 and Bentwich 2005). Direct approaches of target prediction would include either overexpression or knockout of miRNA with subsequent transcriptional analysis at miRNA expression time points to reveal positive miRNA targets (Lim 2005). Additionally, experimental target prediction through the preservation of the miRNA/mRNA target duplex and direct identification of a miRNA target has been performed in *C. elegans* (Andachi *et al*., 2008). In *Ae. aegypti* and *D. melanogaster* there are currently 124 and 426 identified miRNAs, respectively (www.miRbase.org).

miRNAs function as modifiers of translational activity, thus making an ideal tool for *in vivo* manipulation of genes. The sequences of native miRNA can be altered to target the 3'UTR of genes of interest to perform knockdowns. As seen with the synthetic maternal-effect dominant embryonic arrest (MEDEA) system in *Drosophila*, miRNA are used as a "toxin" to drive individuals with an "antidote" through the population. The antidote is a modified miRNA target

gene without the target 3'UTR (Chen *et al*., 2007). When coupled with the appropriate promoter, these artificial miRNAs can cause tissue or stage specific RNAi response (Du *et al*. 2006).

1.3 Sex determination

Sex determination is a developmental process that results in the molecular and morphological sexual identification of organisms (Charlesworth and Mank 2010, Barton and Charlesworth 1998). XY sex determination is but one of the mechanisms of sex determination employed by mammals, insects, and plants (Charlesworth and Mank 2010). In humans, the SRY gene on the Y chromosome initiates the cascade of events leading to the generation of a phenotypically male organism (Sinclair *et al*., 1990). However, there are many other forms of sex determination utilized in other organisms. In eupherian organisms, such as the platypus, sex is determined by multiple sets of sex chromosomes; the males are heterogametic and females are homogametic (Rens *et al*., 2004). All therian mammals have similar XY sex chromosome systems, while all birds have similar ZW sex chromosome systems (Graves 2008). In birds, a ZW sex determination mechanism results in the heterogametic gender being female and the ZZ homogametic gender being male (Bull, 1983). The process in insects varies in complexity depending on the species. The least conserved portions of the pathway are its initiators, but downstream genes such as *tra*, *dsx*, and *fru* are relatively conserved (Saccone *et al*., 2002). Insects have multiple examples of sex chromosome karyotypes including: X:A ratio, XX/XY, ZZ/ZW, multiple X, and haplodiploidy (Marin and Baker 1998, Kaiser and Bachtrog 2010, Rai and Black 1999). In the case of colony-forming social insects, haplodiploidy is used. This process allows for an unfertilized egg to develop into a haploid male organism, whereas the diploid organism is the result of fertilization and usually female. Sterility is often the outcome of

diploid males. This method allows for the queen to select what type of offspring will be produced by either fertilizing or not fertilizing embryos with sperm stored in the spermatheca after mating (Cook 1993). Other organisms such as *C. elegans* are capable of hermaphroditic sexual development where one individual can produce the gametes of both sexes (Hodgkin 1983).

Even though there has been a great deal of research done with mosquitoes their methods of sexual determination are unknown. Manipulating this pathway could be vital to the effort against vector disease transmission since these diseases are only spread by the females. The closest organism to mosquitoes with a well characterized sex determination pathway is *D. melanogaster*.

D. melanogaster utilizes a X:A ratio sex determination system, different than both *Ae. aegypti* and *An. stephensi* (Bridges 1921, Nene 2007, Baker and Sakai 1972). This pathway consists of a complex sequence of alternatively-spliced transcripts that lead to the development of the fly into two sexes, morphologically and behaviorally distinct. The initial signal in the cascade of sex determination is provided by *sex-lethal* (*sxl*); which responds to the ratio of autosomes to X chromosomes. Females have a high X:A ratio relative to males; this difference is acknowledged by the embryo and initiates a cascade which causes the activation of *sxl* (Boggs *et al*. 1987). There are three proteins involved in the initiation of the sex determining cascade in *Drosophila*, *sisterless-a*, a bZip transcription factor; *sisterless-b*; and *daughterless*, a bHLH protein. These proteins interact in concert to sense the quantity of X chromosomes against autosomes and generate a male or female embryonic environment without the direct input of the Y chromosome to initiate the *sxl* cascade (Cronmiller and Cline 1987, Parkhurst and Meneely

1994). *Sxl* then auto-regulates its own transcription and causes the splicing of a functional isoform of *transformer* (*tra*) in females.

Female-specific TRA is capable of forming a complex with constitutively-active TRA2. This complex splices *doublesex* (*dsx*), the protein responsible for morphological sex-specific features, into the female-specific isoform (Amrein *et al*., 1988). In males the X:A ratio is low, thereby not activating *sxl*. When *sxl* is not active a non-functional isoform of *tra* is generated. TRA2, in the presence of non-functional TRA, leads to a male-specific splicing of *dsx* (Robinson *et al*.1987, Gabrieli, *et al*. 2010).

Sex-specific behavior has been linked to *fruitless* (*fru*), which is responsible for the development of musculature and central nervous system structures that vary between the sexes. The regulation of *fru* is also under control of TRA in that TRA promotes sex-specific alternative splicing (Gailey *et al*., 2005).

An. stephensi differs from *Drosophila* in the initiation of this cascade. There is evidence suggesting that *Anopheles* mosquitoes have a male-determining factor on the Y chromosome (Baker and Sakai 1979). No homologs of *sxl* or *tra* have been found in the *An. gambiae* or *An. stephensi* genomes, however, there is a homolog of *dsx* that is alternatively spliced in males and females in both species (Tu lab, unpublished, Scali *et al*., 2005). This suggests the presence of a protein capable of generated sex-specific isoforms of *ds*x similar to TRA. *Tra* and other members of the pathway are conserved in a variety of insects (Saccone *et al*., 2002, Shearman 2002, Verhulst *et al*., 2010). The master regulator of this pathway in *An. stephensi*, the male-determining factor, would function above or in place of *sxl* and initiate the sex determination cascade similar to the post X:A ratio recognition step in *Drosophila*. In most insects with a characterized Y chromosome the identified Y chromosome genes are usually related to male

sexual development, fertility, and behavior (Malone and Oliver 2008, Carvalho *et al*., 2000, Hackstein and Hochstenback 1995, Kennison 1981).

1.4 Y chromosomes

The evolution of the Y chromosome remains a mystery due to the various ways a Y chromosome can be generated and how genes arise on a Y chromosome. Y chromosomes accumulate mutations and become a reservoir of transposable elements, a consequence of their low rate of recombination (Steinemann and Steinemann 2005, Gvozdev *et al*. 2005). It has been theorized that the Y chromosome is in a constant state of flux; periodically being degraded and remade. The sex chromosomes are thought to have originated from a degenerate pair of autosomes (Bachtrog 2006 and Rice 1996). As alleles advantageous to one sex or harmful to the opposite sex accumulate on a chromosome there reaches a point where the pair no longer recombines and natural selection begins to favor Y chromosome retention (Bull 1983 and Rice 1987). The degeneration of the Y chromosome adds new selective pressure for maintaining the genes on the Y chromosome as they provide sex-specific advantageous functions (Charlesworth and Charlesworth 2000). Interestingly, genes from the Y chromosome most often have autosomal derivation, rather than X chromosome derivation. There are a few models for Y chromosome evolution in dipterans: complete independence, where the Y chromosome arises with speciation; partial independence, where the Y chromosomes of some species share a common ancestor; or single origin, where the sex chromosomes all evolved from a common ancestral pair of autosomes (Pease and Hahn 2012). The more widely accepted theory is that the sex chromosomes eveolved from a degenerate pair of autosomes. Most of the genes present on the Y chromosome are involved with spermatogenesis, mating behavior, or general male fitness

(Carvalho *et al*., 2001; Lemos *et al*., 2008). Although Y chromosomes do not always play a direct role in male sex determination, as seen in *Drosophila*, the function of Y chromosome genes still remain beneficial to males (Bridges 1921). With very low conservation of Y chromosome genes between closely-related species of *Drosophila* (Koerich *et al*., 2008), homology-based gene discovery would be ineffective as a method to identify Y chromosome genes in other species.

Methods of assembling Y chromosome data have not been successful in the past and have led to the severe under-representation of Y chromosome sequences in published genomes. Separation of male and female genomic libraries has been hypothesized to be sufficient to identify Y chromosome sequences (Krzywinski *et al*., 2004). However, due to the very nature of the Y chromosome, an area of the genome that is highly heterochromatic, repeat rich, and low in unique gene content, much more processing of the data needs to be done. Due to the high likelihood of Y chromosome genes being duplications of genes from other chromosomes, differentiating between recently-duplicated Y chromosome genes and their autosomal originator based on next generation sequencing data becomes difficult (Marshall-Graves, 2006, Carvalho *et al*., 2003). As of now, no Y chromosome specific genes have been found in any mosquito species, nor has a male-determining factor on the Y chromosome been characterized in any insect.

1.5 Vector control strategies

Currently, the use of pesticides and bed-netting are the most effective strategies for a defense against insect disease vector populations. However, the use of pesticides is a non-specific control method with a large ecological range and regularly needs to be modified due to

the emergence of resistance (Raghavendra *et al*., 2011). New strategies are being developed aimed at the reduction or eradication of mosquito populations and the replacement of the vector population with a genetically-modified strain that is refractory to disease. Strategies aimed at manipulating the sex ratio of a pest population have previously been investigated and shown to be effective methods for decreasing a population (Robinson 1983). The sterile insect technique (SIT), which succeeded in eradicating the screw-worm in North America (Bushland, 1955), and the release of insects carrying a dominant lethal (RIDL) are two strategies aimed at reducing the mosquito population (Thomas *et al*., 2000). In *An. stephensi* for example, there have also been strategies aimed at generating populations that are capable of expressing antisporozoite peptides in their salivary glands and thereby reducing infectivity rates (Sumitani *et al*., 2012).

Current SIT strategies being investigated for vector control rely heavily on the production of large quantities of male mosquitoes, an approach that is both costly and laborious. Manipulation of sex determination towards a male bias could offset these obstacles and increase feasibility. There has yet to be a male-determining factor characterized in any insect and the discovery of this factor would greatly assist in the genetic strategies involved with SIT and RIDL. *An. stephensi* undergoes XY sex determination with the male determining factor on the Y chromosome; however, discovery of this factor would be a challenge caused by the difficulty of Y chromosome assembly and gene identification. Nonetheless, the information gained by studying sex-determination in *An. stephensi* could translate to new genetic strategies in *An. gambiae*, the world's leading malaria vector.

Although sex determination occurs in the early embryo, as seen by the sex-specific splicing of key transcripts involved in the sex determination pathway (Scali *et al*., 2005), sex-specific morphological features become apparent during the larval stage (Suman *et al*., 2008).

The required sorting performed for SIT and RIDL takes advantage of the larger body mass of females during the pupae stage (Papathanos *et al.*, 2009, Marshall and Taylor, 2009). During the late pupae or early adult stage is the time when chemical, genetic, or radioactive sterilization of males occurs, readying them for release to compete for mates with wild type males (Helinsky *et al*., 2009, Catteruccia *et al*., 2009). Mating between wild type females and sterile males results in no offspring and decreased receptivity of females to new mates. When multiple releases are performed this competition leads to a decrease in population size due to a decrease of progeny over the generations. The requirement for multiple releases of large quantities of sterilized male insects makes SIT costly; however, this technique was effective in combatting the screwworm epidemic in North and Central America (Marshall and Taylor, 2009) and has been employed as a population control in mosquitoes (Benedict and Robinson, 2003).

RIDL involves the release of males that carry a gene that would result in death of the female after mating or production of non-viable progeny. This gene must cause lethality while not influencing the competence of the male. RIDL also requires multiple, substantial releases of genetically modified males into the wild-type population. To ensure that lethality occurs in the field and not in the rearing facility the lethality gene could be placed under the control of a conditional promoter, inactive in the presence of a chemical during rearing and active when that chemical is not present. Social, ethical and economical issues are raised in the event that genetically modified insects (GMIs) are released into the environment. The recent release of genetically modified mosquitoes was followed by a volatile public response due to the lack of policy surrounding the release of GMIs; however, policy to support these efforts is being generated (Enserink 2010).

1.6 Research aims

The aims of the work described in this dissertation were to investigate a novel intronic miRNA cluster and its host gene in *Ae. aegyti* and to identify and characterize a Y chromosome gene in *An. stephensi*. The evolutionary conservation of the miRNA cluster and the bZIP1 gene in relationship to other mosquito species was investigated through sequence analysis. The expression of the miRNA cluster was examined through small RNA sequencing and confirmed using Northern blot analysis. Expression of the host gene was examined through RT-PCR. Finally, the function of the gene was probed through the use of siRNA targeting bZIP1 to elicit bZIP1 mRNA knockdown.

Working with *An. stephensi* Y chromosome genes first required the identification of Y chromosome gene candidates discovered through a comparative genomic bioinformatic approach. From this list of 45 candidate genes, one gene, GUY1, was selected and shown through the use of PCR, ddPCR, and single embryo RT-PCR to be Y chromosome-specific. The function of this gene is explored through the use of transient assays and transgenic mosquitoes engineered to ectopically express GUY1.

Chapter 2

A conserved mosquito-specific *bZIP1* gene is maternally and zygotically transcribed, functionally important, and provides the most abundant intronic miRNAs in the *Aedes aegypti* embryo

Chapter 2 was written as a manuscript.

Contributing Authors:

Frank Criscione

Northern blot analysis, bZIP1 gene characterization and expression quantification, bZIP1 phylogenitic inference, siRNA knockdown assays, and writing of manuscript.

Matthew Smith

Sequencing of bZIP1 in species other than *An. stephensi* and *Ae. aegypti*.

Zhijian Tu

Phylogenic inference of miRNA, small RNA bioinformatics including mapping of reads to predicted hairpins, and writing of manuscript.

2.1 Abstract

We are interested in studying the evolutionary contributions of lineage-specific genes in mosquito embryonic development. Here we report the characterization of the *bZIP1* gene which encodes a protein that contains a basic leucine zipper domain (bZip domain). We show that *bZIP1* is conserved and the order of its neighboring genes is maintained in all mosquitoes analyzed, including species in the divergent *Aedes, Culex*, and *Anopheles* genera. *bZIP1* is restricted to mosquitoes and its phylogeny is congruent with the mosquito species phylogeny. We experimentally defined the transcriptional unit of the *bZIP1* gene in *Ae. aegypti* and *An. stephensi* and showed that it is predominantly expressed maternally and zygotically in both species. When the *bZIP1* transcript is knocked down by injection of siRNAs into the *Ae. aegypti* embryo, hatch rate reduced significantly, indicating the functional importance of *bZIP1* during embryonic development. We have previously identified three pre-miRNAs of the miR-2941 family in the first intron of the *Ae. aegypti bZIP1*. Using illumina sequencing and northern blot analysis, here we show that all three pre-miRNAs are transcribed and processed, maintaining an expression pattern similar to that of the *bZIP1* gene. These *Ae. aegypti* miRNAs are by far the most abundant maternally-deposited miRNAs and they are also the most abundant miRNAs in other embryonic stages. They share the same seed sequence and potentially target a number of maternally and zygotically expressed genes which have similarities to genes involved in gene regulation, protein degradation, and the cytoskeleton. The *bZIP1* intron in all four *Anopheles* species studied contains a single predicted miRNA with the same seed sequence as the *Ae. aegypti* miR-2941 family, suggesting that this miRNA family is conserved in all mosquitoes and possibly expanded in the Culicinae lineage. Thus the conserved mosquito-specific *bZIP1* is

critically important to mosquito embryonic development and provides a family of highly abundant intronic miRNAs.

2.2 Introduction

Maternally deposited transcripts and proteins control the initial events during early embryonic development when the early embryo is transcriptionally inactive (Schier, 2007). During the maternal to zygotic transition, maternal RNAs are degraded and zygotic transcription, transcription from the genome of the nascent organism, gradually takes over as the source for new transcripts and proteins (e.g., De Renzis *et al*., 2007; Hooper *et al*., 2007; Walser and Lipshitz, 2011; Wieschaus, 1996). Both maternally contributed factors and early zygotic gene products are important in orchestrating the temporal and spatial specificity of gene activities which defines embryonic development. In many cases, developmentally important genes may be maternally transcribed and loaded into the embryo as well as transcribed in the zygote. One such example is *Zelda* (also known as *Vielfaltig*), a key transcription activator of the *Drosophila melanogaster* zygotic genome (Liang *et al*., 2008).

Small RNAs, including microRNAs (miRNAs), endogenous small interfering RNAs (endo_siRNAs), and piwi-interacting RNAs (piRNAs), are also important players in embryonic development (Pauli *et al*., 2011). miRNAs have been shown to be involved in cellular reprogramming and clearing of maternal transcripts (Giraldez, 2010). MicroRNAs are approximately 22 nucleotide long non-coding RNAs that modulate gene expression by binding to cognate mRNA's for cleavage or translational repression. In *Drosophila*, the primary miRNA is processed by a Drosha-Pasha complex to yield ~ 70 nt stem-loop structures called pre-miRNAs (Saito *et al*., 2005). Following export to the cytosol, the pre-miRNA is further processed to

produce an miRNA:miRNA* duplex (Lau *et al*., 2001; Bartel, 2004; Miska *et al*., 2005), which are then separated to produce the mature miRNA. Some intronic miRNAs bypass Drosha processing and are referred to as mirtrons (Martin *et al*., 2009). Drosophila embryonic miRNAs showed temporal reciprocity to their target mRNAs during the maternal to zygotic transition, indicating their involvement in clearance of maternal transcripts (Bushati *et al*., 2008). Mammalian miR-427 has been shown to mediate deadenylation of maternal mRNA (Lund *et al*., 2009). Moreover, miRNAs are also important in developmental reprogramming, directing cellular differentiation, and regulation of pluripotency of the embryonic cells (Giraldez, 2010; Pauli *et al*., 2011).

The molecular mechanism of mosquito embryonic development is not well understood. Although *Drosophila* and mosquitoes had diverged approximately 250 million years ago (Krzywinski *et al*., 2006), comparative analysis between the two dipteran groups proved to be informative. The spatial and temporal expression profiles of approximately 20 different dorsal-ventral patterning genes suggests that the initial specification of the mesoderm and ventral neurogenic ectoderm is highly conserved in flies and mosquitoes (Goltsev *et al*., 2007). A whole-genome comparative gene expression analysis revealed distinct embryonic patterning mechanisms between *D. melanogaster* and *An. gambiae* (Papatsenko *et al*., 2011). Recent studies showed that siRNA knockdown of the *Ae. aegypti* homologs of the *D. melanogaster Frazzled* and *Semaphorin-1a* genes affect the mosquito central nervous system development, confirming overall functional conservation of these genes, although subtle differences were noted (Clemons *et al*., 2011; Haugen *et al*., 2011). Comparative analysis between the genome of *Drosophila* and the three published mosquito genomes (Holt *et al*., 2002; Nene *et al*., 2007; Arensburger *et al*., 2010) revealed that several genes known to be critical for *Drosophila* development are not found

in one or more of the mosquito species (Behura *et al*., 2011). Furthermore, mosquito developmental genes appear to share common repetitive sequences which may be targets of miRNAs (Behura *et al*., 2011).

We are interested in studying the evolutionary contributions of lineage-specific genes to the maternal and zygotic control of mosquito embryonic development. We have previously shown that gene duplication and subsequent acquisition of new expression patterns gave rise to novel early zygotic genes in mosquitoes (Biedler and Tu, 2010). We also described an embryonically expressed and mosquito-specific miRNA cluster in an intron of a hypothetical gene in *Ae. aegypti* and *Culex quinquefasciatus* (Li *et al*., 2009). This hypothetical gene is predicted to encode a protein that contains a basic leucine zipper domain (bZIP domain). Hence we named the gene *bZIP1*. Here we describe the molecular, transcriptional, and evolutionary analysis of the *bZIP1* gene in divergent mosquito species. We show that the conserved mosquito-specific *bZIP1* provides a protein that is critically important to mosquito embryonic development and a family of highly abundant intronic miRNAs.

2.3 Materials and methods

Mosquito rearing

Ae. aegypti Liverpool strain mosquitoes were reared in incubators at 28°C and 60% relative humidity on a 16 hr light/8 hr dark photoperiod. Larvae were fed Sera Micron Fry Food with brewer's yeast, and Purina Game Fish Chow. Mosquitoes were blood-fed with live female Hsd:ICR (CD-1®) mice (Harlan Laboratories, http://www.harlan.com). Eggs were collected from 72 to 96 hr post blood meal (Biedler and Tu, 2010). The rearing conditions for *An. stephensi* Indian strain were similar to the above except that the temperature was set at 27°C. *Anopheles*

albimanus adults were obtained from MR4 (MRA-156, MR4/ATCC, Manassas, VA, http://www.mr4.org).

Microsynteny analysis, genomic DNA PCR to obtain bZIP1 in An. albimanus, and naming of the mosquito bZIP1 genes

The order of genes in the *bZIP1* genomic region was analyzed using Vectorbase (vectorbase.org) annotations for *An. gambiae*, *Cx. Quinquefasciatus,* and *Ae. aegypti* and by BLAST comparisons to newly obtained genomic sequences. The *An. darlingi* sequence is from a GenBank assembly (EFR25602.1) and sequence of *An. stephensi* is obtained by PCR connecting two contigs (GenBank accession JQ266223). Based on the conservation of gene order, we decided to use degenerate primers designed according to the genes flanking *bZIP1* (See Figure 2.S1 for microsynteny and names of the flanking genes) to amplify *bZIP1* from *An. albimanus* genomic DNA. Degenerate primers were designed according to conserved regions of the two flanking genes, which produced the forward (5'- GCNRTNAAYTTYTGGTAYGAYATG-3') and reverse (5'-YTGRAANGCYTTRTTNGCRTCRTA-3') primers. LA Taq (Takara Bio Inc., Shiga, Japan) and the following cycling conditions were used: denaturation at 95°C for 30 seconds, annealing at 60.5°C for 30 seconds, and extension at 72°C for 4 minutes. PCR products were gel purified and cloned into pGEM-T Easy (Promega, Madison, WI). Sequencing was performed at the Virginia Bioinformatics Institute (Blacksburg, VA) by primer walking. The *An. albimanus* genomic sequence has been deposited in GenBank (accession JQ266224). We use the first letter of the genus name and the first three letters of the species name to specify the source of a *bZIP1* gene or protein. For example, *Aste_bZIP1* refers to *An. stephensi bZIP1*.

RNA isolation for analysis of mRNA and small RNA

Dissections of tissue samples were performed in 50% RNAlater (Ambion, Austin, TX), on individual, glass slides. To reduce the chances of cross tissue contamination the collections were performed in groupings. The first grouping was to obtain the head, thorax and abdomen. The second grouping focused on obtaining parts of the abdomen (Malp. tub., midgut, ovary). Each tissue was pooled and rinsed in 50% RNA later. Excess RNA later was removed from all samples then immediately flash frozen in liquid nitrogen. RNA isolations were performed using the *mir*Vana miRNA Isolation Kit (Ambion) following the total RNA isolation protocol which retains both small RNA and mRNA. Flash frozen sample was homogenized with 1:10 lysis/binding buffer on ice. Homogenate additive equal to 1/10 the volume of homogenate was added, vortexed, and left on ice for 10 mins. One volume of acid-pehonol:chloroform was added and the mixture was vortexed for 60 secs. The aqueous phase was collected after 5 mins centrifugation at 10,000 g. To the lysate 1.25 volumes of room temperature 100% EtOH was added. The lysate/ethanol mixture was passed through the filter cartridge at 10,000g. After washing, RNAs were eluted with 100 µL of 95°C elution solution. In the case of RT-PCR and qRT-PCR, eluted RNA was further treated with the TURBO DNA-*free* kit (Ambion) to remove DNA. All RNA was stored at –80°C if not immediately used.

Reverse transcription PCR (RT-PCR) for Ae. aegypti and An. stephensi samples

Reverse transcription of total RNA was performed using the Superscript III first-strand cDNA synthesis kit (Invitrogen, Carlsbad, CA). Five micrograms of RNA was combined with oligo dT(20) primer, 1 mM dNTPs, and incubated at 65°C for 5 minutes. The chilled mixture was then combined with 10X RT buffer, 200U Superscript III RT, 10mM $MgCl_2$, 40U

RNaseOUT, and 20mM DTT (final concentrations). First-strand synthesis was performed at 50°C for 50 minutes followed by 85°C for 5 minutes. RNase H (2U) was added to the reaction and incubated at 37°C for 20 minutes. Replacement of superscript RT with DEPC-treated H_2O was performed to generate negative controls. 50 ng of cDNA was used per PCR reaction. Primers were designed to flank the first intron of *Ae. aegypti bZIP1* (*Aaeg_bZIP1*) and *An. stephensi* bZIP1 (*Aste_bZIP1*), respectively. The primers for the internal reference control target a region of 40S ribosomal protein S7 (RPS7) in *Ae. aegypti* and RPS4 in *An. stephensi*. The Takara Taq PCR kit was used for 30 cycle PCR reactions with the following cycling conditions: denaturation at 95°C for 30 seconds, annealing at 60.5°C for 30 seconds, and extension at 72°C for 1.5 minutes. The samples were run on a 1.3% agarose gel containing ethidium bromide for visualization. The primer sequences are shown below.

Aaeg_bZIP1 forward 5'-ATGGTTTTCGGATCAAAGG-3',

Aaeg_bZIP1 reverse 5'-GGAATTCGAACGTAACGTCAC-3',

Aaeg_RPS7 forward 5'-GCTTTGACCGTTTTGATCGT-3',

Aaeg_RPS7 reverse 5'-GAGTACCGATCGTTTCGCAT-3',

Aste_bZIP1 forward, 5'-CTCGACGACATCGGCCGAGAA-3',

Aste_bZIP1 reverse, 5'-ACGATCGCACTGGCGCCTTC-3',

Aste_RPS4 forward 5'-GAGTCCATCAAAGGAGAAAGTCTAC-3',

Aste_RPS4 reverse 5'-TAGCTGGCGCATCAGGTAC'-3.

Rapid Amplification of cDNA Ends (RACE)

Total RNA isolated from 0-12 hr was used for the synthesis of both the *Ae. aegypti* and *An. stephensi* 5' and 3' RACE ready cDNA. The SMARTer^{TM} RACE cDNA amplification kit

(Clontech, Mountain View, CA) was used starting with 2 µg of RNA incubated in the reaction mixture for 3 hrs to create libraries of RACE ready cDNA. The final cDNA product was diluted with 100 µL of sterile H_2O. High melting temperature primers (~70°C) were designed to amplify the 5' and 3' ends of AAEL009263 (*Aaeg_bZIP1*) and *Aste_bZIP1* including an overlapping region for assembly and control purposes:

Aaeg_bZIP1 forward, 5'-TATGCCGTGGATTCGCCCCA-3';

Aaeg_bZIP1 reverse, 5'-GCGCACTGTTCAGCTTTTCCGC-3';

Aste_bZIP1 forward, 5'-CTCGACGACATCGGCCGAGAA-3';

Aste_bZIP1 reverse, 5'-ACGATCGCACTGGCGCCTTC-3'.

Touchdown PCR with the Takara Taq PCR Kit was performed at: 5 cycles of 94°C for 30secs, 72°C for 3mins; 5 cycles of 94°C for 30secs, 70°C for 30secs, and 72°C for 3mins; 27 cycles of 94°C for 30secs, 68°C for 30secs, 72°C for 3mins. Products were run on a 1.1% agarose gel and purified with the illustra GFX PCR DNA and gel band purification system (GE Healthcare, Piscataway, NJ). Isolated bands were cloned into pGEM-T Easy (Promega), transformed into Mach T-1 cells (Invitrogen). Positive plasmids were sequenced at the Virginia Bioinformatics Institute.

Quantitative RT-PCR (qRT-PCR) of Ae. aegypti samples and data analysis

Twenty µl triplicate TaqMan® Gene Expression Assays (ABI) were set up utilizing cDNA generated as stated in Section 2.3, TaqMan® Universal PCR master mix, and 20X assay mix. Assay mixes were generated for amplification of both *Aaeg_bZIP1* and RPS7. Cycling condition on the ABI 7300 system were as follows: 50°C for 2 minutes, 95°C for 10 minutes, and 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. The data was analyzed using the

ABI 7300 system software under ddCt relative quantification settings with RPS7 as the endogenous control (Sengul and Tu, 2010). ddCt values are relatively quantitative to the calibrator of each data set. The fold change for each sample measured against the calibrator yields the relative quantification (RQ) value, which was then $\log_{10}$ transformed.

Primers and probe were as follows:

Aaeg_bZIP1:

Forward primer- 5'-CGTCCATTGTTTGTGATACAGGATTG-3'

Reverse primer- 5'-TGGGTGTCTGGATGTAATGAATTGT-3'

Probe- FAM5'-CTGCGAGCTTTAAAGATGGC-3'NFQ

Aaeg_ RPS7:

Forward primer 5'-CGCGCTCGTGAGATCGA-3'

Reverse primer- 5'-GCACCGGGACGTAGATCA-3'

Probe- FAM5'-ACAGCAAGAAGGCTATCG-3'NFQ

Protein identification via mass spectroscopy

Ae. aegypti embryos (500 µg) were sent to ITSI Bio for LC-MS/MS. The sample was homogenized in 10% deoxycholic acid/25mM ammonium bicarbonate buffer to extract the proteins. Protein concentration determined using ITSI Total protein assay kit. Quantified protein was reduced, alkylated and trypsin digested overnight. The digested samples will be processed and then be re-suspended in the appropriate mobile phase for LC-MS/MS analysis. A linear acetonitrile gradient will be used to separate the tryptic peptides based on their hydrophobicity prior to MS analysis on a Thermo Scientific LTQ XL mass spectrometer and the total run time 120 minutes.

Transgenic bZIP1 promoter activity in An. stephensi

An. stephensi were transformed at the Insect Transformation Facility at the University of Maryland, Rockville by piggyBac transposition of a construct containing the 1kb upstream promoter of *An. stephensi* bZIP1 driving luciferase and DsRed driven by a 3xP3/HSP70 promoter. Positive adults were treated (either fed on sugar or blood fed) then flash frozen at the appropriate time points. The luciferase assays were performed following the manufacturer's suggestions for the Dual Luciferase Assay (Promega) were followed with the following modifications: 50 uL of Passive Lysis Buffer was used to homogenize the embryo, then 20 uL of Luciferase Assay Reagent I was added before the reading (See appendix A for detailed results).

Small RNA Illumina sequencing

Small RNA cloning was performed using the small RNA sample prep kit v1.0 (Illumina). Small RNA isolations were performed using the *mir*Vana miRNA Isolation Kit (Ambion). Total RNA was fractionated on a 15% polyacrylamide/urea gel at 200V for 1 hour. The segments corresponding to the 18-30 nucleotide lengths were eluted and purified. After ligating to the 5' and 3' adaptors provided by Illumina, the RNA molecules were reverse transcribed into cDNA. cDNA were amplified via PCR for 11 cycles using adaptor primers. The products were gel purified and comprised the library ready for Illumina sequencing. Prior to Illumina sequencing, the quality of the library was verified by cloning and sequencing a small fraction of the library through conventional cloning using pCR-Blunt II-TOPO vector (Invitrogen) and Sanger sequencing at the Virginia Bioinformatics Institute. Illumina sequencing was performed at the Virginia Commonwealth University Illumina Facility.

Small RNA northern blot analysis

All equipment was rinsed with DEPC-treated water to remove RNases. Samples were denatured at 95°C and run on a 15% polyacrylamide gel at 150V in 1x TBE (Bio-Rad). 19- and 22-nucleotide oligomers were used as size markers. After staining the gel with ethidium bromide for visualization the RNAs were transferred to a Brightstar Plus membrane (Ambion) in 0.5x TBE at 4°C for 1.5 hours at 200mA. The membrane was then crosslinked with a SpectroLinker UV cross-linker at the optimal setting (120 mJ/cm^2). The membrane was then prehybridized at 45°C while rotating with ULTRAhyb™-Oligo hybridization buffer (Ambion) for 30 minutes, after which 5' Dig-labelled antisense LNA probe (Exiqon, Vedbaek, Denmark; aae-miR-2941 5'-TCCGTGGAGTTCTAGCCGTACTA-3', aae-miR-2946 5'-TCCCCATATCTTTTCCGTACTA-3') was added to a final concentration of 0.1 nM and left overnight. Upon completion of hybridization the membrane was washed twice with a stringent wash buffer (2x SSC and 0.1% SDS) for 35 minutes each. The subsequent washes and blocks occurred at 25°C. One wash was for 5 minutes in low stringency wash buffer (0.1M maleic acid, 0.15M NaCl, 0.1%Tween-20, pH 7.5). The membrane was blocked with a solution of 0.5g powdered milk in 10 ml low stringency wash buffer for 30 minutes. Treatment with antibody was performed with 1uL of anti-Dig labeled antibodies in 10 ml blocking solution for 1 hour. Three washes for 15 minutes each with low stringency wash buffer and 2 washes for 5 minutes each with alkaline phosphatase Buffer (0.1M Tris, 50mM $MgCl_2$, 0.1M NaCl, 0.1% Tween, pH 9.5) were performed. Visualization of the membrane was achieved by 5 minutes of room temperature incubation with CDP-star followed by exposure to X-ray film.

siRNA injections, phenotypic observation, and statistical analysis

Silencer siRNAs (Ambion) were designed to target the 3'UTR of *Aaeg_bZIP1* and RFP (non-targeting control). siRNAs were injected into the *Ae. aegypti* embryos 40 minutes after oviposition. Serial dilutions of siRNA concentrations were injected ranging from 10-100 μM to determine the effective dosage. We define the effective siRNA dosage as the concentration of siRNA which yields the highest hatch rate when injected with the RFP targeting siRNA. 20 μM siRNA was selected as the optimal dosage for subsequent injections. During each experiment a handled but un-injected set of embryos was also allowed to hatch. The development of the un-injected embryos was gauged against the injected sets in regard to rates of growth and viability. Physical development was monitored by microscopy during the larval stage to check for abnormalities. qRT-PCR was used to verify transcript knockdown. All data for qRT-PCR were collected in biological and technical triplicates. Hatch rate data was collected from triplicate injection sets. Student t-test was used to evaluate statistical significance between the experimental and control groups. A p-value < 0.05 is used to establish significance.

RNAseq and small RNAseq database analysis

The four embryonic RNAseq databases were described in Biedler and Tu (2010, Geo accession # GSE34480) and the reads are 33 bp in length. To estimate the relative transcription levels of the *Aaeg_bZIP1* gene isoforms, both transcripts (AAEL009263-RA and AAEL-009263-RB) were used as query to BLAST (Altschul *et al*., 1997) against these four RNAseq database and the e-value cutoff was set at 1e-7. Bowtie mapping (Langmead *et al*., 2010) were also performed in addition to the above mentioned BLAST-based analysis and similar results were obtained (not shown). To estimate the relative abundance of the two isoforms in the four

embryonic samples, we used the exon-exon junctions unique to each splice variant as query to compare with the RNAseq data (Figure 2.2 and Table 2.2). 18bp of each side of junction were retrieved to produce a 36bp exon-exon junction. Normalization method is the same as Biedler and Tu (2010) and described in the footnotes of Table 2.2. To investigate the relative expression levels of the three small RNAs in the *Aaeg_bZIP1* intron and to identify their respective miRNA*, the pre-miRNA hairpins were used as query to BLAST (Altschul *et al*, 1997) against the four small RNAseq database (Geo accession # to be provided) using an e-value of 0.01. All known *Ae. aegypti* miRNAs were downloaded from miRBase (mirbase.org, version 17) and used as query to obtain the total miRNA reads in each sample. These totals were used to normalize the comparison between the four embryonic time points (Table 2.3).

Phylogenetic inference

bZIP1 protein sequences (Table 2.1, not shown; Figure 2.S2) were aligned using Clutalw (Thompson *et al*., 1994) with gap opening penalty of 5 and gap extension penalty of 0.05. Phylogenetic relationship of the sequences and the credibility of the groupings were inferred using MrBayes (Huelsenbeck and Ronquist, 2001, 2003) after running 100,000 generations.

Prediction and analysis of miRNA targets

Target mRNAs of aae-miR-2941 and aae-miR-2946 were predicted using miRanda version 3.3a (Enright *et al*., 2003). The 3' UTRs of all annotated *Ae. aegypti* transcripts (Vectorbase transcripts version L1.2) were used as input in the analysis and the cutoff score is set at 150. The predicted targets were searched against the RNAseq databases as mentioned above to determine whether they are transcribed in these embryonic time points. The predicted target

sequences were also searched against the non-redundant protein databases to identify homologous sequences with predicted functions. Gene ontology data associated with these genes were also downloaded from Vectorbase.

2.4 Results and discussion

bZIP1 orthologs are restricted to mosquitoes and the bZIP1 phylogeny is congruent with species phylogeny among divergent mosquitoes

The *bZIP1* gene is annotated as 1:1:1 orthologue in the published genomes of *Ae. aegypti*, *Culex quinquefasciatus*, and *An. gambiae* (AAEL009263, CPIJ000468, and AGAP007767, respectively). We performed multiple BLAST searches using *bZIP1* genes and predicted proteins from the three species as queries against non-redundant nucleotide, protein, genome, expressed sequence tag (EST) and transcriptome shotgun assembly (TSA) databases at NCBI (ncbi.nlm.nih.gov). Similar sequences were only identified in *Anopheles darlingi* (a newly available genome assembly, EFR25602.1), *Anopheles funestus* (TSA, EZ975750.1), and *Armigeres subalbatus* (TSA, EU210875.1 and EU206037.1). No sequences from any non-mosquito species showed a match with an e-value better than 0.3. During this study we obtained a genomic sequence that contains the *bZIP1* gene in *An. stephensi* (submitted to GenBank with accession JQ266223) by PCR connecting two contigs from a preliminary genome assembly. As shown in S1, there is a clear conservation of microsynteny surrounding the *bZIP1* gene in all species with sufficiently long genomic sequences available, including mosquitoes in the divergent *Aedes, Culex*, and *Anopheles* genera. Thus this microsynteny has been maintained for approximately 145-200 million years (Krzywinski *et al*., 2006). We designed degenerate primers according to conserved amino acid sequences in *bZIP1* and the neighboring genes to PCR

amplify the full-length *bZIP1* from *Anopheles albimanus* genomic DNA. A 3.5 kb sequence was obtained (submitted to GenBank with accession JQ266224). As summarized in Table 2.1, *bZIP1* has thus far been identified in five *Anopheles* species as well as *Ae. aegypti*, *Ar. subalbatus*, and *Cx. quinquefasciatus*. It is a single copy gene in all species for which a genome assembly is available.

Predicted protein sequences of all *bZIP1* genes were aligned (Figure 2.S2) and a phylogeny was reconstructed based on the alignment (Figure 2.1). The *bZIP1* phylogeny is congruent with the mosquito phylogeny (Krzywinski and Besansky, 2003; Biedler and Tu, 2007). Within the *Culicinae* lineage, *Ae. aegypti* is grouped with *Ar. subalbatus* and *Culex* is more divergent from *Aedes* and *Armigeres*. Within the *Anopheles* genus, the three species within the Celia subgenus, *An. gambiae, An. funestus*, and *An. stephensi* are grouped together, while the two species within the *Nyssorhynchus* subgenus, *An. albimanus* and *An. darlingi* are grouped together.

Experimental characterization of the bZIP1 cDNA in Ae. aegypti and An. stephensi

To characterize the *Ae. aegypti bZIP1*, we first searched RNAseq databases obtained in a different study (Table 2.2; Biedler and Tu, 2010; Geo accession # GSE34480). The RNAseq was done using *Ae. aegypti* samples from different embryonic stages. As shown in Table 2.2, *Aaeg_bZIP1* appears to have both maternal (0-2 hr embryo, see Biedler and Tu, 2010) and zygotic (later embryonic time points) expression. According to Vectorbase annotation, *Aaeg_bZIP1* has two splice isoforms (AAEL009263-RA and –RB) where the –RA transcript skips a 159 bp internal exon that is found in –RB (Figure 2.2). We compiled short query sequences that span the exon junctions unique to –RA and –RB transcripts (Figure 2.2). The -

RA-specific exon junction (between exons 2 and 3) has 219 total hits in the RNAseq samples. The 5' RB-specific junction (between exons 2 and 3) has 0 hits while the 3' RB-specific junction (between exons 3 and 4) has 2 hits (Table 2.2). It is thus likely that the –RA isoform is the predominant splice variant in this time range. 5' and 3' RACE were performed using cDNA isolated from *Ae. aegypti* 0-12 hr old embryos. Only cDNA derived from –RA was recovered and sequencing of the RACE products confirmed the exon-intron boundaries and revealed the transcription start and poly(A) addition sites (Figure 2.2, Figure 2.S2). Full transcript sequence has been submitted to GenBank with accession # JQ266221. 5' and 3' RACE were also performed using cDNA isolated from *Anopheles stephensi* 0-12hr old embryos. A single cDNA source was recovered and sequencing of the RACE products confirmed the predicted exon-intron boundaries and revealed the transcription start and poly(A) addition sites (Figure 2.2; Figure 2.S2). Full transcript sequence has been submitted to GenBank with accession # JQ266222.

Predominantly maternal and zygotic expression of bZIP1 in Ae. aegypti and An. stephensi

We performed extensive RT-PCR and qRT-PCR to obtain a detailed expression profile of *Aaeg_bZIP1* transcription. Given the small amplicon requirement of TaqMan qRT-PCR and that –RA is the predominant isoform, our primers span the first two exons and the probe covers the exon junction to minimize gDNA contamination. As shown in Figure 2.3, a broad survey by RT-PCR suggests that *Aaeg_bZIP1* transcription is mostly limited to embryonic stages and the adult female ovaries. Further qRT-PCR analysis showed that blood feeding induced *Aaeg_bZIP1* transcription (Figure 2.4A). qRT-PCR is also used to determine the tissue-specific profile of *Aaeg_bZIP1* in adult females 48 hrs after blood feeding (Figure 2.4B). When the three body parts were compared, the abdomen *Aaeg_bZIP1* level was more than 100 and 1000 fold higher

than the thorax and head samples, respectively. In the abdomen, the ovary had the highest level of transcription, more than 10 and 100 fold higher than the malpighian tubles and midgut, respectively. This is consistent with maternal expression of *Aaeg_bZIP1*. To understand the transcription of *Aaeg_bZIP1* during the maternal to zygotic transition, we determined the relative *Aaeg_bZIP1* mRNA levels across the first 12 hrs during embryonic development (Figure 2.4C). The presence *of Aaeg_bZIP1* transcripts in 0-1 hr embryos suggests that the mRNA is maternally deposited and the initial increase in *Aaeg_bZIP1* mRNA levels in subsequent time points suggests new transcripts are made during early zygotic transcription. Therefore, *Aaeg_bZIP1* is both maternally deposited and zygotically transcribed. *Aaeg_bZIP1* protein is present in the early embryo confirming the translation of embryonic transcript (Table 2.2). Zygotic expression of *Aaeg_bZIP1* is also consistent with data showing that the *Aaeg_bZIP1* promoter reporter construct can drive luciferase expression in the early embryos (not shown). The *Anst_bZIP1* promoter driving luciferase in transgenics behaved in accordance with transcriptional data; maternal expression in response to blood feeding (data not shown). According to RT-PCR, the overall transcription profile of *Aste_bZIP1* in *An. stephensi* is the same as that of the *Aaeg_bZIP1 in Ae. aegypti* (Figure 2.3).

bZIP1 is important during Ae. aegypti embryonic development

To investigate the function of *bZIP1* in *Ae. aegypti* embryo, we injected siRNA targeting *Aaeg_bZIP1* into the embryo approximately 40 minutes after egg-laying. Injection of siRNA targeting *bZIP1* reduced the *bZIP1* mRNA in the embryo by more than 10 fold compared to injection of the anti-RFP siRNA control (Figure 2.5). Injection of siRNA targeting *Aaeg_bZIP1* shows an adverse effect on embryonic development. The percent of embryos to hatch into larvae

after injection was 40.9±10.2% and 5.9±2.6% ($p<0.05$) for anti-RFP and anti-*bZIP1* siRNA, respectively. The decrease in hatch rate of embryos injected with anti-*bZIP1* suggests a vital role of *bZIP1* in embryonic development. The anti-RFP group controls for injection trauma or siRNA toxicity. The surviving embryos injected with siRNA targeting either RFP or *bZIP1* appeared normal. The fact that a small fraction of the embryos survived the anti-*bZIP1* siRNA could be explained by either these embryos obtaining sufficient amount of *bZIP1* protein prior to siRNA effect or due to insufficient RNAi knockdown. It is also possible that a small fraction of the embryos simply did not receive a sufficient amount of the siRNA due to the variable nature of microinjection.

Aaeg_bZIP1 harbors an intronic miRNA cluster that shows strong maternal and embryonic expression and provides the most abundant miRNA family in a range of embryonic stages

We previously described a miRNA cluster in the first intron of *Aaeg_bZIP1* (Li *et al*., 2009). This cluster consists of aae-miR-2946, aae-miR-2941-1 and aae-miR-2941-2, the latter two of which have the same mature sequence but slightly different miRNA*. In northern blot analyses, these miRNAs showed strong signals in the embryos but were hardly detectable in larvae, pupae, adult males, or sugar-fed adult females (Li *et al*., 2009). Here we carried out a detailed survey by northern blot of the expression profile of both aae-miR-2941 and aae-miR-2946. As shown in Figure 2.6A, both miRNAs are readily detectable in 0-2 hr, suggesting they are maternally deposited. There appears to be a decline at 2-4 hrs but the levels increase thereafter, indicating new miRNAs are made during early zygotic transcription. Therefore, similar to *Aaeg_bZIP1*, these miRNAs are both maternal and zygotic. Both miRNAs are present throughout embryonic development (data not shown). When tissue samples from blood-fed

females were analyzed, expression of both miRNAs was predominantly in the ovary (Figure 2.6B). Signals for both miRNAs are very weak in sugar-fed adult females but clearly induced after blood feeding (Figure 2.6C). Therefore, it appears that the intronic miRNAs and its "host" gene *Aaeg_bZIP1* share a coordinated expression profile.

Northern analysis described above does not differentiate between aae-miR-2941-1 and aae-miR-2941-2. We thus used the pre-miRNA haipins of the three intronic miRNAs as queries to search the recently obtained small RNA sequencing databases (Table 2.3, Geo accession # to be provided). Illumina sequencing produced reads that match the putative miRNA* strand for all three hairpins (Figure 2.7). There are no other sites matching the aae-miR-2941 and aae-miR-2946 sequences in the *Ae. aegypti* genome assembly. As shown in Figure 2.7, maps of small RNA sequences demark the predicted miRNA and miRNA* boundaries, consistent with the canonical processing of pre-miRNA hairpins. Although the numbers are low compared to the miRNA sequences, the existence of miRNA* which have unique sequences suggest that both miR-2941 hairpins contribute to the generation of miR-2941.

At the four embryonic time points sequenced, the sum of miRNA-2941 and miR-2946 reads represents 40-95% of all identifiable miRNA reads (Table 2.3). Thus the *Aaeg_bZIP1* gene is responsible for the most abundant miRNAs in a broad range of embryonic stages including 0-1 hr after egg-laying when only maternally deposited miRNAs are present (Biedler and Tu, 2010). The ratio of miR-2941 to miR-2946 varies from the expected ~2:1 to a highly biased ~60:1 (Table 2.3). Although biased amplification could alter the miRNA ratio within a given sample, it is unlikely the explanation here as the observed variation is between different samples. Differential processing of the pre-miRNAs and differential stability of miRNAs are known mechanisms to control miRNA levels (Obernosterer *et al*., 2006; Bail *et al*., 2010), both of which

could result in different levels of miRNAs from the same levels of pre-miRNAs. The small RNA northern shown in Figure 2.6 cannot be used to compare the miRNA-2941 and miR-2946 levels as different probes are used and they are from different experiments. Quantitative measurement of the miRNAs in this cluster is needed to explain the variation in the miR-2941/miR-2946 ratios observed in the illumina data.

Predicting and analyzing the targets of miR-2941 and miR-2946 in Ae. aegypti

Forty-seven and 52 *Ae. aegypti* transcripts are predicted to be the targets of aae-miR-2941 and aae-miR-2946, respectively (data not shown). Seventeen of these transcripts are targets of both aae-miR-2941 and aae-miR-2946. As shown in S4, forty of the 47 targets of aae-miR-2941 have significant transcription in the early embryonic time points (at least 30 reads during 0-12 hrs) while 48 of the 52 targets of aae-miR-2946 have significant transcription in the early embryonic time points (at least 30 reads in 0-12 hrs). Gene ontology information about these genes was downloaded from Vectorbase but we could not extract informative insights from these data. Thus we used the predicted targets as queries to search the non-redundant protein database to identify homologs with suggested functions (e-value of 1e-10 or better). For miR-2941, predicted targets comprise 9 transcripts with homology to genes involved in cell division and the cytoskeleton, 8 transcripts with homology to sequences that may be involved in gene regulation either through interaction with DNA or histones or through signal transduction, and 6 transcripts with homology to sequences that may be involved in protein degradation (data not shown). For miR-2946, predicted targets comprise 8 transcripts with homology to genes involved in cell division and cytoskeleton, 10 transcripts with homology to sequences that may be involved in gene regulation either through interaction with DNA or histones or through signal transduction, 6

transcripts with homology to sequences that may be involved in protein degradation, and 8 transcripts that are similar to genes involved in transport (data not shown). These functions are consistent with molecular processes involved in embryonic development.

miRNA hairpin(s) with the same 7-8 bp seed sequence as miR-2941 and miR-2946 are found in bZIP1 of all mosquitoes

We have previously shown that the *Cx. quiquefasciatus bZIP1* gene contains an intron with two aae-miR-2941 hairpins and a third hairpin that encodes an expressed miRNA that shares the same 8bp seed sequence but is different from aae-miR-2941 and aae-miR-2946. No miR-2941 and miR-2946 homolog was found in *An. gambiae* in our previous analysis (Li *et al*., 2009). Searches of the miRbase and Rfam performed during the current study failed to identify any known miRNA or small RNAs that are similar to the miR-2941 and miR-2946 sequences outside mosquitoes. However, there is a single hairpin in the first intron of *Agam_bZIP1* (AGAP007767) and it has a predicted miRNA sequence that starts with the UAGUACG seed sequence shared by aae-miR-2941 and aae-miR-2946 (Figure 2.8). As we obtained the *bZIP1* genomic sequences from *An. stephensi*, *An. darlingi*, and *An. albimanus* (Table 2.1), we were able to compare these *Anopheles* genomic sequences, and identified predicted pre-miRNAs with a single hairpin for each of these species (Figure 2.8). Comparative analysis and RNA folding clearly indicate that all these *Anopheles bZIP1* harbors a single pre-miRNA that could produce similar mature miRNAs with the same 7-8 bp seed sequence as the aae-miR-2941 family. All members of the aae-miR-2941/2946 family across all mosquito species are found in the *bZIP1* intron prior to the start of coding sequence. Therefore these miRNAs that share the same 7-8 bp seed sequences are likely of the same evolutionary origin. Sequence relationships among the

Anopheles miRNAs and pre-miRNAs are consistent with species phylogeny, whereby sequences from species in the same subgenus share higher similarity than those from species in different subgenus (Figure 2.8A). As mentioned above, the *Aedes* and *Culex bZIP1* contain three related and duplicated hairpins (Figure 2.8B). Thus there appears to be an expansion of the miRNA family in the Culicinae lineage. It is noteworthy that the *Anopheles* miR-2941 members would not have been discovered without targeted analysis of the intron of the *bZIP1* gene because only 7-8 bp conservation exist between the *Anopheles* and the known Culicinae miR-2941 members. Therefore knowing orthologous relationships improves annotation of short non-coding sequences residing in conserved genes.

2.5 Concluding remarks

We have presented a molecular, transcriptional, and evolutionary analysis of the *bZIP1* gene in divergent mosquito species. The bZIP1 protein sequences enabled us to infer a phylogeny that is congruent with the species phylogeny and the intronic miRNA family provided additional evolutionary information. The *bZIP1* gene region maintained conserved microsynteny in the three highly divergent mosquito genera. We were able to take advantage of the conserved gene order to amplify *bZIP1* from *An. albimanus* using degenerate primers. Thus it is feasible to obtain the *bZIP1* sequence from diverse mosquitoes for phylogenetic analysis and the *bZIP1* gene may be a good phylogenetic marker to resolve mosquito phylogeny.

Comparison of mosquito and *Drosophila* embryonic development can be informative as there are common genes and shared temporal and spatial expression patterns. However, there are evolutionary innovations in both lineages since they diverged 250 million years ago (Krzywinski *et al*., 2006). The discovery of a functional important *bZIP1* gene that is restricted to but

conserved in all mosquitoes highlighted one such difference between the two dipteran lineages The ability to maternally deposit global effectors into the embryo is a process understood in *Drosophila* but convoluted in mosquitoes. Given the high and predominantly maternal and zygotic expression of *bZIP1* in *Aedes* and *Anopheles* species, and given its potential involvement in DNA-binding and transcriptional regulation, *bZIP1* is likely a critically important gene in mosquito development. This is consistent with the significantly reduced hatch rate when *bZIP1* is knocked down. It will be informative to determine the genes that are regulated by the *bZIP1* protein.

bZIP1 may also contribute to embryonic development by producing a family of mosquito-specific intronic miRNAs. There is experimental evidence for three such miRNAs in both *Ae. aegypti* and *Cx. quinquefasciatus* (Li *et al*., 2009, and this study). Here we have shown that in *Ae. aegypti*, these miRNAs are by far the most abundant maternally-deposited miRNAs and they are also the most abundant in other embryonic stages. They share the same seed sequence and potentially target a number of maternally and zygotically expressed genes which have similarities to genes involved in gene regulation, protein degradation, and the cytoskeleton. All four *Anopheles* species studied contain one predicted miRNA with the same seed sequence as the *Ae. aegypti* miR-2941 family, suggesting that this miRNA family is conserved in all mosquitoes and perhaps expanded in the Culicinae lineage. It will be interesting to determine whether these intronic miRNAs are also expressed in *Anopheles* embryos and to uncover their targets. It will also be interesting to determine whether this miRNA family contributed to the evolutionary difference between flies and mosquitoes and even between mosquito lineages.

From a practical standpoint in vector-borne disease control, bZIP1 provides evidence of critical, maternally inherited, early embryonic transcriptional and translational regulation. The

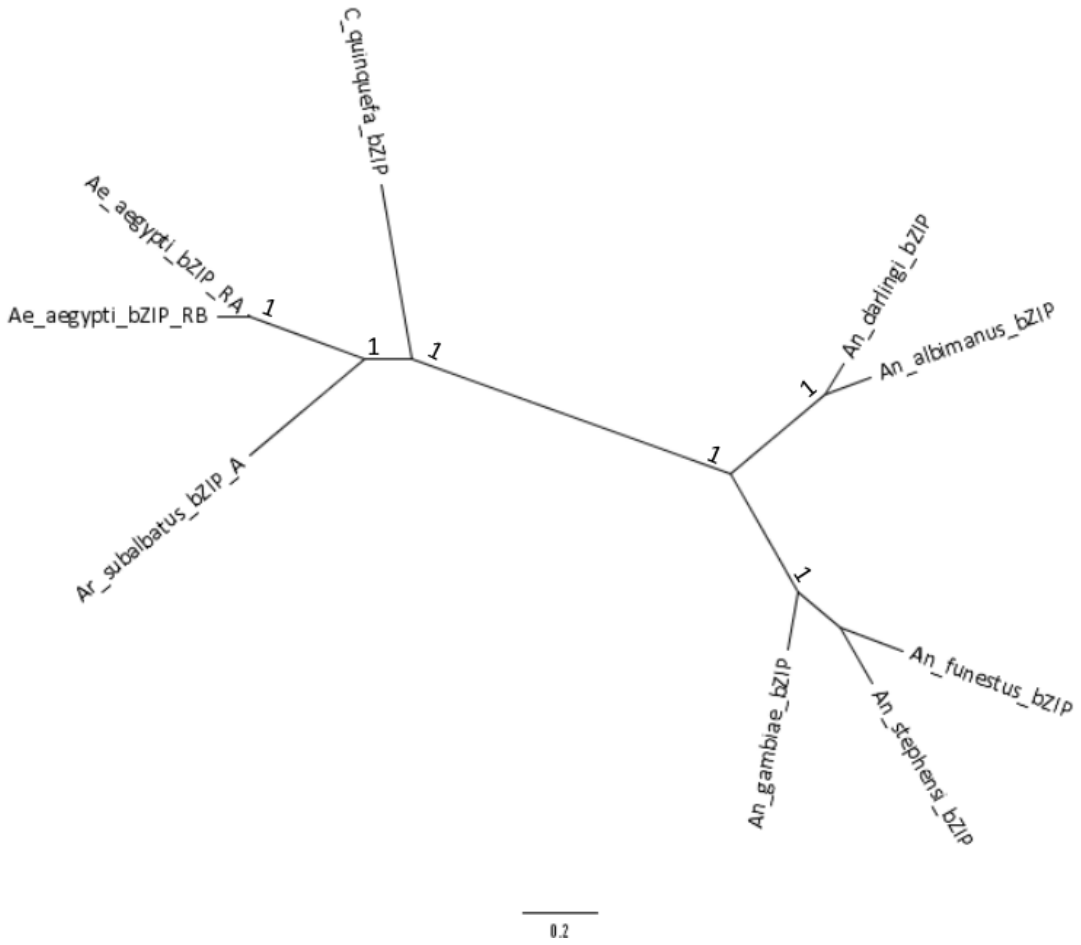
presence of a well-processed intronic miRNA cluster that is maternally inherited and zygotically sustained may provide an approach to express artificial miRNAs to knockdown gene activities (Lin *et al*., 2009). In summary, we have characterized a conserved mosquito-specific *bZIP1* gene in divergent mosquitoes and shown that it provides a transcript that is critically important to mosquito embryonic development and that it produces a family of highly abundant intronic miRNAs.

2.6 Figure legends, tables and supplementary information

Figure 2.S1 Microsynteny of *bZIP1* and surrounding genes in six mosquito species (not drawn to scale).

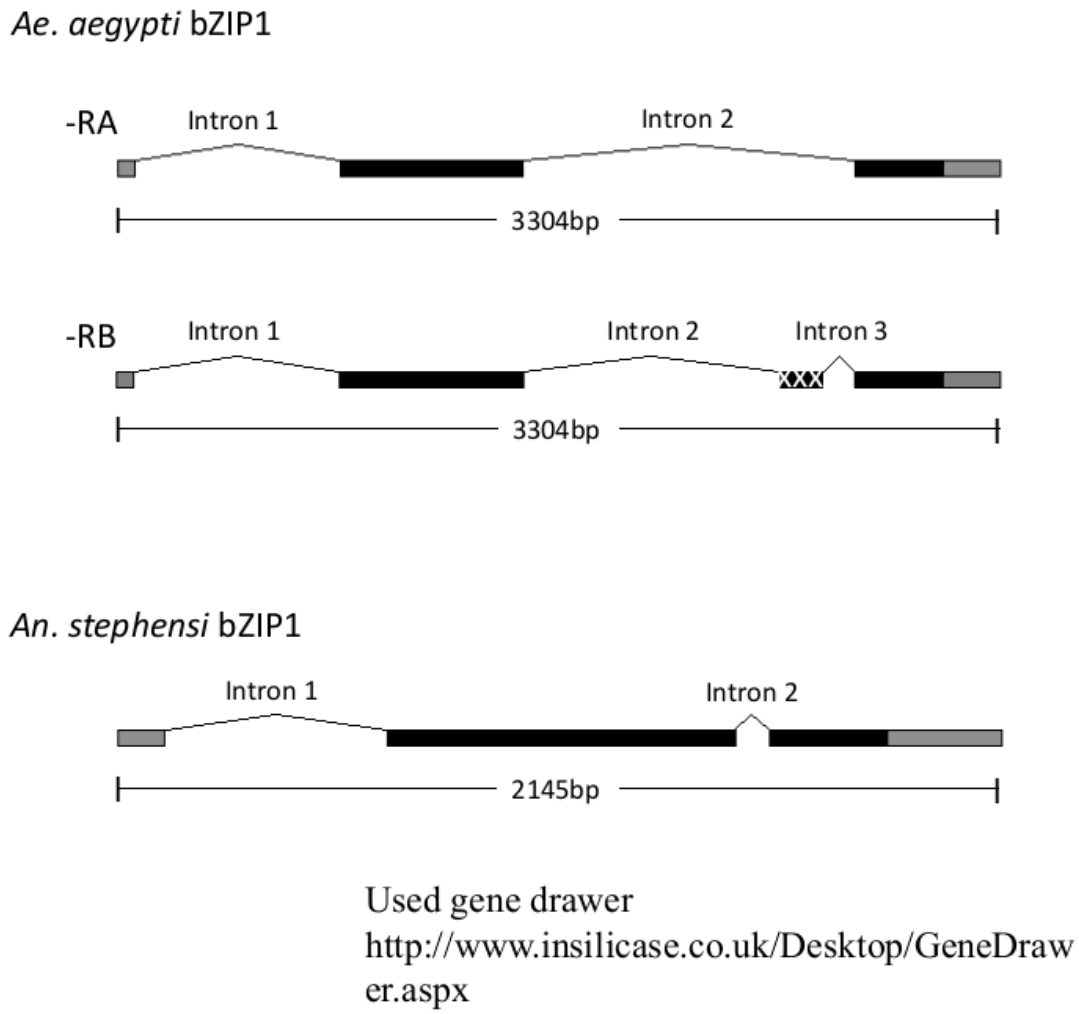
Black box denotes the conserved *bZIP1* gene in each species. Shaded regions denote conserved genes in these genomes. Gene 1 is a putative phospholipase and gene 2 is a putative cytochrome-c oxidase. If we use the Vectorbase (vectorbase.org) annotation of *An. gambiae* as a reference, gene 1 is AGAP007766, *bZIP1* is AGAP007767, and gene 2 is AGAP007768. The open box in *Cx. quinquefasciatus* depicts the fact that there is an additional gene (CPIJ000469) between *bZIP1* and the cytochrome-c oxidase. See Methods for sources of the sequences.

Figure 2.1



Unrooted phylogenetic tree of *bZIP1* amino acid sequences from 8 mosquito species. MrBayes was used to infer the phylogenetic relationships. Confidence values greater than 0.5 after 100,000 generations of calculation are shown at the branch points.

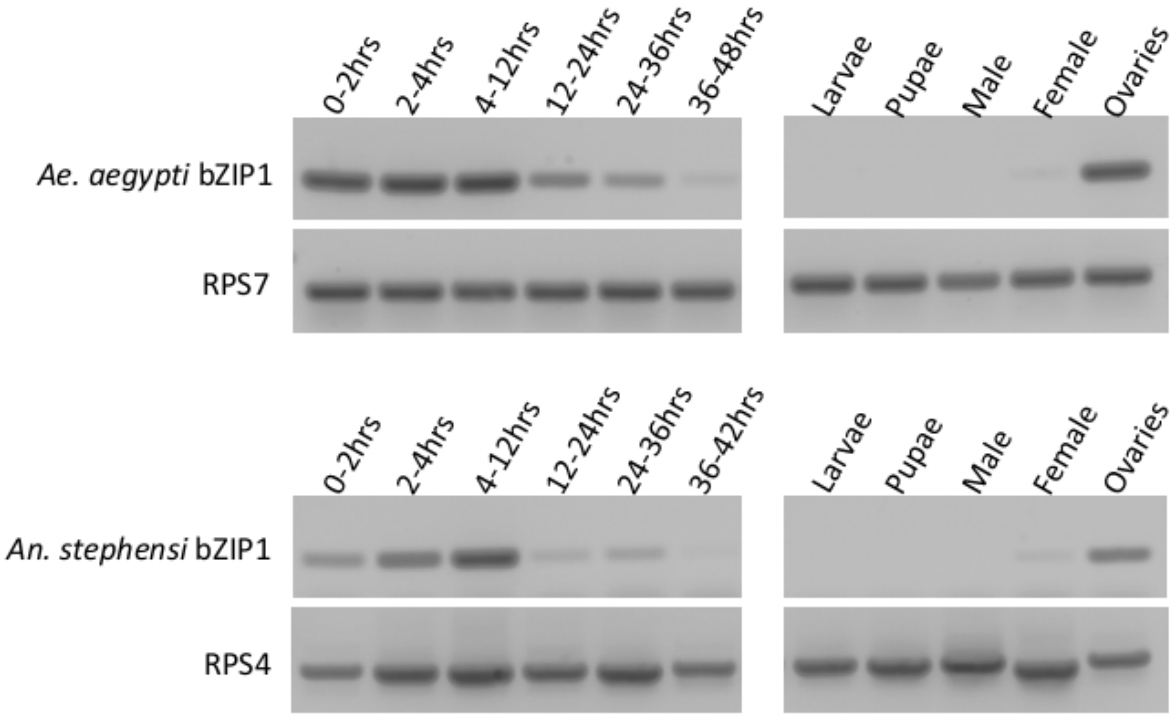
Figure 2.2



Used gene drawer
http://www.insilicase.co.uk/Desktop/GeneDrawer.aspx

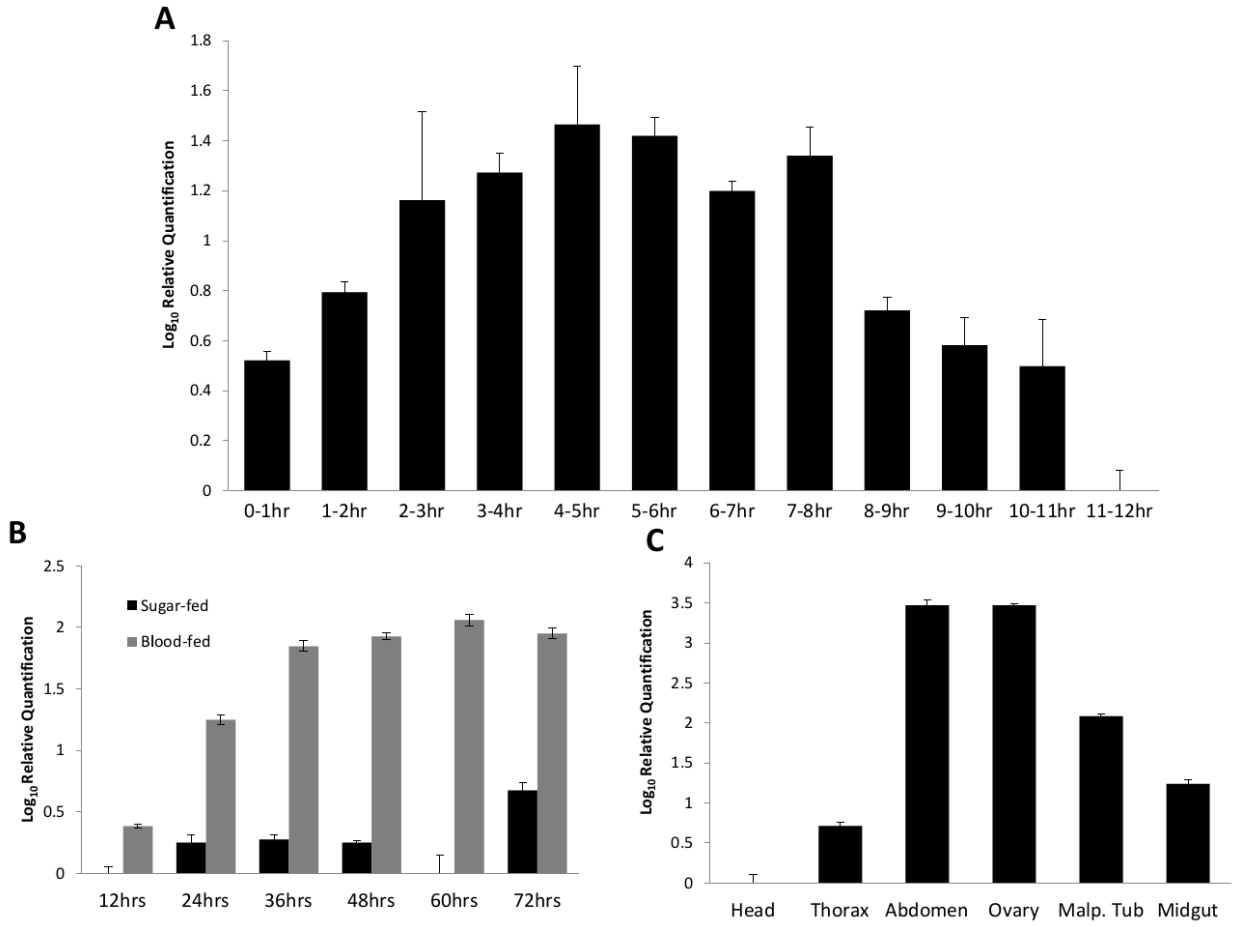
Aedes aegypti and *Anopheles stephensi bZIP1* gene structure. Gene structures of *bZIP1* are determined by 5' and 3' RACE. Solid black boxes indicate coding sequence and shaded boxes indicate UTRs. The extra intron of the –RB splice variant is marked by the cross-hatched black region. * indicates that the miR-2941 family members reside in intron 1 of the *bZIP1* gene.

Figure 2.3



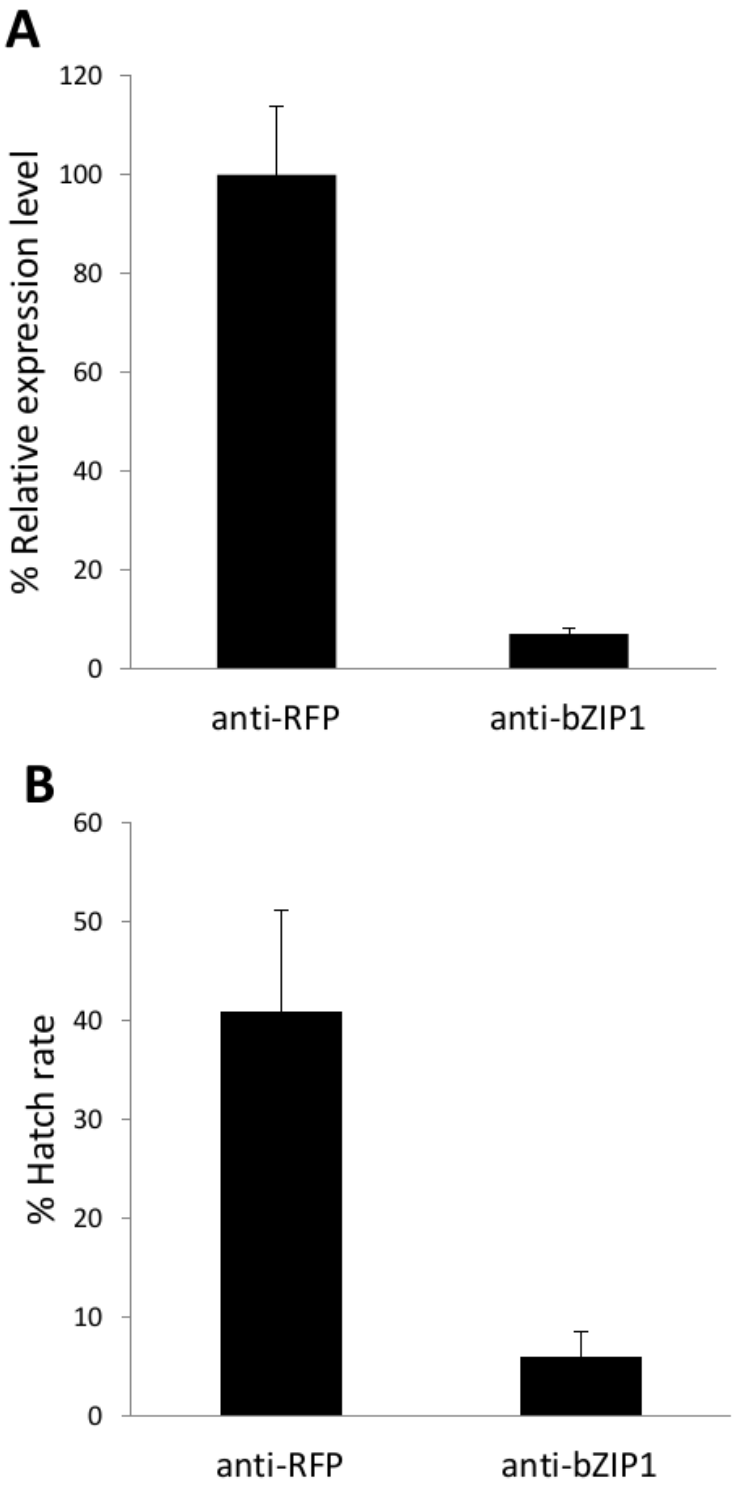
Overall transcription profile of *Aedes aegypti* and *Anopheles stephensi bZIP1* as shown by RT-PCR. RT-PCR targeting *bZIP1* was performed on samples isolated from selected embryonic time points (0-2, 2-4, 4-12, 12-24, 24-36, and 36-48 hrs in *Ae. aegypti* and 0-2, 2-4, 4-12, 12-24, 24-36, and 36-42 hrs in *An. stephensi*), different developmental stages(larvae instars 1-4, early and late pupae, 1-5 day old sugar fed males and females) and ovaries from adult females 48 hrs after blood feeding. RPS7 and RPS4 were used as positive controls in *Ae. aegypti* and *An. stephensi*, respectively.

Figure 2.4



Aedes aegypti bZIP1 expression profile determined by TaqMan® qRT-PCR assays. A) *bZIP1* expression in blood fed and sugar fed females. Newly emerged females were fed on sugar for 3 days at which point they either remained on sugar or were allowed to feed on blood. Time points were then collected as the females aged in parallel. Data is relative to the expression of the 12 hour sugar fed sample. B) *bZIP1* expression in tissues isolated from adult females 48 hours post blood meal. Expression detected is relative to the expression detected in the head. Malp. Tub., the *Malpighian tubule.* C) Embryonic expression of *bZIP1* in 1-hour increments ranging from 0-12 hours. Data is relative to the expression at the 11-12 hour time point.

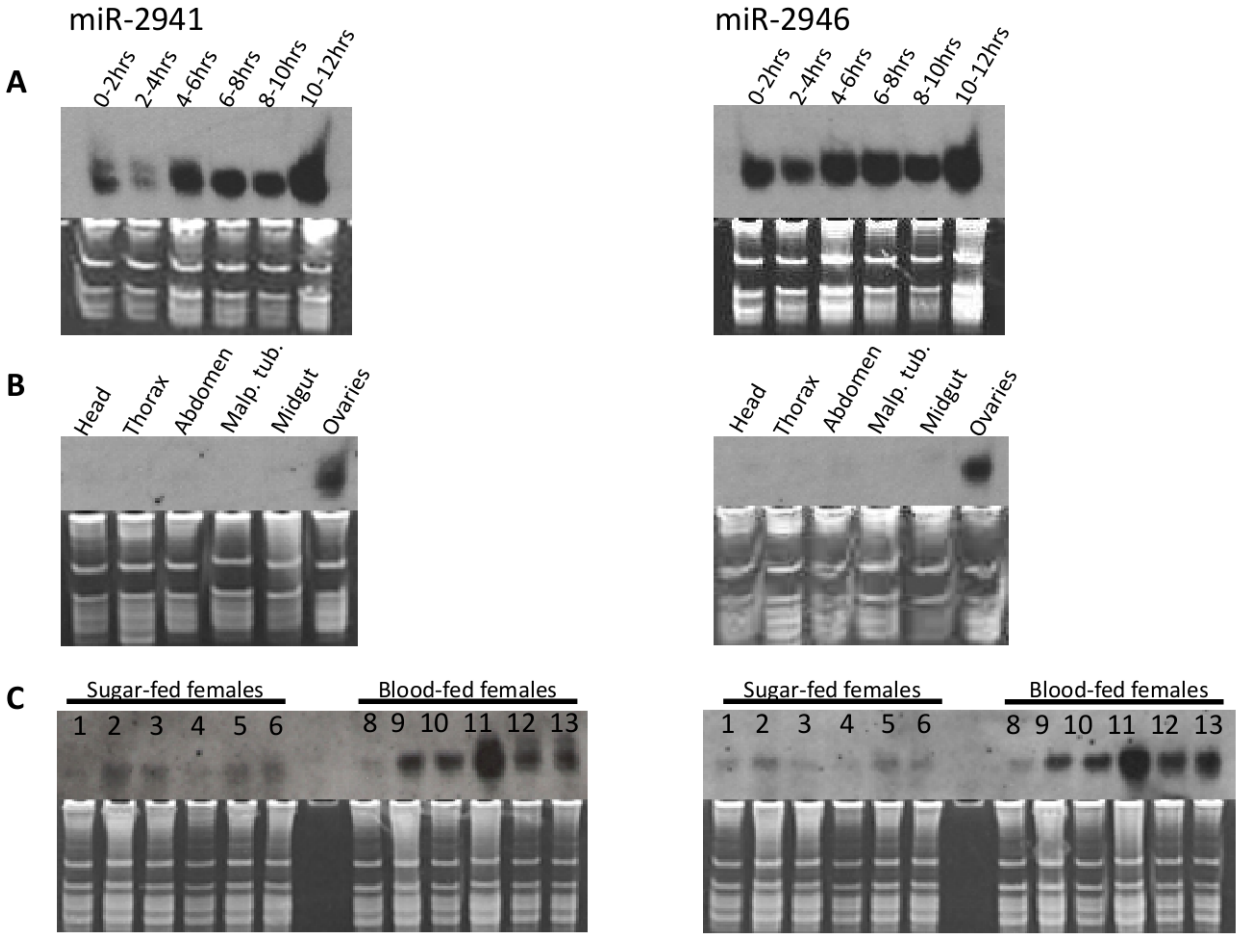
Figure 2.5



Aedes aegypti bZIP1 RNAi knockdown and hatch rate. A) qRT-PCR showing siRNA knockdown of *bZIP1* in *Ae. aegypti*. Total RNA was extracted from embryos 7 hrs after injection

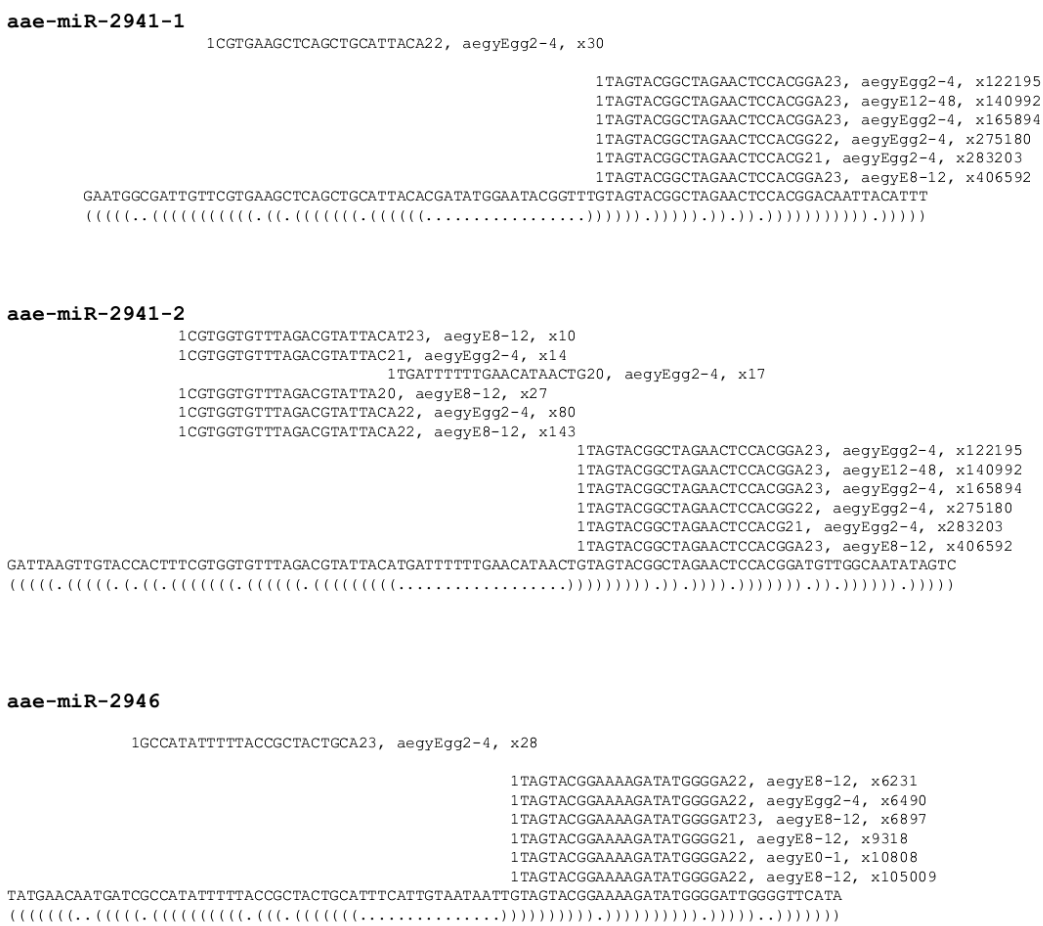
with 20 µM anti-RFP or anti-*bZIP1* siRNA and subjected to Taqman® qRT-PCR. RPS7 was used as an endogenous control to normalize both data-sets. Expression of *bZIP1* with siRNA targeting RFP was set at 100%. Embryos injected with anti-*bZIP1* siRNA showed *bZIP1* mRNA levels to be reduced to 2.6% that of anti-RFP injected embryos ($p < 0.001$). B) Hatch rate differences. Embryos injected with 20 µM anti-RFP (non-targeting control) and anti-*bZIP1* siRNA were placed in water 70 hrs post injection and surviving larvae were counted daily for 11 days. Anti-RFP and anti-*bZIP1* siRNA injected embryos had hatch rates of 40.9±10.2% and 5.9±2.6%, respectively ($p<0.05$).

Figure 2.6



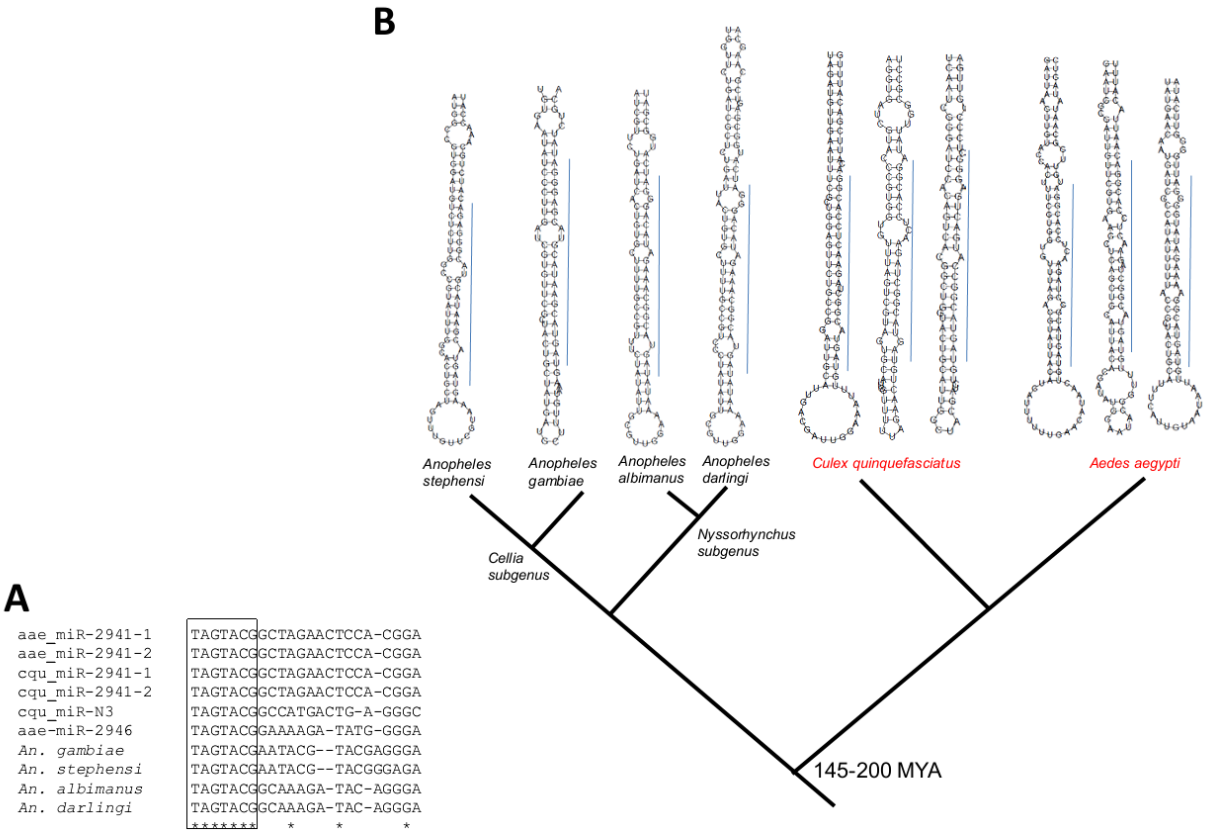
Expression of miRNA-2941 and miR-2946 in *Aedes aegypti* as shown by northern blot. Top panels are northern blot results and bottom panels are corresponding RNA gel images. A) Early embryonic RNA samples in 2 hour increments 0-12 hrs after oviposition. B) RNA samples from the head, thorax, abdomen, Malphigian tubules (Malp. tub.), midgut, and ovaries collected from females 48 hours after blood feeding. C) Whole-body adult females fed on sugar or blood. Three-day old females were kept either on sugar water or allowed to feed on blood, after which samples were taken every 12 hours. Lanes 1-6: 12hrs, 24hrs, 36hrs, 48hrs, 60hrs, 72hrs sugar-fed females. Lanes 8-13: 12hrs, 24hrs, 36hrs, 48hrs, 60hrs, 72hrs post blood feeding.

Figure 2.7



Comparison between the pre-miRNAs and illumina sequence reads showed that all three hairpins in the *Aaeg_bZIP1* intron are transcribed and processed as indicated by the presence of unique miRNA* sequences. The secondary structures of the pre-miRNAs are indicated under the sequences, which were predicted by CentroidFold (Sato *et al*., 2009). aegyEgg0-1, aegyEgg2-4, aegyEgg8-12, and aegyEgg12-48 refer to *Ae. aegypti* egg samples collected at 0-1, 2-4, 8-12 and 12-48 hours after egg deposition, respectively. The numbers after x indicate the occurrence the sequence in the sample. In the case of aae-miR-2941-2, the processed loop sequence TGATTTTTTGAACATAACTG is also recovered. Note that mature miR-2941-1 and miR-2941-2 cannot be distinguished as they are identical. Due to space limitation, not all sequence reads are shown.

Figure 2.8



Conservation and evolution of the miRNA family in the bZIP1 intron. A). Conservation of the known and predicted mature miRNAs in the *bZIP1* intron of divergent mosquitoes. B). Structure of the pre-miRNAs shown in the context of species phylogeny. The vertical lines indicate the actual or predicted mature miRNAs. The figure is not drawn according to the evolutionary time scale.

Table 2.2- Mass spectrometry identification of bZIP1

Protein	Coverage	# PSMs	# Peptides	# AAs	MW [kDa]	calc. pI	Score
bZIP1	70.24	56	42	336	38. 5	6. 70	1. 45
Vitellogenin	13.04	78	24	2139	249.2	7.05	249.13

The MS output for the identification of bZIP1 and vitellogenin (control) for the 8-12hr embryonic time point.

Figure 2.S2 bZIP1 protein sequence alignment

```
An_funestus_bZIP      ------------------------------------------------------------
An_stephensi_bZIP     MTDFDAASALLMLSSGYHQYKNENNNI-----------RTVLKS--VNTATANSSAP-SS
An_gambiae_bZIP       MTDLDAVSALLMLSGG-HQ--------------------PTLVQ--MNNSTANAATSGTS
An_darlingi_bZIP      MADYDAAAALLLLSSGGLS-------------------PQTIDSEKNNNRRSDGATKENM
An_albimanus_bZIP     MADYDAAAALLLLSSGGLS-------------------PQAIDSEKNNNRRSEGATKENI
Ar_subalbatus_bZIP_A  MAPNDVQHSD-------------------------------------DYNNVRPNSDQDP
Ar_subalbatus_bZIP_B  MAPNDVQHSD-------------------------------------DYNNVRPNSDQDP
Ae_aegypti_bZIP_RB    MAPHNSLHPD-------------------------------------THNDP-SSTESDS
Ae_aegypti_bZIP_RA    MAPHNSLHPD-------------------------------------THNDP-SSTESDS
C_pipiens_bZIP        MQDPSKIYPRPIPTRSCPDLSRTSLTACPSASVVFGFSARMAARIVMEYSEEVCKAPMDS


An_funestus_bZIP      ------------------------------------------------------------
An_stephensi_bZIP     PAKKRERKNGSPKQRIVVPQPTAFTIAT--IHENFSQHELLGRLSSET-----VPKGKTP
An_gambiae_bZIP       PPKRRERKNGSPKQR-IVPKST-LRVAS--VHGNIAQHELFTRHAMKRPVISLVPVNKLE
An_darlingi_bZIP      IP---MHTRGSANKAAVAQKQTPALGASGGRQQNVLQNKVIP-PEVNSQTTPLLKVKRTG
An_albimanus_bZIP     IPTGPMHTRGSTNRAAAGQKETPTMMSS-LRQQNALQCKVIA-PEVNLPATPLLTVKRTG
Ar_subalbatus_bZIP_A  NEN-------PPKRIKYDSTYENMIFRAG--YTNMQNQYIYNQQRSQALRMQDDRYSTIS
Ar_subalbatus_bZIP_B  NEN-------PPKRIKYDSTYENMIFRAG--YTNMQNQYIYNQQRSQALRMQDDRYSTIS
Ae_aegypti_bZIP_RB    NSN-------QSKKIKYDDSYQNMMFRAG--FTNMHNEYIYNKNRTHALMMRNDRYSMTS
Ae_aegypti_bZIP_RA    NSN-------QSKKIKYDDSYQNMMFRAG--FTNMHNEYIYNKNRTHALMMRNDRYSMTS
C_pipiens_bZIP        SSSSSISSGQPPKKRKYDSNEVNVMFRMR--YNNMENQYMYDHQRSHAIQKADQSPSYPT


An_funestus_bZIP      ------------------------------------------------------------
An_stephensi_bZIP     RITP------IRMGSSSPDVEK------------------WDSRGSALSSPTLELHPDS-
An_gambiae_bZIP       DTTARNRALAMRMGSSSPDLER------------------WE-RADTLS-PIIVPHPET-
An_darlingi_bZIP      LYTPEREAAISYR-HYHPDIVPPNSS-----SVDG--MIVGPHRQSFQTQRSVSLLPAT-
An_albimanus_bZIP     LYTPDREAAIVKKTRYRP-ILPKREQPTLLREHDSPEQLLGRWDDSRTSSPIIPAIAAIG
```

```
Ar_subalbatus_bZIP_A  SSSSLVSSMDGYSSTPSPDLTL----------------------NGSQASSPLFAVYPQR-
Ar_subalbatus_bZIP_B  SSSSLVSSMDGYSSTPSPDLTL----------------------NGSQASSPLFAVYPQR-
Ae_aegypti_bZIP_RB    TTSSIISSLDGNRSTPSPDFAM---------------------SMGQVSSPLFNGFMQS-
Ae_aegypti_bZIP_RA    TTSSIISSLDGNRSTPSPDFAM---------------------SMGQVSSPLFNGFMQS-
C_pipiens_bZIP        TSS---SGSD--QRTPSPDFAP---------------------SASQTSSPL--------


An_funestus_bZIP      ----------------------------------------K---SQNNQY-TSSPKPPS
An_stephensi_bZIP     -----TTSAENVSS------------MLYASPGMNGFAMLLK---SSGNQY-TSSPEPPA
An_gambiae_bZIP       -----SHGGRSISPRIIETPGNHSPHLLYGSPGMNGFAMMLK---SGGNQY-TSSPEPPS
An_darlingi_bZIP      -----SGCDRNYGSGSDGELERLLHKEVYASPGMHGFAMQLNTAGTEENRFVVSSPEPLM
An_albimanus_bZIP     QSTIPSGHDRNDGSGSDGELERLVHKKVYASPGMHGFAMQLN-AGT-ENRYEVSSPEPLM
Ar_subalbatus_bZIP_A  -----PVDPPP---------------LRYFQFNNQQPPPDHA--QRPEHSYAVDSPQSSP
Ar_subalbatus_bZIP_B  -----PVDPPP---------------LRYFQFNNQQPPPDHA--QRPEHSYAVDSPQSSP
Ae_aegypti_bZIP_RB    -----PGEPSQ---------------SIYDQ-SRVRSPATFN--NPQEHNYAVDSPHA--
Ae_aegypti_bZIP_RA    -----PGEPSQ---------------SIYDQ-SRVRSPATFN--NPQEHNYAVDSPHA--
C_pipiens_bZIP        -----PVLP-------------------------TKPEHNYA--VSSSASSSTDTAPS--


An_funestus_bZIP      VV---T--TFMEDNRKKRFLDNPIA--E-EGASA-----------IVVAAD------KPA
An_stephensi_bZIP     VV---VGGTLLEETNKKRFLGNHLVEDE-EGASA-----------IVVTAD------KPA
An_gambiae_bZIP       ------------DGAS----GN--V--E-HGATAGGS--------LVAIAD------KPA
An_darlingi_bZIP      ITRSMKAETTATDASSISGIHKVLYDDEGDGPSATA---------AVATASHSVSNLRPV
An_albimanus_bZIP     ITRSMKAETTAIDSSSISEIN-TPYDDEGDGASAAATASVPAAVVAVTTGSHSELNFRPV
Ar_subalbatus_bZIP_A  PI--------SNSSEP-----------------------------ENRPAR--------A
Ar_subalbatus_bZIP_B  PI--------SNSSEP-----------------------------ENRPAR--------A
Ae_aegypti_bZIP_RB    ----------LHTPEP-----------------------------IIRPAR--------A
Ae_aegypti_bZIP_RA    ----------LHTPEP-----------------------------IIRPAR--------A
C_pipiens_bZIP        ------------NNEG-----------------------------QRRLMP--------S


An_funestus_bZIP      R----------------SPSDSGVSSILDEIQQPN-LVLQRWNKEMEALDSSIPASVQME
An_stephensi_bZIP     R----------------SPSDSGVSSILDELQQP---VLQRWSKDRDASDTVLPASVQTE
An_gambiae_bZIP       R----------------SPSDSGVSSILDEMLQPN-LVLQRWKK-----DSGIPDCVQKE
An_darlingi_bZIP      RNSIGIGMVKAEEKAQLSPSDSGNSSIHDEILQSG-LIIQAWTRGSD-ITAGLPENVKIE
An_albimanus_bZIP     RNSIGIGMVKAEEKAQLSPSDSGNSSIHDEILQSG-LIIQAWTRGSD-ISTGLPENVKIE
Ar_subalbatus_bZIP_A  R----------------VTSDSGNSTVPSETGEIGPVSMDDWFR--IFQNHEIASKSQRV
Ar_subalbatus_bZIP_B  R----------------VTSDSGNSTVPSETGEIGPVSMDDWFR--IFQNHEI-------
Ae_aegypti_bZIP_RB    R----------------VTSDSGNSSIP-EVDDINPLCMNDFFR--DAVNPELAEKLNSA
Ae_aegypti_bZIP_RA    R----------------VTSDSGNSSIP-EVDDINPLCMNDFFR--DAVNPELAEKLNSA
C_pipiens_bZIP        R----------------AASDSGNSTMS---DEINPLCMTDWYR--DLLKPEYAQRLSTQ


An_funestus_bZIP      LEKINEISAYNKEIFTKAKMTNFSLEMTFHPAKSRMRKKCTNEVDNQDRIRNNEASRRSR
An_stephensi_bZIP     LDEINETSAYNKDIYTKSKLANFPLEMTFNAAKSRIRKECTNEIDHQDRIKNNEASRRSR
An_gambiae_bZIP       LDKIVEISAYNKDLYTKAKLAKFPLEMGFNPNKSRIRKNCENEADNQDRVKNNEASRRSR
An_darlingi_bZIP      LHNILQTSLYNKECYTKEKMAEFPMDVGYHPNKSRLRKEYSNDAEAADRAKNNLASRRSR
An_albimanus_bZIP     LHNILQTSLYNKECYTKEKMAEFPMDVGYHPNKSRLRKEYSNDAEAADRAKNNLASRRSR
Ar_subalbatus_bZIP_A  ------------------------------------------------------------
Ar_subalbatus_bZIP_B  ------------------------------------------------------------
Ae_aegypti_bZIP_RB    LTTIEAKSAQHKERYLEMRMCELPTDLEI---ISSLR----------------LDT----
Ae_aegypti_bZIP_RA    LTTIEAKSAQHKERYLEMRMCELPTDLEYNPNKSRLRKVYESPMEAAERERNNLASRKSR
C_pipiens_bZIP        LDIITNRSDFVKNKILTARIAELPESFDYYPNKSRLRKQYVDPDEAAERERNNLASRRSR


An_funestus_bZIP      HKKKLITHLMNISLEFDRMENRHLYMEERRLEDFILELEEKALSCGIDAQIVKNLRSSCG
An_stephensi_bZIP     HKKKLITHMLNISLEFDRMENRQLYMEERRLTNIIMELEEKALNRGVDAQVVQKLRSSCG
An_gambiae_bZIP       HKKKLMTHMLNTSLEFDRQENRQLYMQERWLTNLICELEEKALNRGIDAQLVRKLRHACG
An_darlingi_bZIP      HKKKMVTQLMNISLEYDRQENRQLFMQERWLTNLIFELEDKALQQGIDTQVLRKLRADCG
An_albimanus_bZIP     HKKKMVTQLMNISLEYDRQENRELFLQERWLTNFIFELEDKALQQGIDTQVLRKLRADCG
Ar_subalbatus_bZIP_A  ------------------------------------------------------------
Ar_subalbatus_bZIP_B  ------------------------------------------------------------
Ae_aegypti_bZIP_RB    ---------TSVRLLVDAGEKA--------------------FSRRRT-------RRPSA
Ae_aegypti_bZIP_RA    FKKKIAQQITNMHLEFDRSESADLYAMQNWMGQVIFELESNCLDRGITPECLADMRQQCG
C_pipiens_bZIP        FKKKRAQLILNMHLEYDRTENAHLYAMQSWIGKIIFQLETQWLERGATAEQMCQLRRDCG


An_funestus_bZIP      FQ--------------
An_stephensi_bZIP     FQ--------------
An_gambiae_bZIP       FQ--------------
An_darlingi_bZIP      FQ--------------
```

```
An_albimanus_bZIP     FQ--------------
Ar_subalbatus_bZIP_A  ----------------
Ar_subalbatus_bZIP_B  ----------------
Ae_aegypti_bZIP_RB    SLRN------------
Ae_aegypti_bZIP_RA    FLRNQNDKVYRARPSF
C_pipiens_bZIP        FPQTPPGAFRLI----
```

Other

An. stephensi and *An. albimanus bZIP1* genomic sequences have been submitted to GenBank (accession numbers JQ266223 and JQ266224, respectively).

Ae. aegypti and *An. stephensi bZIP1* cDNA sequences have been submitted to GenBank (accession numbers JQ266221 and JQ266222, respectively).

Chapter 3

A unique Y gene in the Asian malaria mosquito *Anopheles stephensi* encodes a small lysine-rich protein and is transcribed at the onset of embryonic development

Chapter 3 was published in Insect Molecular Biology: Criscione, F., Qi, Y., Saunders, R., Hall, B., Tu, Z. (2013). A unique Y gene in the Asian malaria mosquito *Anopheles stephensi* encodes a small lysine-rich protein and is transcribed at the onset of embryonic development. *Insect molecular biology*.

Contributing authors:

Frank Criscione

GUY1 gene characterization, GUY1 sequence verification, proof of Y chromosome specificity, luciferase assays, single embryo analysis and writing of manuscript.

Yumin Qi

5' and 3' RACE and GUY1 sequence verification.

Randy Saunders

DAPI staining of embryos.

Brantley Hall

GUY1 Sequence verification.

Zhijian Tu

GUY1 discovery by bioinformatic analysis, GUY1 sequence analysis and writing of manuscript.

3.1 Abstract

In many organisms the Y chromosome initiates sex-determination and regulates male fertility and mating behavior. However, molecular characterization of Y genes is rare outside of a few model species because it is difficult to clone and analyze repeat-rich heterochromatic Y sequences. In insects, Y genes are only well characterized in a small number of *Drosophila* species. Here we report the discovery of *GUY1*, a gene unique to the Y in the Asian malaria mosquito, *Anopheles stephensi*, using an approach that compares Illumina sequences separately obtained from male and female genomic DNA. Experimental evidence confirmed that *GUY1* is a single copy gene found only on the Y chromosome. *GUY1* is transcribed at the very onset of zygotic transcription and encodes a small lysine-rich protein that forms two alpha helices and shows DNA-binding properties. Interestingly, three helix-loop-helix proteins are key factors that determine sex in the early embryo in *D. melanogaster*. Single embryo analysis indicated that *GUY1* is only transcribed in male embryos and the *GUY1* promoter is functional in the early embryos. *GUY1* may be used as a paternally inherited molecular marker. Further investigation of *GUY1* will contribute to the genetic approaches to control mosquito-borne diseases.

3.2 Introduction

Anopheles mosquitoes use the XX/XY sex-determination system. Females possess two X chromosomes and males possess one X and one Y. There is evidence that a dominant male-determining factor(s) on the *Anopheles* Y chromosome initiates sexual differentiation (Baker and Sakai, 1979; Marin and Baker, 1998), which is similar to humans and several species of flies (e.g., Marin and Baker, 1998; Kaiser and Bachtrog, 2010). The common fruit fly *Drosophila melanogaster* also has the XX/XY sex chromosomes. However, it is the X to autosome ratio, not

the mere presence of the Y chromosome that determines the sexual fate of the *D. melanogaster* embryo (Marin and Baker, 1998). Genetic studies show that the *D. melanogaster* Y chromosome is involved in the regulation of male fertility, mating behavior, and global gene expression (e.g., Carvalho *et al*., 2001; Lemos *et al*., 2008; Zhou *et al*., 2012). Similarly, the *Anopheles* Y chromosome has also been shown to control the stenogamy-eurygamy mating behavior (Fraccaro *et al*., 1977). Therefore, it is clear that the Y chromosome is critical to the sexual dimorphism in dipteran insects regardless of its involvement in the initiation of sex determination.

Molecular characterizations of Y genes are rare outside of a few model species, mostly due to the repetitive nature of the Y chromosome and the difficulty of cloning repeat-rich heterochromatic sequences (Carvalho *et al*., 2001). In insects, Y genes are only reported from a few *Drosophila* species (Carvalho *et al*., 2001, 2009). Only a dozen or so Y genes have been reported in *Drosophila*, most of which were discovered though comparison of unmapped sequences to the non-redundant protein database. This method was applied to the African malaria mosquito *An. gambiae* but failed to identify any Y genes (Krzywinski *et al*., 2006). In this study, we report the discovery of a unique Y chromosome gene in the Asian malaria mosquito, *An. stephensi*, using an approach that compares Illumina sequencing results of RNA samples and separately isolated male and female genomic DNAs. We name this gene *GUY1* for Gene Unique to the Y. *GUY1* encodes a small lysine-rich protein and it is expressed at the very beginning of the maternal-to-zygotic transition. The possible function of *GUY1* and the implications to mosquito control by manipulation of sexual dimorphism are discussed. Y genes such as *GUY1* can also provide a novel set of markers for phylogenetic, population and mating behavioral studies. Such markers offer perspectives fundamentally different from the other types of markers as their inheritance is exclusively paternal (e.g., Kopp *et al*., 2006; Underhill and Kivisild, 2007).

3.3 Results

Discovery and molecular characterization of GUY1

Male and female *An. stephensi* genomic DNA sequences were generated via Illumina sequencing (SRA accession SRP013838). The male and female data were compared using blastn with an e-value threshold of 1e-3. The male data was used as the query and the female data was used as the database. All of the male sequences that matched any female sequences were removed, leaving only male-specific sequences. The male-specific sequences were then compared to the RNA-seq databases (SRA accession SRA054605) of all developmental stages to identify transcribed genes. Y genes should neither be expressed in adult females nor the 0-1 hr old embryos. This is because females do not have the Y chromosome and the 0-1 hr embryos are transcriptionally inactive and mainly contain maternally deposited transcripts. Using this strategy, we obtained 45 male-specific Illumina reads (Figure 3.SF1) that had no matches to the transcriptomes of the adult female or 0-1 hr embryos but had at least 3 total matches to the transcriptomes from the adult male, other embryonic time points, larvae or pupae of mixed sexes. These 45 sequences were further analyzed as candidate Y genes. Many of these are similar to reverse transcriptase from retrotransposons.

We focused on a set of paired-end sequences (Figure 3.SF1) that showed embryonic transcription and were not of retrotransposon origin. These sequences were used to design primers (Table 2.1) for PCR and RT-PCR, which subsequently led to the discovery and characterization of the full-length *GUY1*cDNA (Genbank JX174417) and a 5.5 kb genomic fragment that contains the *GUY1* gene (GenBank JX154128). The *GUY1* transcription start site and polyA addition site were determined by 5' and 3' RACE. Figure 3.1A depicts key features

of the *GUY1* gene, which includes a short ORF, 5' UTR, 3' UTR, and the 5' flanking sequences. *GUY1* is intronless, a feature shared with many early zygotic genes (Biedler *et al*., 2012), and it has a short ORF encoding a novel protein of 56 amino acid residues. As shown in Figure 3.1B, this protein is rich in lysine and could form two alpha helices. The second helix contains a stretch of basic lysine residues, a structural feature that has been shown to mediate sequence-specific DNA binding in many transcription factors. The first helix is amphipathic with hydrophobic residues on one side (Figure 3.1C), which could facilitate dimerization through hydrophobic interactions.

Male-specific GUY1 amplification in different strains of An. stephensi.

Although some Y genes have highly similar paralogues on the autosome or X chromosome (e.g., Carvalho *et al*., 2009), our *in silico* subtractive approach focused on the identification of sequences unique to the Y chromosome. Such Y-specific sequences, like *GUY1*, should be found exclusively in the males. Male-specific amplification was demonstrated for all strains tested when *GUY1* primers (Table 2.1) were used to amplify genomic DNA. There are three biological forms or races of *An. stephensi*, namely the "type" form, the mysorensis form and the intermediate form. Shown in Figure 3.2A are PCR results indicating male-specific amplification of GUY1 in three strains (VT, UCI, and JHU) of the "type" biological form. Male-specific amplification of GUY1 is also observed in the IRN and CHB strains of the mysorensis form and the Beech strain of the intermediate form (data not shown). As positive controls, *RPS4* (ribosomal protein subunit 4) was successfully amplified from both male and female genomic DNA sources in all experiments described above.

GUY1 containing DNA is transferred to females during mating.

To further establish the Y chromosome origin of the novel sequences we performed PCR using genomic DNA isolated from females before and after mating (Figure 3.2B). Mature sperm are a mixture of X-containing and Y-containing haploid cells. During mating the sperm are transferred to the female and stored in the spermatheca in the female abdomen. Virgin and mated females were dissected into two parts: head and thorax (together) and abdomen. *GUY1* amplification, after 40 cycles, was detected only with genomic DNA isolated from the abdomens of mated females, the location where sperm are stored.

GUY1 is a single copy sequence in males.

Copy number of *GUY1* was determined as described by Hindson *et al.* (2011) using QX100 Droplet Digital PCR System from Bio-Rad. Setting the copy number of a known single-copy autosomal reference to 2 for the diploid alleles, the copy number of *GUY1* in males and females was determined. The primers and Taqman probes used to compare *GUY1* and the autosomal reference are provided in Table 2.1. Shown in Figure 3.2C are representatives of the QX100 Droplet Digital PCR experiments. Male copy number values (0.96, 0.85, 0.95) and female copy number values (0.01, 0.02, 0.01) are consistent with *GUY1* having a single copy in males and no copies in females. This copy number determination experiment was repeated on 18 male and 18 female individuals and produced similar results.

GUY1 expression is detected at the onset of maternal-to-zygotic transition.

As shown in Figure 3.3, *GUY1* expression appears to be restricted to the embryonic stage as there was no amplification of *GUY1* from larval, pupal, adult male or adult female cDNA.

GUY1 transcripts start to accumulate at 2 hrs after egg deposition and tapers off after 12 hrs (Figure 3.3). RT-PCR results are consistent with RNA-seq analysis and both suggest that *GUY1* is mainly transcribed 2-8 hrs after egg deposition and not detectable in other stages (Table 2.2). RT-PCR on early embryonic samples with tighter time intervals in early embryos showed that *GUY1* is transcribed as early as 2-3 hrs post egg deposition (data not shown). DAPI nuclei staining was used to estimate the timing of the early embryonic development in *An. stephensi*. Under the experimental conditions described in this paper, most of the embryos had 16 to 32 nuclei at 1-2 hrs after egg deposition, representing mitotic cycles 5 and 6. By 2 to 3 hrs after oviposition, most embryos were at cycles 8 and 9. The syncytial blastoderm stage (mitotic cycles 10-13) occurs approximately 3-4 hrs after egg deposition in *An. stephensi* (Figure 3.S1), similar to *An. gambiae* (Juhn and James, 2006). Thus, *GUY1* transcription is initiated at the onset of maternal-to-zygotic transition, prior to syncytial blastoderm formation.

Single-embryo analysis showed that GUY1 is only expressed in males and the GUY1 promoter is able to direct reporter gene expression in both male and female embryos.

By dividing the lysate from a single embryo for RNA and genomic DNA isolation, we were able to sex the single embryo and detect the transcription within the same embryo. In addition to *GUY1*, other Y-chromosome sequences including a Y-specific repeat (Table 3.1) were used as markers to sex the embryo prior to RT-PCR. As expected, *GUY1* is only expressed in the male embryo (Figure 3.4A). The 850 bp upstream promoter of *GUY1* contains an arthropod initiator TCAGT, an upstream TATAA box (GenBank JX154128), and 6 copies of a TAGteam-like motif that was shown to activate early zygotic transcription in *Ae. aegypti* (Biedler *et al*., 2012) and *Drosophila* (Erickson *et al*., 1998; Sefton *et al*., 2000; Liang *et al*.,

2008). This promoter region was used in a dual luciferase reporter assay and shown to be active in *An. stephensi* early embryos (Figure 3.4B). As expected for a promoter active at the very onset of zygotic transcription, the *GUY1* promoter functions in both male and female embryos (Figure 3.4C). However, native *GUY1* transcription was not detected in female embryos (Figure 3.4A) because there is no endogenous *GUY1* gene in the XX embryo.

3.4 Discussion

We have discovered a Y-specific gene, *GUY1,* in *An. stephensi* using a bioinformatic approach comparing high throughput sequences of male and female genomic DNA and RNA samples. Florescent *in situ* hybridization (FISH) can provide direct evidence if *GUY1* signal is observed on the Y chromosome. However, Y-chromosome FISH is challenging in *Anopheles* mosquitoes as we cannot use polytene chromosomes from ovarian nurse cells. In salivary glands, the Y chromosome does not polytenize but becomes a part of the chromocenter where all chromosomes contact each other (Sharakhova *et al*., 2006). Mitotic chromosome FISH is an alternative but the sensitivity is low for mapping short single copy sequences such as *GUY1* onto highly heterochromatic Y. Male-specific amplification is accepted as strong evidence for location on the Y (e.g., Kryzinski *et al*., 2004; Ng'habi *et al*., 2007). In addition to male-specific amplification of *GUY1* in seven strains of all three biological forms of *An. stephensi*, we have also shown other evidence including the transfer of *GUY1*-containing DNA to the abdomen of mated females, *GUY1* being a single copy (single allele) gene in males, and the lack of expression of *GUY1* in females or 0-1 hr embryos. Therefore we have described one of the strongest cases of Y genes in non-*Drosophila* insects.

GUY1 is transcribed at the very onset of zygotic transcription and encodes a protein that contains two alpha helices and has the potential to bind DNA. The *GUY1* protein may belong to either one of the two major families of transcription factors, namely the helix-loop-helix or the helix-turn-helix proteins (Massari and Murre, 2000; Aravind *et al*., 2005). The strength of the predicted structure connecting the two helices in *GUY1* does not allow us to distinguish between these two possibilities with confidence. Interestingly, three helix-loop-helix proteins (*daughterless, deadpan,* and *sisterless-b/scute*) are key transcription factors that trigger the sex determination pathway the early embryos in *D. melanogaster* (flybase.org). The expression pattern of *GUY1* suggests it has a role in very early embryonic development in males. The fact that the *GUY1* promoter functions in XX embryos (Figure 3.4C) suggests that *GUY1* transcription does not require other Y factor(s), which is consistent with the hypothesis that *GUY1* may serve as an initial signal in sex-determination.

GUY1 has no apparent paralogues on the X or autosomes as it is unique to males. Genes on the *Drosophila* Y chromosome often originate from the autosomes through translocation (Carvalho *et al*., 2009). It is not clear whether *GUY1* is a Y-specific innovation or there existed an X- or autosomal- *GUY1* paralogue, which either was lost or took a divergent evolutionary path than *GUY1*. There is no apparent *GUY1* homolog in *An. gambiae*, which is not surprising as there is very low gene content conservation in the Y chromosome of the 12 sequenced *Drosophila* species (Koerich *et al*., 2008).

Y-specific genes such as *GUY1* can provide a novel set of markers for phylogenetic, population and mating behavioral studies. Y chromosome markers lack recombination and offer advantages such as rapid coalescence and reduced interspecific introgression (e.g., Kopp *et al*., 2006; Underhill and Kivisild, 2007). They offer perspectives fundamentally different from the

other types of markers as their inheritance is exclusively paternal. These new markers will help resolve the often intricate phylogenetic relationships of *Anopheles* mosquitoes which often comprise cryptic species complexes. As mentioned above, there are three biological forms of *An. stephensi* that vary in vector capacity. Y-specific markers provide tools for the study of population structure within and between these forms which may offer new insights into the genetic basis of vector capacity and disease transmission. *GUY1* is also a reliable marker to sex *An. stephensi* embryos and early instar larvae, which will be very useful but until now was not feasible. For example, we may now analyze the transcriptomes of male and female embryos and investigate dosage compensation in the early embryos (e.g., Baker and Russell, 2011).

Approximately one-third of the world's population is at risk of infection with *Plasmodium* parasites which cause malaria. In the Middle East and Asia a major malaria vector is *An. stephensi.* Only female mosquitoes feed on blood. Hence, only females transmit disease pathogens. Current genetic strategies for controlling mosquito-borne diseases include the release of sterile males (Sterile Insect Technique, SIT), the release of insects carrying a dominant lethal gene (RIDL), or the replacement or conversion of a pathogen-susceptible population into a pathogen-resistant one (e.g., Benedict and Robinson, 2003; Wise de Valdez *et al*., 2011). A better understanding of the possible functions of Y genes in sexual differentiation will facilitate the production of sterile males for improved SIT. It may also enable the production of highly competitive males to improve SIT and RIDL and to help spread pathogen resistance genes. Genetic manipulation of sex ratios could greatly reduce or crash a local mosquito population (Schliekelman *et al*., 2005) and subsequently control mosquito-borne infectious diseases.

3.5 Experimental procedures

Mosquito rearing and mating

There are three biological forms or races of *An. stephensi* which differ in egg morphology (Subbarao *et al*., 1987; Oshaghi *et al*., 2006). *An. stephensi stephensi* is the "type" form and is an efficient malaria vector in urban areas. *An. stephensi* mysorensis is a rural species and a poor vector. The third form is the intermediate. The Indian "type" form of *An. stephensi stephensi* maintained at Virginia Tech (the VT strain) was used for Illumina sequencing and PCR analysis of genomic DNA and cDNA templates. The VT strain was reared in incubators at 27°C and 60% relative humidity on a 16 hr light/8 hr dark photoperiod. Larvae were fed Sera Micron Fry Food with brewer's yeast, and Purina Game Fish Chow. Adult mosquitoes were fed on a 10% sucrose soaked cotton pad. Mosquitoes were blood-fed on female Hsd:ICR (CD-1®) mice (Harlan Laboratories, http://www.harlan.com). Egg collections occurred approximately 3 days post blood-feeding. Six strains that belong to the three *An. stephensi* forms were included in a PCR survey using genomic DNA as template. They are: two additional *An. stephensi stephensi* strains from the laboratory of Anthony James at the University of California, Irvine (the UCI strain) and the laboratory of George Dimopolous at the Johns Hopkins University (the JHU strain); two *An. stephensi* mysorensis strains collected from Iran (the IRN and the CHB strains provided by Maryam Kamali and Igor Sharakhov at Virginia Tech, and by Mohammad R. Abai and Hamid Basseri at Tehran University of Medical Sciences); and one intermediate form from the London School of Hygiene and Tropical Medicine (the Beech strain). For mating experiments, individual pupae were separated into 50 mL tubes. After emergence, the mosquitoes were sexed based on antennae dimorphism and double caged to mitigate mating through screen. Virgin female mosquitoes were flash frozen 3 days post emergence. Mated female mosquitoes were

obtained by placing males and females in a 3:1 ratio in cages 3 days post emergence. Mating was allowed to occur for 5 days, at which point males and females were separated and flash frozen.

Genomic DNA (gDNA) and RNA sequencing by Illumina

All Illumina sequencing was performed using genomic DNA and RNA isolated from the VT strain of *An. stephensi*. Illumina sequencing was performed separately on *An. stephensi* DNA extracted from virgin adult males and adult females. DNA extraction was performed using DNAeasy from Qiagen (Valencia, CA). The male and female genomic sequence data, each consisting of over 35 million 84 bp paired-end reads, are deposited in the NCBI *Sequence* Read Archive (SRA) with the accession number of SRP013838. Library preparation and Illumina sequencing were performed at the University of California Davis Genome Center. RNA-seq was performed on polyA RNAs across the following embryonic time points and developmental stages: 0-1 hr, 2-4 hr, 4-8 hr, and 8-12 hr embryos, mixed-instar larva, pupa, adult females, and adult males. RNA-seq data of all samples, each consisting of more than 7 million reads of 38-41 bp sequences, are also deposited to SRA and the accession number is SRA054605. RNA-seq library preparation and Illumina sequencing were performed at the DNA facility of the Iowa State University.

gDNA PCR to confirm male-specific amplification

gDNA was isolated from individual adult males and virgin females of various strains. gDNA was also isolated from dissected body parts for the purposes of comparing virgin and mated females. The heads and thoraces of 5 virgin females were pooled together as a sample and the remaining abdomens were pooled together as another sample. The same procedure was

performed for the mated females. All gDNA samples were collected in triplicate and processed using DNAzol® (Molecular Research Center, Inc., Cincinnati, OH). The manufacturer's recommendations were followed for gDNA isolation. After air drying the DNA pellets were re-suspended in 30 uL of nuclease-free water. PCR for the *GUY1* and *RPS4* control genes was performed using the primers listed in Table 3.1. The cycling conditions for all reactions were as follows: 95°C for 5 minutes; 30 cycles of (95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 3 minutes); and final extension at 72°C for 5 minutes. Amplification of *GUY1* from virgin and mated mosquito parts was performed for 40 cycles.

RNA isolation, cDNA synthesis, RT-PCR, and RACE

An. stephensi embryos, larvae, pupae, adult males and adult stage females were collected and flash frozen. RNA was extracted from approximately 100 ug of each of the above mentioned samples following the total RNA extraction protocol for the mirVana RNA Isolation kit (Ambion). Total RNA was eluted into 100 uL of elution buffer and treated with Turbo DNA-free (Applied Biosystems Inc.) stored at -80°C. cDNA synthesis was performed using SuperScript III RT first strand cDNA synthesis kit (Invitrogen). The amount of starting RNA in each reverse transcription reaction was 5 ug. The final cDNA product was diluted 1:5 with nuclease-free water and stored at -20°C. RT-PCR for *GUY1* and *RPS4* control was performed using the primers listed in Table 3.1. The cycling conditions for all reactions were as follows: 95°C for 5 minutes; 30 cycles of 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 3 minutes; and final extension at 72°C for 5 minutes. For 5' and 3' RACE to amplify the *GUY1* cDNA ends, cDNA was made using 1 ug of RNA isolated from 2-8 hr old *An. stephensi* embryos. 5' and 3' RACE cDNA synthesis was carried out using the Firstchoice RLM-RACE kit (Life

Technologies). The manufacturer's recommendations were followed for cDNA synthesis. cDNA synthesis and RACE amplification primers are listed in Table 3.1. Amplification of target template utilized the following PCR cycling conditions: 94°C for 3 minutes, and 35 cycles of 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 30 seconds. The reaction was finished with 72°C for 7 minutes.

DAPI staining of An. stephensi embryos.

We used DAPI staining to estimate the timing of the early embryonic development in *An. stephensi*. Gravid *An. stephensi* females were allowed to lay eggs over a period of 1 hr. Collected eggs/embryos were either processed immediately (for the 0-1 hr time point) or allowed to develop to the desired time period at 27°C. Embryos were treated with 1% sodium hypochlorite briefly and fixed for 1.5 hrs with 9% paraformaldehyde. The chorion was further cracked and peeled according to Goltsev *et al.* (2004). A final concentration of 0.5µg/ml of DAPI (4',6-diamidino-2-phenylindole, MP Biomedicals) in PBS was used to stain the embryos for 10 minutes. Embryos were washed twice with PBS containing 0.1% Tween 20 and photographed at 100x or 200x magnification. For some of the early time points, we counted nuclei by observation at different focal planes within the embryo.

Single embryo gDNA and RNA analysis

Individual 7-8 hr old embryos were placed into 5 uL of Lysis Buffer (0.05M DTT and 10U RNase OUT from Invitrogen) and homogenized. The lysate was flash frozen and stored at -80°C until gDNA preparation or RNA extraction was required. For gDNA isolation 2.5 uL of the lysate was treated with 0.3 mAU of proteinase K for 30 minutes at 28°C, and then inactivated at

95°C for 2 minutes. The reaction was then diluted with 6.5 uL of ddH_2O. For RNA isolation 2.5uL of the lysate was treated with 130U of DNase I (Invitrogen) for 1 hr at 25°C. The reaction was stopped with the addition 1 uL of 25 mM EDTA and incubation at 65°C for 10 minutes. The resulting 4.5 uL reaction was carried over to cDNA synthesis utilizing the previously mentioned SuperScript III reverse transcriptase protocol. The cycling conditions for all reactions were as follows: 95°C for 5 minutes; 40 cycles at 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 3 minutes; and final extension at 72°C for 5 minutes.

Single embryo dual-luciferase assay and sexing of the single embryo.

GUY1 promoter activity was measured using the pGL3 vector containing firefly luciferase. The 847 bp upstream region of *GUY1* was cloned into upstream of the firefly luciferase ORF. The pRL actin plasmid (*Drosophila* actin promoter driving *Renilla* luciferase) was used as a control. The two plasmids were mixed to reach a final concentration of 150 ng/ul and co-injected into 30-35 minute old embryos. Dual-luciferase assay was performed on individual embryos 5 hrs after the injection. The manufacturer's suggestions for the Dual Luciferase Assay (Promega) were followed with a few modifications: 20 uL of Passive Lysis Buffer was used to homogenize the embryo, 20 uL of Luciferase Assay Reagent II was added during the first reading, and 20 uL of Stop and Glow was added during the second reading. This assay was modified from Biedler and Tu (2010). Phenol:chloroform extraction was performed on the single embryo luciferase assay mixture using the Quick-gDNA MicroPrep kit (Zymo Research, Irvine, CA). This gDNA was used as template to determine the sex of the single embryo through PCR targeting *GUY1* and a Y-specific repeat sequences (Table 3.1). As in other PCR assays, *RPS4* was used as positive control to ensure template quality. Genomic DNA of

male embryos should produce positive PCR products for *GUY1*, the Y-specific repeat, and *RPS4*. Genomic DNA of female embryos should only produce positive PCR products for *RPS4*.

Digital Droplet PCR

Copy number of *GUY1* was determined using QX100 Droplet Digital PCR System from Bio-Rad (Hindson *et al*., 2011) with known single copy autosome genes as references. The Bio-Rad Digital Droplet PCR uses microfluidics to convert a single Taqman (Applied Biosystems Inc.) PCR mixture into 20,000 droplets, thus performing and analyzing 20,000 TaqMan assays simultaneously. Results of the Taqman assays are then computed and the copy number of a gene of interest can be determined by comparing to a reference (Hindson *et. al,* 2011). Genomic DNA obtained via the previously mentioned protocol from male and female adult individuals was digested with AvrII (Fermentas, Waltham, MA). The digested DNA was purified with the GFX Gel band purification kit (GE Healthcare, Piscataway, NJ) and used as template in the copy number variation (CNV) experiment. Two probe and primer mixes were used to carry out TaqMan assays: one targeting *GUY1* and the other targeting an autosomal reference sequence (Table 3.1). The two probes are labeled with different fluorescent dyes to allow detection of both in one reaction droplet. Droplet generation and reading was also performed in accordance with the manufacturer's copy number variation protocol.

Bioinformatic analysis of nucleic acid and protein sequences.

Illumina sequences were analyzed mainly using BLAST (Altschul *et al*., 1997) on a 64-bit Thinkmate Linux workstation with 48 processors and 256 GB RAM. For male and female comparisons, male Illumina sequences were used as query and female reads were used as the

database to perform BLASTN with an e-vale cutoff of 1e-3. We chose BLAST over faster short-read analysis programs such as Bowtie (Langmead, 2010) because we are interested in identifying male and female matches that may have insertion or deletions or partial sequence matches. To speed up the BLASTN search, only top match was reported for any given query. Candidate Y sequences were also compared to RNA-seq databases using BLASTN to infer transcription of these candidate sequences. Up to 1 million hits can be reported for each query and an e-value cutoff of 1e-7 was used to minimize hits from paralogous sequences. To search for *GUY1* homologues, BLASTX, BLASTP, and TBLASTN were performed against the NCBI non-redundant databases and the *An. gambiae* PEST genome as well as all trace reads (vectorbase.org). Secondary structure of the deduced amino acid sequence of *GUY1* protein was predicted using Jpred3 (http://www.compbio.dundee.ac.uk/www-jpred/). Helical wheel analysis was performed using the Helical Wheel Projections program (http://rzlab.ucr.edu).

3.6 Acknowledgement

We thank Jim Biedler for help with the dual-luciferase assay. We thank Maryam Kamali, Igor Sharakhov, Mohammad R. Abai, Hamid Basseri, Anthony James, and George Dimopolous for providing various *Anopheles stephensi* strains. We thank Yann Jouvenot and Jack Regan at Bio-Rad for help with digital-droplet PCR. This work is supported by NIH Grants AI77680 and AI105575, the Fralin Life Science Institute, and the Virginia Tech Experimental Station.

3.7 Figure legends, tables and supplemental information

Table 3.1 Primer and probe sequences.

*GUY1*_F [1]	TTTACTCGTCAAAGCTGCCA
*GUY1*_R [1]	GATCCGTTAAAAATTGACACCA
*RPS4*_F [2]	GAGTCCATCAAAGGAGAAAGTCTAC
*RPS4*_R [2]	TAGCTGGCGCATCAGGTAC
*GUY1*_FullcDNA_F [3]	CCTGAAATGATGCTCTGGAAA
*GUY1*_FullcDNA_R [3]	GATAAATAACCCAGACTATGTGAAGG
*GUY1*_RACE_F [4]	CCGGTATAGTCTAATCTCGTGAGT
*GUY1*_RACE_R [4]	GTGGTCTTATGCTATTGTCTGTCA
*GUY1*_RACE_Fnest [4]	GGGAAATTTACTCGTCAAAGC
*GUY1*_RACE_Rnest [4]	CCAATGTCACAGCAGAGTGTT
*GUY1*_promoter_F [5]	TTGAAAAAACACACGGTAAACG
*GUY1*_promoter_R [5]	AATTTGCCTATTTTACAGTGTACGC
AutoRef_ddP_F [6]	ATCACCACTCGTCGTCCGTT
AutoRef_ddP_R [6]	CGAACGAACTCGATTGACCC
AutoRef_ddP_P [6]	HEX-GCAAACACCACAACAGCAGC-BHQ1
*GUY1*_ddP_F [7]	GTCAAAGCTGCCACGGATCT
*GUY1*_ddP_R [7]	TCCAATGTCACAGCAGAGTGTTT
*GUY1*_ddP_P [7]	FAM-TCACAAAGTAGGCGATACAAAAACA-BHQ1
Y-repeat_F [8]	CAAGCCCAAACAAACTCCTC
Y-repeat_R [8]	ATGCAGCATTCAGCTTCCTT

1) Primers used to demonstrate male-specific amplification of *GUY1* and to amplify *GUY1* cDNA.

2) Positive control primers used to perform both genomic DNA and RT-PCR, which produces products of different sizes due to splicing.

3) Primers used to perform RT-PCR to amplify full-length *GUY1* cDNA.

4) Initial and nested primers used to perform 5' RACE (R primers) and 3' RACE (F primers) for *GUY1*.

5) Primers used to clone the *GUY1* promoter to the pGL3 plasmid for the luciferase assay.

6) The autosomal reference primers and probe used for Digital Droplet PCR Taqman assays.

7) The *GUY1* primers and probe used for Digital Droplet PCR Taqman assays.

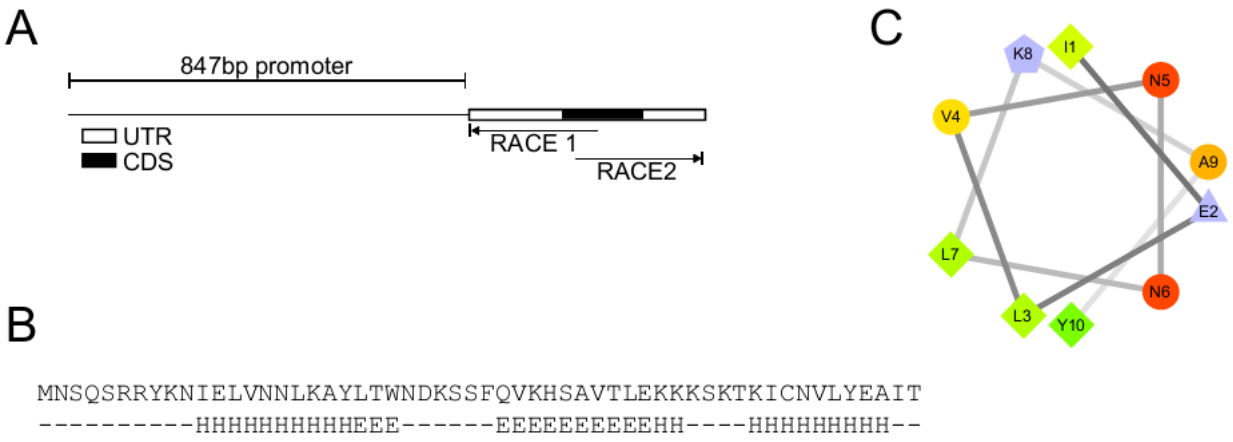
8) Primers used to amplify the Y-specific repeat.

Table 3.2 Number of *GUY1* reads in 11 *An. stephensi* RNA-seq samples.

Sample	**0-2 hr**	**2-4 hr**	**4-8 hr**	**8-12 hr**	**L**	**P**	**M**	**F1**	**F2**	**F3**	**F4**
GUY1 (Raw)	0	54	32	0	0	0	0	0	0	0	0
GUY1 (Normalized)	0	4.15	3.86	0	0	0	0	0	0	0	0

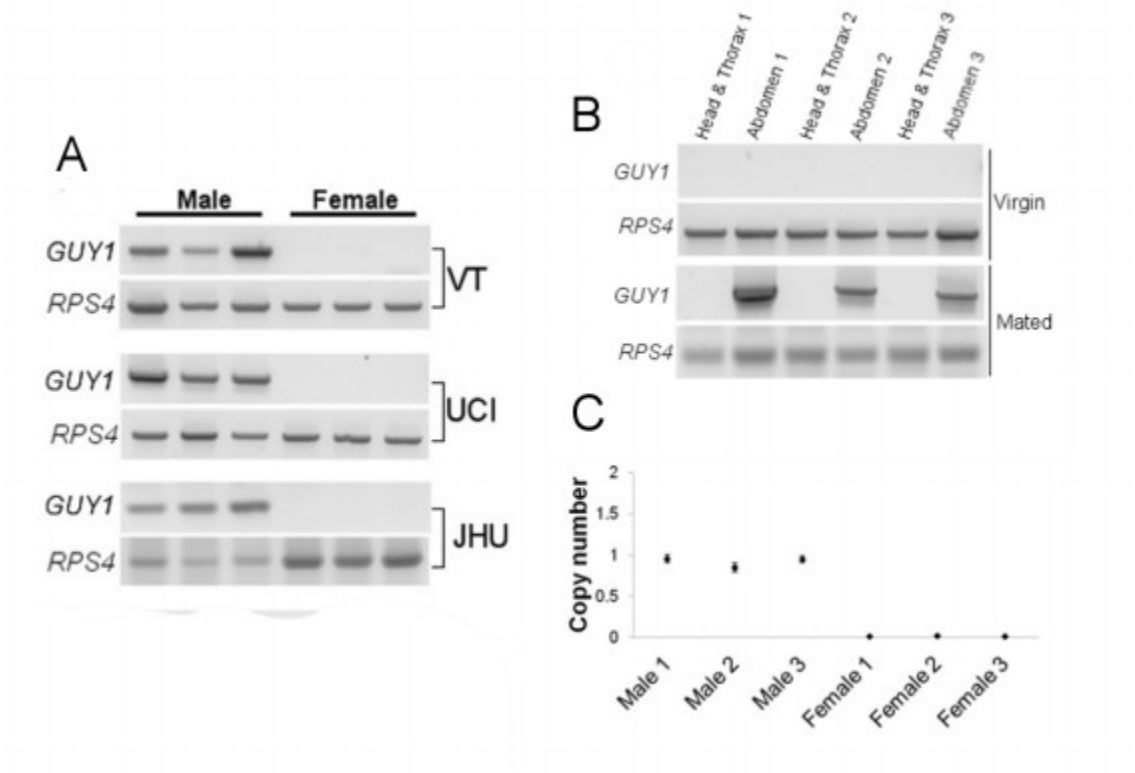
The first four samples are embryos of different time points. L, larvae; P, pupae; M, adult males. F1-4, four adult female samples including ovaries and carcasses from sugar-fed and blood-meal females. Normalized *GUY1* reads are calculated as the number of *GUY1* reads per million of total reads of that sample.

Figure 3.1



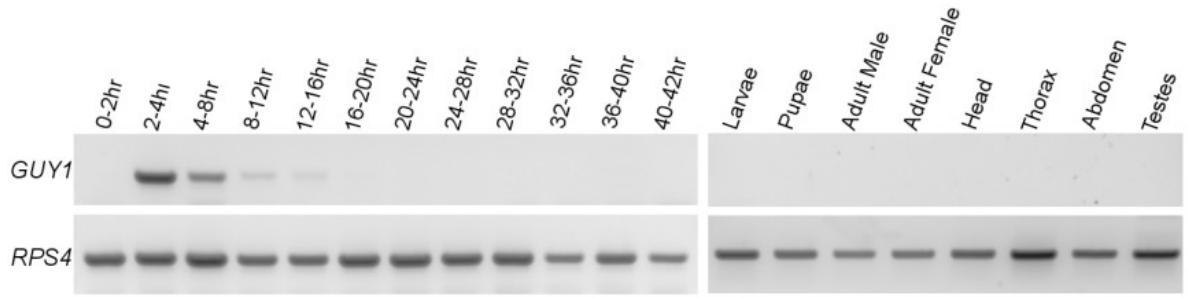
Structure of the *GUY1* gene and its predicted protein. A) *GUY1* gene schematic. The structure is drawn according to the cDNA and genomic DNA sequences that have been submitted to GenBank (JX174417 and JX154128). 5' and 3' RACE were performed to experimentally obtain the full length cDNA sequence. The promoter region shown is the region that was investigated by a dual-luciferase reporter assay as shown in Figure 3.4. B) Predicted secondary structure of the *GUY1* protein. The top line is the sequence and the bottom indicates the secondary structure. H, alpha helix; E, extended. C) The helical wheel depiction of the first alpha helix. Note that hydrophobic residues from V4 to Y10 are located on one side of the helix.

Figure 3.2



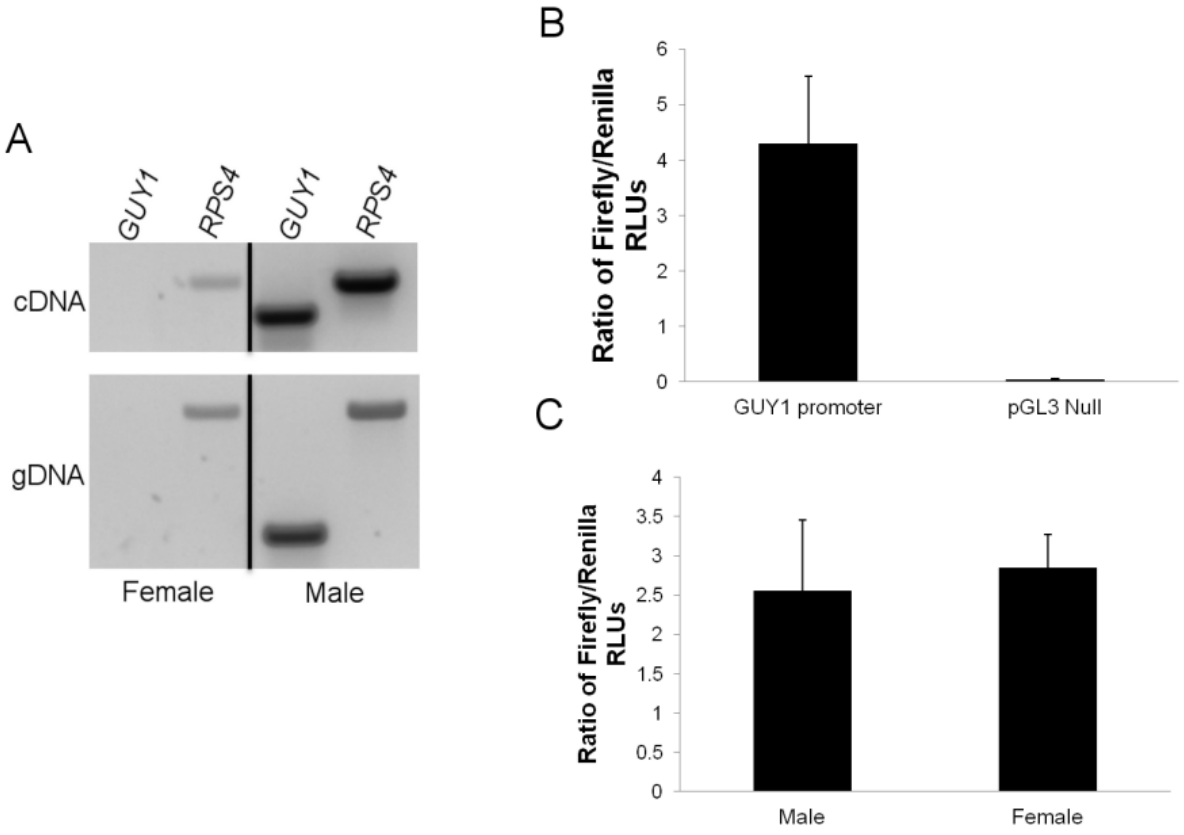
GUY1 is a male-specific and single copy sequence in *An. stephensi*. A) Amplification of *GUY1* from adults of various strains of *An. stephensi*. VT, UCI and JHU are strain names. *RPS4* is amplified from each template as a positive control. B) Transfer of *GUY1*-containing DNA to female abdomen during mating. Pools of virgin and mated females were dissected. DNA was isolated from the collective heads and thoraces as well as the abdomens. *GUY1* was only amplified from the abdomens of mated females. *RPS4* was successfully amplified from all samples. C) Determination of *GUY1* copy number by Digital Droplet PCR (ddPCR). ddPCR was performed on genomic DNA isolated from individual adult male and female mosquitoes. Copy numbers are reported for each individual with Poisson correction.

Figure 3.3



Expression profile of *GUY1*. The cDNA samples included the following embryonic time points: 0-2 hr, 2-4 hr, 4-40 hr in 4 hr increments, and 40-42 hr. Larvae, pupae, whole adult male and female, and adult male tissue or body parts (head, thorax, abdomen, and testis) were also used. *RPS4* was amplified from each template as a positive control.

Figure 3.4



Sexing the single embryo and single embryo luciferase assays. A) gDNA and cDNA derived from the same embryo, at 7-8 hrs post oviposition, was used as template for PCR. Male embryos were shown to express *GUY1* whereas female embryos do not. B) *GUY1* promoter activity measured by a dual-luciferase assay. Embryos were injected with two plasmids: *GUY1* promoter driving the firefly luciferase and the *Drosophila* acting promoter driving the *Renilla* luciferase. Individual embryos were assayed for both luciferase activities and the *Renilla* luciferase activity is used for normalization and the ratio of the relative light unit (RLU) is reported. The averages and standard errors are shown (p-value=0.001 for comparison between the *GUY1* promoter and the pGL3-null). C) *GUY1* promoter activity in male and female embryos. The assays are the same as described in panel B. Here, embryos of known sex were grouped according to gender.

The *GUY1* promoter activity measured as the ratio of RLUs is plotted for each sex. The *GUY1* promoter activity is not significantly different between male and female embryos (p-value=0.39).

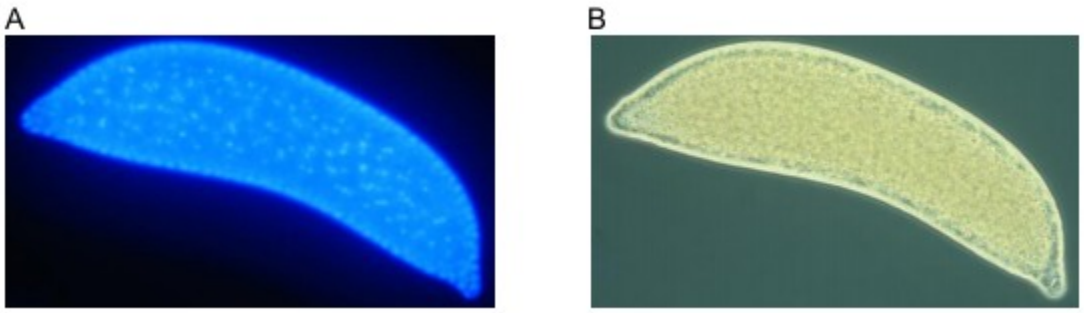
Figure 3.SF1 Male-specific illumina sequences that had three or more total hits in RNA-seq databases.

>SOLEXA3_0001:5:34:17429:5892#0/1
TCCTCGATTCGGGCTGGCAATCTTTGACGGATTTCTACTTCTTTGTCATCCTTCAAGCCGCAGACGAACACGAGACATTTGAGC
>SOLEXA3_0001:5:116:2025:16104#0/1
CTTGGATATGGATTCTTCACGGTAACGTAACTGAATGTGGAGACGTCGGCGTGAGCCCTGGGAGGAGTTCTCTTTTCTTCTTAA
>SOLEXA3_0001:5:107:11877:14157#0/1
ATCGTCCAGATACAACATAGGAACAAGTAACGACGGAATGCCAGCGCATCATCAATTTGAGACACGACAGTGCTATGATCGAGC
>SOLEXA3_0001:5:35:13462:17986#0/1
CGGTGCTCGCAAGCGATTGTTCTTGCGGCGGTTCACCCCCCCCCCCCCCGCCCCATGAAGTCACTCTCCAAAATTTTCCCAATTT
>SOLEXA3_0001:5:86:3889:3621#0/1
TATCTGAGCTTTGGTACTGAGAACCAGAACCTCTTAAGAGGATGATTAATAAGGCTAAATTGAGAAAATTAAGGGCCACTGGCT
>SOLEXA3_0001:5:54:6362:12673#0/1
AACATAGGAACAAGTAACGACGGAATGCCAGCGCATCATCAATTTGAGACACGACAGTGCTATGATCGAGCGTGAGACGACTGA
>SOLEXA3_0001:5:18:12173:16523#0/1
TGAACTTCTGTAATAATAAAATCGGGGCCGTTTAACCAAAAGTGCATATTGTCTTTTCTCACCAATTTGGTAGCCGGAAGCCCT
>SOLEXA3_0001:5:37:12708:17941#0/1
GATCGTTCCAGTCTCGCAGGGGCTGTCGCATTCATCAAGTTGTGTTAAAAATGTGCGGAAAGAACGCGGCGCAATGAAGGTCTC
>SOLEXA3_0001:5:12:18749:4903#0/1
AACACGTCCTAAATCCGCACACAGAGTTCTGCTAGCAATAAGGCAGCTTGTAGTTCCAGTCGTGGTAACGTCAACTCCTTCAAG
>SOLEXA3_0001:5:110:14324:11220#0/1
TTCCCGGAAGCGATGTTGAATAGGGAAAAGTTGGGATTTGTCTAGGGCAAGGGGGCAAGGGTGCTTACCACCCGTACGGGGGAC
>SOLEXA3_0001:5:59:3028:13379#0/1
ACCAAACGAATTGGCCAGGTACAATGGTACACGTCATAAGAATGCCACCATATTGCGGAGCCCAGTACGCAGAGAACCTATCTT
>SOLEXA3_0001:5:82:18840:11671#0/1
CAAGTCCCACTTACTTAAAGGCATCCCTCGGTTCATGGGGGTTCGCTCCTCTCCGGGTCTGAGTGCACGTAGGTCGTTTGGTGT
>SOLEXA3_0001:5:65:11369:7613#0/1
ACCAAAGATAGGTGCTCTGCGTACTGGGCTCCGCAATATGGTGGCATTCTTATGACGTGTACCATTGTACCTGGCCAATTCGTT
>SOLEXA3_0001:5:98:8335:16693#0/1
GGCCATCGTACCGCCCCATAGGTAAGGGAATTATCAACTGAAGTATTTAAGAGTAAAGTTATGACTGAAGAGCATCGGATCATT
>SOLEXA3_0001:5:101:5458:1867#0/1
AACGGCCCCGATTTTATTATTACAGAAGTTCAACCGGAACCTGGCAATGAGGATAGCACAACGCTAGAAAAGACTGCACCACCT
>SOLEXA3_0001:5:112:3370:20312#0/1
TTCTTTACAGGTGTGTTTTACAGTTTCAATGCTAAAAACGTCTGCTTTAAGCCCACCCCACAAACCCACCCCCCCACCTCCCCG
>SOLEXA3_0001:5:49:7640:5186#0/1
CATATTGTCTTTTCTCACCAATTTGGTAGCCGGAAGCCCTCTGGAAACGATATCTGCAGGGTTGAGCTTTGTCGGCACGTACTG
>SOLEXA3_0001:5:102:18878:6018#0/1
CTGAAGATTGCTTCTTGAAAAATCTCATTTTAGTTATTTAGAGGCTTCGATTGGGTAATTTTGGTTTCTCATTTCACCAGAATA
>SOLEXA3_0001:5:96:18347:18382#0/1
GAGTGTTTAACCTGAAAACTTGATTTATCATTCCAAGTCAGATAAGCTTTCAGATTATTCACTAATTCAATGTTTTTGTATCGC
>SOLEXA3_0001:5:120:1419:12543#0/1
TCTGGGAAATTTACTCGTCAAAGCTGCCACGGATCTCGCTCAAGAATTTAGTTTTAAGTTTTAGATAACAGAAAGCGTACACTG
>SOLEXA3_0001:5:75:16977:20196#0/1
AAAAATAAAACAAGAAAAATTACTACGAACCAACTATTGGTTCATATTCATTATACCGGTGATATGAAATTATCGGTCTTATAG
>SOLEXA3_0001:5:64:10127:15111#0/1
GACAAAGCTAAACCCTGCAGATATCGTTTCCAGAGGGCTTCCGGCTACCAAATTGGTGAGAAAAGGTAATATGAACTTTCGGCT
>SOLEXA3_0001:5:114:16984:5716#0/1
TCTGTGGATAACATTCACAAAGAGTTCATCATTGGAAACTTATTCTGGGAAATTTACTCGTCAAAGCTGCCACGGATCTCGCTC
>SOLEXA3_0001:5:30:6020:14001#0/1
GAATGCCAGCGCATCATCAATTTGAGACACGACAGTGCTATGATCGAGCGTGAGACGACTGAACGAGTAAATGCAGTGCAGTAC
>SOLEXA3_0001:5:34:14763:18161#0/1
ATCGTTCCTTTACACCAGCTCTCACTAACCATCTGAAAAACTAAAAAATGCCACCCAGCACGAAATTTCACTTCAAAGCTATCG

>SOLEXA3_0001:5:84:6841:14573#0/1
ATGTGATCGACGACGACGAAGGGCAACAGTCGAGGCTAGTTTGTGCCAAATCAAGAGTTACACCCTTGAAGGAGTTGACGTTAC
>SOLEXA3_0001:5:3:4581:21372#0/1
TTGAAATTTGGCTCCATAGCCTCAGCTTTACTGTCGTTAAACATCAAGCCTGACAATGTATAATTTATCAAAAACTGATTTAAA
>SOLEXA3_0001:5:66:8997:17082#0/1
GGAGGAAATGTGAGTTTTCCACCGCACAAAAAGCCGACAACAACAGTTACCAAGGCTACCTTCTCGTCGTCGTCGCCCTGGCCG
>SOLEXA3_0001:5:97:1789:15813#0/1
ATAATGAATATGAACCAATAGTTGGTTCGTAGTAATTTTTCTTGTTTTATTTTTGGTTGGACACACACAGGAATTTCTTATATT
>SOLEXA3_0001:5:85:17652:4106#0/1
ATTGGCCAGGTACAATGGTACACGTCATAAGAATGCCACCATATTGCGGAGCCCAGTACGCAGAGCACCTATCTTTGGTCAGTC
>SOLEXA3_0001:5:101:12997:8213#0/1
TACAGCGGCGGCGACAATGACTACAACCACCACCACCAGCCGCAGCAAAAACCCCCCCACGTGGGAGGGGAGAAAAGAGGGGGG
>SOLEXA3_0001:5:33:9040:9182#0/1
AAATAGGCAAATTATGAATTCACAAAGTAGGCGATACAAAAACATTGAATTAGTGAATAATCTGAAAGCTTATCTGACTTGGAA
>SOLEXA3_0001:5:12:10315:5493#0/1
TCAGTCGTCTCACGCTCGATCATAGCACTGTCGTGTCTCAAATTGATGATGCGCTGGCATTCCGTCGTTACTTGTTCCTATGTT
>SOLEXA3_0001:5:79:13624:20593#0/1
TGGAAACTTATTCTGGGAAATTTACTCGTCAAAGCTGCCACGGATCTCGCTCAAGAATTTAGTTTTAAGTTTTAGATAACAGAA
>SOLEXA3_0001:5:105:6212:12853#0/1
CCAAAAGTGCATATTGTCTTTTCTCACCAATTTGGTAGCCGGAAGCCCTCTGGAAACGATATCTGCAGGGTTGAGCTTTGTCGG
>SOLEXA3_0001:5:57:14135:15887#0/1
CCAGGACTGACCAAAGATAGGTGCTCTGCGTACTGGGCTCCGCAATATGGTGGCATTCTTATGACGTGTACCATTGTACCTGGC
>SOLEXA3_0001:5:81:11625:9965#0/1
GTACTCTATGAAGCAATCACATAAATAGTTCACAACGAAAAAAAGGCACATTTAGCATATCTGAAAATAAAACGTGACAGACAA
>SOLEXA3_0001:5:60:10809:3700#0/1
CTACCATGGAGGAAATGTGAGTTTTCCACCGCACAAAAAGCCGACAACAACAGTTACCAAGGCTACCTTCTCGTCGTCGTCGTC
>SOLEXA3_0001:5:13:10350:8377#0/1
CTTTTTCTTGAAAGGTGCTGTGTCTGCTTTAGTGGATGACGGCATGGAAATGGACGACGCTCCCGCAGGCGATACGCTGGTGGA
>SOLEXA3_0001:5:117:14039:19939#0/1
CATAATTGCCCATCTCTAGAGCCCCCCTCTAACGTTAAAGCTGTGTTCGTTCCACCTTTACCATCCTCCTCTAGGAGGCAGCCT
>SOLEXA3_0001:5:33:4593:3972#0/1
CTGTTGATAATAAGGGCTTTTGGCCTCAGAAAGCTCACCCACCTCTTTCACCACTCTGATGGTCTCCGATGCCGGTTCCGGAAC
>SOLEXA3_0001:5:75:17000:20197#0/1
AAAAATAAAACAAGAAAAATTACTACGAACCAACTATTGGTTCATATTCATTATACCGGTGATATGAAATTATCGGTCTTATAG
>SOLEXA3_0001:5:50:14832:9419#0/1
GGCCATCGTACCGCCCCATAGGTAAGGGAATTATCAACTGAAGTATTTAAGAGTAAAGTTATGACTGAAGAGCATCGGATCATT
>SOLEXA3_0001:5:1:3895:10610#0/1
CAGCGCATCATCAATTTGAGACACGACAGTGCTATGATCGAGCGTGAGACGACTGAACGAGTAAATGCAGTGCAGTACAAAGGA
>SOLEXA3_0001:5:114:3778:8626#0/1
GCCAAATCAAGAGTTACACCCTTGAAGGAGTTGACGTTACCACGACTGGAACTACAAGCTGCCTTATTGCTAGCAGAACTCTGT
>SOLEXA3_0001:5:79:13624:20593#0/2
GATCCGTTAAAAATTGACACCATAGCCTCTATTTTACTATATTAACACACAAAGTGGTCTTATGCTATTGTCTGTCACGTTTTA

Highlighted in green are the paired-end sequences that were used to design primers (Table 3.1) to uncover *GUY1*. Note that the names of the sequences of the paired-end are nearly identical except they end with either #0/1 or #0/2. Here we only showed the pair of the *GUY1* sequences used for primer design. We only showed one end of the other male-specific sequences.

Figure 3.S1



A typical *Anopheles stephensi* embryo at 3 to 4 hrs after egg deposition. A)DAPI (4',6-diamidino-2-phenylindole) staining of the nuclei; B)Bright field image. Both were taken with 200x magnification.

Chapter 4

The transgenic expression of GUY1 and its implications in female lethality

4.1 Abstract

Anopheles mosquitoes are the main malaria vectors and their sex determination is thought to be controlled by a dominant male-determining factor on the Y chromosome. However, the identification and properties of the Y chromosome factor responsible for the initiation of the sex determination cascade have yet to be elucidated. Moreover, investigation of this pathway would provide strong support to genetic strategies for vector population control. We study a previously discovered Y chromosome specific gene, GUY1, in *An. stephensi*, one of the main mosquito vectors of the malaria parasite in India and the Middle East. Through the use of transgenic mosquitoes we demonstrate the influence of the transgenic construct on the sex of the population. PCR and fluorescence *in situ* hybridization (FISH) confirm the autosomal insertion of the GUY1 containing construct. Lines containing the GUY1 containing transgene are heavily male-biased or completely male. We determine that death of females is caused by the transgene through PCR genotyping of embryo, morphological analysis of larvae, and ddPCR to confirm the number of copies of X chromosomes each individual adult contains. We hypothesize that GUY1 plays a role in sex determination and possesses deleterious qualities towards female embryos. In addition to GUY1's potential function we also propose its use as a method of vector control via population crash through the release of transgenic males.

4.2 Introduction

Y chromosome genes are often difficult to discover due to their localization an area rich in repetitive DNA, transposable elements and heterochromatin (Bachtrog 2006, Barton and Charlesworth 1998). These genes are rarely conserved even amongst very closely related species (Koerich 2008 and Carvalho 2009) due to the Y chromosome's low rate of recombination and accumulation of mutations (Charlesworth *et al*., 2009). The functions of Y chromosome genes have only been investigated in depth in *Drosophila* (Carvalho 2001 and Vibranovski 2008) and are involved in processes such as spermatogenesis, reproductive organ development, and mating competence (Carvalho 2000 and Gepner 1993).

Sex determination systems evolve relatively quickly as seen by the varied incarnations of the pathway (Marin and Baker 1998) and have only been characterized in a few model insects. A master sex determination factor on the Y chromosome has not been characterized in any insect. Within insects there are a number of chromosome compositions such as XY, X:A, ZW, and haplodiploidy (Kaiser and Bachtrog 2010) which govern the fate of an organism's sex. In the case of *Anopheles* mosquitoes, the vectors of malaria, it has been implicated that the male determining factor resides on the Y chromosome (Baker and Sakai, 1979). Although downstream genes of the sex determination pathway such as *tra* and *tra-2* are relatively conserved between different species (Dauwalder *et al*., 1996 and Li *et al*., 2013) the initiator/s of the pathway are highly divergent (Bull 1983).

In vector populations there are a number of genetic strategies being investigated which utilize males to either spread genes capable of rendering the vector refractory to disease or to crash a local vector population (Wilke *et al*., 2009 and Robinson *et al*., 2004). Distortion of the sex ratio of a vector population has thus far not been performed through the use of transgenic manipulation of Y chromosome genes. The current genetic strategies in pest populations include

the sterile insect technique (SIT), which has been proven effective against the screw worm in North America (Bushland *et al*., 1955), and the release of insects carrying a dominant lethal (RIDL). Both strategies rely on the release of males to either crash or replace a population, as in the cases of SIT and RIDL, respectively.

In this study we experimentally characterized the effect of a previously determined Y chromosome gene GUY1 in *Anopheles stephensi*, the Asian malaria vector. GUY1 exhibits a potential secondary structure of a bHLH/bHTH protein, similar to *daughterless* in *Drosophila*. Early embryonic expression of GUY1 is characteristic of a gene potentially being involved with sex determination. The dblwhammy_GUY1 construct was integrated in the genome via piggyBac transposition and resulted in ectopic expression of GUY1, yielding a phenotype that heavily biases the transgenic population to males that is not attributed to Y-linkage.

4.3 Materials and methods

Construction of transgenic cassette and mosquito transformation

The piggyBac transformation backbone plasmid was obtained from Dr. Zachary Adelman and is described in Adelman *et al*., 2002. The EGFP reporter gene was replaced with the coding sequence for the monomeric DsRed protein (Campbell *et al*., 2002). The addition of the attB site was designed as an option for phiC31 recombinase mediated site-specific transformation (not used in these experiments).

Dblwhammy_GUY1

Construction of the transgene was performed by Epoch Life Science, Inc. The transgene utilizes the *An. stephensi* bZIP1 1kb upstream promoter (previously described) which contains the first exon of the gene containing the bZIP1 5' UTR. Following the bZIP1 5' UTR are 168 nts

of the GUY1 coding sequence. The 3' end of the CDS is joined with the final 20bp of the 3' end of exon 1 of bZIP1. Following the end of this exon is the first intron of the bZIP1 gene. The native miRNA cluster within this intron was modified to generate miRNA targeting the GUY1 5' UTR, 3' UTR, and CDS. Following the miRNA containing intron are 20bps of the the 3' end of exon 2 of the bZIP1 gene joined by the reverse compliment of the GUY1 coding sequence. This construct was flanked by AscI and BbvCI restriction enzyme sites and introduction of the transgene was performed by double digest with AscI and BbvCI (Fermentas). Double digests were performed with 2 ug of plasmid DNA for 20 minutes at 37°C in 1X FastDigest® Buffer and 1 uL of each enzyme (following manufacturer's recommendations). Digested samples were analyzed via gel electrophoresis and the bands that corresponded to the appropriate linear product were excised. DNA was purified from the gel slices following the manufacturer's suggestions of the illustra™ GFX PCR DNA and Gel Band Purification Kit and eluted in 20uL of ultra pure H_2O. DNA concentrations were determined via nanodrop and 500ng of cassette and plasmid backbone were combined with 3U T4 DNA ligase and 1x Rapid Ligation Buffer (Promega) for 2 hours at room temperature before being transformed in JM109 cells (Promega). Cells were spread on ampicillin containing plates and colonies were screened via PCR for containing the appropriate plasmid (Table 4.1).

pEE1_GUY1

This plasmid was constructed from the dblwahmmy_GUY1 plasmid by removing the area between AvrII and BbvCI (the dsRNA GUY1 region). The GUY1 protein coding sequence was amplified and purified then adapted with AvrII (5' end) and BbvCI (3' end) restriction enzyme sites via PCR (primers in Table 4.1). Double digest was performed with 500 ug of PCR

product and 2 ug of backbone plasmid, followed by ligation and cloning (as previously described).

pEE2_GUY1

This plasmid was constructed from the dblwhammy_GUY1 plasmid by removing the area between AscI and PacI and introducing the full length GUY1 cDNA sequence (5' UTR, CDS and 3' UTR) driven by the GUY1 native promoter. The full length sequence was amplified and purified then adapted with AscI and PacI sites. Digestion, ligation and cloning were performed as stated previously.

The transformation plasmids were sequence verified by Sanger sequencing (VBI) and transformed into *An. stephensi* at the Insect Transformation Facility at University of Maryland, Rockville. Adult G0's were counted for the number of males and females to observe transient effects of the constructs.

Mosquito rearing, mating and sexing

Transgenic mosquitoes were reared in incubators at 27°C and 60% relative humidity on a 16 hr light/8 hr dark photoperiod. Larvae were fed Sera Micron Fry Food with brewer's yeast, and Purina Game Fish Chow. Adult mosquitoes were fed on a 10% sucrose soaked cotton pad. G0 adult males were individually mated in a 1:5 ratio with wild type virgin females while G0 females were pooled and mated with wild type males in a 1:5 ratio. Transgenic lines predominantly produced all male populations, thus subsequent generations were maintained by supplying transgenic males with wild type Indian strain females. Mosquitoes were blood fed on female Hsd:ICR (CD-1®) mice (Harlan Laboratories, http://www.harlan.com). Egg collections occurred approximately 3 days post blood feeding. Transgenic positive larvae were screened for

DsRed expression during the L-3 stage and non-transgenic individuals were discarded. Sexing of male and female larvae was performed under light microscopy by the detection of larval testes to separate. Competitiveness of transgenic males was determined by combing 10 wild type males with 10 transgenic males and then adding 20 wild type females. Mating was allowed to occur for 48hrs at which point the females were blood fed. Embryos were collected and screened as previously mentioned but the non-transgenic progeny were also reared. The number of males and females for transgenic and non-transgenic populations were counted during the adult stage.

gDNA isolation and PCR

Genomic DNA was isolated from pools of 5 males from transgenic lines and wild type mosquitoes. Isolation of DNA was performed with the Zymo Quick-gDNA™ Mini prep gDNA isolation kit. The manufacturer's instructions were followed and DNA was eluted to a final volume of 50 uL. The cycling conditions for all PCR reactions were as follows: 95°C for 5 minutes; 30 cycles of (95°C for 30 seconds, an annealing temperature of 2°C less than primer Tm for 30 seconds, and 72°C for 3 minutes); and final extension at 72 °C for 5 minutes (See Table 4.1 for primer pairs). All PCR products were analyzed on 1% agarose gels with visualization achieved through the use of GelRed. ddPCR was performed according to the manufacturer's suggestions. Primer and probe mixes were first run through a 10°C temperature gradient to obtain optimal annealing temperatures with the given experimental and reference TaqMan® mixtures. Single embryos from transgenic lines were sexed as previously described (Chapter 2); however, the entire embryo was processed for gDNA preparation and primers targeting DsRed, GUY1 native and RPS4 were used in PCR.

cDNA synthesis and RT-PCR

Embryonic samples of 50 uL were collected for each 2 hr time point. RNA was extracted following the total RNA extraction protocol for the mirVana RNA Isolation kit (Ambion). To generate cDNA 1 ug of DNase treated RNA was used with the SuperScript III First Strand cDNA Synthesis kit (Invitrogen). Standard PCR cycling conditions were used for the detection of transcripts and primers are listed in Table 4.1.

Polytene chromosome preparation and FISH

Polytene chromosome slide preparations were obtained from salivary glands of 4th instar larvae. They were first fixed in Carnoy's solution and then spread in 50% propionic acid. Following preparation the slides were frozen in liquid nitrogen and had the cover slips removed. Slides were then dehydrated in an increasing series of ethanol concentrations (70%, 80%, and 100%) and air dried. Transformation plasmid was used as template for nick translations. DNA was labeled with Cy3 dUTP (GE Healthcare UK Ltd, Buckinghamshire, UK) by nick translation with DNA polymerase I (Fermentas, Inc., Glen Burnie, MD, USA). Fluorescent *in situ* hybridization (FISH) was performed using a protocol described in Timoshevskiy *et al*. 2012. Slides were pretreated with 0.1 mg/ml of pepsin (USB corp., Cleveland, Ohio). Hybridization mix contained 50% formamide, 10% dextransulfate (Sigma, St. Louis, MO, USA), and 200 ng of probe per slide. Slides and probes were denatured simultaneously at 90 ^{o}C. Hybridization on the slide was performed at 37^{o}C in a dark humid chamber over night. After hybridization slides were washed in 0.4x SSC, 0.3% Nanodept-40 at 72^{o}C for 2 min, and 2x SSC, 0.1% Nanodept-40 at RT. Slides were counterstained using 1 μM YOYO-1 iodide (Invitrogen Corporation, Carlsbad, CA, USA) solution in 1x PBS and enclosed under antifade Prolong Gold reagent (Invitrogen

Corporation, Carlsbad, CA, USA) by a cover slip. Slides were analyzed using a Zeiss LSM 510 Laser Scanning Microscope (Carl Zeiss Microimaging, Inc., Thornwood, NY, USA) at 400x magnification.

4.4 Results

The GUY1 containing cassette successfully integrated into the An. stephensi genome

The dblwhammy_GUY1 cassette (Figure 4.1) was successfully integrated into the genome of *An. stephensi*. Three lines with independent insertion events were obtained. During G1 there was an observed phenotypic difference of DsRed expression where some individuals had the expected eye and nerve expression and others had additional thoracic expression which were separated into line "A" and line "B," respectively. Females from line A in the G1 stage were used to generate line Af and males with visibly less expression generated line "C". A single male from line C was used to generate line C2. Individual line B males used to generate line B1, which were progeny without thoracic DsRed expression, and Line B2, those with thoracic DsRed expression. PCR results show amplification of the cassette from transgenic lines and not from WT mosquitoes, with the additional amplification of native GUY1 and RPS4 (ribosomal protein subunit 4) as controls (not shown). Screening of larvae under fluorescence microscopy shows expression of dsRed in the neuronal tissues (in addition to thoracic expression in line B2) consistent with the expression pattern of 3xP3/HSP70 (Figure 4.2B). FISH of polytene chromosomes localizes the three insertions sites to the following positions: line B1, 2R:19; line B2, 3R:37; and line Af, 2R:17(Figure 4.2C). An additional band was observed in all lines that corresponds to the bZIP1 promoter, located in scaffold00003 which spans 3R:29B-29D (Sharakhova *et al*., 2006).

GUY1 cassette ectopically expresses GUY1

Embryos from the transgenic lines B2 and Af were used to generate cDNA at 2-4, 4-6, and 6-8hr time points. In addition to expressing native GUY1 and RPS4, transgenic GUY1 was also being expressed (Figure 4.5). Primers designed in the 5' and 3' UTRs of GUY1 were used to differentiate between native GUY1 and transgenic GUY1. Amplification of the transgene is performed by targeting the 5'UTR of bZIP1 (present at the end of the bZIP1 promoter region) and GUY1 CDS, a product only obtained from the transcription of the transgene. Native and transgenic GUY1 expression is seen in all time points in line B2 embryos but is absent in line Af 6-8hr embryos.

Preliminary results of transient GUY1 expression in G0 embryos

Transient effects of dblwhammy_GUY1, pEE1_GUY1 and pEE2_GUY1 are observed in Table 4.3 and measured against a non-related transgenic microinjection experiment. Comparing the GUY1 containing vectors against the control results in a statistically significant ($p<0.05$) difference between male to female ratios in pEE1_GUY1 G0 and control G0 (Student t-test). Additionally, when all G0 experiments are combined (they all ectopically express GUY1) and compared against the control set of injections for statistical analysis, a greater significant difference in sex ratio is observed.

GUY1 produces a male bias in transgenic lines

Analysis of adults from the screening of 3rd instar larvae expressing DsRed (transgenic positives) and those not expressing DsRed (transgenic negatives) show the positive population

being nearly all male in most generations (Figure 4.3) in lines B1 and B2. However, line Af's sex distribution over the generations is nearly 1:1 male to female. The transgenic negatives from the second generation of each line were allowed to mature to adulthood and sexed. The negatives showed a 1:1 male to female distribution.

Competitiveness of the dblwhammy transgenics was determined through a controlled mating experiment where equal quantities of transgenic and non-transgenic males were mated with wild type females in a 1:1 male to female ratio. The adult progeny from this matings in lines Af, B1 and B2 were counted. Following Mendelian genetics, transgenic males would be equally as competent as wild type males if 14.3% or 25% of the progeny were transgenic in the cases of female killing of transgenic embryos or female to male conversion of transgenic embryos, respectively.

GUY1 transgene containing lines elicit a female-lethal phenotype

The transgenic lines show a heavy bias towards males in the adult stage. Individual adult males from line B2 were subjected to ddPCR (primers in Table 4.1) to determine the number of X chromosomes via CNV. Wild type males and females were used to establish a standard against which the transgenics were measured. All line B2 adult males were true XY males possessing a single X chromosome (Table 4.2). Analysis of the embryos from line B2 was performed using primers specific to DsRed, native GUY1, and RPS4. Male embryos would be positive for GUY1 native and RPS4, whereas females would only be positive for RPS4. Transgenic embryos would be positive for dsRed and non transgenics would be negative for dsRed amplification. Of the 96 line B2 embryos analyzed 68 were used (embryos without RPS4 amplification were not used in this data set) 7% were non-transgenic male, 27% non-transgenic female, 19% transgenic male,

and 47% transgenic female (Figure 4.4A and extrapolated from Figure 4.S3). Screening of line B2 larvae separated transgenic positives from negatives based on expression of DsRed. Positives and negatives were also screened for the presence of testes (Figure 4.S2), those bearing testes are male and those lacking testes are female. At the larval stage 26% and 37% were non-transgenic male and female, respectively. The transgenic males and females contributed to 35% and 2% of the population, respectively (Figure 4.4B).

4.5 Discussion

The dblwhammy_GUY1 construct was successfully integrated into the *An. stephensi* genome based on PCR amplification of the transgene from gDNA isolated from DsRed positive adults. FISH also shows successful single insertion onto the autosomes, leading to the designations line B1, B2, and Af (Figure 4.2). In addition to the signals seen that correspond to the transgene there was an additional band observed that corresponded to the bZIP1 promoter (based on BAC clone mapping of a clone containing bZIP1 sequence). The DsRed negatives were also sexed during the adult stage to confirm that the integration was not Y chromosome linked, as seen by the 1:1 distribution of males and females (Figure 4.S4).

The doublewhammy_GUY1 was designed to knock down GUY1, yet expression of transgenic GUY1 was observed in the early embryo. The transgenic lines display an extreme bias towards producing male progeny when mated with wild type females (Figure 4.3). The transient effects of GUY1 expressing plasmids can be seen in the G0 distribution of male and female adults with each transformation attempt (dblwhammy_GUY1, pEE1_GUY1 and pEE2_GUY1). This bias towards males further rules out Y linkage and potential position effects that would play a role in the transgenic phenotype. This phenomenon of male bias is observed throughout all

generations of lines B1 and B2 in the dblwhammy_GUY1 trangenics. Line Af behaves differently and produces transgenic positive progeny of 1:1 male to female distribution when mated with either other line Af females or wild-type females. This discrepancy between the two lines could be explained by insertion position effects or a potential mutation in line Af that would differ it from lines B1 and B2, as all the lines successfully transcribe the dblwhammy_GUY1 construct. Although this construct was originally designed to knock down GUY1 it appears to potentially be ectopically expressing GUY1, as apparent by the presence of native GUY1 in all expected time points. Additionally, the dblwhammy_GUY1 construct possesses the full coding sequence of GUY1; therefore, if the construct fails to knock down GUY1 mRNA it would in turn be expressing a sequence capable of being translated into full length GUY1 protein.

The male bias in these transgenic lines is an obvious phenotype in the adult stage but lends inquiry into whether females are being converted to males or not maturing and dying before adulthood. Genotypic analysis of the adult stage via ddPCR detected no XX males so far and therefore we have not found evidence for females being converted to males. If conversion was taking place there would be males with an XX karyotype instead of an XY karyotype as determined by the CNV of two in females for the X chromosome marker and a CNV of one in males (Table 4.2). Genotyping via PCR targeting Y chromosome and transgenic markers of individual embryos demonstrates that transgenic XX embryos exist in addition to non transgenic XX, transgenic XY and non transgenic XY embryos. Therefore, female embryos carrying the dblwhammy_GUY1 transgene exists. Sexing of larvae shows that a great majority (95%) of the transgenic larvae are males, suggesting the females are dying between the embryonic stage and larval stage (Figure 4.4).

These studies indicate the ectopic expression of GUY1 causes a shift in the sex of a transgenic population to predominantly male, most likely by causing death of females during embryonic or larval stages. Y chromosome genes tend to either have an advantageous effect on males or a deleterious effect on females, thus their retention on the Y chromosome in species where the heterogametic sex is male (Charlesworth and Charlesworth 2000). The ectopic expression of GUY1 could be deleterious to females if it is playing a role in dosage compensation of the X chromosome. Dosage compensation is a phenomenon where male-specific regulatory factors overact the transcription of X chromosome genes to compensate for only possessing a single X chromosome; this is not necessary in females as they have two X chromosomes (Disteche 2012 and Charlesworth 1996). If GUY1 plays a role in dosage compensation, then having its expression in females would lead to over activation of X chromosome genes which could be lethal to female embryos. Nonetheless, GUY1 displays an effect of being deleterious to females; however, whether this effect is related to being a master regulatory gene in the sex determining pathway is still to be determined.

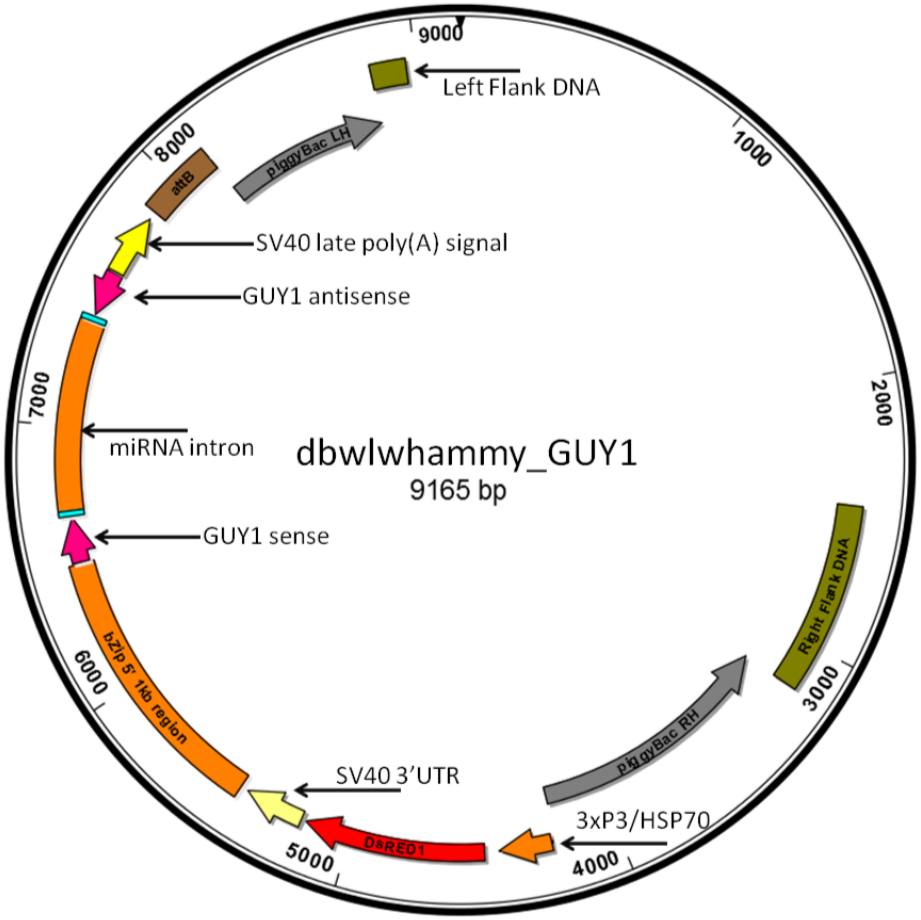
The implications of being able to generate an all male transgenic population of *An. stephensi* are extensive in the field of transgenic control of vector populations. In the case of the sterile insect technique (SIT) or the release of insects carrying a dominant lethal (RIDL) the emphasis on releasing males is directly tied to the success of these efforts. With the addition of ectopic expression of genes like GUY1 the economical expenses of these projects decreases. As females are the important sex in mosquito-borne disease it is of high importance to solve vector control issues without increasing the potential vector population. Therefore, individuals ectopically expressing GUY1 have a skewed bias towards males in their progeny. If the ectopic expression of GUY1 was coupled with a gene drive system or if an individual possessed multiple

insertions of the transgene it may be feasible to cause local population crash based off of the sheer inability to produce female progeny.

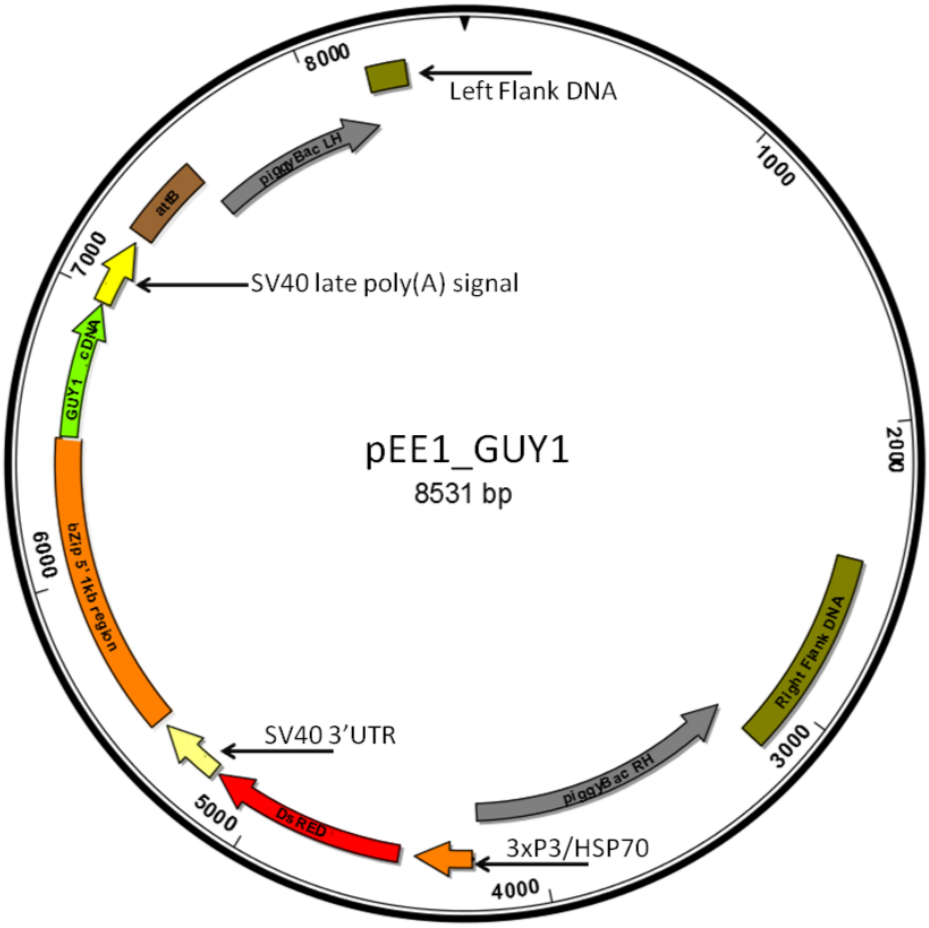
4.6 Figure legend, tables and supplemental information

Figure 4.1 Transformation plasmids

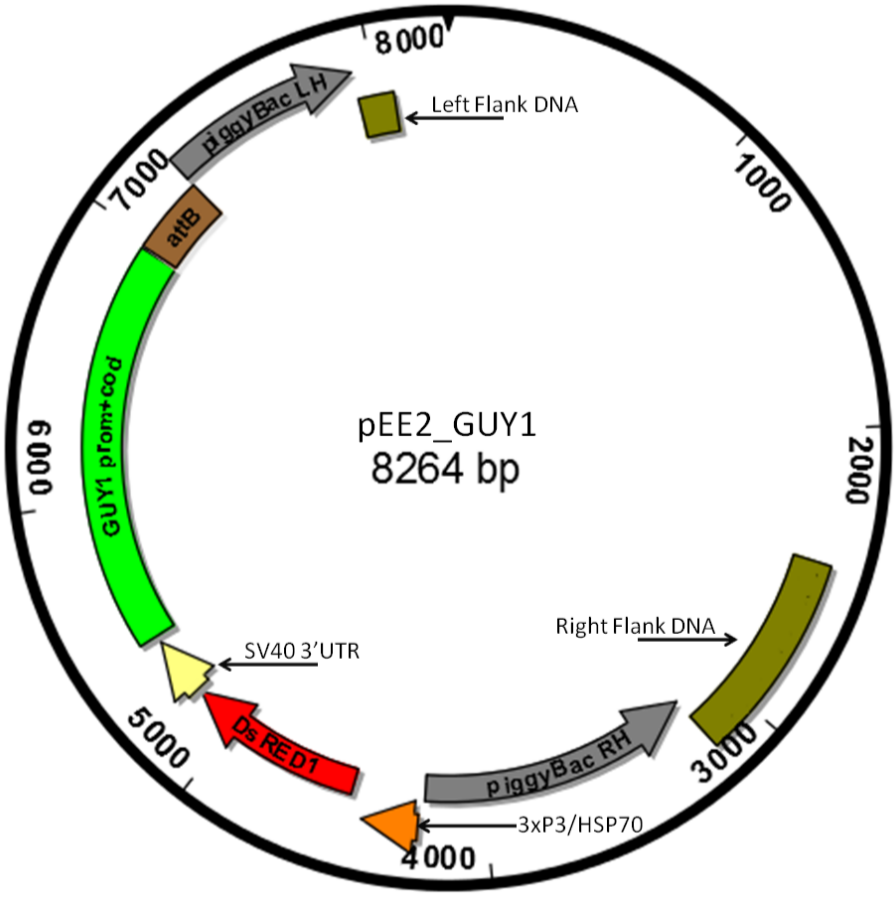
A.



B.

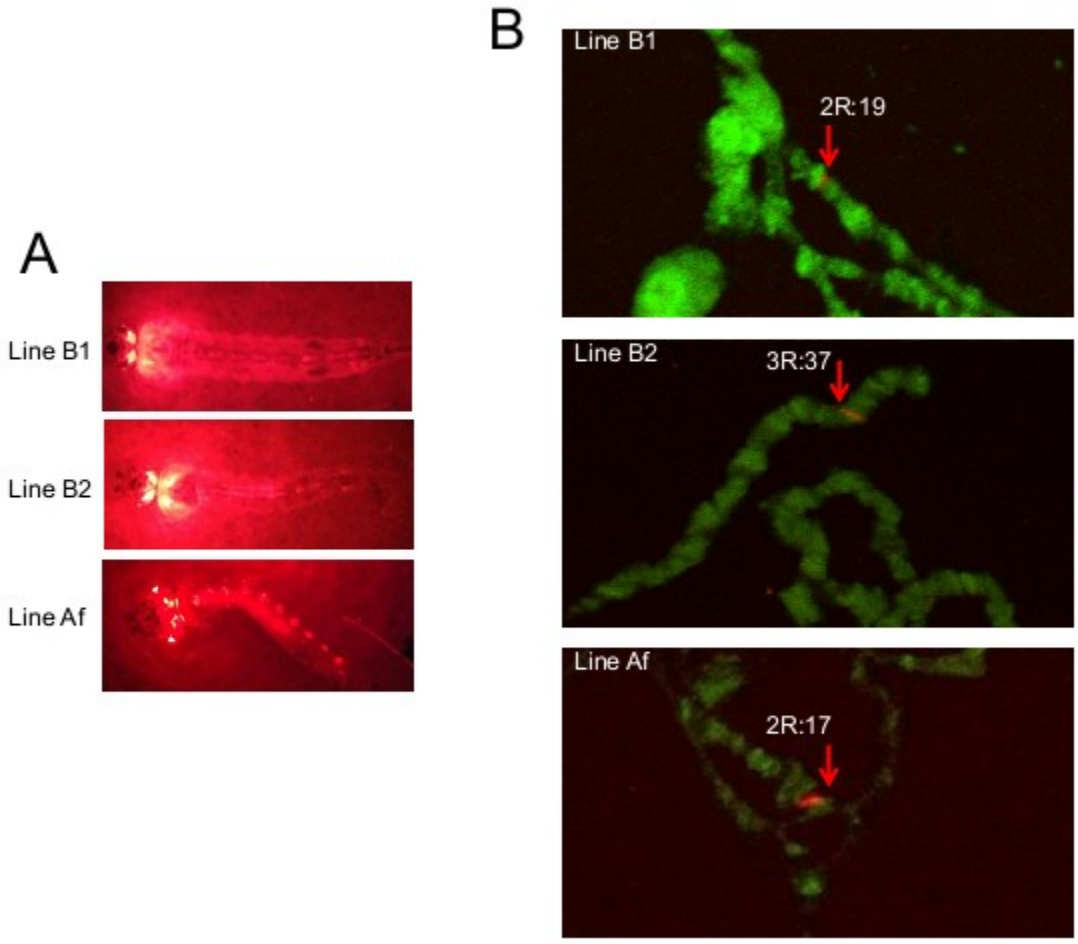

8000
1000
2000
3000
4000
5000
6000
7000
Left Flank DNA
piggyBac LH
attB
SV40 late poly(A) signal
GUY1 cDNA
bZip 5' 1kb region
pEE1_GUY1
8531 bp
Right Flank DNA
piggyBac RH
SV40 3'UTR
Ds RED
3xP3/HSP70

C.



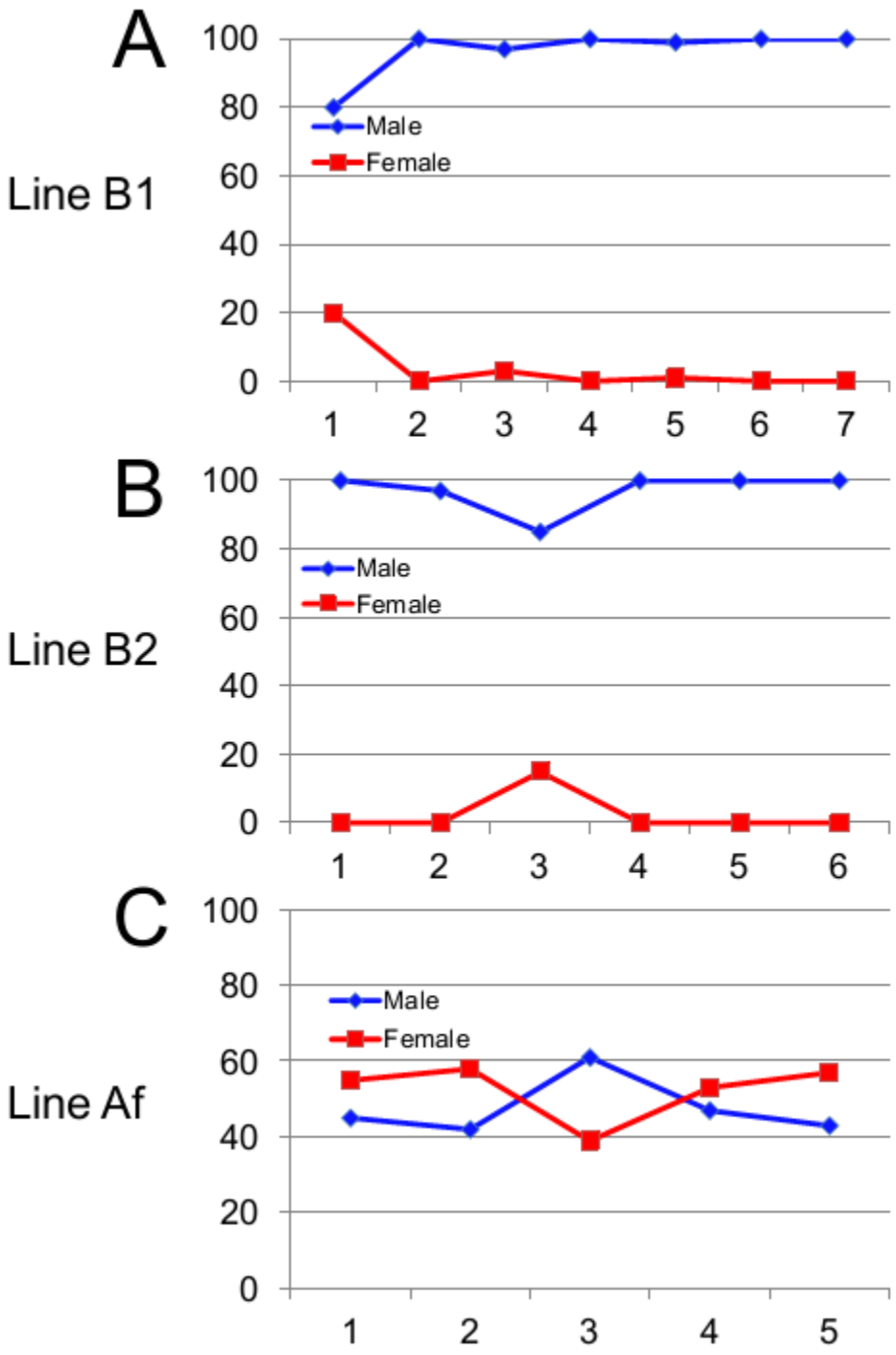
A. The dblwhammy_GUY1 construct was designed with the selectable marker of dsRed driven by a 3xP3/HSP70 promoter. The functional unit of this construct is driven by the *An. stephensi* bZIP1 promoter and contains the sense and anti-sense strand of GUY1 (to produce a dsRNA product upon transcription) separated by a miRNA containing intron. This intron contains three miRNA targeting GUY1's native 3' UTR, 5' UTR and coding sequence. B. The schematic of pEE1_GUY, designed to ectopically express the full length cDNA sequence of GUY1 driven by the bZIP1 promoter. C. A map of pEE2_GUY1, designed to ectopically express GUY1 under its native promoter.

Figure 4.2 Successful transformation of *An. stephensi*



The successful integration of the dblwhammy_GUY1 transgenic construct into the *An. stephensi* genome was determined through PCR, FISH, and fluorescent screening. A. Third instar larvae were screened for the expression of the transgenic marker dsRed using a texas red fluorescent filter. B. FISH results utilizing the dblwhammy transgenic plasmid as probe are shown from larval derived salivary gland polytene chromosomes. Lines B1, B2, and Af hybridized to 2R:19, 3R:37, and 2R:17, respectively.

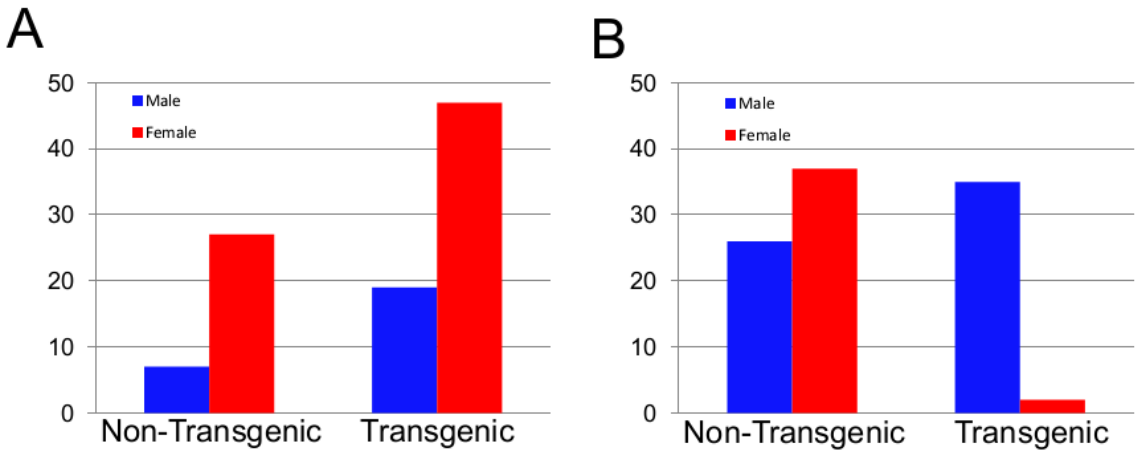
Figure 4.3 Sex of transgenic lines through the generations



The number of males and females over the course of multiple generations was counted during the adult stage for each transgenic line. As shown in lines B1, and B2 (A and B), the number of

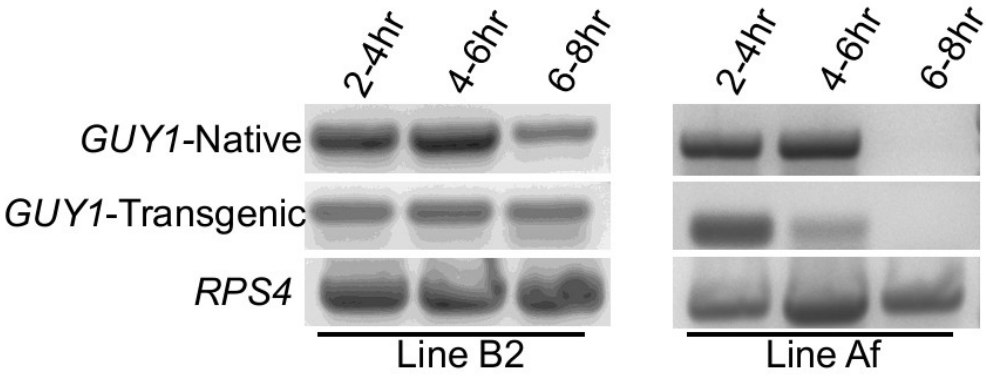
males consistently stays near 100% over seven and 6 generations, respectively. Line Af behaves in a fashion that reflects a 1:1 ratio of males to females as shown over five generations.

Figure 4.4 Determining the time of death of females



The percent of males and females in the embryonic stage (A) and the larval stage (B) of line B2 was acquired via PCR genotyping on individual embryos and phenotypic analysis for the presence of testes in larva. A. Individual embryos (68 total) are shown to be 7% non-transgenic male, 27% non-transgenic female, 19% transgenic male, and 47% transgenic female based on the PCR markers for native GUY1, transgenic GUY1, and RPS4. B. 3rd instar larvae were checked for the presence of testes and screened for expression of DsRed. Of the 113 larva (a single generational cohort), 26% did not express DsRed and possessed testes, 27% did not express dsRed and displayed no testes, 35% expressed DsRed and possessed testes, and 2% expressed DsRed and displayed no testes.

Figure 4.5 Ectopic expression of GUY1 in Line B2



RT-PCR of line B2 (left panel) embryos and line Af (right panel) embryos. Primers were used to amplify native GUY1, transgenic GUY1 and RPS4 from the cDNA templates of 2-4, 4-6, and 6-8hr embryos.

Table 4.1 Primers

Primer Name	Sequence
GUY1_cDNA-F[1]	CCGGTATAGTCTAATCTCGTG
GUY1_cDNA-R[1]	AAAAATTGACACCATAGCCTC
GUY1_cDNA+AvrII-F[1]	CACTCCTAGGCCGGTATAGTCTAATCTCGTG
GUY1_cDNA+BbvCI-R[1]	CTGAGCTGAGGAAAAATTGACACCATAGCCTC
GUY1_Full-F[2]	CCTGAAATGATGCTCTGGAAA
GUY1_Full-R[2]	GAAACGTTTTTCCAACATGTGA
GUY1_Full+AscI-F[2]	CACTGGCGCGCCCCTGAAATGATGCTCTGGAAA
GUY1_Full+PacI-R[2]	TCATGCAATTAATTGAAACGTTTTTCCAACATGTGA
dsRed-F[3]	CCCCGTAATGCAGAAGAAGA
dsRed-R[3]	GGTGATGTCCAGCTTGGAGT
GUY1_Native-F[4]	TTTACTCGTCAAAGCTGCCA
GUY1_Native-R[4]	GATCCGTTAAAAATTGACACCA
GUY1_Trans-F[5]	GCTGAATTGAAGGAAAAGCCTAG
GUY1_Trans-R[5]	TTCCAATGTCACAGCAGAGTG
AnstX_ddP-F[6]	CCTTATGATTTTCTGCCGGA
AnstX_ddP-R[6]	TGTGAACTGGTGATCGCATT
AnstX_ddP-P[6]	FAM-GATTCCGTCACCCTTGGCAT-BHQ1
RPS4-F[7]	GAGTCCATCAAAGGAGAAAGTCTAC
RPS4-R[7]	TAGCTGGCGCATCAGGTAC

AutoRef_ddP-F[8]	ATCACCACTCGTCGTCCGTT
AutoRef_ddP-R[8]	CGAACGAACTCGATTGACCC
AutoRef_ddP-P[8]	HEX-GCAAACACCACAACAGCAGC-BHQ1

Primers used in PCR applications as described previously. 1) Used to amplify the GUY1 full length cDNA sequence and adapt it with the AvrII and BbvCI sites for insertions into the transgenic plasmid backbone. 2) Used to amplify the GUY1 promoter and gene sequence then adapt it with AscI and PacI restriction enzyme sites. 3) Primers targeting DsRed for the detection of transgenic individuals. 4) RT-PCR primers designed to amplify native GUY1. 5) RT-PCR primers designed to amplify transgenic GUY1. 6) TaqMan primer and probe mixture designed to amplify an X chromosome sequence for ddPCR. 7) Primers designed to target ribosomal protein subunit 4 as a positive control. 8) Autosomal reference gene primer and probe mixture for ddPCR CNV.

Table 4.2 CNV of X chromosomes for Wild-type and transgenic mosquitoes

		CNV	PoissonCNVMax	PoissonCNVMin
Wild-type Adults				
	Male 1	1.03	1.07	0.993
	Male 2	0.992	1.02	0.96
	Female 1	2.04	2.13	1.95
	Female 2	1.99	2.05	1.93
Line B2 Adults				
	Male 1	0.849	0.875	0.824
	Male2	0.794	0.822	0.766
	Male 3	0.89	0.919	0.861
	Male 4	0.965	1.01	0.919
	Male 5	0.829	0.855	0.803
	Male 6	0.767	0.791	0.743
	Male 7	0.907	0.94	0.875
	Male 8	0.945	0.978	0.912
	Male 9	0.793	0.82	0.766
	Male 10	0.788	0.812	0.763
	Male 12	0.798	0.825	0.77
	Male 13	0.751	0.777	0.726

Copy number variance (CNV) values were determined for the X chromosomes in wild-type males and females (control) and line B2 adult males. Values for the wild-type are near one and two for males and females, respectively. The line B2 adult males all had CNV values near one, signifying these males only possess one X chromosomes.

Table 4.3 G0 adult male to female ratios

	Male	Female	M/F Ratio
Dblwhammy			
G0-set 1	79	47	1.68
G0-set 2	52	23	2.26
pEE2_GUY1			
G0- set 1 (full service 1)	39	24	1.63
G0- set 2 (full service 2)	49	15	3.27
pEE1_GUY1			
G0-set 1 (picked up)	10	6	1.67
G0- set 2 (picked up)	40	16	2.5
G0- set 3 (full service 1)	66	20	3.3
Control			
DO-04-12	43	50	0.86
DO-05-12	57	55	1.03
DO-07-12	57	50	1.14

P-values	
pEE1 vs DO	0.04
pEE2 vs DO	0.16
dblwhammy vs DO	0.08
dblwhammy+pEE1+pee2 vs DO	0.001

This table accounts for the sex of G0 adults after microinjections performed to obtain transgenic lines. The control experiments were performed with a plasmid that is not known to have an effect on the sex of the organism and the ratios suggest this to be the case by ranging from 0.86 to 1.14 males to females. In the experimental population the ratio of males to females is significantly higher than the control group.

Figure 4.S1 X chromosome sequence

```
>Xchromosome_probe_3
GACCCTCTCGAACATGGCTCACGGTGGTGGCACATGTCGGCAAACCCTCACTGGTGGTGC
CCACTGGTGAGCCGTGTCCTGAGAACGCCGATGTTGCTAGGGAAGCTGCACGCACACCAG
TGTGCCGAAGTGCAATACGGTGACACATTTGCTGTACGCATGCCAGACTAAGGCAGGAGG
GAGAAGTCAAGACATGTCCTTATGATTTTCTGCCGGATTTCCACTGGCAACGCTCGGGAC
CGTTGACCTGGCCGAGGTGTGGAAGGAGGTGAGCTGAATTGTGCTCTACCGCCGATTCCG
TCACCCTTGGCATTACGTACTTAAGGTCCGCTAGGTGTACTGCGATATCTCAGAACTCAA
AAAATGCGATCACCAGTTCACATCAAATCAAATCAACTGGATGGTTTTTGTTATGGTCAA
CTTGCTTGCCAACTGTATCATAGAAACGAGCTCCCGAGAAGGAAGCCTACCAATAGTCTA
CCTCTAGCGTGACACTAGTGACATCTGTAAGCCGCGCTGTTGTTTACGTTCGCACAATTT
TGATAGACTTTTGAGAGAGAGAGAGCGAAAGAGAGAGAGCAAGACAGCAATCGATCGAAT
AGCGTCTCGGGGGGAACCCTTTCCACATTTTCCACAACAATGCGTCCTCTCGTCCGTAGC
CATCTTGATGCCGGCACACAGAAATGTCACGTTCGCCCGGCTAGACTTGAAGTGGACTCA
ATTAAACTCTTCGCTAAACAATTGCCGTTTTCCATGCGTGTGGAAAATGTGCAGATTTTT
CCCCCGTTCGAATTTTCCGTTTTCCATCACATC
```

Primers were designed using the above sequence which has been confirmed to reside on the X chromosome via FISH (Sharakhov Lab). This sequence resides in scaffold 40 of the November 2012 *Anopheles stephensi* assembly.

Figure 4.S2 Sexing of larva

The images of 3rd instar larvae with bright field microscopy depict males (top) and females (bottom) based on the presence of absence of testes (red arrows).

Figure 4.S3 Sexing of line B2 embryos

gDNA Sample #	RPS4	GUY1Native	DsRed	Sex
1				NA
2			X	NA
3			X	NA
4			X	NA
5	X			Female-Non Trans
6	X		X	Female-Trans
7	X		X	Female-Trans
8		X	X	NA
9	X	X	X	Male-Trans
10	X	X		Male-Non Trans
11	X			Female-Non Trans
12				NA
13	X			Female-Non Trans
14	X	X	X	Male-Trans
15	X	X	X	Male-Trans
16	X	X	X	Male-Trans
17	X			Female-Non Trans
18	X			Female-Non Trans
19				NA
20	X		X	Female-Trans
21	X		X	Female-Trans
22			X	NA
23				NA
24	X		X	Female-Trans
25	X			Female-Non Trans
26			X	NA
27	X	X	X	Male-Trans
28				NA
29			X	NA
30	X			Female-Non Trans
31	X			Female-Non Trans
32	X			Female-Non Trans
33	X		X	Female-Trans
34	X	X	X	Male-Trans
35				NA
36				NA
37			X	NA
38	X	X	X	Male-Trans
39	X			Female-Non Trans
40	X	X		Male-Non Trans
41			X	NA
42				NA

43	X	X		Male-Non Trans
44	X			Female-Non Trans
45	X			Female-Non Trans
46	X		X	Female-Trans
47	X		X	Female-Trans
48		X		NA
49	X	X	X	Male-Trans
50	X		X	Female-Trans
51	X		X	Female-Trans
52	X			Female-Non Trans
53	X		X	Female-Trans
54	X		X	Female-Trans
55	X	X	X	Male-Trans
56		X	X	NA
57			X	NA
58	X	X	X	Male-Trans
59	X		X	Female-Trans
60	X		X	Female-Trans
61	X			Female-Non Trans
62			X	NA
63	X		X	Female-Trans
64	X			Female-Non Trans
65	X	X	X	Male-Trans
66	X	X		Male-Non Trans
67	X		X	Female-Trans
68	X		X	Female-Trans
69	X	X	X	Male-Trans
70				NA
71	X		X	Female-Trans
72	X		X	Female-Trans
73				NA
74			X	NA
75				NA
76	X		X	Female-Trans
77	X	X		Male-Non Trans
78	X	X	X	Male-Trans
79	X		X	Female-Trans
80	X		X	Female-Trans
81				NA
82	X			Female-Non Trans
83	X		X	Female-Trans
84	X		X	Female-Trans
85			X	NA

86	X		Female-Non Trans
87	X	X	Female-Trans
88	X	X	Female-Trans
89	X	X	Female-Trans
90		X	NA
91	X	X	Female-Trans
92	X		Female-Non Trans
93	X	X	Female-Trans
94	X	X	Female-Trans
95	X	X	Female-Trans
96	X	X	Female-Trans

Sexing of Line B2 embryos was performed by the PCR targeting RPS4 (control), GUY1 native (Y chromosome marker), and dsRed (transgenic marker). Samples that did not amplify RPS4 were not counted (NA), leaving 68 samples. These samples broke up into 5 non transgenic males, 18 non transgenic females, 13 transgenic males and 32 transgenic females.

Figure 4.S4 Sex distributions for transgenic generations

Line B1

Generation	Male	Female
5	32	8
6	15	0
7	57	2
8	36	0
9	95	1
10	65	0
11	14	0
12	15	6
13	31	0

Line B2

Generation	Male	Female
6	35	0
7	203	6
8	61	11
9	45	0
10	53	0
11	25	0
12	23	0
13	36	0

Line Af

Generation	Male	Female
3	9	11
4	20	42
5	19	12
6	63	72
7	21	54
8	8	12
9	14	12
10	20	54
11	12	17
12	37	31
13	18	25

Line A

Generation	Male	Female
1	86	17
2	105	5
3	124	0
4	119	18

Line B

Generation	Male	Female
1	14	0
2	51	0
3	115	0
4	69	2
5	54	0
6	44	2

Line C

Generation	Male	Female
2	48	0
3	117	0
4	78	5
5	52	0
6	10	2
7	109	3

7	87	5	**8**	60	0
8	43	0			

Non-transgenics

	Generation	Male	Female
Line B1	5	12	7
Line B2	7	103	90
Line Af	3	37	59
Line A	2	35	41
Line B	2	34	20
Line C	3	72	85

Larval stage sex distribution

		Male	Female
Line B1			
	Transgenic	27	0
	Non-transgenic	48	33
Line B2			
	Transgenic	39	2
	Non-transgenic	29	43
Line Af			
	Transgenic	49	61
	Non-transgenic	13	8

This table lists the lines of which were started from individual males (in the case of lines B1 and B2) or a cohort of females (line Af) from the pools A, B or C. Lines B1 and B2 are from individual males derived from line B. Line Af and C were derived from line A, the transgenic line established at G1. Line Af resulted from line A females crossed with wild type males. Line C arose from line A male with faint DsRed expression. The generation number corresponds to the generations of the first pool of transgenic G1 progeny and begins with the generation at which each line was established (for example, line B1 was established during the 5^{th} generation of the cohort from a 4^{th} generation male). Males and females were counted during the adult stage. Larvae of lines B1, B2 and Af were sexed according to the presence of testes (Figure 4.S2) during the G10 to determine if there were female transgenic larvae. In lines B1 there were no

female transgenics and in line B2 there were two. Comparison of these numbers against the 61 transgenic females in line Af suggests female death before the L3 stage.

Figure 4.S5 Competitiveness assay

		Male	Female	Expected %	Observed %
Line B1					
	Transgenic	1	0		
	Non-transgenic	16	15		
				14.3%	**3.13%**
Line B2					
	Transgenic	5	0		
	Non-transgenic	26	11		
				14.3%	**11.90%**
Line C1					
	Transgenic	4	0		
	Non-transgenic	20	17		
				14.3%	**9.76%**

The progeny from a mating of 10 transgenic males, 10 wild type males and 20 wild type females. The expected values listed assume all lines are heterozygous for the transgene and performing female lethality. Lines B2 and C1 are shown to be nearly as competitive as wild type males whereas line B1 performs 4 fold less than the expected.

Figure 4.S6- Sequence of transformation plasmids

>dblwhammy_GUY1

CCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTT
CCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGG
TTATCCACAGAATCAGGGGATAACGCACGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAA
GGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGT
GGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACC
CTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA
TCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCT
TATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGG
ATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAAC
AGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAA
CCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCT
TTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAA
AAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGT
CTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGA
CTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCC
ACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTT

TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAAC
GTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACG
ATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGA
AGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAG
ATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC
CGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGG
CGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC
ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGA
CACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGC
GGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
CGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTT
TCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGG
AGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCA
GATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCC
ATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAA
GGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAG
TGCCAAGCTTTGTTTAAAAATATAACAAAATTGTGATCCCACAAAATGAAGTGGGGCAAAATCAAATAATTAATAGT
GTCCGTAAACTTGTTGGTCTTCAACTTTTTGAGGAACACGTTGGACGGCAAATCCGTGACTATAACACAAGTTGATT
TAATAATTTTAGCCAACACGTCGGGCTGCGTGTTTTTTGCCGACGCGTCTGTGTACACGTTGATTAACTGGTCGATT
AAACTGTTGAAATAATTTAATTTTTGGTTCTTCTTTAAATCTGTGATGAAATTTTTTAAAATAACTTTAAATTCTTC
ATTGGTAAAAAATGCCACGTTTTGCAACTTGTGAGGGTCTAATATGAGGTCAAACTCAGTAGGAGTTTTATCCAAAA
AAGAAAACATGATTACGTCTGTACACGAACGCGTATTAACGCAGAGTGCAAAGTATAAGAGGGTTAAAAAATATATT
TTACGCACCATATACGCATCGGGTTGATATCGTTAATATGGATCAATTTGAACAGTTGATTAACGTGTCTCTGCTCA
AGTCTTTGATCAAAACGCAAATCGACGAAAATGTGTCGGACAATATCAAGTCGATGAGCGAAAAACTAAAAAGGCTA
GAATACGACAATCTCACAGACAGCGTTGAGATATACGGTATTCACGACAGCAGGCTGAATAATAAAAAAATTAGAAA
CTATTATTTAACCCTAGAAAGATAATCATATTGTGACGTACGTTAAAGATAATCATGCGTAAAATTGACGCATGTGT
TTTATCGGTCTGTATATCGAGGTTTATTTATTAATTTGAATAGATATTAAGTTTTATTATATTTACACTTACATACT
AATAATAAATTCAACAAACAATTTATTTATGTTTATTTATTTATTAAAAAAAAACAAAAACTCAAAATTTCTTCTAT
AAAGTAACAAAACTTTTAAACATTCTCTCTTTTACAAAAATAAACTTATTTTGTACTTTAAAAACAGTCATGTTGTA
TTATAAAATAAGTAATTAGCTTAACTTATACATAATAGAAACAAATTATACTTATTAGTCAGTCAGAAACAACTTTG
GCACATATCAATATTATGCTCTCGACAAATAACTTTTTTGCATTTTTTGCACGATGCATTTGCCTTTCGCCTTATTT
TAGAGGGGCAGTAAGTACAGTAAGTACGTTTTTTCATTACTGGCTCTTCAGTACTGTCATCTGATGTACCAGGCACT
TCATTTGGCAAAATATTAGAGATATTATCGCGCAAATATCTCTTCAAAGTAGGAGCTTCTAAACGCTTACGCATAAA
CGATGACGTCAGGCTCATGTAAAGGTTTCTCATAAATTTTTTGCGACTTTGAACCTTTTCTCCCTTGCTACTGACAT
TATGGCTGTATATAATAAAAGAATTTATGCAGGCAATGTTTATCATTCCGTACAATAATGCCATAGGCCACCTATTC
GTCTTCCTACTGCAGGTCATCACAGAACACATTTGGTCTAGCGTGTCCACTCCGCCTTTAGTTTGATTATAATACAT
AACCATTTGCGGTTTACCGGTACTTTCGTTGATAGAAGCATCCTCATCACAAGATGATAATAAGTATACCATCTTAG
CTGGCTTCGGTTTATATGAGACGAGAGTAAGGGGTCCGTCAAAACAAAACATCGATGTTCCCACTGGCCTGGAGCGA
CTGTTTTTCAGTACTTCCGGTATCTCGCGTTTGTTTGATCGCACGGTTCCCACAATGGTTAATTCGAGCTCGCCCGG
GGATCTAATTCAATTAGAGACTAATTCAATTAGAGCTAATTCAATTAGGATCCAAGCTTATCGATTTCGAACCCTCG
ACCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCT
AAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCGGGGTACCGCTAGAGTCGACGGTACCG
CGGGCCCGGGATCCACCGGTCGCCACCATGGTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGT
GCGCATGGAGGGCACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACA
ACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCCAGTAC
GGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTG
GGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCA
TCTACAAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAG
GCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACAAGGCCCTGAAGCTGAAGGACGG
CGGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGG
ACTCCAAGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGCCAC
CACCTGTTCCTGTAGCGGCCGCGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAA
AAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTT
ATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGT
TTGTCCAAACTCATCAATGTATCTTAAAGCTTATCGAATACGCGTACGGCGCGCCGCAATAGAGTTAAATAATGGAA
CTTTACCTTGGCCGTGTGTTCTGTTTTTCTTCAGCACCAAATCTTAGTTCTGCAAAACCACTGCTGTGTGAAAAAAG
ACAAGGATGACAGCTCAACTCAACTAGTAGCGGGTGTAAACGCAGAGGACAGCGGGCACTGACGGGAAAACGATAAG

AAAGCGAATGAGGATTTAAATTTCGCAAGCAAGTTTTAGCAATCAAGCAAGAAATACCGTTAATAAAAATTTCAGTT
AAAATTTTAAATTAATTCCTTCACATGAAGCAATTATTGCACTCCAAAGTCGCTTTGATTTGTAACGCCTACCTGTT
TACCTATATTTCGTTCGTAAAGGAAATTAAGGCCAAAGGGGTAAGAAAATTTTTCATCCATGTTCTGGACCTCCAAA
AAACCGTTAAAATTCATAAAACGATTGTTTCTTTACCCTGGATTGACCTAATAACGAACCGTGTTTTATCTGTTCTC
CCTTTTCGTAACTGGAACGTGAAGGCTTTACCCCAACATTAAATGCCTACGAACGATCCTGTTGGAATTCCAGGTAG
GGAATGCGCAGGATGCGATTCCAGGAACCAGAGCACGCGTATCCACCGGGGCAGATCATCAACAGATGACCGTTAAT
ACCTGGTTCTGCGGTATTCCATTTAACTACCCTAATACCAGCACTAATGGAAAAAAGACACAGCCTTAGCATAATTC
GTGGTGTGTGCAGTTTTTTCTATTATTTGCCGCTGCCACCATGTCACAAACACAGGTAAGCGCTGTATGGATGCTAT
TTACTTGGTCAGATGTTATCTGATGTGCGTAGTAGTACCCGTGTGGGCGCGACGGCGATCGTCGTGTGCGTCGTGGG
TAAAGTTCCTTGCTTGTAGCGATTAGTATATAAATGTGGTCCAGGTCGCCCATTTGCGTTAATTACTGTTCGTCCTT
CTTTCTGCGTAGACCGGATTGTGCGAGTTGCAGAACCGTGTGTGTCCATTGTGTGTGCGTGTTTTTTGAGAAAATAG
CTGAATTGAAGGAAAAGCCTAGGATGAATTCACAAAGTAGGCGATACAAAAACATTGAATTAGTGAATAATCTGAAA
GCTTATCTGACTTGGAATGATAAATCAAGTTTTCAGGTTAAACACTCTGCTGTGACATTGGAAAAAAAGAAAAGTAA
AACAAAAATATGTAACGTACTCTATGAAGCAATCACACTCGAGCTACTCTGCGAGCTTTAAAGGTAAGAAGATCGTT
CGATTAGGCGCTAATTGGGATTTGTGGGAGCAAGAGTGTTGATCTGAAGTGCGAACCGTATGCTGGATGGAAGTTAT
TTGATCGCCATGTCTAAGGAGTCAATATGAACAATGAAACTCACGAGATTAGACTATACATTTCATTGTAATAATTG
TATAGTCTTCTCGTGAGTTTGGGGTTCATAATTTCTACTGAGTCTTTGTGCCTATTGGATTACGCAAGAATGGCGAT
TGTTGTATCGCCTACTTTGTGAACACGATATGGAATACGGTTTGTTCACAAAAGGCGATACAACAATTACATTTGTA
ACTACAGTAACCTGAAACAAGTGATACCTATTGTTATAATACGTGCAGCTAAGAACTGAAAAAATGCCAGTTTTGTT
CAACAACGATACAACGTAATCGACATTTTGTTAAATATGTGATATGAATATTAAGTGATTTCAGTTATGTAAATTAT
AATTGTGATCCAAATACGAGCATTATGAGTGGTTTCTCTATATGTTACAATGGATTAAGTTGTACCACTTTTGTATC
GCCTACTTTGTGAACATGATTTTTTGAACATAACTGTCACAAAGGGCGATACAAATGTTGGCAATATAGTCATCGAG
TTAAACCCATAAATTTGTAAGAAATTTAAATGCATAGCTTTTCTATGAACGTTACATTAATACATTAATTTTTTTTT
CATTCTAGATGGCGCCACACAATTCATTGCGATCGCTGTGATTGCTTCATAGAGTACGTTACATATTTTTGTTTTAC
TTTTCTTTTTTTCCAATGTCACAGCAGAGTGTTTAACCTGAAAACTTGATTTATCATTCCAAGTCAGATAAGCTTTC
AGATTATTCACTAATTCAATGTTTTTGTATCGCCTACTTTGTGAATTCATCCTCAGCGCCGCACCGGCCAGACATGA
TAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGAT
GCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCA
GGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAGCCGGCTTAATTCTTAAG
TTAATTTTAATTAAGTCGACGATGTAGGTCACGGTCTCGAAGCCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTC
CCCGGGCGCGTACTCCACCTCACCCATCTGGTCCATCATGATGAACGGGTCGAGGTGGCGGTAGTTGATCCCGGCGA
ACGCGCGGCGCACCGGGAAGCCCTCGCCCTCGAAACCGCTGGGCGCGGTGGTCACGGTGAGCACGGGACGTGCGACG
GCGTCGGCGGGTGCGGATACGCGGGGCAGCGTCAGCGGGTTCTCGACGGTCACGGCGGGCATGTCGACAGATCTGAC
AATGTTCAGTGCAGAGACTCGGCTACGCCTCGTGGACTTTGAAGTTGACCAACAATGTTTATTCTTACCTCTAATAG
TCCTCTGTGGCAAGGTCAAGATTCTGTTAGAAGCCAATGAAGAACCTGGTTGTTCAATAACATTTTGTTCGTCTAAT
ATTTCACTACCGCTTGACGTTGGCTGCACTTCATGTACCTCATCTATAAACGCTTCTTCTGTATCGCTCTGGACGTC
ATCTTCACTTACGTGATCTGATATTTCACTGTCAGAATCCTCACCAACAAGCTCGTCATCGCTTTGCAGAAGAGCAG
AGAGGATATGCTCATCGTCTAAAGAACTACCCATTTTATTATATATTAGTCACGATATCTATAACAAGAAAATATAT
ATATAATAAGTTATCACGTAAGTAGAACATGAAATAACAATATAATTATCGTATGAGTTAAATCTTAAAAGTCACGT
AAAAGATAATCATGCGTCATTTTGACTCACGCGGTCGTTATAGTTCAAAATCAGTGACACTTACCGCATTGACAAGC
ACGCCTCACGGGAGCTCCAAGCGGCGACTGAGATGTCCTAAATGCACAGCGACGGATTCGCGCTATTTAGAAAGAGA
GAGCAATATTTCAAGAATGCATGCGTCAATTTTACGCAGACTATCTTTCTAGGGTTAAAAAAGATTTGCGCTTTACT
CGACCTAAACTTTAAACACGTCATAGAATCTTCGTTTGACAAAAACCACATTGTGGCCAAGCTGTGTGACGCGACGC
GCGCTAAAGAATGGCAAACCAAGTCGCGCGAGCGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCGTAAT
CATGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGT
GTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGA
AA

>pEE1_GUY1
CCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTT
CCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGG
TTATCCACAGAATCAGGGGATAACGCACGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAA
GGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGT
GGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACC
CTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA
TCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCT
TATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGG

ATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAAC
AGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAA
CCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCT
TTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAA
AAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGT
CTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGA
CTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCC
ACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTT
TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAAC
GTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACG
ATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGA
AGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAG
ATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC
CGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGG
CGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC
ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGA
CACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGC
GGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
CGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTT
TCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGG
AGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCA
GATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCC
ATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAA
GGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAG
TGCCAAGCTTTGTTTAAAAATATAACAAAATTGTGATCCCACAAAATGAAGTGGGGCAAAATCAAATAATTAATAGT
GTCCGTAAACTTGTTGGTCTTCAACTTTTTGAGGAACACGTTGGACGGCAAATCCGTGACTATAACACAAGTTGATT
TAATAATTTTAGCCAACACGTCGGGCTGCGTGTTTTTTGCCGACGCGTCTGTGTACACGTTGATTAACTGGTCGATT
AAACTGTTGAAATAATTTAATTTTTGGTTCTTCTTTAAATCTGTGATGAAATTTTTTAAAATAACTTTAAATTCTTC
ATTGGTAAAAAATGCCACGTTTTGCAACTTGTGAGGGTCTAATATGAGGTCAAACTCAGTAGGAGTTTTATCCAAAA
AAGAAAACATGATTACGTCTGTACACGAACGCGTATTAACGCAGAGTGCAAAGTATAAGAGGGTTAAAAAATATATT
TTACGCACCATATACGCATCGGGTTGATATCGTTAATATGGATCAATTTGAACAGTTGATTAACGTGTCTCTGCTCA
AGTCTTTGATCAAAACGCAAATCGACGAAAATGTGTCGGACAATATCAAGTCGATGAGCGAAAAACTAAAAAGGCTA
GAATACGACAATCTCACAGACAGCGTTGAGATATACGGTATTCACGACAGCAGGCTGAATAATAAAAAAATTAGAAA
CTATTATTTAACCCTAGAAAGATAATCATATTGTGACGTACGTTAAAGATAATCATGCGTAAAATTGACGCATGTGT
TTTATCGGTCTGTATATCGAGGTTTATTTATTAATTTGAATAGATATTAAGTTTTATTATATTTACACTTACATACT
AATAATAAATTCAACAAACAATTTATTTATGTTTATTTATTTATTAAAAAAAAACAAAAACTCAAAATTTCTTCTAT
AAAGTAACAAAACTTTTAAACATTCTCTCTTTTACAAAAATAAACTTATTTTGTACTTTAAAAACAGTCATGTTGTA
TTATAAAATAAGTAATTAGCTTAACTTATACATAATAGAAACAAATTATACTTATTAGTCAGTCAGAAACAACTTTG
GCACATATCAATATTATGCTCTCGACAAATAACTTTTTTGCATTTTTTGCACGATGCATTTGCCTTTCGCCTTATTT
TAGAGGGGCAGTAAGTACAGTAAGTACGTTTTTTCATTACTGGCTCTTCAGTACTGTCATCTGATGTACCAGGCACT
TCATTTGGCAAAATATTAGAGATATTATCGCGCAAATATCTCTTCAAAGTAGGAGCTTCTAAACGCTTACGCATAAA
CGATGACGTCAGGCTCATGTAAAGGTTTCTCATAAATTTTTTGCGACTTTGAACCTTTTCTCCCTTGCTACTGACAT
TATGGCTGTATATAATAAAAGAATTTATGCAGGCAATGTTTATCATTCCGTACAATAATGCCATAGGCCACCTATTC
GTCTTCCTACTGCAGGTCATCACAGAACACATTTGGTCTAGCGTGTCCACTCCGCCTTTAGTTTGATTATAATACAT
AACCATTTGCGGTTTACCGGTACTTTCGTTGATAGAAGCATCCTCATCACAAGATGATAATAAGTATACCATCTTAG
CTGGCTTCGGTTTATATGAGACGAGAGTAAGGGGTCCGTCAAAACAAAACATCGATGTTCCCACTGGCCTGGAGCGA
CTGTTTTTCAGTACTTCCGGTATCTCGCGTTTGTTTGATCGCACGGTTCCCACAATGGTTAATTCGAGCTCGCCCGG
GGATCTAATTCAATTAGAGACTAATTCAATTAGAGCTAATTCAATTAGGATCCAAGCTTATCGATTTCGAACCCTCG
ACCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCT
AAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCGGGGTACCGCTAGAGTCGACGGTACCG
CGGGCCCGGGATCCACCGGTCGCCACCATGGTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGT
GCGCATGGAGGGCACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACA
ACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCCAGTAC
GGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTG
GGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCA
TCTACAAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAG
GCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACAAGGCCCTGAAGCTGAAGGACGG

CGGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGG
ACTCCAAGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGCCAC
CACCTGTTCCTGTAGCGGCCGCGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAA
AAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTT
ATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGT
TTGTCCAAACTCATCAATGTATCTTAAAGCTTATCGAATACGCGTACGGCGCGCCGCAATAGAGTTAAATAATGGAA
CTTTACCTTGGCCGTGTGTTCTGTTTTTCTTCAGCACCAAATCTTAGTTCTGCAAAACCACTGCTGTGTGAAAAAAG
ACAAGGATGACAGCTCAACTCAACTAGTAGCGGGTGTAAACGCAGAGGACAGCGGGCACTGACGGGAAAACGATAAG
AAAGCGAATGAGGATTTAAATTTCGCAAGCAAGTTTTAGCAATCAAGCAAGAAATACCGTTAATAAAAATTTCAGTT
AAAATTTTAAATTAATTCCTTCACATGAAGCAATTATTGCACTCCAAAGTCGCTTTGATTTGTAACGCCTACCTGTT
TACCTATATTTCGTTCGTAAAGGAAATTAAGGCCAAAGGGGTAAGAAAATTTTTCATCCATGTTCTGGACCTCCAAA
AAACCGTTAAAATTCATAAAACGATTGTTTCTTTACCCTGGATTGACCTAATAACGAACCGTGTTTTATCTGTTCTC
CCTTTTCGTAACTGGAACGTGAAGGCTTTACCCCAACATTAAATGCCTACGAACGATCCTGTTGGAATTCCAGGTAG
GGAATGCGCAGGATGCGATTCCAGGAACCAGAGCACGCGTATCCACCGGGGCAGATCATCAACAGATGACCGTTAAT
ACCTGGTTCTGCGGTATTCCATTTAACTACCCTAATACCAGCACTAATGGAAAAAAGACACAGCCTTAGCATAATTC
GTGGTGTGTGCAGTTTTTTCTATTATTTGCCGCTGCCACCATGTCACAAACACAGGTAAGCGCTGTATGGATGCTAT
TTACTTGGTCAGATGTTATCTGATGTGCGTAGTAGTACCCGTGTGGGCGCGACGGCGATCGTCGTGTGCGTCGTGGG
TAAAGTTCCTTGCTTGTAGCGATTAGTATATAAATGTGGTCCAGGTCGCCCATTTGCGTTAATTACTGTTCGTCCTT
CTTTCTGCGTAGACCGGATTGTGCGAGTTGCAGAACCGTGTGTGTCCATTGTGTGTGCGTGTTTTTTGAGAAAATAG
CTGAATTGAAGGAAAAGCCTAGGCCGGTATAGTCTAATCTCGTGAGTTTTACTTCTAAAGGTTCTGTGGATAACATT
CACAAAGAGTTCATCATTGGAAACTTATTCTGGGAAATTTACTCGTCAAAGCTGCCACGGATCTCGCTCAAGAATTT
AGTTTTAAGTTTTAGATAACAGAAAGCGTACACTGTAAAATAGGCAAATTATGAATTCACAAAGTAGGCGATACAAA
AACATTGAATTAGTGAATAATCTGAAAGCTTATCTGACTTGGAATGATAAATCAAGTTTTCAGGTTAAACACTCTGC
TGTGACATTGGAAAAAAAGAAAAGTAAAACAAAAATATGTAACGTACTCTATGAAGCAATCACATAAATAGTTCACA
ACGAAAAAAAGGCACATTTAGCATATCTGAAAATAAAACGTGACAGACAATAGCATAAGACCACTTTGTGTGTTAAT
ATAGTAAAATAGAGGCTATGGTGTCAATTTTTCCTCAGCGCCGCACCGGCCAGACATGATAAGATACATTGATGAGT
TTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTA
ACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTG
GGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAGCCGGCTTAATTCTTAAGTTAATTTTAATTAAGTCG
ACGATGTAGGTCACGGTCTCGAAGCCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCAC
CTCACCCATCTGGTCCATCATGATGAACGGGTCGAGGTGGCGGTAGTTGATCCCGGCGAACGCGCGGCGCACCGGGA
AGCCCTCGCCCTCGAAACCGCTGGGCGCGGTGGTCACGGTGAGCACGGGACGTGCGACGGCGTCGGCGGGTGCGGAT
ACGCGGGGCAGCGTCAGCGGGTTCTCGACGGTCACGGCGGGCATGTCGACAGATCTGACAATGTTCAGTGCAGAGAC
TCGGCTACGCCTCGTGGACTTTGAAGTTGACCAACAATGTTTATTCTTACCTCTAATAGTCCTCTGTGGCAAGGTCA
AGATTCTGTTAGAAGCCAATGAAGAACCTGGTTGTTCAATAACATTTTGTTCGTCTAATATTTCACTACCGCTTGAC
GTTGGCTGCACTTCATGTACCTCATCTATAAACGCTTCTTCTGTATCGCTCTGGACGTCATCTTCACTTACGTGATC
TGATATTTCACTGTCAGAATCCTCACCAACAAGCTCGTCATCGCTTTGCAGAAGAGCAGAGAGGATATGCTCATCGT
CTAAAGAACTACCCATTTTATTATATATTAGTCACGATATCTATAACAAGAAAATATATATATAATAAGTTATCACG
TAAGTAGAACATGAAATAACAATATAATTATCGTATGAGTTAAATCTTAAAAGTCACGTAAAAGATAATCATGCGTC
ATTTTGACTCACGCGGTCGTTATAGTTCAAAATCAGTGACACTTACCGCATTGACAAGCACGCCTCACGGGAGCTCC
AAGCGGCGACTGAGATGTCCTAAATGCACAGCGACGGATTCGCGCTATTTAGAAAGAGAGAGCAATATTTCAAGAAT
GCATGCGTCAATTTTACGCAGACTATCTTTCTAGGGTTAAAAAAGATTTGCGCTTTACTCGACCTAAACTTTAAACA
CGTCATAGAATCTTCGTTTGACAAAAACCACATTGTGGCCAAGCTGTGTGACGCGACGCGCGCTAAAGAATGGCAAA
CCAAGTCGCGCGAGCGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCGTAATCATGTCATAGCTGTTTCC
TGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCT
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAA

>pEE2_GUY1 *Note: The PacI site has been abolished in this construct
CCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTT
CCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGG
TTATCCACAGAATCAGGGGATAACGCACGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAA
GGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGT
GGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACC
CTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA
TCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCT
TATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGG
ATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAAC

AGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAA
CCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCT
TTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAA
AAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGT
CTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGA
CTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCC
ACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTT
TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAAC
GTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACG
ATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGA
AGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAG
ATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC
CGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGG
CGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC
ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGA
CACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGC
GGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
CGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTT
TCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGG
AGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCA
GATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCC
ATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAA
GGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAG
TGCCAAGCTTTGTTTAAAAATATAACAAAATTGTGATCCCACAAAATGAAGTGGGGCAAAATCAAATAATTAATAGT
GTCCGTAAACTTGTTGGTCTTCAACTTTTTGAGGAACACGTTGGACGGCAAATCCGTGACTATAACACAAGTTGATT
TAATAATTTTAGCCAACACGTCGGGCTGCGTGTTTTTTGCCGACGCGTCTGTGTACACGTTGATTAACTGGTCGATT
AAACTGTTGAAATAATTTAATTTTTGGTTCTTCTTTAAATCTGTGATGAAATTTTTTAAAATAACTTTAAATTCTTC
ATTGGTAAAAAATGCCACGTTTTGCAACTTGTGAGGGTCTAATATGAGGTCAAACTCAGTAGGAGTTTTATCCAAAA
AAGAAAACATGATTACGTCTGTACACGAACGCGTATTAACGCAGAGTGCAAAGTATAAGAGGGTTAAAAAATATATT
TTACGCACCATATACGCATCGGGTTGATATCGTTAATATGGATCAATTTGAACAGTTGATTAACGTGTCTCTGCTCA
AGTCTTTGATCAAAACGCAAATCGACGAAAATGTGTCGGACAATATCAAGTCGATGAGCGAAAAACTAAAAAGGCTA
GAATACGACAATCTCACAGACAGCGTTGAGATATACGGTATTCACGACAGCAGGCTGAATAATAAAAAAATTAGAAA
CTATTATTTAACCCTAGAAAGATAATCATATTGTGACGTACGTTAAAGATAATCATGCGTAAAATTGACGCATGTGT
TTTATCGGTCTGTATATCGAGGTTTATTTATTAATTTGAATAGATATTAAGTTTTATTATATTTACACTTACATACT
AATAATAAATTCAACAAACAATTTATTTATGTTTATTTATTTATTAAAAAAAAACAAAAACTCAAAATTTCTTCTAT
AAAGTAACAAAACTTTTAAACATTCTCTCTTTTACAAAAATAAACTTATTTTGTACTTTAAAAACAGTCATGTTGTA
TTATAAAATAAGTAATTAGCTTAACTTATACATAATAGAAACAAATTATACTTATTAGTCAGTCAGAAACAACTTTG
GCACATATCAATATTATGCTCTCGACAAATAACTTTTTTGCATTTTTTGCACGATGCATTTGCCTTTCGCCTTATTT
TAGAGGGGCAGTAAGTACAGTAAGTACGTTTTTTCATTACTGGCTCTTCAGTACTGTCATCTGATGTACCAGGCACT
TCATTTGGCAAAATATTAGAGATATTATCGCGCAAATATCTCTTCAAAGTAGGAGCTTCTAAACGCTTACGCATAAA
CGATGACGTCAGGCTCATGTAAAGGTTTCTCATAAATTTTTTGCGACTTTGAACCTTTTCTCCCTTGCTACTGACAT
TATGGCTGTATATAATAAAAGAATTTATGCAGGCAATGTTTATCATTCCGTACAATAATGCCATAGGCCACCTATTC
GTCTTCCTACTGCAGGTCATCACAGAACACATTTGGTCTAGCGTGTCCACTCCGCCTTTAGTTTGATTATAATACAT
AACCATTTGCGGTTTACCGGTACTTTCGTTGATAGAAGCATCCTCATCACAAGATGATAATAAGTATACCATCTTAG
CTGGCTTCGGTTTATATGAGACGAGAGTAAGGGGTCCGTCAAAACAAAACATCGATGTTCCCACTGGCCTGGAGCGA
CTGTTTTTCAGTACTTCCGGTATCTCGCGTTTGTTTGATCGCACGGTTCCCACAATGGTTAATTCGAGCTCGCCCGG
GGATCTAATTCAATTAGAGACTAATTCAATTAGAGCTAATTCAATTAGGATCCAAGCTTATCGATTTCGAACCCTCG
ACCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCT
AAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCGGGGTACCGCTAGAGTCGACGGTACCG
CGGGCCCGGGATCCACCGGTCGCCACCATGGTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGT
GCGCATGGAGGGCACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACA
ACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCCAGTAC
GGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTG
GGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCA
TCTACAAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAG
GCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACAAGGCCCTGAAGCTGAAGGACGG
CGGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGG

ACTCCAAGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGCCAC
CACCTGTTCCTGTAGCGGCCGCGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAA
AAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTT
ATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGT
TTGTCCAAACTCATCAATGTATCTTAAAGCTTATCGAATACGCGTACGGCGCGCCCCTGAAATGATGCTCTGGAAAA
ACCAAAACAGGCTGTTGGCTTGCTTTAGGTTGGGTTTAAGTACGTTGGGTGTAAGCAGTGCGCACGAGACGCGGTTT
TTCAATTGCGGTAGTCAGAAAGAGAAAAAAAATATCCTTGAAAAAACACACGGTAAACGAATCGCAAGGGGAGGTGT
AGCACGAGTATGCTTACAAAGCGTCTGGACGGGTAATCAATTTTGGCATATGCACAAACACTTCCATGCCTGCGGAG
CCTAACGCGTGAGTGCGAGGGAGCGCGATATACATTTTGCCATCGAAGTACATAAGGTCATCCTCCAAAATGTTTAC
TTGGGAAGAAGTTTCGTATCCATACCTTACCAGTGAAGTGTTGCTACCTCACCGGCTTACAGACCGAACCAGATAAA
CAAGATGTACAATTCTGGAAACAAAACAATGATCTACAAAGAAGCAACACACCAAGGAAGAGAAACTTTGTTCAAAA
ACCAAGATCGAATCCTTATGGTGTAGTATTCTAACTATTAGAGGAATTTCATATAAAAATATAACTGCATATGCTTC
AGTGCAGCAGAAATAAATGTTGGCATTACTATTATAAAATATGACTTTTATCTCGTACCAATCATCCGAATCACACA
ACGAGGTTCGACTCACCGACGCGAAGAAACAAACTTTTTTTAATCTAAAACAGTACCTAGAATCGAGGCGTAAATTA
GGTAGGTAGCAAAACGAAATGCCATAAATGCAATTATCGTAAAATATGAAAACAAGTAATAATGCGGTACAGGCCTT
CGTTGGAAAATTAGCTTTTTGACAACAAACCTACAACGGAAAATATGTGGGGCTAGCCTCGAGACCGTAAACGTACC
CGAATAAGAATAAATGAGATCAAGGTCATGCTATTTCACATGAGTTTACCACATACATTTAGTATATAAGACCAAAA
TTTTGTGCATAAACTTCAGTTGCATTTTGAACCTTAACCGGTATAGTCTAATCTCGTGAGTTTTACTTCTAAAGGTT
CTGTGGATAACATTCACAAAGAGTTCATCATTGGAAACTTATTCTGGGAAATTTACTCGTCAAAGCTGCCACGGATC
TCGCTCAAGAATTTAGTTTTAAGTTTTAGATAACAGAAAGCGTACACTGTAAAATAGGCAAATTATGAATTCACAAA
GTAGGCGATACAAAAACATTGAATTAGTGAATAATCTGAAAGCTTATCTGACTTGGAATGATAAATCAAGTTTTCAG
GTTAAACACTCTGCTGTGACATTGGAAAAAAAGAAAAGTAAAACAAAAATATGTAACGTACTCTATGAAGCAATCAC
ATAAATAGTTCACAACGAAAAAAAGGCACATTTAGCATATCTGAAAATAAAACGTGACAGACAATAGCATAAGACCA
CTTTGTGTGTTAATATAGTAAAATAGAGGCTATGGTGTCAATTTTTAACGGATCGACTTAGTTTTTTTTTTGTAAAA
TTTAGTTTCACATGTTGGAAAAACGTTTCAATTAATTGCATGATTAAGTCGACGATGTAGGTCACGGTCTCGAAGCC
GCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACCTCACCCATCTGGTCCATCATGATGA
ACGGGTCGAGGTGGCGGTAGTTGATCCCGGCGAACGCGCGGCGCACCGGGAAGCCCTCGCCCTCGAAACCGCTGGGC
GCGGTGGTCACGGTGAGCACGGGACGTGCGACGGCGTCGGCGGGTGCGGATACGCGGGGCAGCGTCAGCGGGTTCTC
GACGGTCACGGCGGGCATGTCGACAGATCTGACAATGTTCAGTGCAGAGACTCGGCTACGCCTCGTGGACTTTGAAG
TTGACCAACAATGTTTATTCTTACCTCTAATAGTCCTCTGTGGCAAGGTCAAGATTCTGTTAGAAGCCAATGAAGAA
CCTGGTTGTTCAATAACATTTTGTTCGTCTAATATTTCACTACCGCTTGACGTTGGCTGCACTTCATGTACCTCATC
TATAAACGCTTCTTCTGTATCGCTCTGGACGTCATCTTCACTTACGTGATCTGATATTTCACTGTCAGAATCCTCAC
CAACAAGCTCGTCATCGCTTTGCAGAAGAGCAGAGAGGATATGCTCATCGTCTAAAGAACTACCCATTTTATTATAT
ATTAGTCACGATATCTATAACAAGAAAATATATATATAATAAGTTATCACGTAAGTAGAACATGAAATAACAATATA
ATTATCGTATGAGTTAAATCTTAAAAGTCACGTAAAAGATAATCATGCGTCATTTTGACTCACGCGGTCGTTATAGT
TCAAAATCAGTGACACTTACCGCATTGACAAGCACGCCTCACGGGAGCTCCAAGCGGCGACTGAGATGTCCTAAATG
CACAGCGACGGATTCGCGCTATTTAGAAAGAGAGAGCAATATTTCAAGAATGCATGCGTCAATTTTACGCAGACTAT
CTTTCTAGGGTTAAAAAAGATTTGCGCTTTACTCGACCTAAACTTTAAACACGTCATAGAATCTTCGTTTGACAAAA
ACCACATTGTGGCCAAGCTGTGTGACGCGACGCGCGCTAAAGAATGGCAAACCAAGTCGCGCGAGCGTCGACTCTAG
AGGATCCCCGGGTACCGAGCTCGAATTCGTAATCATGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAA
TTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATT
GCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAA

Chapter 5

Summary and Future Prospect

5.1 Summary of research

Mosquito-borne illness is a global pandemic that motivated the work within this dissertation. In these experiments two mosquitoes, *Ae. aegypti* and *An. stephensi* (vectors of dengue virus and malaria, respectively) served as the experimental organisms tested in an effort to contribute to the fundamental knowledge of the vectors and potential control strategies. In *Ae. aegypti*, a novel, mosquito-specific miRNA cluster was found in the intron of an essential maternal and early embryonic gene. The efforts in *An. stephensi* lead to the discovery of a Y-chromosome gene with implications of playing a role in sex determination. These studies investigate the potential evolutionary importance of these systems and the implications of their contribution to both current and novel vector control strategies.

In *Ae. aegypti*, a miRNA cluster containing two copies miR-2941 and a single copy of miR-2946 was discovered through Illumina small RNA sequencing obtained from embryonic RNA samples. These miRNA within the cluster were shown to be maternally deposited into the embryo, accounted as the most abundant miRNAs in the embryonic stage and inducible by blood feeding. This cluster is housed within the first intron of a mosquito specific transcription factor named bZIP1. Analysis of the intron in six other mosquito species show the evolutionary conservation of one miRNA in the cluster, however, it seems the expansion of the cluster is specific to the Culicinae lineage. The conservation of the host gene, bZIP1, was also investigated in *Aedes*, *Culex*, and *Anopheles* species and found to be mosquito specific following the phylogeny of these species and displaying conserved microsynteny.

Expression of the bZIP1 gene was characterized in both *Ae. aegypti* and *An. stephensi* with the miRNA being characterized in *Ae. aegypti*. The gene and miRNA are shown to have similar expression patterns; strong embryonic expression and post blood feeding induction in females. Transient RNAi experiments were performed and demonstrated that the knocking down of bZIP1 mRNA has dramatic adverse effects on embryonic survival.

The bZIP1 promoter showed very strong maternal and zygotic expression, a characteristic that has been exploited in our lab in numerous transient and transgenic applications. The use of the bZIP1 promoter not only allows for maternal and zygotic expression, but also for the potential deposition of transcript into the early embryo. Future unrelated works could benefit from the use of the bZIP1 promoter properties in studies by enabling early embryonic gene manipulation and could allow for a better experimental understanding of the role of some maternally deposited genes. In addition to the promoter that was characterized, the miRNA cluster has been shown to be effectively manipulated to target genes of interest in transgenic applications.

The work with *An. stephensi* focuses on the Y chromosome and its role in sex determination. Our goal was to first develop a method to identify Y chromosome genes *in silico* and demonstrate their Y chromosome specificity via molecular techniques. Using a comparative genomic approach with male and female gDNA databases the discovery of gene unique to the Y 1 (GUY1) as well as other potential Y chromosome genes was possible. GUY1 showed Y chromosome specificity, early embryonic expression, and a potential coding sequence resembling a bHLH/HTH protein with possible DNA binding capability.

In order to prove GUY1 was indeed on the Y chromosome various molecular techniques were employed. PCR was performed on gDNA obtained from differing strains of *An. stephensi*

targeting GUY1; demonstrating male-specific amplification of this gene was not an anomaly of our lab's strain and ruled out sample contamination. As males are the only sex to possess a Y chromosome we also exploited the transfer of sperm bearing both X and Y chromosomes in through the use of PCR with template derived from virgin and mated female mosquitoes. Y chromosome material was successfully amplified from mated female abdomens, where sperm are stored in the spermatheca post copulation, and not from virgin female samples. In order to show GUY1 is a single copy gene, as there is only one Y chromosome in males and zero in females, digital droplet PCR was utilized for its highly quantitative results and capability to effectively determine copy number variation. It was shown that GUY1 is a single copy gene in males and females possess no copies of the gene. As GUY1 is expressed in the early embryo, single embryo PCR and RT-PCR was performed to demonstrate the expression profile of GUY1 was attributed solely to male embryos.

Understanding the function of GUY1 was performed through three transient assays. The plasmids designed for this assay were engineered to: knock down GUY1 with dsRNA and miRNA targeting GUY1 driven by the bZIP1 promoter (dblwhammy_GUY1), ectopically express GUY1 under the bZIP1 promoter (pEE1_GUY1), or ectopically express GUY1 under its native promoter (pEE2_GUY1). The results of transient assays show a significant propensity towards a male bias in the adult stage. The dblwhammy_GUY1 plasmid was used for the generation of transgenic *An. stephensi* via piggyBac transposition and resulting in three defined insertion lines. RT-PCR was performed on these lines and showed that GUY1 was not being knocked down, but rather ectopically expressed. The transgenic progeny from two of the three described lines always have a high bias towards males while the remaining line behaves similar to wild type mosquitoes in its sex distribution.

In order to determine the reasoning behind the lack of females in the adult stage ddPCR was utilized to genotype transgenic adults in a search for males of XX karyotype, suggestive of female to male conversion. All adults that were analyzed were indeed true XY males and not XX individuals displaying a male phenotype due to ectopic expression of the transgene. Single embryo PCR was performed to determine if transgenic males were only producing Y containing sperm, thereby producing only male embryos. It was shown that in the embryonic stage there are males and females which are either transgenic or non transgenic. Phenotypic analysis of the larvae revealed the presence of testes in all transgenic organisms suggesting the GUY1 transgene is playing a lethal role to females between the embryonic and larval stages.

5.2 Future directions

The miR-2941 and miR-2946 cluster represent the most abundant miRNAs expressed in the embryonic stage and therefore likely perform an essential developmental role. Target analysis of miRNA was performed based on the complementary binding of miRNA to the 3' UTRs of certain genes and confirmation of these targets could reveal the function of these miRNAs. In the case of the bZIP1 intronic miRNA some experimental analysis would be possible. The candidate target list could be used to generate a set of plasmids that expresses luciferase and replaces the luciferase 3' UTR with the 3' UTR of the candidate gene. A dual luciferase experiment could be performed by co-injecting experimental plasmid (firefly luciferase) with a modified 3' UTR and a luminescence normalizing plasmid (*Renilla* luciferase) into embryos and incubating. Analyzing the sample at the time point which greatly expresses miR-2941 or miR-2946 would allow for the experimental testing of knockdown efficacy in a dual luciferase assay. Control experiments for this assay would utilize a luciferase expressing

construct with no 3' UTR or a UTR that has no known miRNA targets. The 3' UTRs of genes that displayed a knockdown of luciferase expression, demonstrated by a decrease in RLUs, would have experimental evidence of being a potential miRNA target. This experiment does not take into account the possibility of other miRNA being present at the same time and performing the knockdown.

The bZIP1 gene was shown to drastically effect embryonic survival when knocked down and further characterization of this gene could reveal its mosquito-specific importance in early embryonic development. Based on the peptide sequence analysis, it is hypothesized that this gene acts as a transcription factor. Further experimentation characterizing the activity of this protein could lend to understanding why it is so vital in the embryonic stage. First, protein expression and antibody generation would need to be performed to determine when the protein is being expressed through the use of western blotting. Possessing antibodies to bZIP1 would then allow for a ChIP assay to be performed and determine the DNA binding domain, if there is any. Additionally, antibodies against bZIP1 could be used to perform a pull down assay to determine if there are any other proteins that interact with bZIP1. When coupled with mass spectrometry, the pull down assay could lead to peptide identification. These experiments would help to explain why bZIP1 is necessary to embryonic development and potentially which pathways are being affected by its knockdown.

Demonstrating the molecular function of GUY1 remains an important task; however, experimental evidence has shown it is likely involved in sex determination. Determining whether GUY1 is a protein has failed via mass spectrometry and western analysis, however, it still remains to be an important step in characterization of the gene. A possibility as to why these experiments failed would be that the protein itself is not very highly expressed and the gene is

only transcribed in a short window of time. Potentially the protein is not abundant enough to be effectively detected by previously attempted methods. Enriching for smaller proteins from a large amount of raw material could aid in the detection of GUY1 protein. Enrichment could be performed through size separation chromotograhpy or through affinity chromatography in conjunction with the GUY1 antibodies that have previously been generated. The GUY1 antibodies could also be flawed in their recognition of native epitopes. The antibodies were generated from synthesized partial GUY1 peptides; a better strategy might be to express GUY1 and generate antibodies from the full length peptide.

Understanding the relationship GUY1 has with the environment of the early embryo could help to solidify the proposed involvement in sex determination and attaining antibodies that can successfully identify GUY1 would be vital to this endeavor. Working antibodies would allow for the detection of potential DNA binding sites as well as identification of other proteins that could interact with GUY1. Recombinant GUY1 protein would be used as antigen for antibody generation to the full length GUY1 protein. The current antibodies were generated from synthesized antigen and have yet to yield any results with western blots, possibly due to non recognition of native GUY1 epitopes. ChIP analysis would be possible with antibodies that are shown to identify GUY1 proteins, thereby identifying the DNA binding sequence of GUY1. Additionally, the use of GUY1 antibodies in a co-immuniprecipitation experiment in combination with mass spectrometry would be able to identify if any proteins interact with GUY1.

Sexing of embryos and performing sex-specific early embryonic transcriptomes would be possible due to the capability of identifying the sex of single embryo and obtaining RNA from the same embryo. In this regard, it would be possible to obtain sex specific RNA pools. In the

early embryo this would be highly beneficial as it could be used to elucidate genes that are alternatively spliced or differentially regulated between males and females. Additionally, this work could be used to solve how X compensation would be working in the *An. stephensi* male embryo, similar to work performed in *Drosophila* (Lott *et al*., 2011).

It is currently hypothesized that ectopic expression of GUY1 peforms female lethality by activating downstream genes which are deleterious in females. Performing transcriptome analysis on male and female transgenic embryos could also prove to be useful in identifying how the GUY1 transgene in the dblwhammy_GUY1 lines is affecting other genes and pinpoint the potential cause of female embryonic arrest. The same strategies as previously described to obtain RNA from single embryos would be used to create RNA libraries from transgenic embryos. The comparison between transgenic and non transgenic transcriptome profiles would display genes that might be alternatively regulated due to the presence of the transgene. The current ectopic expression lines that have been developed show a very high male bias. This bias proves to be problematic due to the constant necessity of mating transgenic males with wild type females, forcing all transgenic progeny to be heterozygous for the transgene. Studies in which a homozygous line would be used could assist in understanding the molecular mechanism by which ectopic expression of GUY1 is causing female lethality. Using a homozygous line would simplify the step in embryo genotyping to obtain male and female transgenics as non transgenic embryos would not be present in the progeny. Obtaining a homozygous line could be done through the use of the GAL4-UAS system. In this system the transgene would be under regulation of the UAS promoter, a promoter that is activated by GAL4. By generating lines that express GAL4 with either the bZIP1 or GUY1 promoter and a line expressing GUY1 under the UAS promoter we could observe GUY1 ectopic expression by crossing of the two lines.

Homozygous line generation would be possible because the GUY1 cassette would only be activated in the presence of GAL4, which would be absent, and allow for male and female transgenic progeny. When the UAS-GUY1 line would be crossed with the bZIP1/GUY1-GAL4 line the resulting progeny would express GAL4 and in turn express GUY1, giving more control over when the phenotype is observed.

The ectopic expression addresses the potential influence of GUY1 in female individuals. However, the knockdown lines that were generated do not have a clear sex bias and the only implications of successful integration are from DsRed expression and luciferase activity. To characterize the insertion, FISH and PCR should be performed to determine potential insertion position effects or mutations. Sequencing of the transgenic lines would also provide information on the site of integration. Performing ddPCR on adult males and females of these knockdown lines would confirm if there are pseudo males or females present in the population as they might be present at a low frequency. Confirming the knockdown construct is performing effectively would be validated by qRT-PCR of transgenic embryos against wild type embryo levels of GUY1.

As the efficacy of the knockdown constructs has yet to be determined an alternate strategy to knocking down GUY1 would be to use TALEN technology to knock out GUY1 from the Y chromosome. Even if the knockdown of GUY1 via shRNA expression is working it is possible that only small amounts of transcript are required to carry out normal embryonic development. Knocking out GUY1 would remove any transcriptional activity and help differentiate between GUY1 functioning as the Y chromosome M factor, resulting in XY females, or functioning as a female lethal gene, in which case there would be no change in the sex ratio when compared to wild type.

A final experiment to demonstrate the efficacy of using the dblwhammy_GUY1 lines as a tool for population crash would be to perform cage studies. These experiments would include a "release" of transgenic males into a wild type population and observing how the population behaved over the generations. Increasing or decreasing the initial amount of males released should lend valuable information on how the introduction of the dblwhammy_GUY1 lines into a population affects the distribution of sex and viability of future generations.

Appendix A

Abstract

The GUY1 ectopic expression transgenic lines analyze the effect of adding a Y chromosome gene to an autosome for expression in XX individuals. An adverse effect on female embryos was observed in all lines except for one that ectopically express GUY1. Observing the knockdown of GUY1 in a transgenic environment would help determine the role that GUY1 is performing in the early embryo. In this study, transgenic lines were generated that express shRNA targeting GUY1 and the luciferase ORF for measuring the activity of the construct. Two different plasmids were used to generate transgenic lines independently; both constructs express shRNA targeting GUY1 but under the control of different promoters. The bZIP1 promoter is used in pKD2_GUY1 and the maternal promoter 143874 is used in pKD1_GUY1. The transgenic progeny do not show a sex bias as seen in the GUY1 ectopic expression lines. Luciferase activity has been observed and characterized in the adult stage. The greater expression of luciferase in females, especially after blood feeding, suggests the promoters are behaving as hypothesized.

Materials and Methods

Generation of transgenic lines

The pKD2_GUY1 transformation plasmid was generated from the dblwhammy_GUY1 plasmid with the purpose of knocking down GUY1 with maternally deposited shRNA that target GUY1. The shRNA targeting GUY1 were designed according to Ni *et al*., 2011 and synthesized as oligomers. The area between the bZIP1 promoter and the attB site was removed by double

digest with SpeI and XhoI, then replaced with the sense orientated sequence of shRNA targeting GUY1 followed by luciferase cloned from the pGL3 plasmid (Promega).

The pKD1_GUY1 construct replaces the bZIP1 promoter with the maternal promoter 143874 and maintains the shRNA targeting GUY1 followed by luciferase. See Figure 4.1 for plasmid maps.

Mosquito rearing and mating

Transgenic mosquitoes were reared in incubators at 27°C and 60% relative humidity on a 16 hr light/8 hr dark photoperiod. Larvae were fed Sera Micron Fry Food with brewer's yeast, and Purina Game Fish Chow. Adult mosquitoes were fed on a 10% sucrose soaked cotton pad. Injected embryos for each transgenic construct were obtained from UMD Insect Transformation Facility and reared to adulthood. G0 adult males were mated individually mated in a 1:5 ratio with wild type females while G0 females were pooled and mated with wild type males in a 1:5 ratio. Mosquitoes were blood-fed on female Hsd:ICR (CD-1®) mice (Harlan Laboratories, http://www.harlan.com). Egg collections occurred approximately 3 days post blood-feeding. Transgenic positive larvae were screened for dsRed expression during the L-3 stage and non-transgenic individuals were discarded. G1 males and females were segregated after emergence and mated with wild type in a 1:5 ratio. Progeny from this segregation are denoted as “Md” or “Fd”, meaning male derived or female derived, respectively. Subsequent generations were allowed to in-breed for the purpose of producing homozygous lines.

Luciferase activity assays

Luciferase activity was measured from adult transgenic mosquitoes flash frozen at the following time points: 48 hrs post emergence males, 48 hrs post emergence females, 24 hrs post blood fed females, and 72 hrs post blood fed females. Adults were dissected into heads, thoraces, and abdomens, and processed individually. Samples were processed using the Luciferase Assay Kit (Promega) and homogenized in 100 uL 1X Passive Lysis Buffer. Luciferase activity detection was performed by combining 50 uL of homogenate with 20 uL of Luciferase Assay Reagent I (LARI) and measuring firefly luciferase (adapted from Biedler and Tu 2010).

Results

Classification of transgenic lines

Transgenic lines were successfully generated via piggyBac transposition as determined by DsRed expression in nerve tissue of larvae (as per the 3xP3/HSP70 promoter). Transgenic positives for both pKD1_GUY1 and pKD2_GUY1 were separated into the Md and Fd designations at the G1 and G2 adult stage, respectively. Both transgenic lines show a 1:1 male to female sex ratio, consistent with wild type *An. stephensi* and not having a significant bias towards either sex (Table 4.1A). There also appear to be no X or Y linked insertions based on the 1:1 male to female ration in the non-transgenic population (Table 4.1B).

Expression of luciferase

Luciferase activity is measured as a confirmation that the construct is being transcribed in line with the promoters driving the construct and processed efficiently. Expression of luciferase based on the characterization of the bZIP1 and 143874 promoters should be maternal and elevated in response to blood feeding. Expression in males is outlined in Table 4.2A and shows a

group of individuals with near background levels of expression and others with 100 fold increase over background. Despite this large difference, these values are considerably less than what is seen in the female both pre and post blood feeding. Notably, in the KD1_GUY1-Md line there is a 12 fold increase in luciferase activity in response to blood feeding. In the KD2_GUY1 line there is ~194 fold increase in luciferase activity in response to blood feeding. Comparing the promoters of the KD1 and KD2 lines show that based on luciferase activity the bZIP1 promoter is approximately 3 times stronger after blood feeding that the 143874 maternal promoter.

Discussion

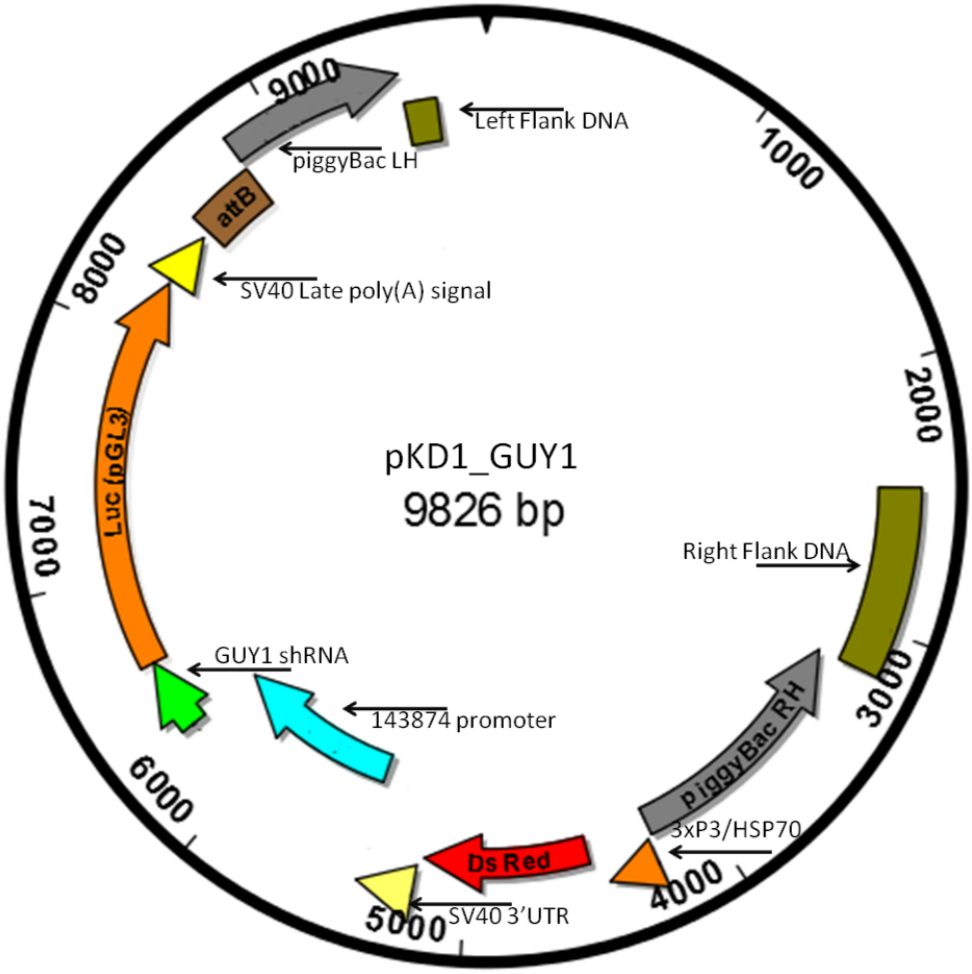
The purpose of generating these knock down lines was to add more experimental data to the potential function of GUY1. Previously, it had been shown that the ectopic expression of GUY1 leads to a male bias. Successful integrations of two knock down constructs have been observed based on DsRed expression. In these knock down experiments there are no observed biases to either sex.

Based on luciferase activity, the 143874 maternal promoter and the bZIP1 promoter appear to be behaving as previously characterized. However, there is expression, albeit low, in some males and in areas of the organism where activity from these promoters has not been previously observed. Although it has been shown that we have successful integration of the construct into the genome based on DsRed expression and luciferase activity, lines KD1 and KD2 are not “pure” in the sense that there could be multiple insertions in each line. Further characterization of the lines could reveal position effects on the transgene or mutations in the construct. Additional analysis of the GUY1 mRNA levels also would be required to determine the efficacy of the shRNA construct.

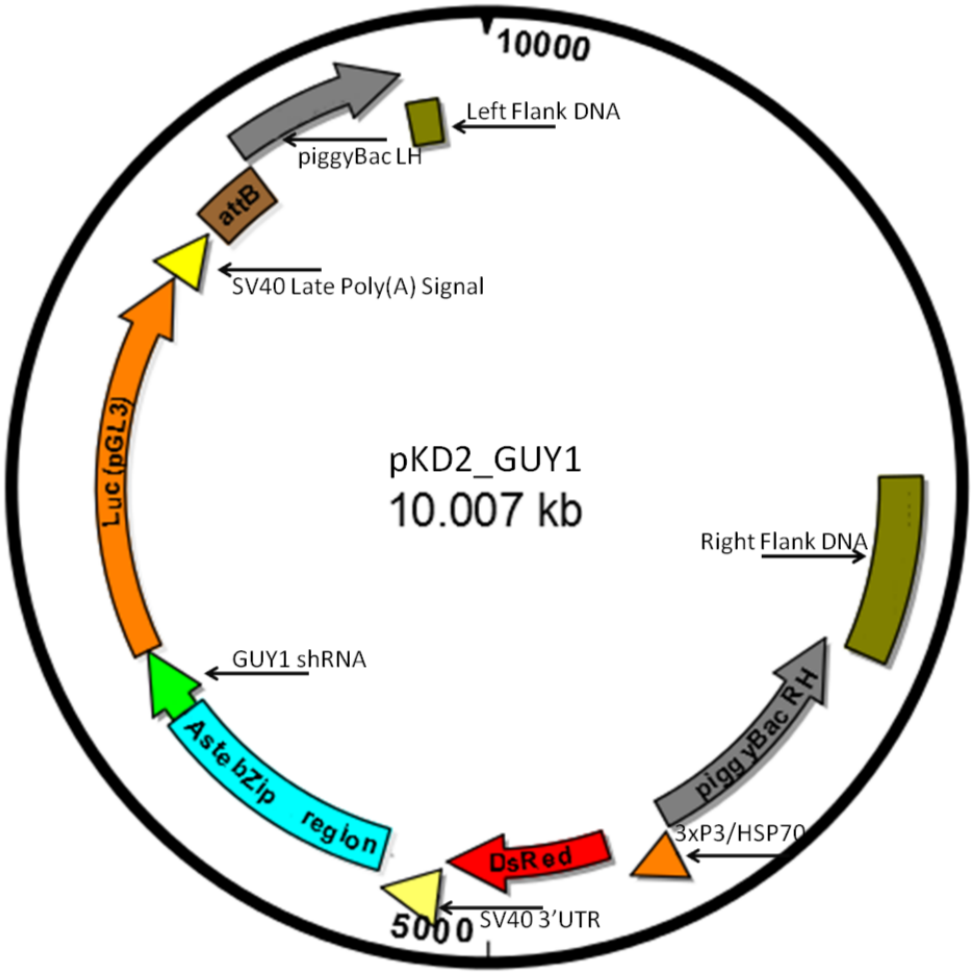
Figure Legends, Tables, and Supplemental Information

Figure A.1 Plasmid maps of pKD1_GUY1 and pKD2_GUY1

A.



B.



A. This map shows pKD1_GUY1 to have the 143874 maternal promoter driving GUY1 targeting shRNA and luciferase. B pKD2_GUY1 is shown to have the bZIP1 promoter driving GUY1targeting shRNA and luciferase

Table A.1 Sex distribution of transgenic progeny by generation

A.

pKD1_GUY1

Generation	Male	Female
0	33	29
1	11	23

pKD1_GUY1-Md

	Male	Female
2	22	24

pKD2_GUY1

Generation	Male	Female
0	26	9
1	7	6
2	57	45

pKD2_GUY-Md

	Male	Female

3	77	55
4	31	28
5	41	20

pKD1_GUY1-Fd

	Male	Female
2	17	24
3	35	40
4	24	27
5	4	2

3	6	9

pKD2_GUY-Fd

	Male	Female
3	23	26

B.

	Male	Female
KD1_GUY1-Fd (Generation 2)	4	4
KD1_GUY1-Md (Generation 2)	12	16
KD2_GUY1 (Generation 2)	55	38

Table 4.1A lists the numbers of adult males and females observed in each generation. The generation numbers also describe when the lines were split to form the "Md" and "Fd" lines.

Table 4.1B shows the distribution of the non-transgenic adult to rule out X or Y linkage.

Table A.2 Luciferase data

A.

KD1_GUY1-Fd(Male)

Sample	Head	Thorax	Abdomen
1	240	116	62
2	79	328	150
3	488	281	212
AVG	269	241.6667	141.3333

KD1_GUY1-Md(Male)

Sample	Head	Thorax	Abdomen
1	39728	39410	36506
2	225	170	71
3	289	166	73
AVG	13414	13248.67	12216.67

KD2_GUY1(Male)

Sample	Head	Thorax	Abdomen

1	6975	18145	884
2	156	181	136
3	23924	14565	448
AVG	10351.67	10963.67	489.3333

B.

KD1_GUY1-Fd(Sugar Fed Female)

Sample	Head	Thorax	Abdomen
1	492	109	123
2	2562	137	101
3	1277	222	494
AVG	1443.667	156	239.3333

KD1_GUY1-Fd(72hrs PBM)

Sample	Head	Thorax	Abdomen
1	1051	64	16995
2	1233	55	8067
3	9492	65	7457
4	7167	60	2817
AVG	4735.75	61	8834

KD1_GUY1-Md(Sugar Fed Female)

Sample	Head	Thorax	Abdomen
1	482	59	72
2	141239	12205	276107
3	122223	12415	431694
AVG	87981.33	8226.333	235957.7

KD1_GUY1-Md(72hrs PBM)

Sample	Head	Thorax	Abdomen
1	156314	12376	27632
2	188096	69	9727909
3	787	77	219
4	1285629	57	7982946
5	58326	136	172646
6	136020	82	49622
AVG	304195.3	2132.833	2993496

KD2_GUY1(Sugar Fed Female)

Sample	Head	Thorax	Abdomen
1	1715	63	6270
2	313	1054	4994
3	10054	3566	121202
AVG	4027.333	1561	44155.33

KD2_GUY1(72hrs PBM Female)

Sample	Head	Thorax	Abdomen
1	144872	38228	815494
2	53867	1359	12835243
3	133943	2414	12123968
AVG	110894	14000.33	8591568

Luciferase activity assay RLU units from KD1 and KD2 lines adult male and female (sugar fed and blood fed). Table 4.2A lists the male luciferase activity for the head, thorax, and abdomen for lines KD1_GUY1-Md, KD1_GUY1-Fd and KD2_GUY1. Table 4.2B compares sugar fed females and 72hr post blood fed female luciferase activity in the head, thorax and abdomen of individual females from lines KD1_GUY1-Md, KD1_GUY1-Fd, and KD2_GUY1.

Figure A.S1 Sequence of pKD1_GUY1 and pKD2_GUY1

>pKD1_GUY

CCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTT
CCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGG
TTATCCACAGAATCAGGGGATAACGCACGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAA
GGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGT
GGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACC
CTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA
TCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCT
TATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGG
ATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAAC
AGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAA
CCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCT
TTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAA
AAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGT
CTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGA
CTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCC
ACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTT
TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAAC
GTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACG
ATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGA
AGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAG
ATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC
CGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGG
CGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC
ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGA
CACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGC
GGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
CGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTT
TCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGG
AGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCA
GATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCC
ATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAA
GGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAG
TGCCAAGCTTTGTTTAAAAATATAACAAAATTGTGATCCCACAAAATGAAGTGGGGCAAAATCAAATAATTAATAGT
GTCCGTAAACTTGTTGGTCTTCAACTTTTTGAGGAACACGTTGGACGGCAAATCCGTGACTATAACACAAGTTGATT
TAATAATTTTAGCCAACACGTCGGGCTGCGTGTTTTTTGCCGACGCGTCTGTGTACACGTTGATTAACTGGTCGATT
AAACTGTTGAAATAATTTAATTTTTGGTTCTTCTTTAAATCTGTGATGAAATTTTTTAAAATAACTTTAAATTCTTC
ATTGGTAAAAAATGCCACGTTTTGCAACTTGTGAGGGTCTAATATGAGGTCAAACTCAGTAGGAGTTTTATCCAAAA
AAGAAAACATGATTACGTCTGTACACGAACGCGTATTAACGCAGAGTGCAAAGTATAAGAGGGTTAAAAAATATATT
TTACGCACCATATACGCATCGGGTTGATATCGTTAATATGGATCAATTTGAACAGTTGATTAACGTGTCTCTGCTCA
AGTCTTTGATCAAAACGCAAATCGACGAAAATGTGTCGGACAATATCAAGTCGATGAGCGAAAAACTAAAAAGGCTA
GAATACGACAATCTCACAGACAGCGTTGAGATATACGGTATTCACGACAGCAGGCTGAATAATAAAAAAATTAGAAA
CTATTATTTAACCCTAGAAAGATAATCATATTGTGACGTACGTTAAAGATAATCATGCGTAAAATTGACGCATGTGT
TTTATCGGTCTGTATATCGAGGTTTATTTATTAATTTGAATAGATATTAAGTTTTATTATATTTACACTTACATACT
AATAATAAATTCAACAAACAATTTATTTATGTTTATTTATTTATTAAAAAAAAACAAAAACTCAAAATTTCTTCTAT
AAAGTAACAAAACTTTTAAACATTCTCTCTTTTACAAAAATAAACTTATTTTGTACTTTAAAAACAGTCATGTTGTA
TTATAAAATAAGTAATTAGCTTAACTTATACATAATAGAAACAAATTATACTTATTAGTCAGTCAGAAACAACTTTG
GCACATATCAATATTATGCTCTCGACAAATAACTTTTTTGCATTTTTTGCACGATGCATTTGCCTTTCGCCTTATTT
TAGAGGGGCAGTAAGTACAGTAAGTACGTTTTTTCATTACTGGCTCTTCAGTACTGTCATCTGATGTACCAGGCACT
TCATTTGGCAAAATATTAGAGATATTATCGCGCAAATATCTCTTCAAAGTAGGAGCTTCTAAACGCTTACGCATAAA
CGATGACGTCAGGCTCATGTAAAGGTTTCTCATAAATTTTTTGCGACTTTGAACCTTTTCTCCCTTGCTACTGACAT
TATGGCTGTATATAATAAAAGAATTTATGCAGGCAATGTTTATCATTCCGTACAATAATGCCATAGGCCACCTATTC
GTCTTCCTACTGCAGGTCATCACAGAACACATTTGGTCTAGCGTGTCCACTCCGCCTTTAGTTTGATTATAATACAT
AACCATTTGCGGTTTACCGGTACTTTCGTTGATAGAAGCATCCTCATCACAAGATGATAATAAGTATACCATCTTAG
CTGGCTTCGGTTTATATGAGACGAGAGTAAGGGGTCCGTCAAAACAAAACATCGATGTTCCCACTGGCCTGGAGCGA
CTGTTTTTCAGTACTTCCGGTATCTCGCGTTTGTTTGATCGCACGGTTCCCACAATGGTTAATTCGAGCTCGCCCGG
GGATCTAATTCAATTAGAGACTAATTCAATTAGAGCTAATTCAATTAGGATCCAAGCTTATCGATTTCGAACCCTCG
ACCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCT

AAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCGGGGTACCGCTAGAGTCGACGGTACCG
CGGGCCCGGGATCCACCGGTCGCCACCATGGTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGT
GCGCATGGAGGGCACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACA
ACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCCAGTAC
GGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTG
GGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGaCCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCA
TCTACAAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAG
GCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACAAGGCCCTGAAGCTGAAGGACGG
CGGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGG
ACTCCAAGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGCCAC
CACCTGTTCCTGTAGCGGCCGCGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAA
AAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTT
ATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGT
TTGTCCAAACTCATCAATGTATCTTAAAGCTTATCGATACGCGTACGGCTGAGCGGTTTGGTGCAGGGCTCTCAACC
AAACCACCAAAACCGCCGGGTGATCAGACCCCGCAGAACTGACAATTATACTATTACTAGAGAGGATCGCCAGCGGA
AAGATCTACGGCACGTTTAATTTATCGGTTTATCAATAATAGCACCTAGTTACCATAAGCTAGCATCTCCAAAGTAG
TGGCTCTTTAGTTTCTACCGCTTCTGTTTTTTATTTGCACAGTTTTCCAACAGCATAATTCGAACCGATTGGCGGTA
GTAGCAGCAGCAGCAACACCAAAACATACAATTTATGCAGACAAAGGTAATAGGCAATAGCGAGCAGGAGGAGGGGA
CTAATCTTTAAGTGTAAATTCTGAAGATTTTTTTTTCGGTGGGAAAAGCTTTCAAAACTATAATTGTAACTTAATGA
TGGCAGCATCATTAGAATAGGGCACAATATTCATCAACTCAGGCGTCACACACGTCAGGCCTATGAAGGCATATTTA
GTATTTGTATAGTAATATGCTAGGGCATAAGTTACTCTAGTCAGTAAATGCACACAAAAATTTAAACATTTGTTTCT
TTTTTGGTTTTAAAGTTAAGAAGCACGAAAGAAATTCACCACTATTCGCTTACACCGCTTCGATCGTTCGACCCAAC
TGCGAAAACCAAACCCGGCTGTGGCCCTGCTGCTTCGCAATGGCGGTAGACTCCTACCGGTTTCCTGTCCTCCATTT
TGTACTCCACAAACACGCTACTAATCCTGGCGTGGCTTGGTGTGTTGCGGTGCAGCAGCTAGCAATTAATTATCTGT
TTTGTGTCGTTTTTTTTTGTTTGTGTGTCTGTCTAATTGTGTGTTTTAATTGTTCTTGATCATATAaagGTAAGAAG
ATCGTTCGATTAGGCGCTAATTGGGATTTGTGGGAGCAAGAGTGTTGATCTGAAGTGCGAACCGTATGCTGGATGGA
AGTTATTTGATCGCCATGTCTAAGGAGACTAGTtCGAAAGTACCACTTTGTGTGTTAATATATAGTGATTCTGAAAG
CATATATATTAACACACAAAGTGGTGCGCTCGgCTCGAGCAATATAGTCATCGAGTTAAACCCATAAATTTGTAAGA
AATTTAAATGCATAGCTTTTCTATGAACGTTACATTAATACATTAATTTTTTTTTCATTCTAGATGGAAGACGCCAA
AAACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAAGATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGA
AGAGATACGCCCTGGTTCCTGGAACAATTGCTTTTACAGATGCACATATCGAGGTGGACATCACTTACGCTGAGTAC
TTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGGGCTGAATACAAATCACAGAATCGTCGTATGCAG
TGAAAACTCTCTTCAATTCTTTATGCCGGTGTTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACA
TTTATAATGAACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACCGTGGTGTTCGTTTCCAAAAAGGGGTTG
CAAAAAATTTTGAACGTGCAAAAAAAGCTCCCAATCATCCAAAAAATTATTATCATGGATTCTAAAACGGATTACCA
GGGATTTCAGTCGATGTACACGTTCGTCACATCTCATCTACCTCCCGGTTTTAATGAATACGATTTTGTGCCAGAGT
CCTTCGATAGGGACAAGACAATTGCACTGATCATGAACTCCTCTGGATCTACTGGTCTGCCTAAAGGTGTCGCTCTG
CCTCATAGAACTGCCTGCGTGAGATTCTCGCATGCCAGAGATCCTATTTTTGGCAATCAAATCATTCCGGATACTGC
GATTTTAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACACTCGGATATTTGATATGTGGATTTCGAG
TCGTCTTAATGTATAGATTTGAAGAAGAGCTGTTTCTGAGGAGCCTTCAGGATTACAAGATTCAAAGTGCGCTGCTG
GTGCCAACCCTATTCTCCTTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTTACACGAAATTGC
TTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCCAAGAGGTTCCATCTGCCAGGTATCAGGCAAG
GATATGGGCTCACTGAGACTACATCAGCTATTCTGATTACACCCGAGGGGGATGATAAACCGGGCGCGGTCGGTAAA
GTTGTTCCATTTTTTGAAGCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTTAATCAAAGAGGCGAACT
GTGTGTGAGAGGTCCTATGATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATTGACAAGGATG
GATGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAACACTTCTTCATCGTTGACCGCCTGAAGTCTCTG
ATTAAGTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCTTGCTCCAACACCCCAACATCTTCGACGC
AGGTGTCGCAGGTCTTCCCGACGATGACGCCGGTGAACTTCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACGA
TGACGGAAAAAGAGATCGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAGTTGTGTTT
GTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGCAAGAAAAATCAGAGAGATCCTCATAAAGGCCAAGAA
GGGCGGAAAGATCGCCGTGTAAGCGGCCGCACCGGTCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACA
ACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTG
CAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAA
GCAAGTAAAACCTCTACAAATGTGGTAGCCGGCTTAATTCTTAAGTTAATTTTAATTAAGTCGACGATGTAGGTCAC
GGTCTCGAAGCCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACCTCACCCATCTGGT
CCATCATGATGAACGGGTCGAGGTGGCGGTAGTTGATCCCGGCGAACGCGCGGCGCACCGGGAAGCCCTCGCCCTCG
AAACCGCTGGGCGCGGTGGTCACGGTGAGCACGGGACGTGCGACGGCGTCGGCGGGTGCGGATACGCGGGGCAGCGT

CAGCGGGTTCTCGACGGTCACGGCGGGCATGTCGACAGATCTGACAATGTTCAGTGCAGAGACTCGGCTACGCCTCG
TGGACTTTGAAGTTGACCAACAATGTTTATTCTTACCTCTAATAGTCCTCTGTGGCAAGGTCAAGATTCTGTTAGAA
GCCAATGAAGAACCTGGTTGTTCAATAACATTTTGTTCGTCTAATATTTCACTACCGCTTGACGTTGGCTGCACTTC
ATGTACCTCATCTATAAACGCTTCTTCTGTATCGCTCTGGACGTCATCTTCACTTACGTGATCTGATATTTCACTGT
CAGAATCCTCACCAACAAGCTCGTCATCGCTTTGCAGAAGAGCAGAGAGGATATGCTCATCGTCTAAAGAACTACCC
ATTTTATTATATATTAGTCACGATATCTATAACAAGAAAATATATATATAATAAGTTATCACGTAAGTAGAACATGA
AATAACAATATAATTATCGTATGAGTTAAATCTTAAAAGTCACGTAAAAGATAATCATGCGTCATTTTGACTCACGC
GGTCGTTATAGTTCAAAATCAGTGACACTTACCGCATTGACAAGCACGCCTCACGGGAGCTCCAAGCGGCGACTGAG
ATGTCCTAAATGCACAGCGACGGATTCGCGCTATTTAGAAAGAGAGAGCAATATTTCAAGAATGCATGCGTCAATTT
TACGCAGACTATCTTTCTAGGGTTAAAAAAGATTTGCGCTTTACTCGACCTAAACTTTAAACACGTCATAGAATCTT
CGTTTGACAAAAACCACATTGTGGCCAAGCTGTGTGACGCGACGCGCGCTAAAGAATGGCAAACCAAGTCGCGCGAG
CGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCGTAATCATGTCATAGCTGTTTCCTGTGTGAAATTGTT
ATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAA
CTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAA

>pKD2_GUY1
CCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTT
CCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGG
TTATCCACAGAATCAGGGGATAACGCACGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAA
GGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGT
GGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACC
CTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA
TCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCT
TATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGG
ATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAAC
AGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAA
CCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCT
TTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAA
AAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGT
CTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGA
CTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCC
ACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTT
TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAAC
GTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACG
ATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGA
AGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAG
ATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC
CGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGG
CGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC
ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGA
CACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGC
GGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
CGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTT
TCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGG
AGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCA
GATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCC
ATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAA
GGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAG
TGCCAAGCTTTGTTTAAAAATATAACAAAATTGTGATCCCACAAAATGAAGTGGGGCAAAATCAAATAATTAATAGT
GTCCGTAAACTTGTTGGTCTTCAACTTTTTGAGGAACACGTTGGACGGCAAATCCGTGACTATAACACAAGTTGATT
TAATAATTTTAGCCAACACGTCGGGCTGCGTGTTTTTTGCCGACGCGTCTGTGTACACGTTGATTAACTGGTCGATT
AAACTGTTGAAATAATTTAATTTTTGGTTCTTCTTTAAATCTGTGATGAAATTTTTTAAAATAACTTTAAATTCTTC
ATTGGTAAAAAATGCCACGTTTTGCAACTTGTGAGGGTCTAATATGAGGTCAAACTCAGTAGGAGTTTTATCCAAAA
AAGAAAACATGATTACGTCTGTACACGAACGCGTATTAACGCAGAGTGCAAAGTATAAGAGGGTTAAAAAATATATT
TTACGCACCATATACGCATCGGGTTGATATCGTTAATATGGATCAATTTGAACAGTTGATTAACGTGTCTCTGCTCA
AGTCTTTGATCAAAACGCAAATCGACGAAAATGTGTCGGACAATATCAAGTCGATGAGCGAAAAACTAAAAAGGCTA
GAATACGACAATCTCACAGACAGCGTTGAGATATACGGTATTCACGACAGCAGGCTGAATAATAAAAAAATTAGAAA

CTATTATTTAACCCTAGAAAGATAATCATATTGTGACGTACGTTAAAGATAATCATGCGTAAAATTGACGCATGTGT
TTTATCGGTCTGTATATCGAGGTTTATTTATTAATTTGAATAGATATTAAGTTTTATTATATTTACACTTACATACT
AATAATAAATTCAACAAACAATTTATTTATGTTTATTTATTTATTAAAAAAAAACAAAAACTCAAAATTTCTTCTAT
AAAGTAACAAAACTTTTAAACATTCTCTCTTTTACAAAAATAAACTTATTTTGTACTTTAAAAACAGTCATGTTGTA
TTATAAAATAAGTAATTAGCTTAACTTATACATAATAGAAACAAATTATACTTATTAGTCAGTCAGAAACAACTTTG
GCACATATCAATATTATGCTCTCGACAAATAACTTTTTTGCATTTTTTGCACGATGCATTTGCCTTTCGCCTTATTT
TAGAGGGGCAGTAAGTACAGTAAGTACGTTTTTTCATTACTGGCTCTTCAGTACTGTCATCTGATGTACCAGGCACT
TCATTTGGCAAAATATTAGAGATATTATCGCGCAAATATCTCTTCAAAGTAGGAGCTTCTAAACGCTTACGCATAAA
CGATGACGTCAGGCTCATGTAAAGGTTTCTCATAAATTTTTTGCGACTTTGAACCTTTTCTCCCTTGCTACTGACAT
TATGGCTGTATATAATAAAAGAATTTATGCAGGCAATGTTTATCATTCCGTACAATAATGCCATAGGCCACCTATTC
GTCTTCCTACTGCAGGTCATCACAGAACACATTTGGTCTAGCGTGTCCACTCCGCCTTTAGTTTGATTATAATACAT
AACCATTTGCGGTTTACCGGTACTTTCGTTGATAGAAGCATCCTCATCACAAGATGATAATAAGTATACCATCTTAG
CTGGCTTCGGTTTATATGAGACGAGAGTAAGGGGTCCGTCAAAACAAAACATCGATGTTCCCACTGGCCTGGAGCGA
CTGTTTTTCAGTACTTCCGGTATCTCGCGTTTGTTTGATCGCACGGTTCCCACAATGGTTAATTCGAGCTCGCCCGG
GGATCTAATTCAATTAGAGACTAATTCAATTAGAGCTAATTCAATTAGGATCCAAGCTTATCGATTTCGAACCCTCG
ACCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCT
AAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCGGGGTACCGCTAGAGTCGACGGTACCG
CGGGCCCGGGATCCACCGGTCGCCACCATGGTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGT
GCGCATGGAGGGCACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACA
ACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCCAGTAC
GGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTG
GGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGaCCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCA
TCTACAAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAG
GCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACAAGGCCCTGAAGCTGAAGGACGG
CGGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGG
ACTCCAAGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGCCAC
CACCTGTTCCTGTAGCGGCCGCGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAA
AAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTT
ATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGT
TTGTCCAAACTCATCAATGTATCTTAAAGCTTATCGATACGCGTACGGCGCGCCGCAATAGAGTTAAATAATGGAAC
TTTACCTTGGCCGTGTGTTCTGTTTTTCTTCAGCACCAAATCTTAGTTCTGCAAAACCACTGCTGTGTGAAAAAAGA
CAAGGATGACAGCTCAACTCAACTAGTAGCGGGTGTAAACGCAGAGGACAGCGGGCACTGACGGGAAAACGATAAGA
AAGCGAATGAGGATTTAAATTTCGCAAGCAAGTTTTAGCAATCAAGCAAGAAATACCGTTAATAAAAATTTCAGTTA
AAATTTTAAATTAATTCCTTCACATGAAGCAATTATTGCACTCCAAAGTCGCTTTGATTTGTAACGCCTACCTGTTT
ACCTATATTTCGTTCGTAAAGGAAATTAAGGCCAAAGGGGTAAGAAAATTTTTCATCCATGTTCTGGACCTCCAAAA
AACCGTTAAAATTCATAAAACGATTGTTTCTTTACCCTGGATTGACCTAATAACGAACCGTGTTTTATCTGTTCTCC
CTTTTCGTAACTGGAACGTGAAGGCTTTACCCCAACATTAAATGCCTACGAACGATCCTGTTGGAATTCCAGGTAGG
GAATGCGCAGGATGCGATTCCAGGAACCAGAGCACGCGTATCCACCGGGGCAGATCATCAACAGATGACCGTTAATA
CCTGGTTCTGCGGTATTCCATTTAACTACCCTAATACCAGCACTAATGGAAAAAAGACACAGCCTTAGCATAATTCG
TGGTGTGTGCAGTTTTTTCTATTATTTGCCGCTGCCACCATGTCACAAACACAGGTAAGCGCTGTATGGATGCTATT
TACTTGGTCAGATGTTATCTGATGTGCGTAGTAGTACCCGTGTGGGCGCGACGGCGATCGTCGTGTGCGTCGTGGGT
AAAGTTCCTTGCTTGTAGCGATTAGTATATAAATGTGGTCCAGGTCGCCCATTTGCGTTAATTACTGTTCGTCCTTC
TTTCTGCGTAGACCGGATTGTGCGAGTTGCAGAACCGTGTGTGTCCATTGTGTGTGCGTGTTTTTTGAGAAAATAGC
TGAATTGAAGGAAAAGaagGTAAGAAGATCGTTCGATTAGGCGCTAATTGGGATTTGTGGGAGCAAGAGTGTTGATC
TGAAGTGCGAACCGTATGCTGGATGGAAGTTATTTGATCGCCATGTCTAAGGAGACTAGTtCGAAAGTACCACTTTG
TGTGTTAATATATAGTGATTCTGAAAGCATATATATTAACACACAAAGTGGTGCGCTCGgCTCGAGCAATATAGTCA
TCGAGTTAAACCCATAAATTTGTAAGAAATTTAAATGCATAGCTTTTCTATGAACGTTACATTAATACATTAATTTT
TTTTTCATTCTAGATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAAGATGGAACCG
CTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCCCTGGTTCCTGGAACAATTGCTTTTACAGATGCACATATC
GAGGTGGACATCACTTACGCTGAGTACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGGGCTGAA
TACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTTATGCCGGTGTTGGGCGCGTTATTTATCG
GAGTTGCAGTTGCGCCCGCGAACGACATTTATAATGAACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACC
GTGGTGTTCGTTTCCAAAAAGGGGTTGCAAAAAATTTTGAACGTGCAAAAAAAGCTCCCAATCATCCAAAAAATTAT
TATCATGGATTCTAAAACGGATTACCAGGGATTTCAGTCGATGTACACGTTCGTCACATCTCATCTACCTCCCGGTT
TTAATGAATACGATTTTGTGCCAGAGTCCTTCGATAGGGACAAGACAATTGCACTGATCATGAACTCCTCTGGATCT
ACTGGTCTGCCTAAAGGTGTCGCTCTGCCTCATAGAACTGCCTGCGTGAGATTCTCGCATGCCAGAGATCCTATTTT
TGGCAATCAAATCATTCCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACAC

TCGGATATTTGATATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTGTTTCTGAGGAGCCTTCAG
GATTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCTATTCTCCTTCTTCGCCAAAAGCACTCTGATTGACAAATA
CGATTTATCTAATTTACACGAAATTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCCAAGA
GGTTCCATCTGCCAGGTATCAGGCAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTGATTACACCCGAGGGG
GATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAAGCGAAGGTTGTGGATCTGGATACCGGGAAAAC
GCTGGGCGTTAATCAAAGAGGCGAACTGTGTGTGAGAGGTCCTATGATTATGTCCGGTTATGTAAACAATCCGGAAG
CGACCAACGCCTTGATTGACAAGGATGGATGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAACACTTC
TTCATCGTTGACCGCCTGAAGTCTCTGATTAAGTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCTT
GCTCCAACACCCCAACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGACGATGACGCCGGTGAACTTCCCGCCGCCG
TTGTTGTTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGATCGTGGATTACGTCGCCAGTCAAGTAACAACCGCG
AAAAAGTTGCGCGGAGGAGTTGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGCAAGAAAAAT
CAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGTGTAAGCGGCCGCACCGGTCAGACATGATAAGA
TACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTAT
TGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTC
AGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAGCCGGCTTAATTCTTAAGTTAAT
TTTAATTAAGTCGACGATGTAGGTCACGGTCTCGAAGCCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGG
GCGCGTACTCCACCTCACCCATCTGGTCCATCATGATGAACGGGTCGAGGTGGCGGTAGTTGATCCCGGCGAACGCG
CGGCGCACCGGGAAGCCCTCGCCCTCGAAACCGCTGGGCGCGGTGGTCACGGTGAGCACGGGACGTGCGACGGCGTC
GGCGGGTGCGGATACGCGGGGCAGCGTCAGCGGGTTCTCGACGGTCACGGCGGGCATGTCGACAGATCTGACAATGT
TCAGTGCAGAGACTCGGCTACGCCTCGTGGACTTTGAAGTTGACCAACAATGTTTATTCTTACCTCTAATAGTCCTC
TGTGGCAAGGTCAAGATTCTGTTAGAAGCCAATGAAGAACCTGGTTGTTCAATAACATTTTGTTCGTCTAATATTTC
ACTACCGCTTGACGTTGGCTGCACTTCATGTACCTCATCTATAAACGCTTCTTCTGTATCGCTCTGGACGTCATCTT
CACTTACGTGATCTGATATTTCACTGTCAGAATCCTCACCAACAAGCTCGTCATCGCTTTGCAGAAGAGCAGAGAGG
ATATGCTCATCGTCTAAAGAACTACCCATTTTATTATATATTAGTCACGATATCTATAACAAGAAAATATATATATA
ATAAGTTATCACGTAAGTAGAACATGAAATAACAATATAATTATCGTATGAGTTAAATCTTAAAAGTCACGTAAAAG
ATAATCATGCGTCATTTTGACTCACGCGGTCGTTATAGTTCAAAATCAGTGACACTTACCGCATTGACAAGCACGCC
TCACGGGAGCTCCAAGCGGCGACTGAGATGTCCTAAATGCACAGCGACGGATTCGCGCTATTTAGAAAGAGAGAGCA
ATATTTCAAGAATGCATGCGTCAATTTTACGCAGACTATCTTTCTAGGGTTAAAAAAGATTTGCGCTTTACTCGACC
TAAACTTTAAACACGTCATAGAATCTTCGTTTGACAAAAACCACATTGTGGCCAAGCTGTGTGACGCGACGCGCGCT
AAAGAATGGCAAACCAAGTCGCGCGAGCGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCGTAATCATGT
CATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAA
GCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAA

References

Altschul, SF, TL Madden, AA Schäffer, J Zhang, Z Zhang, W Miller, DJ Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25:3389-3402.

Ambros, V. 2004. The functions of animal microRNAs. Nature 431:350-355.

Amrein, H, M Gorman, R Nöthiger. 1988. The sex-determining gene tra-2 of Drosophila encodes a putative RNA binding protein. Cell 55:1025-1035.

Andachi, Y. 2008. A novel biochemical method to identify target genes of individual microRNAs: Identification of a new Caenorhabditis elegans let-7 target. RNA 14:2440-2451.

Antonio J, G. 2010. microRNAs, the cell's Nepenthe: clearing the past during the maternal-to-zygotic transition and cellular reprogramming. Current Opinion in Genetics & Development 20:369-375.

Aravind, L, V Anantharaman, S Balaji, MM Babu, LM Iyer. 2005. The many faces of the helix-turn-helix domain: Transcription regulation and beyond⋆. FEMS microbiology reviews 29:231-262.

Arensburger, P, K Megy, RM Waterhouse, et al. 2010. Sequencing of Culex quinquefasciatus Establishes a Platform for Mosquito Comparative Genomics. Science 330:86-88.

Bachtrog, D. 2006. A dynamic view of sex chromosome evolution. Current Opinion in Genetics & Development 16:578-585.

Bail, S, M Swerdel, H Liu, X Jiao, LA Goff, RP Hart, M Kiledjian. 2010. Differential regulation of microRNA stability. RNA 16:1032-1039.

Baker, DA, S Russell. 2011. Role of Testis-Specific Gene Expression in Sex-Chromosome Evolution of Anopheles gambiae. Genetics 189:1117-1120.

Baker, RH, RK Sakai. 1979. Triploids and male determination in the mosquito, Anopheles culicifacies. Journal of Heredity 70:345-346.

Barton, NH, B Charlesworth. 1998. Why Sex and Recombination? Science 281:1986-1990.

Behura, SK, M Haugen, E Flannery, J Sarro, CR Tessier, DW Severson, M Duman-Scheel. 2011. Comparative Genomic Analysis of <italic>Drosophila melanogaster</italic> and Vector Mosquito Developmental Genes. PLoS ONE 6:e21504.

Benedict, MQ, AS Robinson. 2003. The first releases of transgenic mosquitoes: an argument for the sterile insect technique. Trends in Parasitology 19:349-355.

Bentwich, I. 2005. Prediction and validation of microRNAs and their targets. FEBS Letters 579:5904-5910.

Bernardo Carvalho, A, LB Koerich, AG Clark. 2009. Origin and evolution of Y chromosomes: Drosophila tales. Trends in Genetics 25:270-277.

Bernardo Carvalho, A, LB Koerich, AG Clark. 2009. Origin and evolution of Y chromosomes: Drosophila tales. Trends in Genetics 25:270-277.

Biedler, J, Z Tu. 2010. Evolutionary analysis of the kinesin light chain genes in the yellow fever mosquito Aedes aegypti: gene duplication as a source for novel early zygotic genes. BMC evolutionary biology 10:206.

Biedler, JK, W Hu, H Tae, Z Tu. 2012. Identification of Early Zygotic Genes in the Yellow Fever Mosquito <italic>Aedes aegypti</italic> and Discovery of a Motif Involved in Early Zygotic Genome Activation. PLoS ONE 7:e33933.

Biedler, JK, Z Tu. 2007. The Juan non-LTR retrotransposon in mosquitoes: genomic impact,

vertical transmission and indications of recent and widespread activity. BMC evolutionary biology 7:112.
Biemar, F, R Zinzen, M Ronshaugen, V Sementchenko, JR Manak, MS Levine. 2005. Spatial regulation of microRNA gene expression in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America 102:15907-15911.
Boggs, RT, P Gregor, S Idriss, JM Belote, M McKeown. 1987. Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the transformer gene. Cell 50:739-747.
Boyd, MF. 1938. On the comparative susceptibility of certain strains of neaarctic and neotropical anophelines to certain strains of *P. vivax* and *P. falciparum* from the same regions. Am J Trop Med:157-168.
Bridges, C. 1921. Triploid intersexes in Drosophila. Science NS:252-254.
Bull, JJ. 1983. Evolution of sex determining mechanisms: The Benjamin/Cummings Publishing Company, Inc.
Bushati, N, A Stark, J Brennecke, SM Cohen. 2008. Temporal Reciprocity of miRNAs and Their Targets during the Maternal-to-Zygotic Transition in Drosophila. Current Biology 18:501-506.
Bushland, R, A Lindquist, E Knipling. 1955. Eradication of screw-worms through release of sterilized males. Science 122:287-288.
Campbell, RE, O Tour, AE Palmer, PA Steinbach, GS Baird, DA Zacharias, RY Tsien. 2002. A monomeric red fluorescent protein. Proceedings of the National Academy of Sciences 99:7877-7882.
Carvalho, AB, BA Dobo, MD Vibranovski, AG Clark. 2001. Identification of five new genes on the Y chromosome of Drosophila melanogaster. Proceedings of the National Academy of Sciences 98:13225-13230.
Carvalho, AB, BP Lazzaro, AG Clark. 2000. Y chromosomal fertility factors kl-2 and kl-3 of Drosophila melanogaster encode dynein heavy chain polypeptides. Proceedings of the National Academy of Sciences 97:13239-13244.
Carvalho, AB, M Vibranovski, J Carlson, S Celniker, R Hoskins, G Rubin, G Sutton, M Adams, E Myers, A Clark. 2003. Y Chromosome and Other Heterochromatic Sequences of the Drosophila Melanogaster Genome: How Far can we go? Genetica 117:227-237.
Catteruccia, F, A Crisanti, EA Wimmer. 2009. Transgenic technologies to induce sterility. Malar J 8:S7.
CDC. Centers for Disease Control.

Charlesworth, B, A Betancourt, V Kaiser, I Gordo. 2009. Genetic recombination and molecular evolution. Cold Spring Harbor symposia on quantitative biology: Cold Spring Harbor Laboratory Press. p. 177-186.

Charlesworth, B, D Charlesworth. 2000. The degeneration of Y chromosomes. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 355:1563-1572.
Charlesworth, D, JE Mank. 2010. The Birds and the Bees and the Flowers and the Trees: Lessons from Genetic Mapping of Sex Determination in Plants and Animals. Genetics 186:9-31.
Chen, C-H, H Huang, CM Ward, JT Su, LV Schaeffer, M Guo, BA Hay. 2007. A synthetic

maternal-effect selfish genetic element drives population replacement in Drosophila. Science 316:597-600.
Clements, AN. 1992. The biology of mosquitoes. London: Chapman & Hall.
Clemons, A, M Haugen, C Le, A Mori, M Tomchaney, DW Severson, M Duman-Scheel. 2011. siRNA-Mediated Gene Targeting in *Aedes aegypti* Embryos Reveals That *Frazzled* Regulates Vector Mosquito CNS Development. PLoS ONE 6:e16730.
Collins, WE, JS Sullivan, D Nace, T Williams, JJ Sullivan, GG Galland, KK Grady, A Bounngaseng. 2002. Experimental infection of Anopheles farauti with different species of Plasmodium. Journal of Parasitology 88:295-298.
Cook, JM. 1993. Sex determination in the Hymenoptera: a review of models and evidence. HEREDITY-LONDON- 71:421-421.
Cronmiller, C, TW Cline. 1987. The Drosophila sex determination gene daughterless has different functions in the germ line versus the soma. Cell 48:479-487.
Dauwalder, B, F Amaya-Manzanares, W Mattox. 1996. A human homologue of the Drosophila sex determination factor transformer-2 has conserved splicing regulatory functions. Proceedings of the National Academy of Sciences 93:9004-9009.
David P, B. 2004. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell 116:281-297.
De Renzis, S, O Elemento, S Tavazoie, EF Wieschaus. 2007. Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the <italic>Drosophila</italic> Embryo. PLoS Biol 5:e117.
Du, G, J Yonekubo, Y Zeng, M Osisami, MA Frohman. 2006. Design of expression vectors for RNA interference based on miRNAs and RNA splicing. FEBS Journal 273:5421-5427.
Enright, A, B John, U Gaul, T Tuschl, C Sander, D Marks. 2003. MicroRNA targets in Drosophila. Genome Biology 5:R1.
Enserink, M. 2010. GM mosquito trial alarms opponents, strains ties in Gates-funded project. Science 330:1030-1031.
Erickson, JW, TW Cline. 1998. Key aspects of the primary sex determination mechanism are conserved across the genus Drosophila. Development 125:3259-3268.
Fraccaro, M, L Tiepolo, U Laudani, A Marchi, SD Jayakar. 1977. Y chromosome controls mating behaviour on Anopheles mosquitoes. Nature 265:326-328.
Gabrieli, P, A Falaguerra, P Siciliano, L Gomulski, F Scolari, A Zacharopoulou, G Franz, A Malacrida, G Gasperi. 2010. Sex and the single embryo: early deveopment in the Mediterranean fruit fly, Ceratitis capitata. BMC developmental biology 10:12.
Gailey, DA, J-C Billeter, JH Liu, F Bauzon, JB Allendorfer, SF Goodwin. 2006. Functional Conservation of the fruitless Male Sex-Determination Gene Across 250 Myr of Insect Evolution. Molecular Biology and Evolution 23:633-643.
Gaunt, MW, MA Miles. 2002. An Insect Molecular Clock Dates the Origin of the Insects and Accords with Palaeontological and Biogeographic Landmarks. Molecular Biology and Evolution 19:748-761.
Gepner, J, TS Hays. 1993. A fertility region on the Y chromosome of Drosophila melanogaster encodes a dynein microtubule motor. Proceedings of the National Academy of Sciences 90:11132-11136.
Goltsev, Y, N Fuse, M Frasch, RP Zinzen, G Lanzaro, M Levine. 2007. Evolution of the dorsal-ventral patterning network in the mosquito, Anopheles gambiae. Development 134:2415-2424.

Goltsev, Y, W Hsiong, G Lanzaro, M Levine. 2004. Different combinations of gap repressors for common stripes in Anopheles and Drosophila embryos. Developmental Biology 275:435-446.
Graves, JAM. 2006. Sex Chromosome Specialization and Degeneration in Mammals. Cell 124:901-914.
Graves, JAM. 2008. Weird Animal Genomes and the Evolution of Vertebrate Sex and Sex Chromosomes. Annual Review of Genetics 42:565-586.
Gubler, DJ. 1998. Dengue and dengue hemorrhagic fever. Clinical microbiology reviews 11:480-496.
Gvozdev, VA, GL Kogan, LA Usakin. 2005. The Y chromosome as a target for acquired and amplified genetic material in evolution. BioEssays 27:1256-1262.
Gwadz, R. 1972. Neuro-hormonal regulation of sexual receptivity in female< i> Aedes aegypti</i>. Journal of Insect Physiology 18:259-266.
Hackstein, JHP, R Hochstenbach. 1995. The elusive fertility genes of Drosophila: the ultimate haven for selfish genetic elements. Trends in Genetics 11:195-200.
Haugen, M, E Flannery, M Tomchaney, A Mori, SK Behura, DW Severson, M Duman-Scheel. 2011. Semaphorin-1a Is Required for <italic>Aedes aegypti</italic> Embryonic Nerve Cord Development. PLoS ONE 6:e21694.
Helinski, M, AG Parker, BG Knols. 2009. Radiation biology of mosquitoes. Malar J 8:S6.
Hindson, BJ, KD Ness, DA Masquelier, et al. 2011. High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number. Analytical Chemistry 83:8604-8610.
Hodgkin, J. 1983. Two types of sex determination in a nematode.
Holt, RAGM SubramanianA Halpern, et al. 2002. The Genome Sequence of the Malaria Mosquito Anopheles gambiae. Science 298:129-149.
Hooper, SD, S Boue, R Krause, LJ Jensen, CE Mason, M Ghanim, KP White, EEM Furlong, P Bork. 2007. Identification of tightly regulated groups of genes during Drosophila melanogaster embryogenesis. Mol Syst Biol 3.
Huelsenbeck, JP, F Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.
Isik, M, HC Korswagen, E Berezikov. 2010. Expression patterns of intronic microRNAs in Caenorhabditis elegans. Silence 1:5.
James, S, W Nicol, P Shute. 1932. A study of induced malignant tertian malaria. Proceedings of the Royal Society of Medicine 25:1153.
Jost, E., Mameli, M. 1972. DNA Content in Nine Species of Nematocera with Special Reference to the Sibling Species of the Anopheles Maculipennis Group and the Culex Pipiens Group. Chromosoma 37.2: 201-08.
Juhn, J, AA James. 2006. oskar gene expression in the vector mosquitoes, Anopheles gambiae and Aedes aegypti. Insect Molecular Biology 15:363-372.
Kaiser, VB, D Bachtrog. 2010. Evolution of Sex Chromosomes in Insects. Annual Review of Genetics 44:91-112.
Kennison, JA. 1981. THE GENETIC AND CYTOLOGICAL ORGANIZATION OF THE Y CHROMOSOME OF DROSOPHILA MELANOGASTER. Genetics 98:529-548.
Koerich, LB, X Wang, AG Clark, AB Carvalho. 2008. Low conservation of gene content in the Drosophila Y chromosome. Nature 456:949-951.
Kopp, A, A Frank, J Fu. 2006. Historical biogeography of Drosophila simulans based on Y-

chromosomal sequences. Molecular Phylogenetics and Evolution 38:355-362.
Krzywinski, J, NJ Besansky. 2003. MOLECULAR SYSTEMATICS OF ANOPHELES: From Subgenera to Subpopulations. Annual Review of Entomology 48:111-139.
Krzywinski, J, M Chrystal, N Besansky. 2006. Gene Finding on the Y: Fruitful Strategy in Drosophila does not Deliver in Anopheles. Genetica 126:369-375.
Krzywinski, J, DR Nusskern, MK Kern, NJ Besansky. 2004. Isolation and Characterization of Y Chromosome Sequences From the African Malaria Mosquito Anopheles gambiae. Genetics 166:1291-1302.
Kurane, I, F Ennis. 1992. Immunity and immunopathology in dengue virus infections. Seminars in immunology. p. 121.

Kutter, C, P Svoboda. 2008. miRNA, siRNA, piRNA: Knowns of the unknown. RNA Biology 5:181-188.
Langmead, B. 2002. Aligning Short Sequencing Reads with Bowtie. Current Protocols in Bioinformatics: John Wiley & Sons, Inc.
Lau, NC, LP Lim, EG Weinstein, DP Bartel. 2001. An Abundant Class of Tiny RNAs with Probable Regulatory Roles in Caenorhabditis elegans. Science 294:858-862.
Lee, Y, M Kim, J Han, K-H Yeom, S Lee, SH Baek, VN Kim. 2004. MicroRNA genes are transcribed by RNA polymerase II. EMBO J 23:4051-4060.
Lemos, B, LO Araripe, DL Hartl. 2008. Polymorphic Y Chromosomes Harbor Cryptic Variation with Manifold Functional Consequences. Science 319:91-93.
Li, F, SP Vensko II, EJ Belikoff, MJ Scott. 2013. Conservation and Sex-Specific Splicing of the transformer Gene in the Calliphorids Cochliomyia hominivorax, Cochliomyia macellaria and Lucilia sericata. PLoS ONE 8:e56303.
Li, S, E Mead, S Liang, Z Tu. 2009. Direct sequencing and expression analysis of a large number of miRNAs in Aedes aegypti and a multi-species survey of novel mosquito miRNAs. BMC Genomics 10:581.
Liang, H-L, C-Y Nien, H-Y Liu, MM Metzstein, N Kirov, C Rushlow. 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature 456:400-403.
Lim, LP, NC Lau, P Garrett-Engele, A Grimson, JM Schelter, J Castle, DP Bartel, PS Linsley, JM Johnson. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433:769-773.
Lin, S-S, H-W Wu, SF Elena, K-C Chen, Q-W Niu, S-D Yeh, C-C Chen, N-H Chua. 2009. Molecular Evolution of a Viral Non-Coding Sequence under the Selective Pressure of amiRNA-Mediated Silencing. PLoS Pathog 5:e1000312.
Lott, SE, JE Villalta, GP Schroth, S Luo, LA Tonkin, MB Eisen. 2011. Noncanonical Compensation of Zygotic X Transcription in Early <italic>Drosophila melanogaster</italic> Development Revealed through Single-Embryo RNA-Seq. PLoS Biol 9:e1000590.
Lund, E, M Liu, RS Hartley, MD Sheets, JE Dahlberg. 2009. Deadenylation of maternal mRNAs mediated by miR-427 in Xenopus laevis embryos. RNA 15:2351-2363.
Maher, C, L Stein, D Ware. 2006. Evolution of Arabidopsis microRNA families through duplication events. Genome Research 16:510-519.
Malone, JH, B Oliver. 2008. The sex chromosome that refused to die. BioEssays 30:409-411.
Marín, I, BS Baker. 1998. The Evolutionary Dynamics of Sex Determination. Science 281:1990-

1994.
Marshall, JM, CE Taylor. 2009. Malaria Control with Transgenic Mosquitoes. PLoS Med 6:e1000020.
Martin, R, P Smibert, A Yalcin, DM Tyler, U Schäfer, T Tuschl, EC Lai. 2009. A Drosophila pasha Mutant Distinguishes the Canonical MicroRNA and Mirtron Pathways. Molecular and Cellular Biology 29:861-870.
Marygold, S*ea*. 2013. FlyBase: Improvements to the bibliography. Nucleic Acids Research:D751-757.
Massari, ME, C Murre. 2000. Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms. Molecular and Cellular Biology 20:429-440.
Megy, K*ea*. 2012. VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. Nucleic Acids Research:D729-734.
Miller, BR, RS Nasci, MS Godsey, HM Savage, JJ Lutwama, RS Lanciotti, CJ Peters. 2000. First field evidence for natural vertical transmission of West Nile virus in Culex univittatus complex mosquitoes from Rift Valley province, Kenya. The American Journal of Tropical Medicine and Hygiene 62:240-246.
miRbase. miRbase.

Miska, EA. 2005. How microRNAs control cell division, differentiation and death. Current Opinion in Genetics & Development 15:563-568.
Nene, V, JR Wortman, D Lawson, et al. 2007. Genome Sequence of Aedes aegypti, a Major Arbovirus Vector. Science 316:1718-1723.
Ng'habi, KR, A Horton, BGJ Knols, GC Lanzaro. 2007. A New Robust Diagnostic Polymerase Chain Reaction for Determining the Mating Status of Female Anopheles gambiae Mosquitoes. The American Journal of Tropical Medicine and Hygiene 77:485-487.
Ni, J-Q, R Zhou, B Czech, L-P Liu, L Holderbaum, D Yang-Zhou, H-S Shim, R Tao, D Handler, P Karpowicz. 2011. A genome-scale shRNA resource for transgenic RNAi in Drosophila. Nature methods 8:405-407.
Obernosterer, G, PJF Leuschner, M Alenius, J Martinez. 2006. Post-transcriptional regulation of microRNA expression. RNA 12:1161-1167.
Oshaghi, MA, F Yaaghoobi, MR Abaie. 2006. Pattern of mitochondrial DNA variation between and within Anopheles stephensi (Diptera: Culicidae) biological forms suggests extensive gene flow. Acta Tropica 99:226-233.
Papatsenko, D, M Levine, Y Goltsev. 2011. Clusters of Temporal Discordances Reveal Distinct Embryonic Patterning Mechanisms in <italic>Drosophila</italic> and <italic>Anopheles</italic>. PLoS Biol 9:e1000584.
Parkhurst, SM, PM Meneely. 1994. Sex determination and dosage compensation: lessons from flies and worms. Science-AAAS-Weekly Paper Edition-including Guide to Scientific Information 264:924-932.
Patz, JA, PR Epstein, TA Burke, JM Balbus. 1996. Global climate change and emerging infectious diseases. JAMA: the journal of the American Medical Association 275:217-223.
Pauli, A, JL Rinn, AF Schier. 2011. Non-coding RNAs as regulators of embryogenesis. Nat Rev Genet 12:136-149.
Pease, JB, MW Hahn. 2012. Sex chromosomes evolved from independent ancestral linkage groups in winged insects. Molecular Biology and Evolution 29:1645-1653.

Pillai, RS, SN Bhattacharyya, W Filipowicz. 2007. Repression of protein synthesis by miRNAs: how many mechanisms? Trends in Cell Biology 17:118-126.
Raghavendra, K, TK Barik, P Sharma, RM Bhatt, HC Srivastava, U Sreehari, AP Dash. 2011. Chlorfenapyr: a new insecticide with novel mode of action can control pyrethroid resistant malaria vectors. Malar J 10:16.
Rai, KS, WC Black IV. 1999. 1 Mosquito Genomes: Structure, Organization, and Evolution. Advances in genetics 41:1-33.
Rajewsky, N. 2006. microRNA target predictions in animals. Nat Genet.
Raminani, L, E Cupp. 1975. Early embryology of< i> Aedes aegypti</i>(L.)(Diptera: Culicidae). International Journal of Insect Morphology and Embryology 4:517-528.
Rens, W, F Grützner, PCM O'Brien, H Fairclough, JAM Graves, MA Ferguson-Smith. 2004. Resolution and evolution of the duck-billed platypus karyotype with an X1Y1X2Y2X3Y3X4Y4X5Y5 male sex chromosome constitution. Proceedings of the National Academy of Sciences of the United States of America 101:16257-16261.
Rice, WR. 1987. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. Genetics 116:161-167.
Rice, WR. 1996. Evolution of the Y sex chromosome in animals. Bioscience 46:331-343.
Robinson, A, V Pham. 1987. Cytological, linkage and insecticide studies on a genetic sexing line in Anopheles stephensi Liston. Heredity 58:95-101.
Robinson, AS. 1983. Sex-ratio manipulation in relation to insect pest control. Annual Review of Genetics 17:191-214.
Ronquist, F, JP Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.
Saccone, G, A Pane, L Polito. 2002. Sex determination in flies, fruitflies and butterflies. Genetica 116:15-23.
Saito, K, A Ishizuka, H Siomi, MC Siomi. 2005. Processing of Pre-microRNAs by the Dicer-1–Loquacious Complex in <italic>Drosophila</italic> Cells. PLoS Biol 3:e235.
Sato, K, M Hamada, K Asai, T Mituyama. 2009. CentroidFold: a web server for RNA secondary structure prediction. Nucleic Acids Research 37:W277-W280.
Scali, C, F Catteruccia, Q Li, A Crisanti. 2005. Identification of sex-specific transcripts of the Anopheles gambiae doublesex gene. Journal of experimental biology 208:3701-3709.
Schier, AF. 2007. The Maternal-Zygotic Transition: Death and Birth of RNAs. Science 316:406-407.
Schliekelman, P, S Ellner, F Gould. 2005. Pest Control by Genetic Manipulation of Sex Ratio. Journal of Economic Entomology 98:18-34.
Sefton, L, JR Timmer, Y Zhang, F Beranger, TW Cline. 2000. An extracellular activator of the Drosophila JAK/STAT pathway is a sex-determination signal element. Nature 405:970-973.
Sengul, MS, Z Tu. 2010. Expression Analysis and Knockdown of Two Antennal Odorant-Binding Protein Genes in Aedes aegypti. Journal of Insect Science 10:1-18.
Sharakhova, MV, A Xia, SI McAlister, IV Sharakhov. 2006. A Standard Cytogenetic Photomap for the Mosquito Anopheles stephensi (Diptera: Culicidae): Application for Physical Mapping. Journal of Medical Entomology 43:861-866.
Shearman, D. 2002. The evolution of sex determination systems in dipteran insects other than Drosophila. Genetica 116:25-43.
Simmons, CP, M Wolbers, MN Nguyen, J Whitehorn, PY Shi, P Young, R Petric, VVC Nguyen,

J Farrar, B Wills. 2012. Therapeutics for Dengue: Recommendations for Design and Conduct of Early-Phase Clinical Trials. PLoS Negl Trop Dis 6:e1752.
Sinclair, AH, P Berta, MS Palmer, JR Hawkins, BL Griffiths, MJ Smith, JW Foster, A-M Frischauf, R Lovell-Badge, PN Goodfellow. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346:240-244.
Southwood, T, G Murdie, M Yasuno, RJ Tonn, P Reader. 1972. Studies on the life budget of Aedes aegypti in Wat Samphaya, Bangkok, Thailand. Bulletin of the World Health Organization 46:211.
Steinemann, S, M Steinemann. 2005. Y chromosomes: born to be destroyed. BioEssays 27:1076-1083.
Subbarao, SK, K Vasantha, T Adak, VP Sharma, CF Curtis. 1987. Egg-float ridge number in Anopheles stephensi: ecological variation and genetic analysis. Medical and Veterinary Entomology 1:265-271.
Suman, DS, B Parashar, S Prakash. 2008. Morphological Sexual Dimorphism in Three Species of Anopheline Mosquito Larvae. Journal of the American Mosquito Control Association 24:308-310.
Sumitani, M, K Kasashima, DS Yamamoto, K Yagi, M Yuda, H Matsuoka, S Yoshida. 2013. Reduction of malaria transmission by transgenic mosquitoes expressing an antisporozoite antibody in their salivary glands. Insect Molecular Biology 22:41-51.
Tanzer, A, PF Stadler. 2004. Molecular Evolution of a MicroRNA Cluster. Journal of Molecular Biology 339:327-335.
Thomas, DD, CA Donnelly, RJ Wood, LS Alphey. 2000. Insect population control using a dominant, repressible, lethal genetic system. Science 287:2474-2476.
Thompson, JD, DG Higgins, TJ Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22:4673-4680.
Timoshevskiy, V., Sharma, A., Sharakhov, I., Sharakhova, M., 2012. Fluorescent in situ Hybridization on Mitotic Chromosomes of Mosquitoes. Journal of visualized experiments: JoVE. 67
Underhill, PA, T Kivisild. 2007. Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations. Annual Review of Genetics 41:539-564.
Verhulst, EC, L van de Zande, LW Beukeboom. 2010. Insect sex determination: it all evolves around transformer. Current Opinion in Genetics & Development 20:376-383.
Vibranovski, MD, LB Koerich, AB Carvalho. 2008. Two new Y-linked genes in Drosophila melanogaster. Genetics 179:2325-2327.
Walser, CB, HD Lipshitz. 2011. Transcript clearance during the maternal-to-zygotic transition. Current Opinion in Genetics & Development 21:431-443.
WHO. World Health Organization.

Wieschaus, E. 1996. Embryonic Transcription and the Control of Developmental Pathways. Genetics 142:5-10.
Wilke, ABB, AdC Gomes, D Natal, MT Marrelli. 2009. Control of vector populations using genetically modified mosquitoes. Revista de Saúde Pública 43:869-874.
Wise de Valdez, MR, D Nimmo, J Betz, H-F Gong, AA James, L Alphey, WC Black. 2011. Genetic elimination of dengue vector mosquitoes. Proceedings of the National Academy

of Sciences 108:4772-4775.
Xia, A, MV Sharakhova, SC Leman, Z Tu, JA Bailey, CD Smith, IV Sharakhov. 2010. Genome Landscape and Evolutionary Plasticity of Chromosomes in Malaria Mosquitoes. PLoS ONE 5:e10592.
Yang, J-S, MD Phillips, D Betel, P Mu, A Ventura, AC Siepel, KC Chen, EC Lai. 2011. Widespread regulatory activity of vertebrate microRNA* species. RNA 17:312-326.
Ying, S-Y, S-L Lin. 2004. Intron-derived microRNAs—fine tuning of gene functions. Gene 342:25-28.
Zhou, J, TB Sackton, L Martinsen, B Lemos, TH Eickbush, DL Hartl. 2012. Y chromosome mediates ribosomal DNA silencing and modulates the chromatin state in Drosophila. Proceedings of the National Academy of Sciences 109:9941-9946.
Zhou, Q, D Bachtrog. 2012. Sex-specific adaptation drives early sex chromosome evolution in Drosophila. Science 337:341-345.