Identifying Novel Contributors to RNA Interference in *Aedes aegypti*

Angela Paige Saadat

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

Master of Science in Life Sciences
In
Entomology

Zach N. Adelman, Chair
Ina Hoeschele
Kevin M. Myles

August 11, 2015
Blacksburg, VA

Keywords: RNAi, siRNA, *Aedes aegypti*, RNA-seq, QTL, fluorescence microscopy

Identifying Novel Contributors to RNA Interference in *Aedes aegypti*

Angela Paige Saadat

Abstract

Aedes aegypti is an important vector of human pathogens including the viruses yellow fever, dengue and chikungunya. The small interfering RNA (siRNA) pathway is a critical immune response for controlling viral replication in *Aedes aegypti*. The goal of this research is to identify components of the *Aedes aegypti* genome that influence this pathway.

A transgenic mosquito strain that reports the status of the siRNA pathway via enhanced green fluorescent protein (EGFP) intensity was employed to differentiate silencing abilities among individuals. Extreme EGFP expression phenotypes, representing efficient and poor silencing abilities, were enriched over five generations.

Transcriptome sequencing and analyses were performed from pools of individuals from each enriched phenotype, revealing potential RNAi contributors. 1,120 transcripts were significantly different (FDR<0.0001) among the extreme phenotypes.

Four genes were chosen, amplified, sequenced for SNP analysis. These analyses were performed on samples obtained by crossing enriched, extreme phenotype F_0 individuals, intercrossing their progeny, then selecting individuals representing the extreme phenotypes from the F_2 population. Though further verification is needed, findings from these analyses imply the regions of *Aedes aegypti*, Liverpool strain (AAEL) gene identifiers AAEL005026, AAEL013438 and AAEL011704 amplified do not contribute to the two extreme, opposite RNAi silencing in the sensor strain used here. SNP analyses of AAEL000817 indicate this gene either influences

extreme RNAi phenotypes or is closely linked to a gene(s) that contributes to RNAi in *Aedes aegypti*.

The 1,120 genes identified can be validated or eliminated as potential targets in the quest to mitigate the impact of *Aedes aegypti*.

Acknowledgements

The work within was a collaborative effort of Dr. Zach N. Adelman, Dr. Kevin Myles, Dr. Ina Hoeschele, Michelle Anderson, Dr. Sanjay Basu, Heather Eggleston and other members of the Adelman laboratory. Matthew O'Keefe and J.T. Thomas, members of Virginia Tech's Laboratory for Interdisciplinary Statistical Analysis (LISA) program, suggested fitting the enrichment values into an ANCOVA model and demonstrated how to do so in JMP Pro 11; Matthew O'Keefe suggested the scatter plot graph of the EGFP data as well.

This work was funded by NIH grant AI085091.

Special thanks to Esmat Ansari-Lari and Shokrollah Saadat for their infinite kindness, support and countless cups of *chai*.

Table of Contents

List of Figures

List of Tables

Chapter 1

Literature Review

Aedes aegypti

The *Aedes aegypti* mosquito is a significant species of study. *Aedes aegypti* is the primary vector of yellow fever virus, dengue virus serotypes and chikungunya virus (CDC, Dengue and the *Aedes aegypti* mosquito, 2012). *Aedes aegypti* is also capable of vectoring other human pathogens such as West Nile virus, Zika virus and certain filarial worm species. These pathogens are perpetuated in tropical and subtropical regions in conjunction with *Aedes aegypti* populations and cause preventable suffering, mortality and economic loss (Reviewed by Torres and Castro, 2007). A deeper understanding of the relationship between this disease vector and the pathogens it transmits is integral to disease control.

Aedes aegypti females preferentially feed on humans; this preference likely evolved through a series of events including habitat desiccation, domestication and transplantation. The *Aedes aegypti* mosquito is believed to have originated in the once sylvatic African Sahara thousands of years ago. This ancestral mosquito likely lived in tree holes and acquired blood meals from non-human vertebrates. Commensalism with humans likely transpired as reliable water sources were sought and found in human settlements. As a result of this proximity, a preference for human blood meals evolved. Oscillation between non-human and human blood meals likely allowed viral zoonoses to develop. The domesticated subspecies, *Aedes aegypti* aegypti, was then likely inadvertently deposited in North and South America with trade and human trafficking. Genetic analyses indicate that *Aedes aegypti* was then spread, again through trade, in the nineteenth century to tropical and subtropical regions of Asia and Australia. With this domestication, worldwide spread and colonization, the *Aedes aegypti* mosquito evolved a

relationship between the mosquito and the pathogens it vectors to humans. Thus, today *Aedes aegypti* can be found on every continent, save Antarctica, inadvertently perpetuating arboviral disease cycles (Brown et al., 2011).

Aedes aegypti is known commonly as the yellow fever mosquito. Walter Reed and his colleagues demonstrated that this mosquito transmitted yellow fever in the early 1900s (Reed et al., 1900). The yellow fever virus attacks the liver and gets its namesake from the resulting jaundice. Symptoms can include myalgia, fever and hemorrhaging from eyes, nose and mouth. Even today the prognosis following infection by the virus is bleak; indeed, a five percent mortality rate has been estimated in endemic areas (Davis, 2009). Before the development of an effective vaccine, yellow fever epidemics were notoriously deadly, with mortality rates in the tens of thousands. After many failed attempts by many researchers an effective vaccine, 17D, was developed in the 1930s (Reviewed by Frierson, 2010).Though the established vaccine is efficacious for the lifespan of the recipient, it is not without significant cost and other limitations, including contraindication with pregnancy. Furthermore, there is no treatment available to date, other than palliative measures, against the virus. This virus continues to significantly impact Africa and South America (Black et al., 2002). Interestingly, yellow fever cases have not been reported in Asia nor the Pacific Islands.

Dengue is the most significant arbovirus worldwide and its four serotypes are transmitted primarily by the *Aedes aegypti* mosquito. In their 2012 strategy for dengue prevention, the World Health Organization (WHO) estimated that 50-100 million dengue infections occur each year, seeing a 30-fold increase in fifty years (WHO Global strategy for dengue prevention and control, 2012). The dengue burden extends to all continents except Antarctica (WHO, Global strategy for dengue prevention and control, 2012). Dengue fever is caused by any of four virus serotypes and

presents painful symptoms similar to those of yellow fever. A second dengue virus infection in an individual can manifest with encephalitis and hemorrhagia, where infection is exacerbated by antibody-dependent enhancement (Ruklanthi et al., 2014). Davis estimates that of the tens of millions of people infected each year, about five percent present with brain and lung hemorrhages (Davis, 2009). Shepard et al developed a model that estimates dengue illnesses cost 2.1 billion USD per year in the Americas alone, excluding even the costs of insect control (Shepard et al., 2011). Lack of an effective vaccine, its prevalence, cost and mortality make dengue virus serotypes the most important group of arboviruses worldwide (WHO Global strategy for dengue prevention and control, 2012).

Aedes aegypti is also recognized as a primary vector of chikungunya virus, along with *Aedes albopictus* (WHO Chikungunya, 2014). Chikungunya generally manifests symptoms such as fever and severe arthralgia, or joint pain, after around a week following a bite. de Andrade et al reported that the joint pain from a chikungunya infection can persist for months, finding a mean duration of 89 days +/- 2 days (de Andrade et al., 2010). There is no vaccine or antiviral treatment for chikungunya to date. Chikungunya is of growing concern to the United States, as autochthonous cases, or disease contracted locally, of chikungunya have been reported in Florida (CDC Chikungunya, 2014).

Countries around the world invest significant funds into vaccinations, drug therapies, education and other measures to combat these diseases. Unfortunately, no one control measure alone has proven to be a complete solution; a consortia of control measures is predicted to be the optimal approach to limiting disease transmission (Reviewed by Alphey, 2014).

Relationship between *Aedes aegypti* & the viruses it transmits

As only female mosquitoes bloodfeed, only female *Aedes aegypti* mosquitoes vector these viruses to humans. Both male and female mosquitoes have fluid feeding mouthpart morphologies, but male mosquitoes feed exclusively on plant sugars. Female mosquitoes facultatively feed on nectar, requiring components of vertebrate blood for vitellogenesis, or to produce yolk for their eggs (Christophers, 1960). It is during such bloodmeals that female mosquitoes can acquire and transmit arboviruses, or viruses that infect arthropods and are then vectored to vertebrate hosts.

A bloodmeal is a multi-step process involving complex mouthparts. First, a female mosquito detects a favorable site on a prospective host with her labellum, a fleshy pad at base of her proboscis. Next, maxillary stylets of her proboscis cut into the source, and allow the labrum, hypopharynx, mandibles and maxillae to enter the wound. The labrium folds back to allow the other structures to further probe for a bloodsource. A salivary duct runs down the hypopharynx and salivary secretions containing anticoagulant and anesthetic are inserted in the wound; it is at this point that viral pathogens are transferred from the mosquito to the new host. The female then draws host blood through her hypopharynx and into her alimentary canal (Gordon and Lumsden, 1939). From here, proteins from the bloodmeal are processed and utilized in vitellogenesis. In addition, if the blood acquired was from a host infected with an arbovirus, such as yellow fever or dengue, viral particles can propogate within the mosquito and be vectored to a new host in a future bloodmeal.

In order to vector viruses acquired from viremic bloodmeals, viral titers must be high enough in the mosquito's salivary glands to adequately infect a new host. Viral particles imbibed by a female mosquito must infect the midgut , penetrate the midgut epithelia, circulate in the

hemocoele and then penetrate and infect the mosquito's salivary glands. A mosquito with adequately infected salivary glands is able to transmit a virus through salivary secretions that are part of the bloodfeeding process (Reviewed by Black et al., 2002).

Caveats to arboviral transmission include the extrinsic incubation period and conditions of the infected host. Viral titers within the mosquito are dependent upon conditions such as time, temperature and initial inoculation levels. Given this, a mosquito can theoretically be infected but not transmit enough virus to infect a vertebrate host. Also, viral pathogenesis in the secondary host is subject to that individual's immune response (Reviewed by Black et al., 2002). Ancillary cycles may maintain arboviruses in nature. It has been demonstrated that female *Aedes aegypti* mosquitoes can transmit dengue serotypes 1-4 and yellow fever transovarially to their offspring (Aitken et al., 1979; Fontenille et al., 1997; Hull et al., 1994; Joshi et al., 1996; Khin and Than,1983; Rosen et al., 1983;). Attempts to demonstrate vertical transmission in *Aedes aegypti* with chikungunya virus were unsuccessful (Mourya et al., 1987); however, Thavara et al. noted the presence of the virus in males (Thavara et al., 2009). Since viral titers in embryos are comparably low, it could be that the methods used were not sensitive enough to detect the virus. Mavale et al. demonstrated, though, that chikungunya virus can be transmitted venereally in *Aedes aegypti* (Mavale et al., 2010).

Aedes aegypti preferentially feeds on a human host; indeed, it has been reported that over 90% of *Aedes aegypti* blood meals are acquired from humans (Scott and Takken, 2012). Exclusive feedings from human hosts is also associated with increased fitness in *Aedes agypti* (Scott et al., 1997). Furthermore, *Aedes aegypti* mosquitoes prefer to feed indoors and during the daytime, and are known to bite indoors after sundown in artificial light (CDC Dengue and the *Aedes aegypti* mosquito, 2012). *Aedes aegypti* females typically need at least one blood meal per

gonotropic cycle, or cycle of feeding and egg laying, but have been shown to feed multiple times per cycle (Scott et al., 1993).

The siRNA pathway modulates viral titers in *Aedes aegypti*

Mosquitoes depend on innate resources for immunity. Unlike some animals that develop antibodies against pathogens, mosquitoes and other invertebrates are not known to acquire immunity. However, Garver et al. found that anopheline mosquitoes express a rudimentary antibody though it is decidedly nonspecific and no immune role was determined (Garver et al., 2008). Instead, mosquitoes rely on innate immune responses for protection against pathogens. Mosquito innate immune responses to viral pathogens include RNA interference (RNAi) and other cellular defenses such as JAK/STAT, Toll and Imd pathways (Reviewed byVijayendran et al., 2013). The research presented herein focuses on viral defense by RNAi.

RNA interference and its synonyms describe a variety of mechanisms that essentially regulate gene expression. These mechanisms generally act to degrade transcribed messenger RNA (mRNA). RNA interference is noted in many eukaryotic organisms and therefore accrued multiple synonyms. RNAi was initially described in petunias and entitled post-transcriptional gene silencing (Napoli et al., 1990). Analogous pathways in fungi and mammals are referred to as quelling and the interferon response respectively (Hammond et al., 2001).

Three categories of small RNAs have been characterized in eukaryotes in general as well as in mosquitoes. These categories are determined by function and include small interfering RNAs (siRNA), microRNAs (miRNA) and piwi-interacting RNAs (piRNA). In invertebrates, the miRNA pathway has been shown to play a role in gene regulation and bacterial immunity (Reviewed byVijayendran et al., 2013). The piRNA innate immune pathway protects the mosquito against transposable element integration with some indication that the piRNA pathway

plays a role in antiviral defense (Reviewed byVijayendran et al., 2013). The siRNA pathway plays an integral role in virus defense in *Aedes aegypti* mosquitoes, as well as protection from transposable elements. Known arboviruses like yellow fever virus, the dengue virus serotypes and chikungunya virus are all RNA viruses and generate dsRNA during replication (Reviewed by Vijayendran et al., 2013). This dsRNA prompts the RNAi response through the siRNA pathway and degrades the viral RNA, hence lowering viral titers. The research presented herein focuses on the siRNA viral defense pathway.

DsRNA triggers the siRNA response and sources of dsRNA are classified as either endogenous or exogenous. Endogenous sources of dsRNA include native gene and transposable element RNA naturally existing within a eukaryotic organism and RNA from genes artificially inserted into the genome. Exogenous sources include viral RNA and injected dsRNA (Reviewed by Vijayendran et al., 2013).

It is generally agreed that the relationship between vector mosquitos and the arboviruses they transmit evolved tofoster viral titers high enough for effective viral transmission but low enough to not negatively affect the mosquito's ability to transmit the virus (Reviewed by Forrester et al., 2014). Not surprisingly, the longer a mosquito has been associated with an arbovirus, the greater its competence of vectoring that virus (Reviewed by Powell and Tabachnik, 2013). The siRNA pathway has been demonstrated as the mechanism that maintains this balance (Sanchez-Vargas et al., 2009, Myles et al., 2008). Most of the known viruses vectored by *Aedes aegypti* are RNA viruses, which replicate mostly in the cytoplasm of the host cell. Double-stranded RNA is generated as the virus replicates, attracting the dicer nuclease and prompting the degradation of viral mRNA by components of the siRNA pathway. Viruses continually evolve and circumvent vector immune responses. Examples include virus encoded

RNAi suppressors, decoy messenger RNA from non-essential parts of the viral genome and relatively rapid mutation rates of viral genomic regions that are targeted more frequently by RNAi (Vijayendran et al., 2013).

Multiple genes are known to influence the siRNA pathway in *Aedes aegypti*. These genes include those for Dicer-2 and Argonaute-2 proteins (Adelman, 2008; Campbell, 2008). The siRNA pathway is induced by the presence of dsRNA. Dicer, a double stranded ribonuclease, "dices" this dsRNA into siRNAs. Argonaute-2, an endonuclease, degrades the passenger strand of these siRNAs, leaving the guide strand. The RISC complex forms and the remaining guide strand is loaded within. The guide strand binds its mRNA complement and the mRNA is then degraded, or silenced, by the RISC complex (Reviewed by Meister and Tuschl, 2004).

Transgenic organisms in disease control

Transgenic organisms are organisms whose genome has been altered to include transgenes, or genes from other organisms. Objectives for generating a transgenic organism are wide ranging, with goals from drug development to improving agricultural efficiency and include organisms from bacteria to fungi, to animals and plants.

A variety of methods can be used to develop a transgenic organism. New genes can be inserted randomly into the genome with viral vectors and transposons or at specific sites in the genome with tools like zinc finger nucleases, transcription activator-like effector nucleases and cas9 systems (Reviewed by Doudna and Charpentier, 2014; Reviewed by Ivics and Izsvak, 2010). Development of transgenic mosquito lines for the purposes of mosquito population and subsequent disease control has been proposed and investigated for over 20 years. This is a disease control angle with numerous creative solutions proposed and in some cases demonstrated *in natura*. Genetic control is considered a supplementary measure, best used in conjunction with

traditional control measures to ensure stable, sustainable control. Mosquito genomes have been altered to produce inviable offspring, flightlessness, less competent vectors and others (Reviewed by Alphey, 2014).

Transgenic mosquitoes, as with other transgenic organisms, can be used to study specific pathways. Adelman et al. developed transgenic *Aedes aegypti* lines to study the siRNA pathway, a demonstrated antiviral defense in this animal (Adelman et al., 2008). The transgenic cassette includes genes for EGFP as well as an inverted repeat of the EGFP gene; when transcribed, the inverted repeat generates dsRNA, mimicking the creation of dsRNA in arboviral replication within a mosquito and triggering mRNA degradation through the siRNA pathway. Thus, the resulting amounts of EGFP translated correlates with how well or poorly a mosquito's siRNA pathway functions. Variation in siRNA knockdown was seen in these transformed mosquitoes and a qualitative scale was used to document this variation.

Detecting genetic contributors of phenotypic variation

Detecting the genetic causes of phenotypic differences has wide-ranging applications in disease management, personalized medicine and agriculture. Studies investigating phenotypic variation have identified modifications in the genetic sequence that correlate with human disease states like cancer and cystic fibrosis. Genetic analyses can also be used to predict how an individual may respond to certain drugs. For example, Cooper et al demonstrated that a single nucleotide polymorphism (SNP), or one base substitution in an analogous location in a sequence, can determine how one may respond to the blood-thinning drug warfarin (Cooper et al., 2008). In yet another application, the agriculture industry has long been dedicated to identifying genetic components that yield optimal phenotypes in livestock and produce, such as large size, disease resistance and milk production.

A common method used to detect genetic differences among two phenotypes is single nucleotide polymorphism, or SNP, analysis. Here, DNA is compared at the nucleotide level between phenotypes of interest to reveal base-pair changes. Such comparisons have led to the discovery of SNPs in the human genome that result in disease states like cystic fibrosis, breast cancer, Huntington and Alzheimer's diseases. These base changes can occur anywhere in a given genome, such as regions between genes, within coding sequences and in intronic regions, or regions of a gene that get transcribed then spliced out before translation. SNPs in the exon, or protein coding, regions can lead to changes in the amino acid sequences and potentially disrupt protein function. An obstacle in SNP analyses is how to exclude the many arbitrary SNPs without inadvertently excluding the causative SNP(s); indeed, SNPs that contribute in a small way to the phenotypic variation of interest can be inadvertently discarded (Bush and Moore, 2012).

Technological advances continue to improve our ability to detect genomic characteristics that contribute to phenotypic variation. Stephen Fodor and colleagues developed the microarray technology in the late 1980s and early 1990s, allowing for an explosion of studies, mostly aimed at identifying SNPs associated with disease phenotypes, across entire genomes (Fodor et al., 1991). These genome-wide association studies, or GWAS, are studies that compare whole-genome DNA sequences between phenotypes and identify possible SNPs associated with each phenotype. Preceding microarray technology, DNA comparisons could only be performed on genome fragments with technologies such as restriction fragment length polymorphisms (RFLPs), DNA fingerprinting with microsatellites and Sanger sequencing. To summarize the microarray method, a chip is produced to represent DNA sequences of a genome of a known phenotype. Next, fragments of an alternative phenotype are introduced to the sequences on the

chip; complimentary sequences hybridize and fluoresce. Negative fluorescence data indicates areas of potentially causative sequence changes like SNPs.

Over time, whole genome sequencing has become economically feasible in GWAS studies. Whole genome sequencing for SNP identification boasts a few advantages over the microarray method. For example, microarrays can introduce bias from sample amplification, can fail to detect low frequency alleles and require well characterized molecular markers like tagSNPs (Hurd and Nelson, 2009). To date, there is a dearth of such identifier SNP information in *Aedes aegypti* (Bonizonni et al., 2013). Microarrays require a much larger working volume than whole-genome sequencing; therefore, PCR is often employed to amplify samples for microarrays and disproportionate amplification introduces bias. Tagging, or the use of known SNPs to measure the variability between SNPs of interest is not necessary with whole genome sequencing techniques. There is also more variation in the methodology and analysis of microarray data (Hurd and Nelson, 2009). These advantages and the decreasing price of Next-generation sequencing are driving the chip-based microarray technique for GWAS out of popularity. However, the microarray method is still relevant, as it is better characterized (Bush and Moore, 2012).

In many cases, GWAS studies aimed at identifying causative SNPs failed to do so. Aside from possible environmental and epigenetic factors, a likely reason for this is that a single SNP within a single gene as the sole causal factor for phenotypic variation is rare. Rather, multiple genes often work in concert to produce phenotypic variation.

A common method for identifying the multiple genes involved in a quantitatively inherited phenotype is by quantitative trait loci mapping. Quantitative traits are influenced by multiple loci that collectively influence a phenotype, vary continuously and are generally in high

linkage disequilibrium and recombine together during meiosis. Phenotypes of quantitative traits vary continuously, or show a range of phenotypes. Linkage describes the proximity of SNPs to one another on a chromosome. For example, SNPs that have high linkage are more closely linked, or physically closer on a chromosome with SNPs that have low linkage disequilibrium, or linkage equilibrium.

Another means of characterizing genetic differences between phenotypes are transcriptome comparisons. A transcriptome is the complete set of transcripts and other RNAs in an organism at a given time. Methods to quantify gene expression include RNA-seq, CHIP-seq, microarrays and qPCR. This data can illuminate genes differentially expressed between the two populations/individuals. Transcriptome data with next-generation sequencing, like RNA-seq, reveals an additional layer of information: the sequences of the transcripts; this information can reveal SNPs within the coding regions of genes.

As RNA-seq has grown in favor over microarray technology, so have the methods to optimize RNA-seq data. While microarray analyses are well-developed, RNA-seq analytical methods are flourishing. The workflow of a typical RNA-seq project as this begins with RNA extraction and conversion to cDNA libraries, high-throughput sequencing, transcriptome assembly, mapping the reads to a reference genome and bioinformatics processing to calculate transcript abundance and differential expression. Mapping methods range from aligning the sequencing reads to a reference transcriptome, as is the case here, to assembling and mapping transcripts *de novo*. Bowtie2 is high-speed, sensitive and accurate alignment method for mapping RNA-seq data to a reference genome (Langmead and Salzberg, 2013). The number of total reads vary from sample to sample; therefore, normalizing the number of reads is important in

calculated relative transcript abundance. The edgeR program has been demonstrated to produce quality differential expression results with RNA-seq experiments (Nikolayeva et al, 2014).

The research herein utilizes genome and transcriptome comparisons in pooled individuals representing impaired RNA interference against those representing efficient RNA interference capabilities.

Epigenetic causes of phenotypic variation

Genotype differences influence phenotypic variation through changes in the DNA sequences, like SNPs, gene deletions, insertions and other mutations. Genotype differences are heritable in mechanisms well studied and characterized. During DNA replication, the genome containing these differences is replicated, and passed on to progeny. However, phenotypic presentation can also be influenced by epigenetic means.

Conrad Waddington, who is credited for the term, described epigenetics as "the causal interactions between genes and their products, which bring the phenotype into being" (Waddington, 1942). There are many known mechanisms for epigenetic phenotypic variation, including promoter methylation, interruption of polymerase activity, transcript feedback loops, imprinting and noncoding RNAs. Promoter methylation effects phenotypic presentation by interfering with gene transcription; here, a methyl group binds to a promoter sequence, thereby blocking transcription factor binding and action (Reviewed by Greer and Shi, 2012). Interruption of RNA polymerase results in incomplete transcripts that cannot be used to synthesize a full protein. Messenger RNA can itself function as a positive feedback stimulus and contribute to phenotypic variation (Reviewed by Heard and Martienssen, 2014). Acetylation inhibits histone interactions with DNA, and "loosens" the DNA for easy access to transcription factors; gene expression can be down regulated if chromatin is more tightly wound (Reviewed by Berger,

2007). Non-coding RNAs like siRNAs, piRNAs and miRNAs regulate gene expression by triggering a cascade which degrades transcripts, as in the siRNA pathway, or inhibiting translation, as with the miRNA pathway (Reviewed by Heard and Martienssen, 2014). Indeed, the siRNA pathway itself is an epigenetic mechanism. Long non-coding RNAs have demonstrated involvement in methylation and demethylation, where small non-coding RNAs are involved in chromatin modification and methylation levels. Imprinting is where a phenotype is determined by only one allele, from a specific parent. Such is the case with the human diseases Angelman and Prader-Willi syndromes, where mutations in the allele that is not silenced lead to disease phenotypes (Driscoll et al, 1992).

Epigenetics a promising field for disease control. If the epigenetic influences for an unfavorable phenotype, such as increased vector competence, those influences could be controlled. Furthermore, characterizing epigenetic influences can possibly reverse epigenomic defects in disease states (Dummer et al., 2012). It is important to note that the field of epigenetics is nascent (Reviewed by Greer and Shi, 2012).

Summary

Aedes aegypti is an important arboviral vector with a preference for human bloodmeals. *Aedes aegypti* mosquitoes vector yellow fever and chikungunya fever, dengue fever and other diseases. Dengue virus serotypes are of particular concern due to a lack of an effective vaccine and associated hemorrhagic and shock syndromes. Chikungunya is a rising threat to the United States, as locally contracted cases have recently been reported.

Arthropod vectors are vessels that viruses exploit to replicate to high titers necessary to infect and cause disease in vertebrate hosts. It is generally agreed that a relationship evolved to balance the need for viruses to replicate to titers high enough to effectively infect vertebrate

hosts though not harming the mosquito vector enough to negatively impact its ability to vector the virus. These viral titers are modulated in *Aedes aegypti* by the small interfering RNA pathway. The double stranded RNA produced in viral replication triggers a cascade of the mosquito's cellular immunity and leads to the degradation of viral RNA. This innate pathway varies in its efficiency between individual mosquitoes and herein lies a useful target for disease control.

Transgenic organisms, among other applications, can be used to study genes and pathways in detail. Such a transgenic mosquito was developed by the Adelman and Myles laboratories to study the siRNA pathway in *Aedes aegypti*. Variation in siRNA knockdown was seen in these transformed mosquitoes; indeed, a qualitative scale was used to document this variation.

Detecting genetic influences on phenotypic variation, such as the variation seen in silencing abilities in sensor mosquitoes, can be accomplished with methods including transcriptome comparisons, SNP analysis and Next-generation sequencing. Epigenetic influence is important to consider when studying phenotypic variation and demonstrate a wide variety of mechanisms.

Chapter 2

Developing a homozygous sensor mosquito & enriching for extreme siRNA knockdown capabilities

Abstract

The siRNA pathway is critical in controlling viral titers in the disease vector *Aedes aegypti*. Despite the global health impact of these arboviruses vectored by this mosquito, few contributors to the siRNA pathway have been identified in *Aedes aegypti*. Challenges in identifying such contributors in *Aedes aegypti* include a large, poorly mapped genome containing many repeats and transposable element sequences. Adelman et al. previously developed a tool for studying the siRNA in *Aedes aegypti*: a transgenic mosquito line that reports the status of the siRNA pathway via EGFP intensity. However, the previously developed sensor mosquito cannot be made homozygous for the sensor construct and lacks a genetic background that can be compared to the published *Aedes aegypti* genome. Here, a mosquito strain homozygous for the sensor construct was developed from a Liverpool strain genetic background was developed. Using the phenotypic marker EGFP to differentiate silencing abilities among individuals, extreme EGFP expression phenotypes, representing efficient and poor silencing abilities, were enriched over five generations. The studies described here generated two pools of mosquitoes of opposite, extreme siRNA knockdown capabilities that are theoretically homozygous at loci correlated with siRNA knockdown capabilities but arbitrary at other loci.

Introduction

Aedes aegypti is a vector of arboviruses that impact human health, including dengue viruses, chikungunya and yellow fever viruses. Within the mosquito vector, viral particles

replicate and infect various organs, eventually infecting the salivary glands, from which the mosquito inadvertently transmits virus to vertebrate hosts through bloodmeals necessary for vitellogenesis. The siRNA pathway is an innate immune defense pathway shown to modulate concentrations of these viruses in *Aedes aegypti* (Sanchez-Vargas et al., 2009, Myles et al., 2008).

The sensor mosquito has been used to demonstrate a correlation between quality of silencing the EGFP transgene and disruption of viral replication (Adelman et al., 2013). Indeed, it was shown that an increase in EGFP expression was associated with higher viral titers, where low levels of EGFP expression were shown to be correlated with lower viral titers. The EGFP inverted repeat in the sensor construct allows the formation of a dsRNA hairpin; this hairpin complements the EGFP sequence, inducing the small interfering RNA pathway and leading to the degradation of the EGFP mRNA (Adelman et al; 2008; see Figure 2.1). EGFP fluorescence is variable despite DsRED fluorescence remaining constant; therefore, bright green eye fluorescence indicates impairment of the siRNA pathway, where dim green fluorescence indicates robust silencing of the EGFP transgene (Adelman et al., 2008). Therefore, the use of this transgenic mosquito to represent potential viral infection titers is a convenient means of analyzing the genetic contributors to viral competence. Use of this sensor mosquito to study the siRNA pathway can minimize the risk, time and cost associated with working with viral pathogens.

Despite controlling for known environmental influences, EGFP expression is consistently variable among generations of sensor mosquitoes. This suggests a possible genetic influence in the function of the small interfering RNA pathway. Therefore, we hypothesize that genetic factors contribute to RNAi effectiveness and can explain the variation in EGFP noted even when

environmental conditions are controlled. To demonstrate heritability of RNA interference capabilities in sensor mosquitoes, a strain homozygous for the transgene is necessary. Previously developed sensor mosquito lines were either hemizygous, or having only one copy, for the transgene or too feeble for rigors of this study. Furthermore, the genome of the parent strain, kh^{w}, is not sequenced, published and readily available like the *Aedes aegypti* Liverpool strain genome. Therefore, a primary objective for the research herein is to create an *Aedes aegypti* strain homozygous for sensor construct with robust Liverpool strain genetic background.

Materials and methods

Germline transformation

Germline transformation was accomplished by microinjecting embryos at their posterior poles with a helper plasmid containing the *Mos*1 mariner transposase and a donor plasmid containing the sensor construct. A Fempto-Jet microinjector and micromanipulator were employed to inject embryo pole cells as described in Coates et al. 1998. The transformation plasmid solution contained the *Mos*1 transposon arms encompassing the genes for DsRED, EGFP and the EGFP inverted repeat at 500 ng/uL. Each of the fluorescent protein genes are preceded by the 3xP3 eye-specific promoter and followed by the 3' untranslated region from simian virus 40 (Sheng et al., 1997). The EGFP inverted repeat comprises the first 505 nucleotides of the EGFP sequence followed by the inverse of that sequence and is followed by the hepatitis delta ribozyme for end cleavage (Perrotta and Been, 1991). The helper plasmid solution used contained the *Mos*1 mariner transposase in a pGL_3 PUb-*Mos*1 plasmid vector at 300 ng/uL (Carpenetti et al., 2012). Construction and maintenance of the sensor plasmid was

performed by Michelle Anderson; the transformation plasmid was not altered from the original sensor injections (Adelman et al., 2008).. Microinjections were performed by Dr. Sanjay Basu.

Parent strain

The white-eyed *Aedes aegypti* Liverpool strain used here was previously developed by the Adelman laboratory using a TALEN system to disrupt the kynurenine 3-monooxygenase gene, which is necessary for eye pigmentation. (Aryan et al., 2013).

Standard rearing conditions

Mosquitoes were reared and maintained in a room warmed to 28^0C +/- 4^0C, holding ~30-60% relative humidity and a 15 hour light, nine hour dark diurnal cycle. Larvae were hatched in pans of reverse osmosis-purified water and maintained on flaked Tetra fish food. Adults were maintained on raisins and reverse osmosis-purified water-soaked cottons. Mated mosquitoes were maintained on cottons soaked with reverse osmosis-purified water dyed red with food coloring to discourage oviposition on the cotton. Females were offered defibrinated Sheep's blood for vitellogenesis about every five days, as needed. Embryos were collected on chromatography paper placed inside 15-30 mL plastic beakers lined filled roughly halfway with reverse osmosis- purified water.

Controlled rearing conditions

Environmental conditions were strictly controlled in generations where EGFP expression was assessed. Mosquitoes were reared under these conditions only until EGFP expression was observed and were transferred to standard conditions thereafter.

Embryos were vacuum hatched to synchronize development, counted and aliquoted to pre-acclimated rearing pans at ~100 larvae/L densities. All rearing up to EGFP observation was

contained to environmental chambers at 28^0C +/- 1^0C and ~90% relative humidity with a 15 hour light, 9 hour dark diurnal cycle. Pan water was changed daily to prevent the buildup of juvenile hormone and overgrowth of microorganisms. Larvae were maintained on flaked fish food and adults on raisins and reverse osmosis-purified water-soaked cotton.

Marker gene screening & homozygous challenge

Larvae and pupae positive for at least one copy of the transgene were identified by screening for the DsRED marker gene with a Leica MZ-16FL stereofluorescent dissecting microscope. Larvae and pupae were gently pipetted onto an iced glass petri for these observations.

Mated females were transferred to laying tubes one to two days after bloodfeeding. Laying tubes were prepared by placing a 3.81 cm disc of chromatography paper onto a reverse osmosis-purified water-soaked cotton ball within a 50 mL Drosophila tubes with a foam stopper. The tubes were gently transferred to a cardboard box to promote oviposition

EGFP assessments

All EGFP assessments were performed on mosquitoes that emerged from the pupal casing three days prior unless otherwise stated. Mosquitoes were anesthetized with carbon dioxide and placed in an iced glass petri dish. Thereafter mosquitoes were gently manipulated to the supine position and observed with a Leica MZ-16FL stereofluorescent dissecting microscope at 45x total magnification. EGFP expression was scored using a one to four qualitative scale previously developed (Adelman et al., 2008). The scale was later modified to include a “five” to capture more intense EGFP expression.

DNA extraction & Southern analysis

Adult mosquitoes each of the dim cohort, bright cohort and parent strain were collected into microcentrifuge tubes, snap frozen and stored at -80^{0}C until use. Dim and bright cohort mosquitoes were both collected from generation 14, or the progeny of the fifth and final enrichment. Ten male mosquito bodies were crushed with pestles in 200 uL of Bender buffer (0.1M NaCl, 0.2M Sucrose, 0.1M Tris pH 9, 0.05M EDTA , 0.5M SDS), treated with Proteinase K and incubated at 50^{0}C overnight. Genomic DNA was purified twice with 25:24:1 phenol-chloroform-isoamyl and precipitated with ispopropanol. Pellets were washed with 75% ethanol and resuspended in 200 uL DEPC-treated water.

Genomic DNA was digested with restriction enzymes EcoRI, EcoRV and SpeI and buffers recommended by New England Biolabs. DNA was precipitated with sodium acetate and ethanol and loaded onto a 0.8% agarose gel. The DNA was blotted to a nylon membrane overnight. The membrane was then hybridized to a ^{32}P-labeled probe labeled with fragments of the sensor construct containing sequences for the right hand flank, three copies of the eye-specific promoter, hsp70 promoter, DsRED gene, and SV40, washed and exposed to film. The probe is a mixture two DNA fragments, both roughly 1.2 kb, produced from digestion with HindIII, containing sequences for the right hand flank, 3xP3 promoter, Hsp70 core promoter, DsRED transgene and SV40 untranslated region. The probe was randomly primed and labeled with deoxyadenosine triphosphate containing ^{32}P at the alpha phosphate group from PerkinElmer using the Amersham Megaprime DNA labeling system, then purified with an Illustra Nick column. The probe was hybridized to the nylon membrane overnight at 65°C, then the membrane was washed and exposed to Kodak BioMax maximum sensitivity film at -80°C. Hybridization, wash steps and film exposures were performed by Michelle Anderson.

Fluorescence photography

Photographs were taken with a Canon PowerShot S3 IS digital camera using a Leica MZ-16FL stereofluorescent dissecting microscope. ZoomBrowser EX software settings throughout include AE program, fluorescent, no flash, and -1 exposure. To circumvent the subjective nature of the software zoom, head size on the laptop view was normalized by ensuring head width was ~7.57 cm for females or ~7.19 cm for males across a full view each day on the same Dell laptop. These values were obtained by averaging head width, at its widest point, from photos of the unenriched, starting population of females and males each.

Generations nine – 13 were photographed with an 81 North light source, where generation 14 photos were taken using a mercury light source.

Photograph analysis

Photos were analyzed with Image J software by taking measurements from left eye, right eye and background gates. The measurements of the left and rights eyes were averaged and the background measurement was subtracted from that average to give the corrected value. The corrected DsRED and EGFP values were entered in JMP statistical software along with generation, gender, enrichment phenotype (dim, bright or all phenotypes). These data were fitted to an ANCOVA model with corrected EGFP values as the dependent variable and DsRED, gender, generation and phenotype data and their interactions entered as model effects. Other data such as order of emergence, background values and uncorrected values were also collected, but found to have an insignificant impact in the model and were thus excluded in the analysis.

Results

Developing an Aedes aegypti strain homozygous for the sensor construct

An *Aedes aegypti* line homozygous for the sensor construct was developed in nine generations; the major steps in this process are outlined in Figure 2.2. First, 1,369 *Aedes aegypti* Liverpool *kmo* embryos were injected with a helper plasmid containing the *Mos*1 mariner transposase and a donor plasmid containing the sensor construct. The microinjected generation zero (G_0) embryos were hatched under standard rearing conditions; 186 males and 134 females survived, giving a 23% survival rate. These G_0 mosquitoes were combined into eight experimental pools, E1-E8, and outcrossed with the parent strain. As the *Mos*1 mariner transposable element system integrates essentially randomly into the genome, a variety of integration sites were assessed and compared to choose optimal pool to pursue developing into a homozygous line. E1-E5 pools contained microinjected females, and E6-E8 contained males. G_1 embryos were collected, hatched and larvae were screened for presence of the marker gene. Two of the eight pools, E4 and E6, showed no mosquitoes positive for the marker gene. For the remaining pools, positive G_1 males and females were placed into pool-specific cages and outcrossed to the parent strain; embryos from these crosses were collected, hatched and screened for DsRED expression. To avoid female mosquitoes with the transgene inserted into the female sex-determining locus, positive G_2 males only, still segregated by pool, were out-crossed with Liverpool *kmo* females. These G_3 embryos, all assumed heterozygous for the transgenic cassette, were hatched under environmentally controlled conditions outlined above to control against factors that are known to influence EGFP intensity. Sex ratio and EGFP score data were collected from each pool and are shown in Table 2.1. The sex ratio data revealed insertion of the construct within or near the female sex determining locus for the majority of mosquitoes in pool

E1, with 98% of the females positive for the marker gene but only 1% of the males were positive. The remaining pools with transgene positive progeny, E2, E3, E5, E7 and E8, showed an even distribution of marker gene positive individuals among males and females. EGFP assessments revealed a variety of EGFP profiles among the pools; lines E1, E2 and E3 were overall dimmer in EGFP expression than lines E5, E7 and E8. The DsRED positive mosquitoes from pools E2, E3, E5, E7 and E8 were intercrossed and DsRED positive and negative data from these progeny were fit to a chi-square goodness-of-fit tests. Tested were scenarios where 100% of the homozygotes survive, which should yield 25% negative, 75% positive, or none of the homozygotes survive, which should yield 33% negative, 66% positive; the results of that analysis are summarized in Table 2.1. Variable DsRED expression was noted among the pools and documented (see Table 2.1); these data are significant to consider in choosing pools to pursue in developing a homozygous sensor. For example, overall dim transgene expression could increase false negatives in DsRED screenings and make differentiating EGFP silencing abilities more difficult.

Based on the projected homozygote survivability results, EGFP silencing profiles and the quantity of mosquitoes assessed, pools E7 and E8 were chosen to further pursue in the development of a homozygous sensor line. To enrich for homozygous individuals in these pools, DsRED positive individuals from these lines were intercrossed. Progeny from these individual crosses that were 100% DsRED positive, ten subgroups for E7 and nine for E8, were again intercrossed. Progeny from two E7 subgroups, with 342 and 408 individuals, were all positive for the marker gene, but no groups from E8 had all positive progeny. Embryos from the two E7 subgroups were pooled to decrease inbreeding effects, intercrossed and screened to verify all progeny were positive for the transgene, indicating homozygosity. Pooling the subgroups

compounded the genetic diversity, helping to combat the negative effects from potential genetic bottlenecking effects in the homozygous challenge. Next, an EGFP profile was obtained for the finalized line at generation nine, showing variation in EGFP silencing among the analyzed cohort (see Figure 2.4).

A final concern was the variable DsRED expression noted in E7 in earlier generations (see Table 2.1). This variation was concerning in that it could indicate multiple transgene sites among the population, thus confounding future siRNA knockdown assessments. To test if this bimodality in DsRED expression was successfully bred out or persists in the homozygous sensor line, corrected DsRED value distribution analyses were analyzed; Figure 2.5 shows these distribution analyses to be devoid of bimodality in DsRED readings in both male and female mosquitoes. Genders were analyzed separately due the size difference between male and female mosquitoes.

Extreme phenotype enrichments

The goal of the enrichments was to obtain two populations homozygous for loci involved in the siRNA pathway by enriching the phenotypes representing siRNA silencing abilities. Five compounding crosses of mosquitoes with the brightest and dimmest EGFP expression each were performed to enrich the extreme, opposite silencing ability phenotypes. To begin the enrichment crosses, mosquitoes with the dimmest and brightest EGFP expression from the newly developed line homozygous for the sensor construct (G_9) were selected and placed in cages segregated by their EGFP phenotype and intercrossed. These mosquitoes were vacuum hatched and reared as described in the methods section and assessed at a consistent age post emergence from the pupal casing to control for factors such as temperature and age which could confound EGFP comparisons. For the next four generations, embryos from the previous dimmest and brightest

crosses were hatched and reared to the same age, then the brightest or dimmest mosquitoes selected and placed in a cage to mate and produce adequate numbers of embryos. Throughout the enrichments, values derived from fluorescence photography analyses were collected to measure the change of EGFP intensity. Figure 2.6 demonstrates a scatterplot of these EGPF values with a schematic of how these values were derived. To determine if the difference in dim and bright cohort corrected EGFP means was significant and if this difference increased significantly with each enrichment cross, the corrected EGFP data was fitted into an ANCOVA, or analysis of covariance;

Figure **2.**7 shows the results of this analysis and a schematic of the enrichment crosses. Effects fitted into the final model were generation, phenotype cohort, corrected DsRED and gender. The generation and phenotype cohorts were fitted to the model in order to confirm or refute an increasing gap in corrected EGFP means. Since males are smaller than females, and the corrected EGFP values between the two are around 30% higher in females due to their increased size, gender was entered into the model. Corrected DsRED was entered into the model as it serves as a control. Changes in DsRED may reflect the gradual dimming of a the metal-halide bulb in over the course of generations nine to 13 and the change to a mercury light source in generation 14; entering DsRED into the model helps to correct for these issues. A Tukey's test was run to test the significance of the differences between the corrected EGFP means. The results of this analysis showed statistically significant different means between the extreme phenotypes in all generations of enrichment. Contrasts were performed to determine if the differences between the means was significantly increasing from one generation of enrichments to the next. The results from the contrasts show that the differences between the extreme

phenotype corrected EGFP means increased significantly in all generations except between generation 12 to generation 13; the difference between the means of the dim and bright cohorts at generation 10 and the difference between the means of the dim and bright cohorts at generation 14 showed a p-value of 0e0.

Confirmation of transgene insertion site and copy number in homozygous E7 mosquitoes

Genomic DNA extracted from male, generation 14 mosquitoes each of the dimmest and brightest cohort as well as male mosquitoes from the parent strain, Liverpool *kmo* was digested with a restriction enzyme that does not cut within the transgene as well as two enzymes that cut within the construct and Southern analysis performed. Two bands are seen in the samples digested with these enzymes (see Figure 2.8). The thicker bands represent strong binding, likely the probe binding the target sequence. The smaller size bands in both cases are thinner than the larger size bands; these thinner bands likely represent the probe partially hybridizing to repetitive sequences in the construct. Bands among the experimental groups appear roughly the same size, indicating the insertion site location is equivalent between the extreme, enriched phenotypes.

Discussion

The development of a sensor mosquito with a Liverpool genetic background and homozygous transgene insertion has multiple implications. The EGFP profile of the finalized, homozygous "sensor" strain demonstrates EGFP variability among individuals in a given cohort despite controlling for environmental factors (see Figure 2.4). This illustrates that, as with the kh^{w} sensor mosquitoes, EGFP expression is variable among a cohort of mosquitoes, despite tightly controlling environmental conditions and viewing at a consistent day post emergence (Adelman et al., 2008). Furthermore, the overall EGFP profile for the homozygous line is

skewed bright (see Figure 2.4); this exaggeration in EGFP expression makes differentiating siRNA silencing abilities relatively easier to detect. Another benefit to the newly developed sensor line is its genetic background. The Liverpool background allows molecular data to be compared to the published *Aedes aegypti* genome readily available in VectorBase. A third benefit to this newly developed sensor strain is its homozygosity for the sensor construct. The development of a line of mosquitoes homozygous for the sensor construct allowed for the determination that siRNA capability phenotypes are heritable. Early in development for homozgysity, E7 fit neither the 100% survivability scenario nor the 0% survivability scenario for homozygotes in the chi-square challenge (see Table 2.1.). The p-values for both scenarios were significantly different at a 98% confidence interval, with p-value of 0.008 for the scenario for all homozygotes surviving and 0.022 for none of the homozygotes surviving. Rather, it appears that a portion of homozygotes in this pool are surviving.

This homozygous sensor strain can be used to study the siRNA pathway in projects outside of the research herein. Use of this sensor mosquito to study the siRNA pathway can minimize the risk, time and cost associated with working with viral pathogens. In addition to studying viral infection, this sensor strain can also be used to study RNAi in other biological processes like apoptosis, development and survivability. The Liverpool genetic background allows such studies to reference the published VectorBase genome. Homozygosity is another added benefit to this strain for the studies outlined above. In contrast to working with a sensor strain with sex-linked transgene inheritance, this line does not need to be sorted and outcrossed, reducing labor, costs and time.

Heritability of siRNA silencing abilities was demonstrated with this homozygous sensor in the enrichment crosses (see

Figure **2.**7). This finding may help explain at least some of the variation in vector competence, or the ability for a disease vector to transmit disease, seen in natural populations of *Aedes aegypti*. The heritability of RNAi knockdown phenotypes shown here could explain that variation in vector competence can be genetically determined by silencing abilities. For example, genetic proclivities of RNAi-based silencing in individual geographic colonies are established in founding individuals and thereby contribute to the vector competence profile that colony. Indeed, variation in vector competence among *Aedes aegypti* populations has been demonstrated in studies in including those by Chepkorir et al., Bennett et al., Gubler et al. and Knox et al (Chepkorir et al., 2014; Bennett et al., 2002; Gubler et al., 1979; Knox et al. 2003 and Adelman et al., 2013). A potential study to validate this hypothesis could be to compare phenotype-associated SNPs extreme, opposite RNAi silencing phenotypes in wild strains of *Aedes aegypti* against the corresponding opposite phenotypes in this homozygous sensor strain.

The products of the phenotype enrichment crosses can be utilized for transcriptome comparisons as well as a bulked segregant analysis for SNPs in genomic regions. The five-times enriched, extreme opposite siRNA knockdown ability phenotypes demonstrate that at least some genetic contributors influencing the siRNA pathway are heritable. The difference between dim and bright cohort means increased tenfold over five generations of enrichments. The goal for exagerating the genetic contributors to the extreme phenotypes was to make each line homozygous at genomic regions that contribute to siRNA knockdown abilities. In the next chapter, F_0 parents representing the extreme phenotypes are crossed, their offspring intercrossed and the extreme phenotypes rescued from the F_2 generation to identify possible causative SNPs for the extreme phenotypes.

It is important that the copy number of sensor constructs in each the dimmest and brightest enrichment cohorts be equal, so that the poor and efficient knockdown capabilities be enriched for, not merely selecting mosquitoes that are brighter in EGFP expression because they have more copies of the transgene. Southern blotting is a technique commonly employed to identify the copy number of transgenes in a given genome (Reviewed by Southern, 2006). Equivalent band numbers and sizes were seen in both enriched lines with multiple restriction enzymes, which is expected with equivalent transgene copies that are roughly in analogous genome location between both enriched cohorts. This indicates that the brightest and dimmest mosquito selection for the enrichment crosses were not merely chosen by differences in transgene copy number. Indeed, the objective of the enrichments was to segregate and enrich genomic characteristics of extreme siRNA knockdown capabilities and discrepancies in sensor construct copy number could allow for false positive individuals in the phenotype enrichments. Two or more copies of the transgene could manifest in brighter eye fluorescence; more copies of EGFP should yield more EGFP protein which should equate to brighter green fluorescence. Therefore, unequal copy numbers between the enriched cohorts would defeat the purpose of the phenotype enrichments.

The lines produced from the phenotype enrichments are useful in applications outside of the research herein. The enriched lines can be used in knockdown and overexpression studies investigating specific genes. Either line can also be crossed with transgenic organisms to create mosquito lines with multiple transgenic cassettes, a feat difficult to achieve with microinjection alone. Either line can also be crossed with *Aedes aegypti* gene knockout lines to create new lines to study the effects of gene knockout on siRNA knockdown.

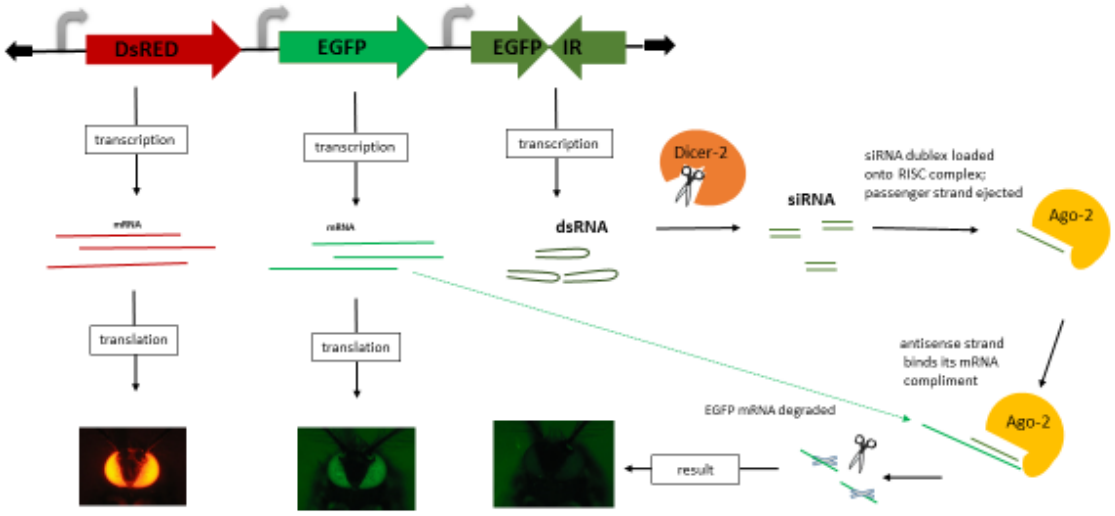


Figure 2.1 Mechanism through which the sensor construct reports the status of the siRNA pathway. The sensor construct is displayed at the top and the transcription and translation products are shown underneath. The EGFP inverted repeat transcript folds in on itself, creating dsRNA that is processed by Dicer-2 endonuclease activity; Dicer-2 cleaves the dsRNA into small, interfering RNAs which are loaded onto Argonaute-2. The passenger strand of the loaded siRNA duplex is ejected and the RISC loading complex binds complimentary sequences of the EGFP mRNA. The EGFP mRNA is degraded, hence silencing EGFP expression and resulting in a dim, or low, phenotype. Scissor images courtesy of clker.com, used under fair use, 2015.

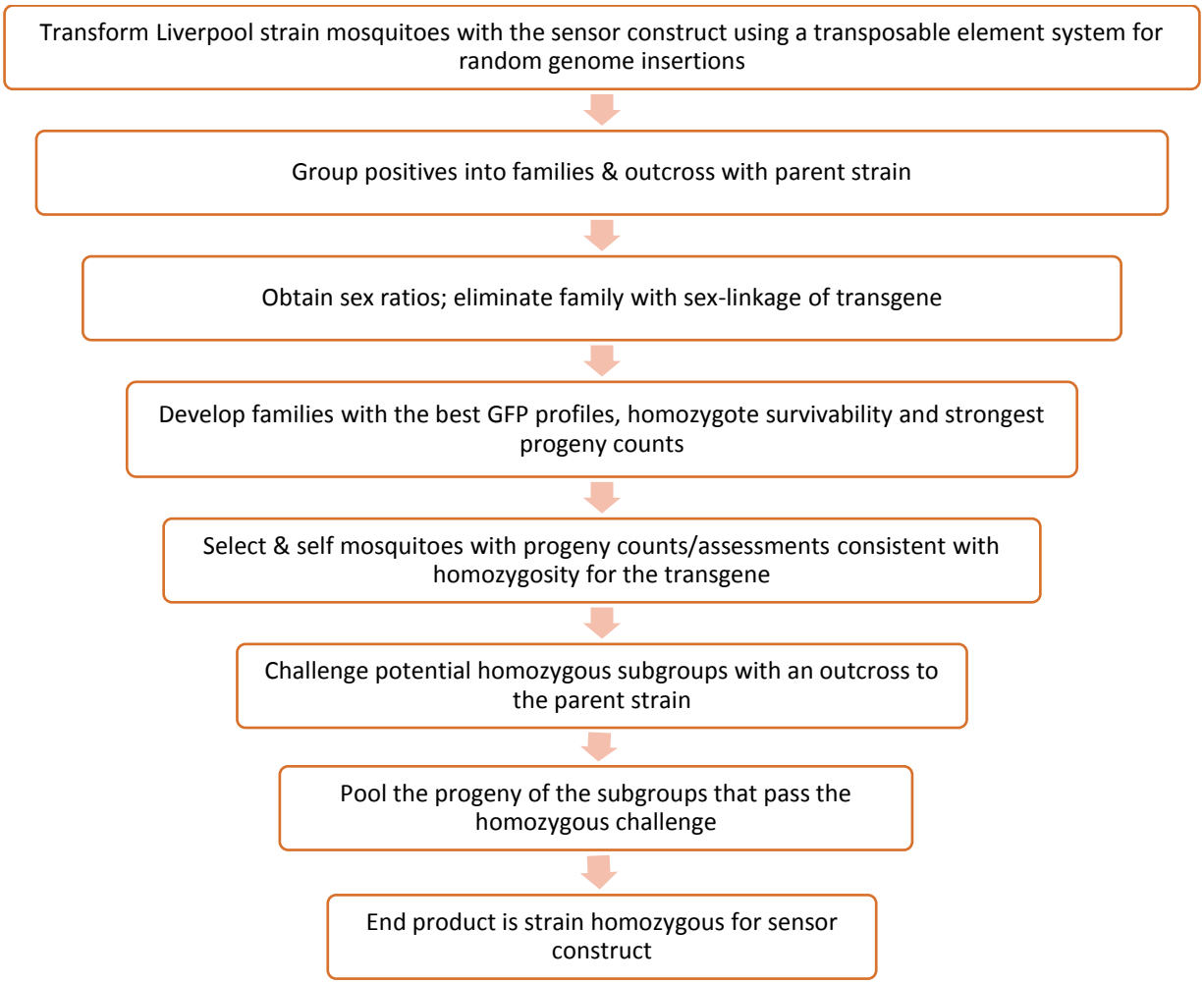


Figure 2.2 Outline of the major steps taken in developing a homozygous sensor mosquito.

Major milestones are listed and their overall filtration effects visually represented.

Table 2.1 Transformation details and individual pool characteristics in the development of a homozygous sensor line. Listed are information pertaining to the microinjection details, DsRED data through generations one to four, sex ratio data, DsRED expression characteristics and homozygote survivability by pool.

Transformation details				% Transgene positive by DsRED detection						Chi-square goodness of fit p-values	
	G_0							sex ratio		G_4	
# embryos injected	# G_0 survivors	Pools		G_1	G_2	G_3	G_3 ♂	G_3 ♀	G_4 DsRED characteristics	75/25 All homozygotes surviving	66/33 No homozygous surviving
1,369	186 ♂ 134 ♀	E1	♀ ß parent	4 +/ 418 - 0.9%	79 +/ 81 - 49%	321 +/ 337 - 49%	2 1%	319 98%	Integration site linked to female sex determining locus	not assessed, as sex-linked and not valid in homozygous stain development	
	surviv-ability = 320+/ 1369 - = 23%	E2		4 +/ 1222 - 0.3%	56/ 94 - 37%	624 +/ 567 - 52%	336 55%	288 50%	VERY dim DsRED	0.102	<0.001
		E3		28 +/ 948 - 2.9%	244 +/ 281 - 46%	530 +/ 543 - 49%	122 54%	408 48%	Variable; some dim, some bright, in both male and females	0.281	0.09
		E4		0 +/ 1085 - 0.0%				not assessed, as no positive progeny detected			
		E5		2 +/ 888 - 0.2%	17 +/ 47 - 27%	319 +/489 - 39%	140 37%	179 42%	Slightly fainter than E7	0.269	<0.001
		E6	♂ ß parent	0 +/ 3385 - 0.0%				not assessed, as no positive progeny detected			
		E7		57 +/ 5021 - 1.1%	33 +/ 30 - 52%	379 +/ 394 - 49%	200 48%	179 50%	Variable; some dim, some bright, in both male and females; ~30% dim	0.008	0.022
		E8		17 +/ 4719 - 0.4%	145 +/ 194 - 43%	618 +/ 533 - 54%	325 55%	293 52%	dim	0.156	<0.001

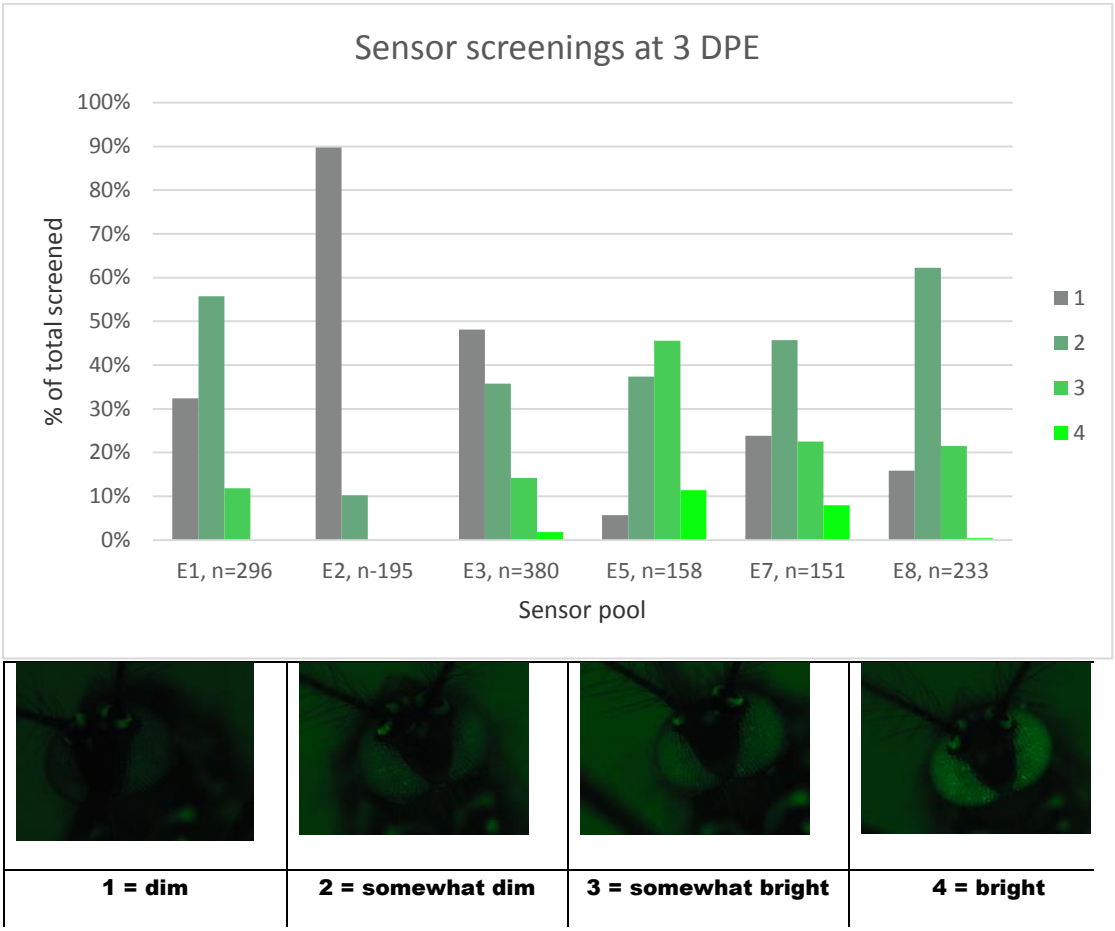
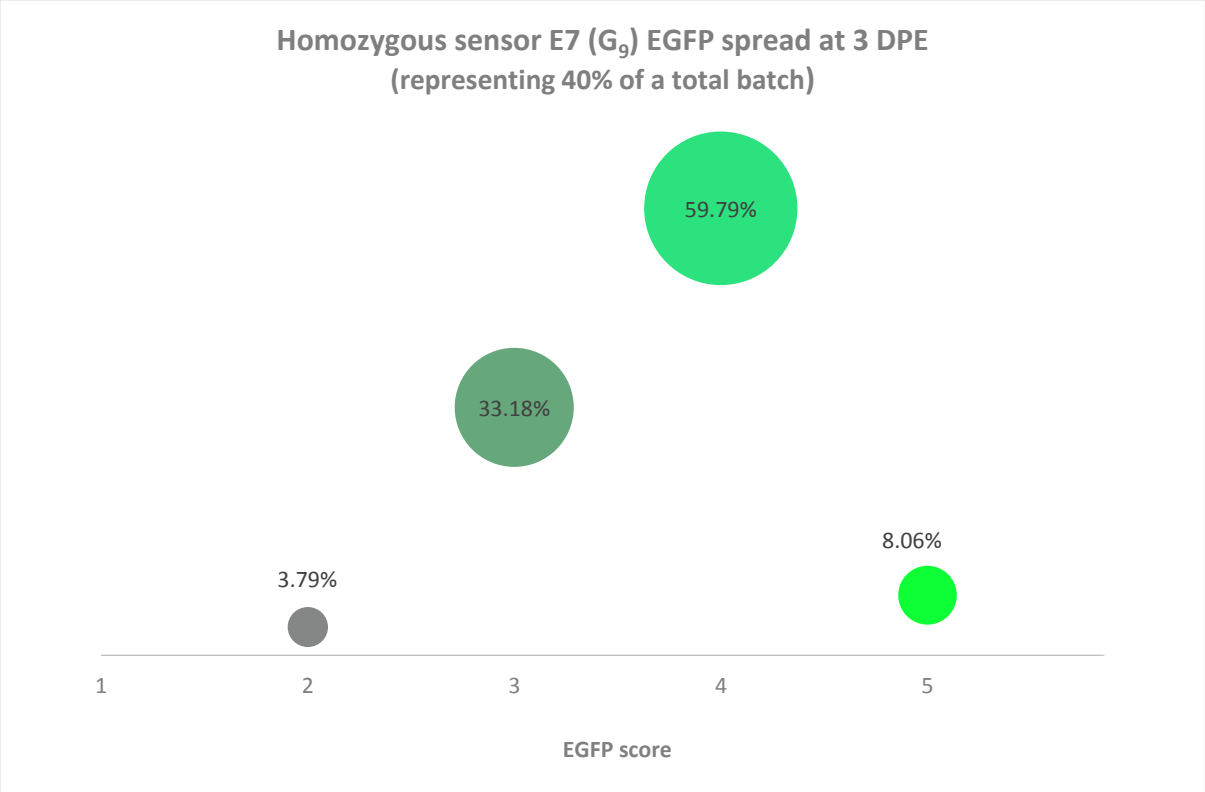


Figure 2.3 Assessments at G_3 reveal a variety of EGFP profiles for each pool. Top: EGFP profiles of each experimental pool at three days post emergence from pupal casings. Only females were assessed for continuity. Bottom: Qualitative scale used for assessing EGFP profiles.



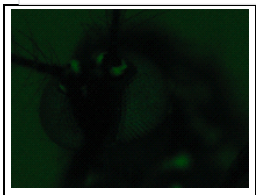
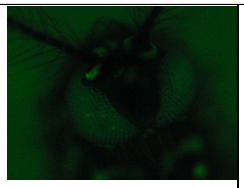
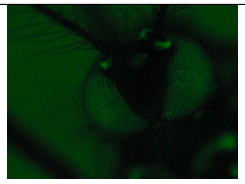
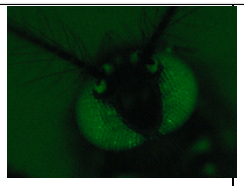
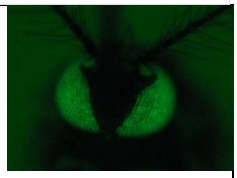
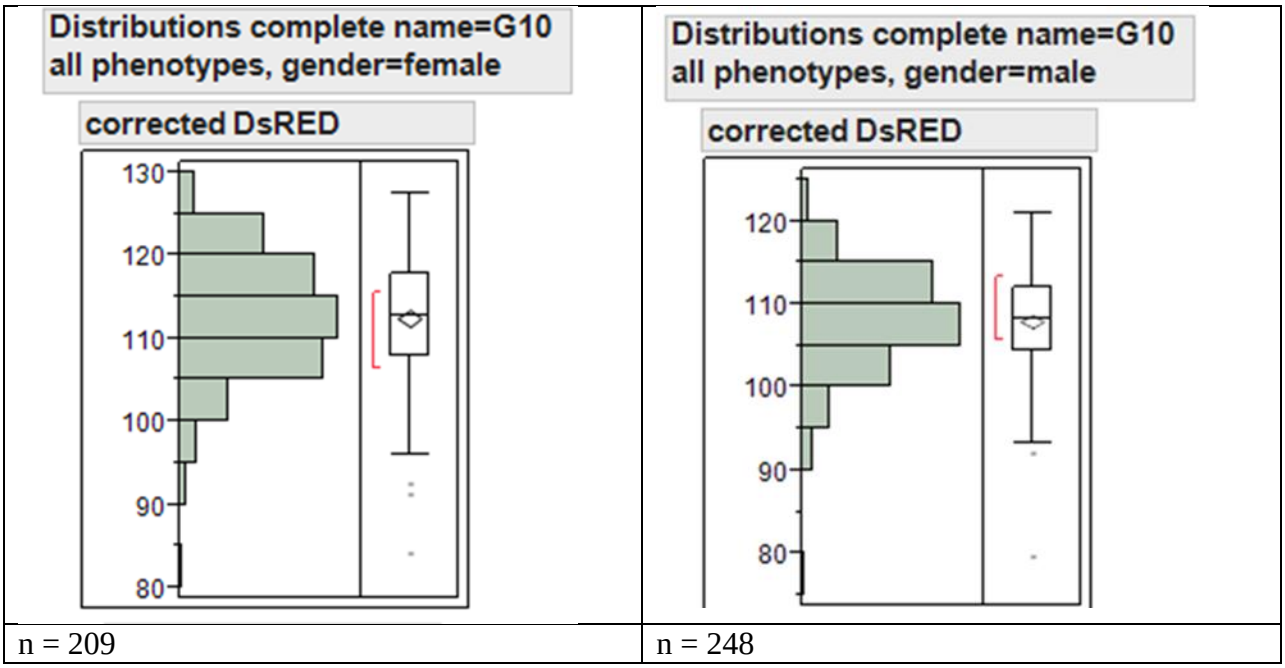
				
1 = dim	2 = somewhat dim	3 = somewhat bright	4 = bright	5 = very bright

Figure 2.4 EGFP profile of the developed homozygous sensor line. Top: EGFP profile of the finalized homozygous sensor at three days post emergence from pupal casings, generation three (n=211). Bottom: Qualitative 1-5 scale used in assessing EGFP profiles.



n = 209	n = 248

Figure 2.5 DsRED expression in the final homozygous line is normally distributed.

Corrected image J intensity measurements of DsRED expression in female and male E7 homozygous sensor mosquitoes at generation ten, four days post-emergence validates consistent control gene expression.

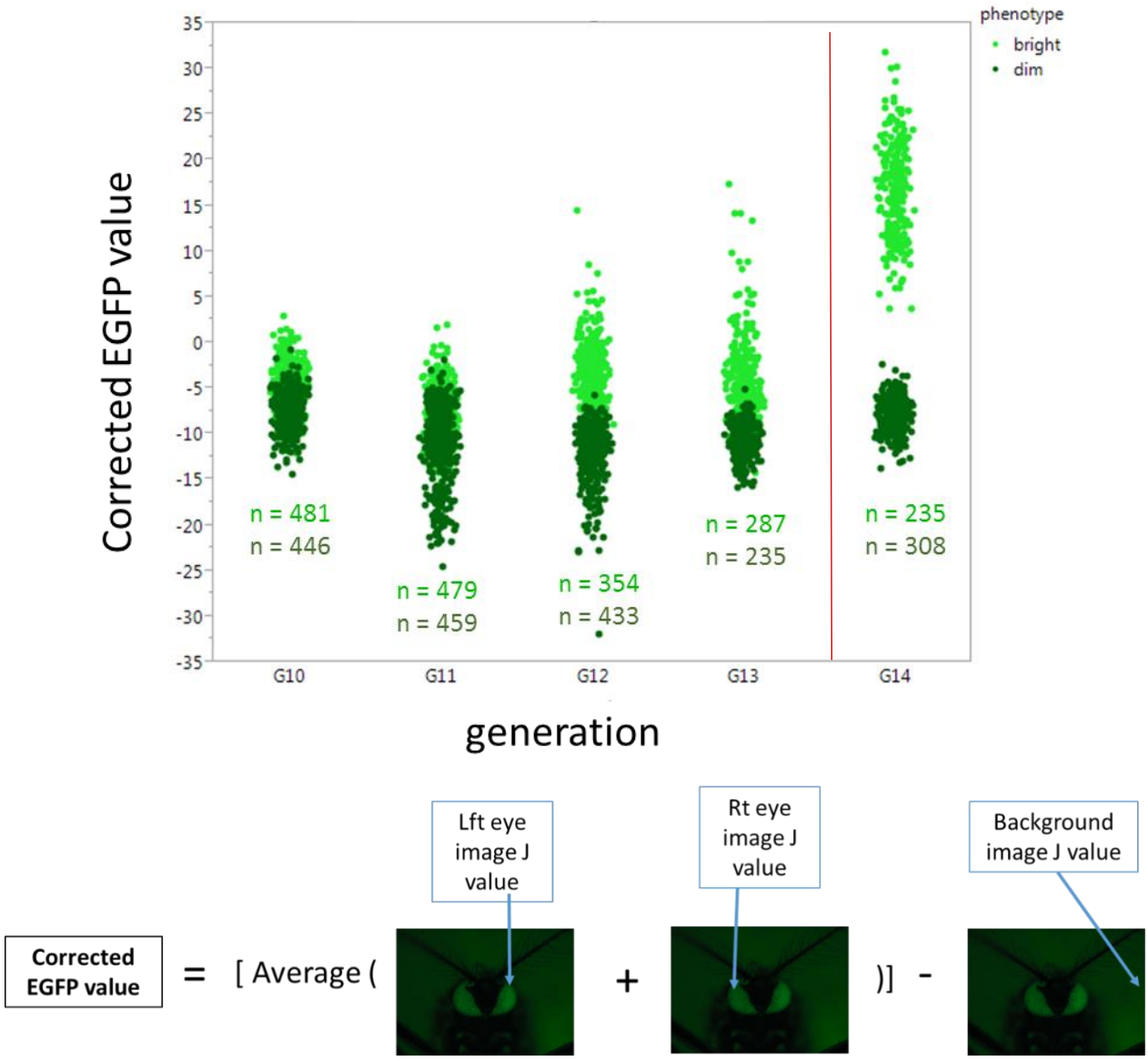
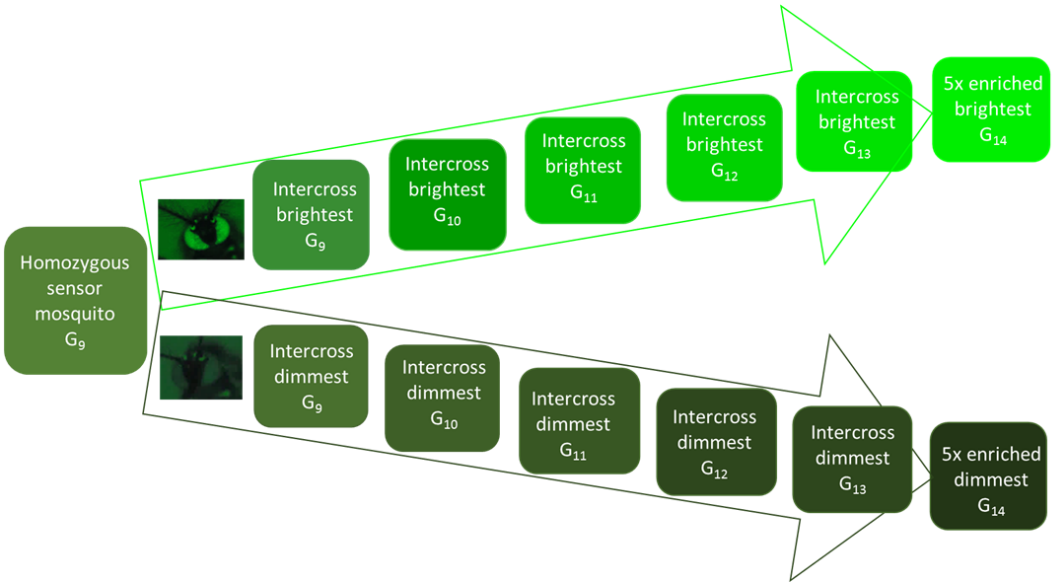


Figure 2.6 Phenotypic marker intensity shows increasing separation over each enrichment cross. Plotted are corrected EGFP values of each enrichment type by generation; a schematic of how EGFP intensities were corrected for background fluorescence is shown. The red line demarcates the light source change.

A.



B.

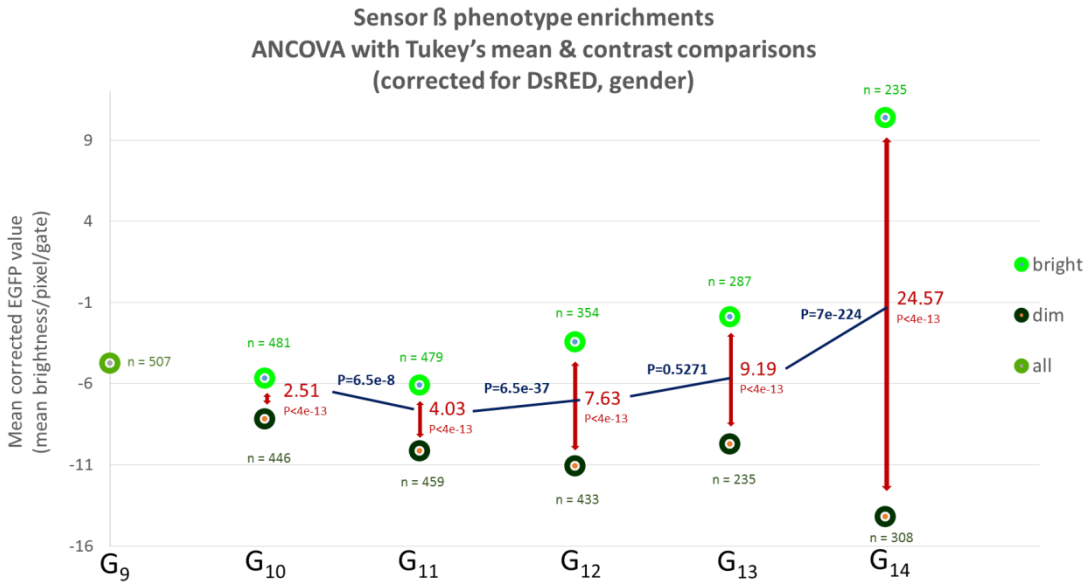


Figure 2.7 Extreme siRNA knockdown abilities are heritable and were enriched over five generations. Panel A: schematic of the enrichments. Panel B: corrected Image J EGFP photo analysis results were fitted to an ANCOVA model corrected for generation, phenotype cohort, gender and corrected DsRED value effects. Means generated from Tukey's tests for each bright,

dim and the starting population are plotted here for each generation; differences between the means are shown in red with their p-values below. Contrast comparisons of the mean differences are indicated by blue lines; contrast p-values are shown in blue.

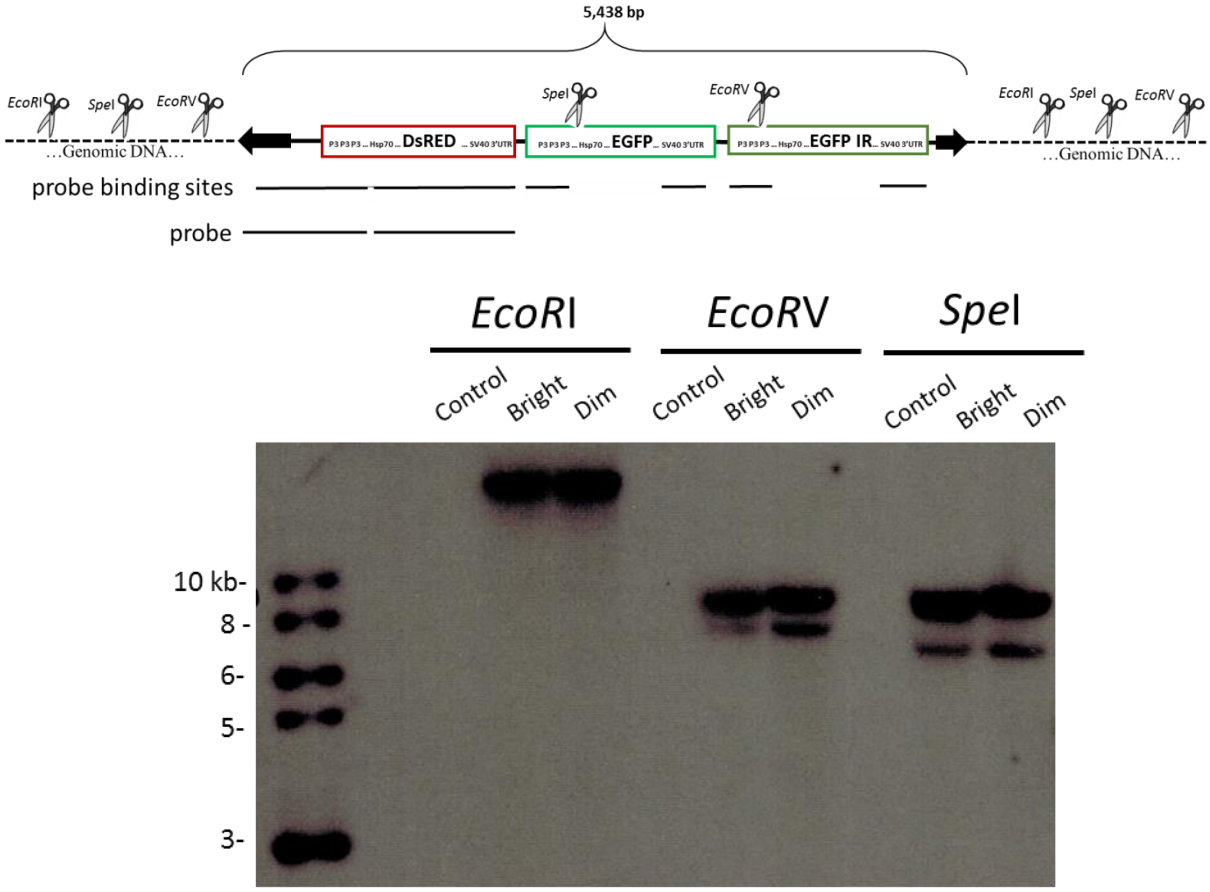


Figure 2.8 Southern analysis verifies sensor construct copy numbers are equivalent among the enriched, extreme phenotypes. A schematic of a theoretical insertion site and restriction enzyme cut sites is depicted. EcoRV and SpeI both cut once within the sensor cassette, giving two pieces containing portions of the cassette, theoretically yielding one DNA fragment containing the sequence complimentary to the probe and one fragment without the probe. Dotted lines indicate genomic DNA. Components common to each gene in the construct are listed within each gene box. Equivalent fragment sizes and numbers are seen in both extreme phenotype pools. Non-injected LVP *kmo* was used as a control. Scissor images from clker.com and used under free use, 2015.

Chapter 3

Transcriptome & SNP analyses

Abstract

Aedes aegypti is an important vector of human pathogens including yellow fever, dengue and chikungunya viruses. The small interfering RNA pathway is a critical immune response for controlling viral replication in *Aedes aegypti*. The goal of this research is to identify components of the *Aedes aegypti* genome that influence this innate immune pathway. Transcriptome sequencing and analyses were performed from pools of individuals from extreme opposite, enriched phenotypes of efficient and poor RNA silencing abilities, revealing potential RNAi contributors. 11,562 quality-filtered transcripts mapped to the published transcriptome. From this, 1,120 transcripts were significantly different (FDR<0.0001) among the extreme phenotypes, with 590 transcripts significantly upregulated in the bright cohort and 530 downregulated. Several of these significantly different transcripts were targeted for DNA sequencing and SNP analysis in both enriched extreme phenotype representatives and in samples obtained by crossing enriched, extreme phenotype individuals, intercrossing their progeny and selecting individuals representing the extreme phenotypes. The results from these studies show many hundreds of potential RNAi contributors for future investigations.

Introduction

The small interfering RNA pathway is an innate immune response to RNA virus infection in *Aedes aegypti* (Reviewed by Vijayendran et al., 2013). The siRNA pathway is triggered by dsRNA. Double-stranded RNA is generated during viral replication, attracting Dicer-2, which cleaves, or "dices," the long dsRNA into small, still double-stranded, interfering RNAs around

21 base pairs in length (Lee et al., 2004). Viral inhibition is furthered by sequence-specific silencing, where one strand of the siRNA duplex binds complementary RNA and advances its degradation (Doench et al., 2003). In *Drosophila*, this silencing is made possible when the siRNA duplex is loaded onto Argonaute-2 from the RNA-induced silencing complex (RISC)-loading complex (RLC), comprising Dicer-2 and R2D2 (Liu et al., 2003). Hsp90 facilitates a conformational change in Argonaute-2 that allows it to accept the siRNA duplex from the RLC (Miyoshi et al, 2010). The siRNA strands are unwound and the passenger strand is ejected by aid of C3P0 and degraded, leaving a mature RISC complex comprising Ago-2 and the guide RNA strand (Miyoshi et al, 2010; Liu et al., 2009). This complex then binds the RNA complement of the guide strand, and Ago-2 "slices," or degrades, the complimentary strand, silencing the mRNA, or in the case of viral replication, the positive sense genome. The rate of viral silencing by the siRNA pathway within an individual mosquito impacts infection titers in the mosquito and therefore the concentration of viral particles available for transmission (Adelman et al., 2013). The research in this chapter investigates the genetic differences in mosquitoes that demonstrate high rates of RNA knockdown against mosquitoes with low rates of knockdown.

Phenotypes that are influenced by multiple traits are continuous (Miles and Wayne, 2008). siRNA knockdown abilities in *Aedes aegypti* are presumed to be quantitative traits, with contributing loci in at least genes for Argonaute-2, Dicer-2 and R2D2 (Adelman et al., 2008; Haac et al., 2015). The varying EGFP expression seen in a given cohort of sensor mosquitoes demonstrates that knockdown by the siRNA pathway varies continuously. In the previous chapter, extreme phenotypes representing efficient and poor knockdown were enriched for by crossing the individuals with the desired extreme phenotype over five generations. This gives two substrains with opposite phenotypes, each theoretically homozygous at genomic regions

contributing to or regulating the success of the siRNA pathway. A potential off-target effect of these enrichments could be inadvertent enrichment of genetic material having no role in RNAi, an artifact of bottlenecking with each enrichment cross. A QTL, or quantitative trait loci, cross can be employed to circumvent such confounding artifacts by evenly distributing off-target genetic variations while the variations of interest segregate together (Reviewed by Miles and Wayne, 2008). By cross mating the two, opposite, enriched cohorts and self-crossing their F_1 offspring, the resulting F_2 progeny should give two sets of opposite, genetically distinct cohorts homozygous in genomic regions contributing to the siRNA pathway and heterozygous at genomic regions not contributing to the siRNA pathway. Michelmore et al. employed a similar cross strategy, which they termed bulked segregant analysis, to reliably isolate SNPs from genes of interest in lettuce (Michelmore et al., 1991). The individuals homozygous at each extreme of the siRNA knockdown abilities continuum can then be selected by their extremely bright and extremely dim green eye fluorescence.

The goal of the research herein is to identify novel contributors to the siRNA pathway in *Aedes aegypti*, an important disease vector. Transcriptome analysis of the extreme, enriched groups was used to identify genomic regions of interest. F_0 and F_2 extremely dim and bright green fluorescing mosquitoes from a quantitative trait loci cross were sequenced at these regions for SNP analysis.

Materials and Methods

Mosquito collection for transcriptome samples

The G_{14} mosquitoes not selected for the fifth and final enrichment for each extreme phenotype were placed in cages segregated by their enrichment line of dim or bright green

fluorescence. The G_{15} embryos from these crosses were vacuum hatched to synchronize development, counted and evenly distributed to pre-acclimated rearing pans at ~100 larvae/L densities. To avoid confusing the samples by human error, the enriched bright line was hatched four days prior to hatching the enriched dim line. Larvae were reared in environmental chambers at 28^0C +/- 1^0C and ~90% relative humidity with a 15 hour light, 9 hour dark diurnal cycle. Pan water was changed daily to prevent the buildup of juvenile hormone and overgrowth of microorganisms. Larvae were maintained on flaked fish food and adults on raisins and reverse osmosis-purified water-soaked cotton balls. At three days post-emergence, females were snap frozen and stored at -80C daily in the afternoon (between 2-2:30 p.m.) to minimize changes in diurnal cycle gene expression. Females were kept virgin to minimize any off-target expression data from mating.

RNA isolation & library preparation

Total RNA from females from each enriched line was extracted with TRIzol and chloroform. mRNA was purified from the total RNA from each line and converted to cDNA; cDNA libraries from each pooled, enriched line were prepared in tetraplicate with Illumina's TruSeq kit, using the low sample protocol. The paired-end, cDNA libraries were submitted to the Virginia Bioinformatics Institute at Virginia Tech, where sample concentrations were measured with Qubit® fluorometry, verified with an Agilent BioAnalyzer and sequenced with Illumina's HiSeq 2500. RNA extraction, purification and conversion to cDNA libraries was performed by Michelle Anderson.

RNA-seq analysis

A laboratory server was used to store and to analyze the RNA sequencing data of all samples. The read quality of all the sequencing files was evaluated with FastQC before further analysis was performed. The sequencing reads for each sample were mapped to a reference transcriptome (VectorBase, Aedes-aegypti-Liverpool_TRANSCRIPTS_AaegL3.3.fa.gz) and to the transposon library (tefam.biochem.vt.edu) with Bowtie2 (Langmead and Salzberg, 2012); this work was performed by Dr. Zach N. Adelman. SAMtools (Li et al., 2009) was used to sort the resulting BAM files and to filter out low quality reads that did not map to the reference file, or mapped to more than one transcript. The data was then filtered with edgeR, where reads below a stringency factor of three (one per one million in three or less replicates) were eliminated from further analysis. Library sizes were normalized and abundance data calculated in edgeR. Differentially expressed transcripts were assessed with edgeR with the dim cohort arbitrarily set as the control. P-values were corrected for multiple testing using the false discovery rate (FDR) option, with an FDR threshold of 0.05, and a cutoff FDR of less than 0.0001 was used to determine significance (Robinson et al., 2010). RNA-seq data was mapped to the reference transcriptome, sorted and analyzed by Heather Eggleston. Gene names, descriptions and other attributes were obtained from VectorBase's BioMart version 0.7 using the Aaegl3.3 dataset, with non-coding RNA information eliminated. Perl version 5.18.2 was used to merge the gene attributes data with the differential expression results to create a final results file; this work was also performed by Heather Eggleston.

DNA extraction

Genomic DNA was isolated from previously frozen mosquitoes with either a phenol/chloroform/Bender buffer assay Macherey-Nagel NucleoSpin® Tissue kit.

Phenol/chloroform/Bender buffer samples were only used in the initial AAEL011708 and AAEL011704 intergenic region amplifications. The kit tissue protocol five was modified at the first step only as follows: frozen microcentrifuge tubes of six female or ten male corpses were removed from -80°C storage and submerged in liquid nitrogen; kit lysis buffer and Proteinase K were added as directed, then a micropestle affixed to a Kontes Pellet Pestle® motor was used to homogenize the samples. DNA concentrations were measured with a Nanodrop 1000. F_2 bright and dim pools were created by combining five ug from each individual pool (six female or ten male corpses).

Primer design, PCR, DNA sequencing & alignments

DNA sequences were exported from the published *Aedes aegypti* genome on VectorBase to Lasergene's SeqBuilder, where primers were generated using the primer design tool (Nene et al., 2007). Primers for AAEL000817 were designed to amplify the intronic regions, exonic regions for reference, between exons two and five. Primer pairs for AAEL05026 and AAEL013438 were designed to amplify the intronic region between exons three and four. Primer pairs for AAEL011708 and its paralogs were designed to amplify the upstream regions these genes; primers for AAEL011708 were designed to also 100% match AAEL014845 and primers for AAEL011704 were designed to also 100% match AAEL014843. BLAST hits of primer sequences were analyzed to decrease the likelihood of off-target primer binding; primer sequences within exons were verified with ten or less hits with no more than 90% similarity and genomic sequences were verified with less than fifty BLAST hits with no more than 70% of the sequence 90% homology. Sequence chromatograms were downloaded with Chromas Lite software. Primer pairs were first verified in the enriched dim and cohorts and the optimal pair for each gene selected before PCR was performed with the QTL samples. PCR was performed on

Bio-Rad thermocyclers with Platinum® Pfx or Phusion® DNA polymerase and their complimentary buffers. Pfx PCR conditions used were 95°C for two minutes, 95°C for 30 seconds, then a variable annealing temperature equal to the lowest primer melting temperature minus at least 5°C, then 68°C for three minutes for elongation; denaturation, annealing and elongation steps were repeated 30-34 times, 68°C for 10 minutes, the held at 4°C. Phusion PCR conditions were adapted from the manufacturer's recommendations to 98°C for 30 seconds, then 98°C for ten seconds, 59°C for 30 seconds, 72°C for 1.5 minutes, repeat denaturation, annealing and elongation steps 34 times, a final extension at 10 minutes and a 4°C hold until use. PCR products were stored at -20°C overnight when necessary. Sanger sequencing of PCR products was conducted at the Virginia Bioinformatics Institute at Virginia Tech and Eton Scientific. Sequence alignments were executed with Lasergene's SeqMan Pro. Sequences were verified to match their target areas by BLAST analysis.

QTL crosses

F_0

All rearing up to EGFP observation was contained to environmental chambers at 28^{0}C +/- 1^{0}C and ~90% relative humidity with a 15 hour light, nine hour dark diurnal cycle. G_{14} embryos of each four times enriched extreme, opposite enrichment crosses were vacuum hatched to synchronize development, counted and aliquoted to pre-acclimated rearing pans at ~100 larvae/L densities. Pan water was changed daily to prevent the buildup of juvenile hormone. Larvae were maintained on flaked fish food and adults on raisins and reverse osmosis-purified water-soaked cotton dyed red to discourage oviposition on the cotton. To avoid confusing the samples by human error, the enriched bright line was hatched seven days prior to hatching the

enriched dim line. Females were kept virgin. At three days post-emergence, G_{14} mosquitoes with the brightest EGFP expression from the brightest enriched line were selected with fluorescence microscopy as described previously. These virgin mosquitoes were carefully placed in individual cages placed to wait for a dim mate of the opposite sex. The following week the G_{14} mosquitoes with the dimmest EGFP expression from the dimmest enriched line were selected with fluorescence microscopy; these mosquitoes were placed in a cage with a waiting bright mate of the opposite sex. Bloodmeals were offered daily until taken. Embryos from two gonotropic cycles were collected on moist strips of chromatography paper, sealed in plastic and stored at room temperature until hatched. Males were snap frozen as soon as the first eggs were lain and females were snap frozen and stored at -80°C as soon as a second cycle of eggs were lain.

F_1

F_1 mosquitoes were hatched and reared in a room warmed to 28°C +/- 5°C at 30-60 % relative humidity with a 15 hour light to nine hour dark diurnal cycle. F_1 embryos were hatched in pans with two liters of reverse osmosis-purified water, segregated by their mate pair ancestry; larvae were maintained on flaked Tetra® fish food. Pan water was changed as needed to purge accumulated juvenile hormone. F_1 pupae were collected and placed in cages segregated by their F_0 lineage and allowed to emerge and mate. Adults were maintained on raisins and cottons soaked with reverse osmosis-purified water dyed red with food coloring to discourage oviposition on the cotton. Bloodmeals of defibrinated sheep's blood were offered approximately every other day. F_2 embryos were collected on moist strips of chromatography paper and stored in sealed plastic at room temperature until use.

F_2

Three, F_2 families were vacuum hatched and reared under controlled environmental conditions stated in previously. Family hatchings were staggered by one week to distribute the experimental workload. Adult females were kept virgin. All mosquitoes were screened at three days post emergence from their pupal casings, in the supine position on an iced petri dish. Individuals with the dimmest and brightest EGFP expression were collected into microfuge tubes in groups of six (female) or ten (male), snap frozen and stored at -80°C until use.

Fluorescence microscopy, photography & analysis

A Canon EOS Rebel digital camera with an adapter to fit the Leica MZ-16FL stereofluorescent dissecting microscope was used to photograph at least 100 males and 100 females of each of three F_2 progeny. A DsRED and EGFP photograph was taken for each mosquito sampled. Default ZoomBrowser EX version 6.7 software settings were modified to "large" and "fluorescent" and used throughout. Head sizes were normalized by adjusting microscope magnification until the width of the most central segment of the ZoomBrowser display grid was filled with the mosquito's head. The same Dell laptop was used for continuity. A fluorescent mercury light source was used for all F_2 progeny photographs.

Fluorescence intensity was assessed with Image J software by averaging measurements from the left and right eye in each photograph. Image J averages for each DsRED and EGFP photograph were entered into JMP Pro 11statistical software for ANOVA analysis for each mate pair group.

Results

RNA-seq comparison of pools of extreme phenotypes

G_{15} progeny from each extreme, opposite, enriched phenotype line were reared under controlled environment conditions to minimize environmental impacts on gene expression. Females from each enriched line were collected and snap frozen at three days post emergence; from the enriched bright cohort, 124 females were collected and from the dim line, 83 females were collected. Only female mosquitoes are known to vector disease; therefore, only female mosquitoes were collected for this analysis. Total RNA from female mosquitoes of each phenotype was extracted for four replicates each to yield eight total libraries. Due to the disparity in sample numbers, each bright sample represents about 30 mosquitoes and each dim sample about 24. The RNA was converted to paired-end, cDNA libraries and sequenced with an Illumina HiSeq2500. Reads that didn't map to the published transcriptome as well as reads very low in abundance and found in three or less replicates were removed from further analysis, leaving 11,562 quality-filtered transcripts. Library sizes ranged from 4.21e7 to 2.50e7 and were normalized for differential expression and abundance assessments with edgeR. Correcting for multiple testing error left 1,120 transcripts significantly different (FDR < 0.0001) between the extreme phenotypes. Of these significantly different transcripts, 590 were significantly upregulated and 530 downregulated in the bright cohort relative to the dim cohort. Log2 transformed differential expression data was plotted against log10 transformed abundance data, revealing roughly 100 transcripts dramatically outlying the general shape of the plotted data (see Figure 3.1). Fifty transcripts of the highest and lowest log2 fold-change values are shown in Tables 3.1 and 3.2, respectively. EGFP expression was 2.86 log2 fold-change higher in the bright substrain (FDR = 1.15e-84), verifying the disparity in EGFP expression in the two, opposite

extreme phenotypes (see Table 3.2). Merging BioMart data with the edgeR data showed that roughly half of the significant, differentially expressed genes have not been characterized (48%). The complete dataset of the filtered transcript and BioMart data is available upon request (Appendix A).

Known RNAi genes in *Aedes aegypti* as well as orthologs to RNAi genes in *Drosophila* were investigated in the abundance and differential expression data (see Table 3.3). A common denominator among the known RNAi genes was high transcript abundance, from 4.09 – 9.57 log10 counts per million. Of this gene set only Argonaute-1 and the Hsp90 ortholog were significantly different between the extreme phenotypes (FDR = 2.31e8 and 0, respectively). Argonaute-1 was upregulated in the poor silencing phenotype, where Hsp90 was dramatically downregulated in the poor silencing phenotype (0.502 and -6.14 log2 fold-change, respectively). Because of this dramatic downregulation and its integral role in RNAi, the Hsp90 ortholog AAEL011708 was singled out for further investigations. Investigations of AAEL011708 in VectorBase showed that this gene has three paralogs, AAEL011704, AAEL014845 and AAEL014843 that are 100% identical in amino acid sequence; their differential expression and abundance data is reported in Table 3.4. AAEL011704 expression was not significantly different (FDR = 0.0627), AAEL014845 was significantly upregulated in the bright group (FDR = 3.25e-40) and AAEL014845 was not found in the final data set. Common throughout AAEL11708 and its paralogs found in the final data set is high abundance (5.52 – 7.98 log counts per million).

QTL crosses & SNP analyses

Crosses were performed with the goal of obtaining two substrains of opposite, genetically distinct cohorts homozygous in genomic regions contributing to the siRNA pathway and heterozygous at genomic regions not contributing to the siRNA pathway. Of the 25 mate pair F_0

crosses attempted, 12 laid the two requisite gonotropic cycles of embryos. Of these 12 crosses, 11 F_1 families of embryos were hatched and intercrossed, segregated by their mate pair lineage. F_2 embryos from families 15, 16 and 20 were chosen and hatched based on criteria to optimize both the quality and amount of DNA obtained as well as represent the two cross possibilities. These criteria were that both parents were snap frozen alive, the number of surviving F_1 progeny relatively high and at least one F_2 family has a five-time-enriched dim paternal grandparent and five-time-enriched bright maternal grandparent and at least one family has the converse lineage. Information for each of these criteria for each of the successful mate pair cross is reported in Table 3.5. Between 797 and 1,033 F_2 individuals were screened for EGFP expression in each family and the brightest 12-13% and dimmest 10-11% female mosquitoes were selected, snap frozen and stored (see Table 3.6). To estimate and document the quality of the brightest and dimmest EGFP expression selections, photographs were taken and analyzed from samples of ~100 females for each family (see Figure 3.2). This data verifies that the representative individuals selected where phenotypically different, though some of the best qualifying mosquitoes were not selected. Female mosquito DNA was extracted, measured and pooled by extreme phenotype for each family to create an average F_2 phenotype pool each for dim and bright EGFP expression.

To demonstrate SNP phenotype associations in genes not known to influence RNAi, regions within significant, highly abundant genes not known to influence RNAi silencing abilities sequenced for SNP analysis. The genes selected meet four criteria: similarity in abundance and fold change to the AAEL011708 transcript, having a known function that is unlikely to contribute to the siRNA pathway, having few paralogs and having low percent similarity to those paralogs. The genes chosen were AAEL000817, AAEL005026 and

AAEL013438, which have one paralog with the highest percent similarity being 50, 27 paralogs with the highest percent similarity 56% and seven paralogs at most 29% similar, respectively. Since intronic regions generally show a high frequency of SNPs these areas were targeted for amplification and SNP analysis. To capture SNPs within introns of these genes, primer pairs were designed for DNA sequencing results that would yield enough exon sequences to match the data as well as the intronic sequences. Primers were designed to bind within exons of these genes, where sequences are generally conserved and relatively unique for optimal target priming. Primers within exons should also yield sequencing data that can be verified to the VectorBase transcript sequence. The primer pairs for each gene and other relevant details are compiled in Table 3.7. Sequencing data uncovered distinct SNPs between the F_0s and F_2s in all three families 15, 16 and 20; AAEL000817 showed two distinct SNPs, AAEL005026 showed three and AAEL013438 showed four (see Tables 3.8-3.11). Interestingly, the two SNPs in the AAEL000817 sequence data show an inheritance pattern consistent with segregation with RNAi genetic components; the F_2 pools show homozygosity for bases consistent with the F_0 grandparent that matches their phenotype. Inheritance of the SNPs in the AAEL005026 and AAEL013438 data is roughly consistent that these SNPs are independent of the siRNA pathway phenotypes; F_2 pools show heterozygosity for bases in both F_0 grandparents. The first SNP presented in each gene table shows that the base for the VectorBase sequence is most often conserved in the dim F_0 grandparent.

Due to the dramatic downregulation of the Hsp90 ortholog in the pool representing poor silencing abilities and its known integral role in RNAi in *Drosophila*, AAEL011708 and its paralogs were singled out for SNP analyses. Since Hsp90 was dramatically downregulated in the group with impaired RNA silencing, it was theorized that this dearth in Hsp90 could explain the

impaired function of the siRNA pathway. According to the *Drosophila* model, low levels of Hsp90 may result in decreased binding of Argonaute-2 to the siRNA duplex and decreased unwinding of the dsRNA and impairment of sequence specific mRNA knockdown through the siRNA pathway; thus resulting in lower levels of RNAi (Myoshi et al., 2010). To investigate the regulatory regions of these genes, primer pairs were designed to amplify the upstream sequences of AAEL011708, AAEL011704, AAEL014845 and AAEL014843; these primer sequences and other relevant details are compiled in Table 3.7. PCR with primer pairs designed to amplify AAEL011708 and AAEL14845 sequences were unsuccessful. PCR with primers pairs designed to amplify AAEL011704 and AAEL014843 sequences showed bands consistent with the predicted length of the Vectorbase AAEL011704 sequence; these bands were purified and sequenced. Only 189 bases of this sequence data was useable and three SNPs were detected within. The consensus sequence of this useable portion was BLASTed and matched most closely to the upstream region of AAEL011704 (98.7% match, E-value = 1e-88) but also closely matched AAEL014843 (97.9% match, E-value = 2e-87). The inheritance patterns of these three SNPs do not demonstrate strong linkage to genes influencing the extreme, opposite silencing phenotypes (see Table 3.9). The first SNP shows strong fluorescence for adenine in the bright F_0 parent and cytosine in the dim F_0 parent in all three families. If the presence of adenine at this locus is associated with poor RNA silencing abilities, then this base should be dominantly present in the F_2 bright pool sequence data; instead a mixture of adenine and cytosine was shown in the bright F_2 pool. The second and third SNPs show similar inheritance patterns, where the base seen in the like-phenotype F_0 is not conserved in the corresponding F_2 phenotype pool sequence data, rather, the F_2 pools show a mixture of both F_0 grandparent bases.

Discussion

Transcriptome data gives a snapshot of gene expression at a given time in development (Reviewed by Wolf, 2013). Here enriched, extreme, opposite EGFP expression phenotypes indicating either a predisposition for efficient viral knockdown, observed by low, or dim EGFP expression or poor viral knockdown, demonstrated by high, or bright EGFP expression in sensor mosquitoes were compared at adulthood. This comparison revealed hundreds of potential contributors to these RNA silencing abilities, with 1,120 transcripts significantly different between the two extreme, opposite phenotypes. Here, only four of these genes were investigated, leaving hundreds more to be studied further.

Interestingly, gene AAEL000817, which has no known link to RNAi in *Aedes aegypti*, demonstrated SNP inheritance patterns that suggest that these bases segregate with siRNA silencing abilities. This gene is characterized as a rhomboid protein and its ortholog in *Drosophila* is associated with development and the epidermal growth factor signaling pathway (Sturtevant et al., 1993; Ruohola et al., 1993). This finding may indicated that AAEL000817 is relatively closely linked to genomic regions that influence RNAi silencing abilities or that this protein does play a role in siRNA silencing in *Aedes aegypti*, though there is no evidence in the literature supporting this postulate. Relevance of this gene in the siRNA can be tested by monitoring EGFP fluorescence intensity changes in sensor mosquitoes where AAEL000817 expression is exaggerated or knocked down.

Investigations into AAEL011708 and its paralogs uncovered information that indicates possible gene duplication events. The organization and sequences of the regions between AAEL011708 and AAEL011704 and between AAEL014845 and AAEL014843 are similar. In both pairs, the genes are arranged in opposite directions and AAEL11708 and AAEL014845

both contain a very large intron proximal to the 5' untranslated region. Furthermore, the intergenic regions between these genes is rich in repeats, which may indicate gene duplication facilitated by selfish elements. This highly repetitive region could have also confounded the assembly of the reference genome, creating paralogs that are artifacts. Further validating this speculation is that sequences among AAEL011708 and its paralogs are 100% similar in amino acid sequence *(*VectorBase *Aedes aegypti* genes (AaegL3 (AaegL3.3)). However, it is important to note that the genetic drift accumulated from the last common ancestor of the reference *Aedes aegypti* Liverpool genome and our laboratory Liverpool strain is unknown. Genetic bottlenecking events from previous crosses as well as the crosses described herein could have resulted in a loss or gain in gene copy number.

Among the siRNA proteins known to contribute to the siRNA pathway in *Drosophila*, a highly significant 6.14 log2 fold-change deficit of an Hsp90 ortholog was detected in pool with impaired silencing abilities. The three SNPs detected in the upstream region the AAEL011708 paralog successfully amplified demonstrated inheritance patterns that suggest that these SNPs do not segregate with RNAi silencing ability phenotypes. Since this disparity in AAEL011708 expression was demonstrated only once, additional expression comparisons are needed to confirm the dramatic downregulation of Hsp90 seen here associated with poor RNAi silencing abilities. Such studies could include ddPCR, RT-PCR and additional RNA-seq comparisons of individuals representing the extreme, opposite silencing phenotypes. SNP analysis can be performed in comparisons of sequences of this transcript between the two extreme, opposite RNAi silencing phenotypes; SNPs in these regions could result in mutated protein products explaining an impaired state, as is the case with human disease sickle cell anemia (Gilman and Huisman, 1985). The relevance of this gene in RNAi silencing abilities can be further

investigated in experiments where AAEL011708 is either overexpressed or knocked down in sensor mosquitoes. Differences in EGFP expression in such studies could elucidate an effect in siRNA silencing abilities in the important disease vector, *Aedes aegypti*.

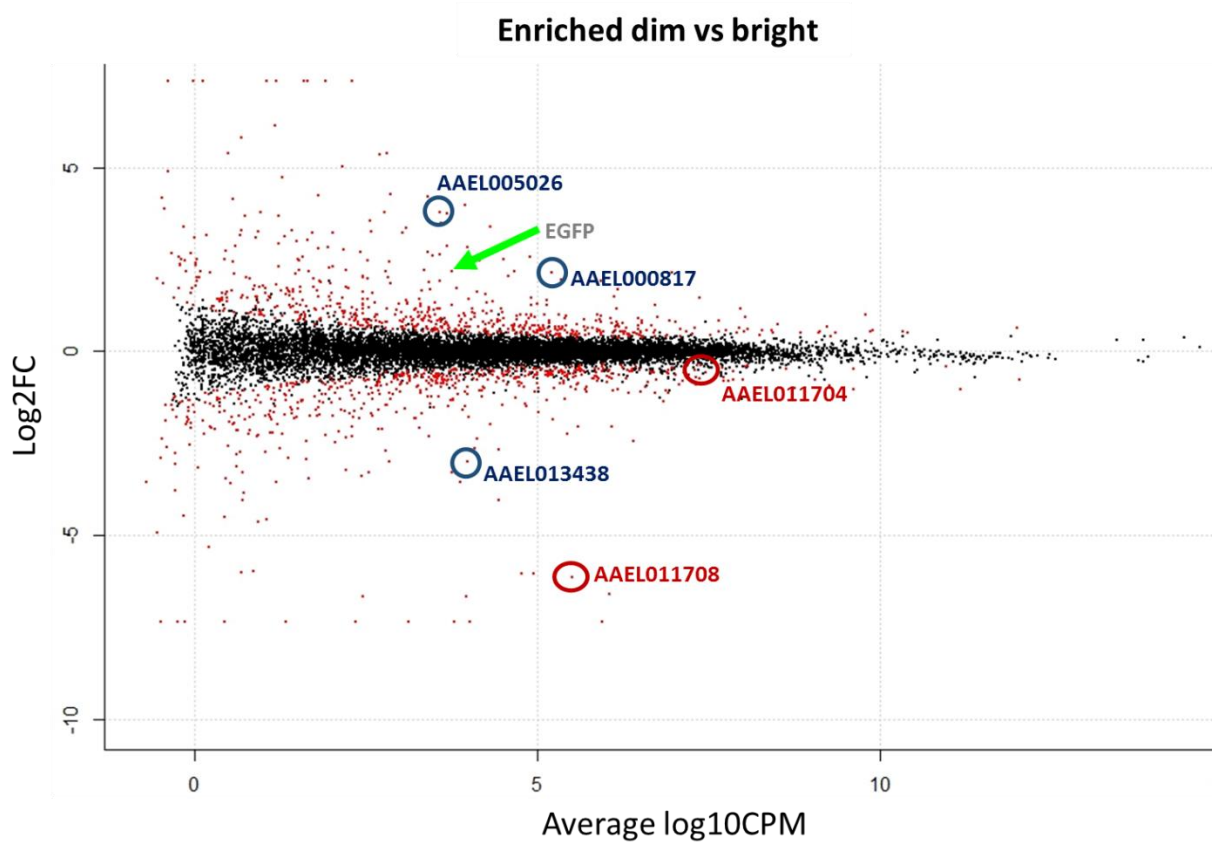


Figure 3.1 RNA-seq comparisons of the enriched, extreme phenotypes reveals 1,120 genes that are significantly different. Shown is a plot of bright cohort differential expression against abundance, relative to the dim cohort. Transcripts in red are significantly different between the cohorts (FDR < 0.0001) while black indicates transcripts that are not significantly different. The green arrow demarcates the EGFP transcript. AAEL000817, AAEL005026, AAEL013438, AAEL0011708 and AAEL00111704 data points are circled and labeled.

Table 3.1 The fifty most significantly downregulated transcripts in the group representing poor silencing abilities. Shown are transcripts in the top 50 lowest log2 fold-change values. Listed are transcript identity, protein description, location within and scaffold number annotations as well as the abundance (log10 counts per million values), differential expression (log2 fold-change values) and FDR values. Highlighted in gold is the Hsp90 ortholog; highlighted in green are AAEL013438 and EGFP.

Transcript ID	Log2 FC	Log10 CPM	FDR	Gene description	Gene start (bp)	Gene end (bp)	Scaffold
AAEL011514-RA	-12.84	5.96	0	cytochrome c oxidase subunit VIIC putative [Source:VB External Description;Acc:AAEL011514]	394152	394797	supercont 1.586
AAEL007774-RA	-12.24	3.12	3.35E-150		342469	355080	supercont 1.279
AAEL005212-RA	-10.91	4.02	0		29618	30566	supercont 1.145
AAEL016974-RA	-10.47	1.34	5.67E-89		605708	606262	supercont 1.215
AAEL015285-RB	-9.26	2.35	3.82E-130		68186	71673	supercont 1.1707
AAEL014880-RA	-8.95	-0.14	5.62E-42		92407	102477	supercont 1.1322
AAEL012189-RA	-8.88	-0.24	1.65E-34	multidrug resistance protein 1 (ATP-binding cassette C1) [Source:VB External Description;Acc:AAEL012189]	689083	689967	supercont 1.664

(continued)

Table 3.1 *(cont)*

Transcript ID	Log2 FC	Log10 CPM	FDR	Gene description	Gene start (bp)	Gene end (bp)	Scaffold
AAEL014326-RA	-8.60	-0.49	6.76E-35	out at first protein [Source:VB External Description;Acc:AAEL014326]	7362	24110	supercont 1.1072
AAEL001419-RA	-7.34	0.46	8.63E-58	nocturnin [Source:VB External Description;Acc:AAEL001419]	38309	67288	supercont 1.32
AAEL013638-RA	-6.85	3.79	8.28E-199		288279	291228	supercont 1.888
AAEL006323-RA	-6.67	3.97	1.72E-76		1521988	1524686	supercont 1.200
AAEL014896-RA	-6.66	2.46	3.17E-139		117763	119619	supercont 1.1328
AAEL011708-RA	-6.14	5.52	0	heat shock protein [Source:VB External Description;Acc:AAEL011708]	5416	14915	supercont 1.604
AAEL002612-RA	-6.05	4.96	5.66E-87		532168	532798	supercont 1.61
AAEL014897-RA	-5.99	0.86	1.11E-60	metalloprotease [Source:VB External Description;Acc:AAEL014897]	3358	33667	supercont 1.1328
AAEL013743-RA	-5.33	0.22	3.28E-37		151781	152602	supercont 1.908
AAEL000348-RA	-4.91	-0.53	2.93E-30		629049	652044	supercont 1.6
AAEL009036-RA	-4.64	0.95	1.22E-44		112195	113403	supercont 1.363
AAEL012976-RA	-4.56	1.06	1.32E-45		45719	79334	supercont 1.770
AAEL004965-RA	-4.48	-0.14	5.95E-33		1526244	1527300	supercont 1.136

(continued)

Table 3.1 *(cont)*

Transcript ID	Log2 FC	Log10 CPM	FDR	Gene description	Gene start (bp)	Gene end (bp)	Scaffold
AAEL006586-RA	-4.03	4.45	1.21E-38	serine protease [Source:VB External Description;Acc:AAEL006586]	715330	717211	supercont 1.211
AAEL017422-RA	-3.84	0.73	2.36E-18		2424415	2424959	supercont 1.20
AAEL007394-RB	-3.59	0.54	1.48E-26		1259775	1260660	supercont 1.253
AAEL010578-RA	-3.54	-0.69	4.86E-17	mixed-lineage leukemia protein mll [Source:VB External Description;Acc:AAEL010578]	808310	838011	supercont 1.488
AAEL005124-RA	-3.54	1.19	4.73E-24		1572996	1611726	supercont 1.142
AAEL008884-RA	-3.54	3.87	8.51E-140		501638	523625	supercont 1.354
AAEL015072-RA	-3.45	0.47	1.56E-27	carboxylesterase [Source:VB External Description;Acc:AAEL015072]	59249	60765	supercont 1.1449
AAEL004850-RA	-3.40	2.44	1.69E-84		1199075	1215618	supercont 1.132
AAEL012838-RA	-3.29	3.76	2.27E-131		662479	665121	supercont 1.747
AAEL009002-RA	-3.28	2.56	3.84E-50		462892	511764	supercont 1.361
AAEL008193-RA	-3.28	0.72	1.28E-19		870820	875546	supercont 1.305
AAEL013438-RA	-3.00	3.99	3.67E-94	LIM domain-binding protein 3 putative [Source:VB External Description;Acc:AAEL013438]	427655	440361	supercont 1.846

(continued)

Table 3.1 *(cont)*

Transcript ID	Log2 FC	Log10 CPM	FDR	Gene description	Gene start (bp)	Gene end (bp)	Scaffold
AAEL015595-RA	-2.94	1.64	3.87E-34		206	1959	supercont 1.3596
AAEL010671-RA	-2.91	1.38	9.77E-31	oxidoreductase [Source:VB External Description;Acc:AAEL010671]	595733	596717	supercont 1.495
AAEL013799-RB	-2.89	-0.49	1.30E-16		160004	258093	supercont 1.924
AAEL005960-RA	-2.79	0.62	2.83E-11		1351322	1364250	supercont 1.182
AAEL015294-RA	-2.78	-0.28	2.89E-17	serine-type enodpeptidase [Source:VB External Description;Acc:AAEL015294]	19920	21021	Supercont 1.1737
AAEL015414-RA	-2.74	0.67	3.74E-21		43670	44505	supercont 1.2126
AAEL002641-RA	-2.74	1.05	1.15E-27		816953	819283	supercont 1.62
AAEL010139-RA	-2.67	4.45	2.84E-10	serine protease putative [Source:VB External Description;Acc:AAEL010139]	51590	67346	supercont 1.456
AAEL012305-RA	-2.65	4.10	8.98E-102	hect E3 ubiquitin ligase [Source:VB External Description;Acc:AAEL012305]	525999	536695	supercont 1.677
AAEL005697-RA	-2.62	1.25	6.85E-33		1136254	1189099	supercont 1.170
AAEL011227-RA	-2.57	0.51	1.18E-14		167933	202027	supercont 1.556

(continued)

Table 3.1 *(cont)*

Transcript ID	Log2 FC	Log10 CPM	FDR	Gene description	Gene start (bp)	Gene end (bp)	Scaffold
AAEL012911-RA	-2.46	2.07	1.24E-26	leucine-rich immune protein (Coil-less) [Source:VB Community Annotation;Acc:AAEL012911]	409613	419272	supercont 1.758
AAEL014932-RA	-2.44	6.41	1.49E-129	transcription factor btf3 [Source:VB External Description;Acc:AAEL014932]	102121	103266	supercont 1.1345
AAEL001674-RA	-2.44	1.63	7.03E-11	serine-type enodpeptidase [Source:VB External Description;Acc:AAEL001674]	1642121	1643095	supercont 1.39
AAEL010999-RA	-2.39	-0.47	2.29E-11		436092	436508	supercont 1.532
AAEL015684-RA	-2.38	4.14	1.19E-79		2285107	2290958	supercont 1.94
AAEL003052-RA	-2.34	2.22	1.73E-37		1347759	1456908	supercont 1.76
AAEL009178-RA	-2.31	2.30	3.24E-11	Gram-Negative Binding Protein (GNBP) or Beta-1 3-Glucan Binding Protein (BGBP). [Source:VB Community Annotation;Acc:AAEL009178]	525277	526577	supercont 1.376

Table 3.2 The fifty most significantly upregulated transcripts in the group representing poor silencing abilities. Listed are transcript identity, protein description, location within and scaffold number annotations as well as the abundance (log10 counts per million values), differential expression (log2 fold-change values) and FDR values. EGFP, AAEL005026 and AAEL000817 information is highlighted.

Transcript ID	logFC	Log CPM	FDR	Gene description	Gene start (bp)	Gene end (bp)	Scaffold
AAEL014262-RA	11.33	2.3	9.2E-124	translation initiation factor eif-2b alpha subunit [Source:VB External Description;Acc:AAEL014262]	116425	120235	supercont 1.1051
AAEL012961-RA	10.66	1.65	1.4E-108	vacuolar protein sorting 26 vps26 [Source:VB External Description;Acc:AAEL012961]	215195	267134	supercont 1.768
AAEL017517-RA	10.61	1.6	1.2E-49		82872	83392	supercont 1.1082
AAEL006758-RA	10.04	1.06	1.5E-61		703286	704050	supercont 1.220
AAEL017181-RA	9.05	0.13	8.68E-38	GPCR Muscarinic Acetylcholine Family [Source:VB Community Annotation;Acc:AAEL017181]	277986	372439	supercont 1.834

(continued)

Table 3.2 *(cont)*

Transcript ID	logFC	Log CPM	FDR	Gene description	Gene start (bp)	Gene end (bp)	Scaffold
AAEL006574-RA	8.91	-0.01	1.1E-42	nuclear transcription factor x-box binding 1 (nfx1) [Source:VB External Description;Acc:AAEL006574]	1359901	1370864	supercont1.211
AAEL009602-RB	8.52	-0.38	3.67E-31		198175	199424	supercont1.412
AAEL014881-RA	8.5	1.92	2.1E-108		34452	58633	supercont1.1322
AAEL003981-RA	6.89	1.2	4.59E-50		155941	158330	supercont1.103
AAEL008191-RA	6.12	1.18	8.51E-63	predicted protein [Source:VB External Description;Acc:AAEL008191]	1262958	1271752	supercont1.305
AAEL010997-RA	5.82	0.7	2.83E-49	myosin light chain kinase [Source:VB External Description;Acc:AAEL010997]	40712	77106	supercont1.532
AAEL013301-RA	5.38	0.49	3E-45		250878	251470	supercont1.818
AAEL012423-RA	5.38	2.81	5E-113		212258	213073	supercont1.695
AAEL010133-RA	5.03	2.17	4.6E-41		588036	596313	supercont1.455
AAEL005612-RA	4.88	-0.39	7.43E-29		1605760	1616882	supercont1.164
AAEL015058-RA	4.73	1.28	6.91E-68		78224	96522	supercont1.1433
AAEL004031-RA	4.7	4.18	1.3E-237		2026968	2035502	supercont1.105
AAEL011190-RA	4.26	2.87	3.7E-130		229091	230021	supercont1.550

(continued)

Table 3.2 *(cont)*

Transcript ID	logFC	Log CPM	FDR	Gene description	Gene start (bp)	Gene end (bp)	Scaffold
AAEL000325-RA	4.23	1.81	5.8E-57	cytochrome P450 (CYP325S2) [Source:VB Community Annotation;Acc:AAEL000325]	399772 8	401409 9	supercont 1.6
AAEL008506-RA	4.2	3.41	7.6E-163	stromal antigen [Source:VB External Description;Acc:AAEL008506]	127654	136172	supercont 1.330
AAEL005883-RA	4.16	-0.46	2.02E-21		109656 4	110225 7	supercont 1.178
AAEL012196-RA	4.13	0.57	3.85E-37		100700	129720	supercont 1.665
AAEL005673-RA	3.97	3.96	8E-145	Serine Protease Inhibitor (serpin) likely cleavage at K/F. [Source:VB Community Annotation;Acc:AAEL005673]	286950	289171	supercont 1.169
AAEL006568-RB	3.88	-0.43	7.93E-18	serine protease [Source:VB External Description;Acc:AAEL006568]	742487	743767	supercont 1.211
AAEL015039-RA	3.76	3.7	1.9E-158		11727	25890	supercont 1.1414
AAEL007040-RA	3.76	2.78	1.56E-63	lozenge [Source:VB External Description;Acc:AAEL007040]	134095 9	139434 0	supercont 1.232
AAEL000521-RA	3.76	0.97	2.37E-32		402409 8	402847 6	supercont 1.9
AAEL006349-RA	3.69	1.24	2.37E-28		163474 4	163663 9	supercont 1.201

(continued)

Table 3.2 *(cont)*

Transcript ID	logFC	Log CPM	FDR	Gene description	Gene start (bp)	Gene end (bp)	Scaffold
AAEL006376-RA	3.69	0.74	2.05E-22	trypsin putative [Source:VB External Description;Acc:AAEL006376]	1099685	1100733	supercont 1.202
AAEL005026-RA	3.5	3.6	5.26E-60	ATP-dependent bile acid permease [Source:VB External Description;Acc:AAEL005026]	1168407	1184363	supercont 1.139
AAEL013352-RA	3.42	3.54	2.64E-85	lethal(2)essential for life protein l2efl [Source:VB External Description;Acc:AAEL013352]	117019	118025	supercont 1.826
AAEL012978-RA	3.34	3.09	9E-117		229538	247690	supercont 1.770
AAEL005322-RA	3.27	1.43	1.17E-33	unknown wavelength sensitive opsin [Source:VB Community Annotation;Acc:AAEL005322]	1942356	1943586	supercont 1.151
AAEL000463-RA	3.26	0.94	7.16E-27		524640	525381	supercont 1.8
AAEL005110-RA	3.23	1.73	4.62E-38		671971	672980	supercont 1.142
AAEL010016-RA	3.23	2.03	5.21E-46		107110	107750	supercont 1.447
AAEL014500-RB	3.22	3.04	1.77E-31	cation chloride cotransporter [Source:VB External Description;Acc:AAEL014500]	124160	170371	supercont 1.1137
AAEL014714-RA	3.17	0.19	1.53E-20		134958	135803	supercont 1.1233

(continued)

Table 3.2 *(cont)*

Transcript ID	logFC	Log CPM	FDR	Gene description	Gene start (bp)	Gene end (bp)	Scaffold
AAEL006578-RA	3.15	0.67	5.93E-21		1278752	1293844	supercont 1.211
AAEL012771-RA	3.14	1.33	5.13E-39	leucine-rich immune protein (Coil-less) [Source:VB Community Annotation;Acc:AAEL012771]	448592	449769	supercont 1.738
AAEL005702-RA	3.09	0.01	1.34E-18		1866063	1867266	supercont 1.170
AAEL001380-RA	3.08	0.9	1.49E-32	lim homeobox protein [Source:VB External Description;Acc:AAEL001380]	1643977	1775230	supercont 1.31
AAEL009335-RA	3.03	1.65	1.02E-44	adhesion regulating molecule 1 (110 kda cell membrane glycoprotein) [Source:VB External Description;Acc:AAEL009335]	337000	350541	supercont 1.390
AAEL006109-RA	2.99	1.49	1.94E-05	odorant binding protein OBP23 [Source:VB Community Annotation;Acc:AAEL006109]	216817	217969	supercont 1.189
AAEL013346-RA	2.86	0.6	1.01E-26	lethal(2)essential for life protein l2efl [Source:VB External Description;Acc:AAEL013346]	100916	101824	supercont 1.826
EGFP505	2.86	3.70	1.15E-84	Enhanced green fluorescent protein	N/A		

(continued)

Table 3.2 *(cont)*

Transcript ID	logFC	Log CPM	FDR	Gene description	Gene start (bp)	Gene end (bp)	Scaffold
AAEL005693-RA	2.83	3.98	6.67E-33	mitochondrial NADH:ubiquinone oxidoreductase B16.6 subunit putative [Source:VB External Description;Acc:AAEL005693]	42881	43684	supercont 1.170
AAEL003114-RA	2.78	2.18	3.26E-42		2842954	2843830	supercont 1.78
AAEL000496-RB	2.78	1.82	2.08E-44		1759252	1766133	supercont 1.9
AAEL005687-RA	2.73	0.46	1.08E-22	protein serine/threonine kinase putative [Source:VB External Description;Acc:AAEL005687]	1568687	1569736	supercont 1.170
AAEL005974-RA	2.71	3.41	4.79E-28	homeobox protein dbx [Source:VB External Description;Acc:AAEL005974]	275908	335192	supercont 1.183
AAEL012985-RA	2.66	-0.33	1.04E-10		165295	166093	supercont 1.772
AAEL000817-RA	2.14	5.22	4.48E-71	rhomboid [Source:VB External Description;Acc:AAEL000817]	1462477	1486386	supercont 1.17

Table 3.3 RNA-seq data pertaining to known RNAi genes. *Drosophila* siRNA and miRNA protein orthologues in *Aedes aegypti* and their abundance, differential expression and FDR data from the transcriptome comparisons.

Drosophila siRNA protein	***Ortholog in Aedes aegypti***	***logFC compared to Dim control***	***Abundance***	***FDR***
Argonaute-2	AAEL017251	0.0686	7.53	0.619
Argonaute-1	AAEL015246*		Not found in dataset	
	AAEL012410*	0.502	5.94	2.31e-8
Dicer-2	AAEL006794	0.136	7.59	0.294
Dicer-1	AAEL001612	-0.0199	5.39	0.905
Loqs	AAEL008687	-0.114	6.63	0.326
R2D2	AAEL011753	-0.176	4.09	0.180
Hsp90 (Hsp83)	AAEL011708	-6.14	5.32	0
Hsc70-4		No gene model found		
Droj2	AAEL005165	-0.129	9.57	0.305
Hop	AAEL012553	-0.2292	6.85	0.884
P23	AAEL014943	0.763	7.94	0.549
	*Identical paralogs; data collapsed to one transcript, AAEL012410.			

Table 3.4 RNA-seq and annotation data regarding the hsp90 ortholog and its paralogs.

RNA-seq results and BioMart details of AAEL011708 and its known paralogs. Listed are annotation data such as transcript identity, scaffold and location as well as edgeR output abundance (log10 counts per million), differential expression (log2 fold change) and FDR values.

ID	Log2 FC	Log10 CPM	FDR	Gene start (bp)	Gene end (bp)	Gene description	Scaffold
AAEL011708-RA	-6.14	5.516	0	5416	14915	heat shock protein [Source:VB External Description;Acc:AAEL011708]	SC 1.604
AAEL011704-RA	-0.195	7.51	0.0627	22562	25978	heat shock protein [Source:VB External Description;Acc:AAEL011704]	SC 1.604
AAEL014845-RA	1.17	7.98	3.25e-40	48934	55836	heat shock protein [Source:VB External Description;Acc:AAEL014845]	SC 1.1292
AAEL014843	Not found in dataset						

Table 3.5 Mate pair cross details. F_0 extreme phenotype mate pair crosses details and F_1 population sizes are listed. The phenotype of the female parent is denoted by a "B" for bright and "d" for dim. Whether the parents were snap frozen alive is listed, along with the number of embryos, number of pupae and calculated percent of embryos that hatched and survived to the pupal stage.

F_0 female parent	Both F_0 frozen alive?	cross ID #	# embryos	# F_1 pupae	% hatched
d	no	3	F_1s not hatched, since acquired late in study		
d	yes	4	88	55	63%
d	yes	6	94	18	19%
B	yes	7	125	41	33%
B	no	9	158	84	53%
B	yes	10	139	92	66%
d	yes	13	107	61	57%
d	yes	15	98	81	83%
d	yes	16	137	105	77%
d	yes	17	114	70	61%
B	yes	20	139	89	64%
B	yes	22	147	72	49%

Table 3.6 F_2 extreme phenotype pool selection summary. Listed by family are the total number of mosquitoes in each family screened for EGFP expression, the numbers of bright and dim female mosquitoes selected and the amount of DNA extracted.

	Total # screened	females	# brightest females	DNA extracted (ug)	# dimmest females	DNA extracted (ug)
Mate pair cross 15	1033	671	78	521	65	526
	% of total screened	65%	12%		10%	
Mate pair cross 16	965	906	112	537	98	690
	% of total screened	94%	12%		11%	
Mate pair cross 20	797	622	81	389	67	310
	% of total screened	78%	13%		11%	

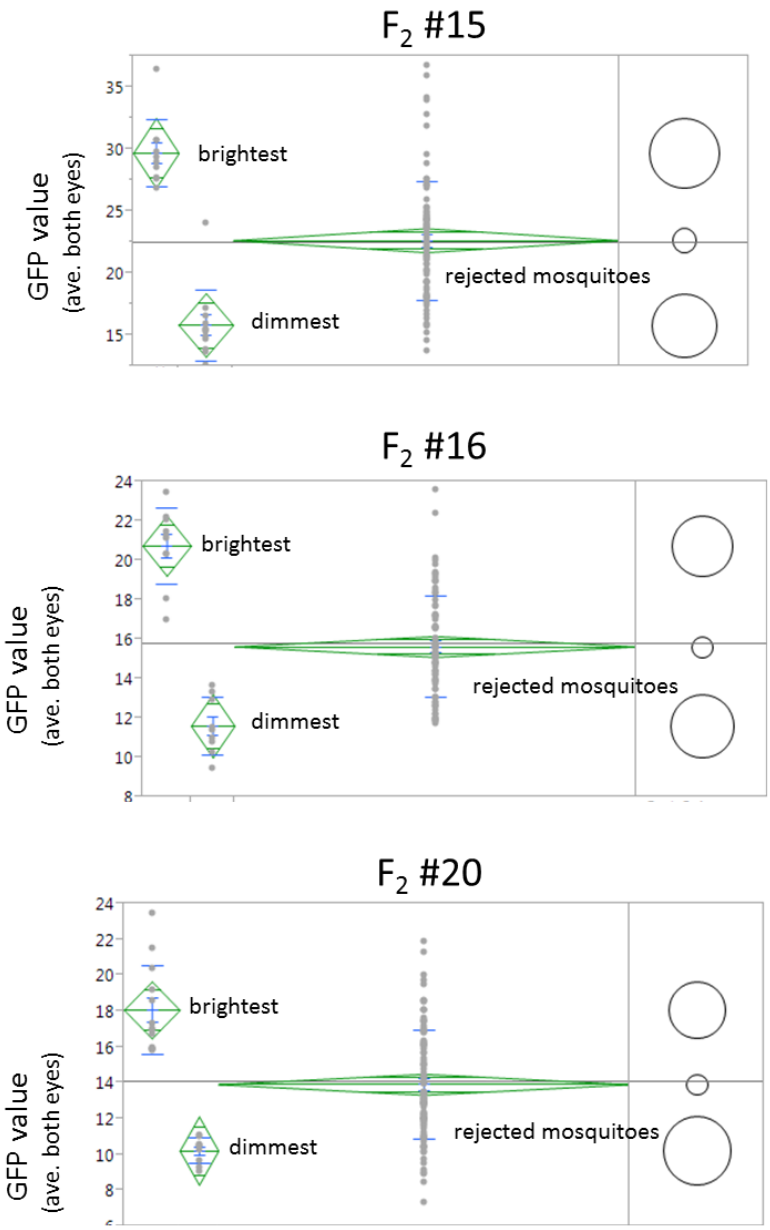
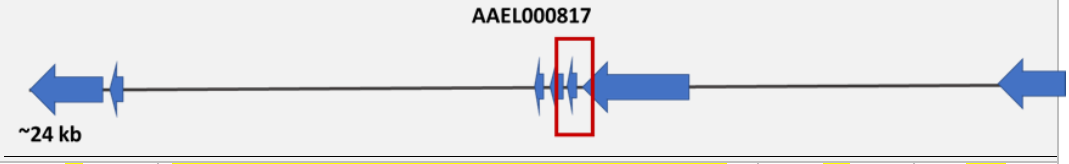
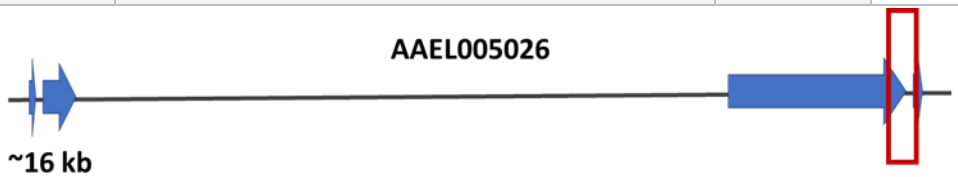
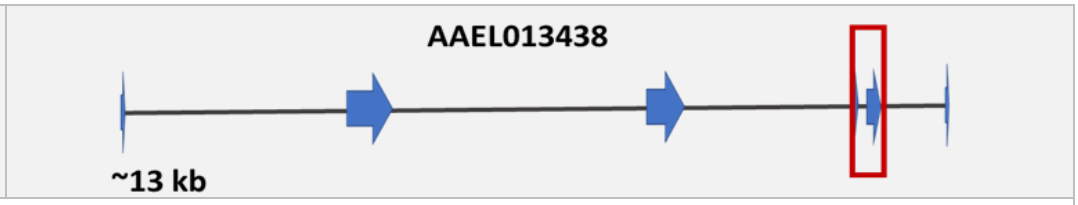
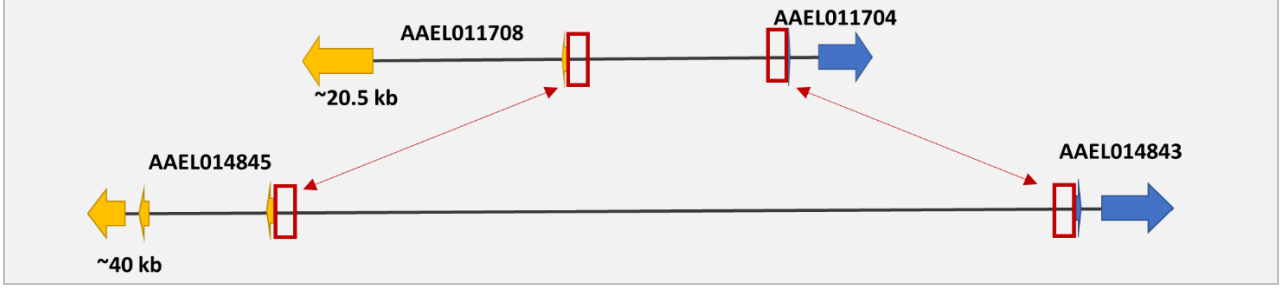


Figure 3.2 Quality of samples in F_2 extreme phenotype selections for each family. EGFP intensity measurement comparisons in samples of brightest and dimmest selections from F_2 pools. Each of three replicates is shown: family 15, 16 and 20. Shown are ANOVA analyses of Image J EGFP photo values for each mate pair family, grouped by the brightest and dimmest selections as well as mosquitoes not collected. Sample sizes are 107, 102 and 108 female

mosquitoes for each family, respectively. Student's t-tests are shown in Venn diagrams to the right of each ANOVA with alpha at 0.05.

Table 3.7 Target gene amplification regions and primer details. Primer sequences, their number of matches in the genome, expected band sizes and orientations are listed. Target regions within each gene are depicted below the primer pairs: exons are denoted by block arrows and red boxes indicate the region the primers listed target; red arrows indicate target regions where primer pairs were designed to be 100% complimentary to both genes. The primer pairs used to sequence the intronic regions in the control genes and upstream regions for Hsp90 are highlighted.

Target gene/s	Forward or reverse	Primer sequence, 5'-3'	BLAST scaffold matches	Expected band size
AAEL000817	F	GGCGATAAAACTGATCTGGATATTCACTGGCG	10	895
	R	GTCGGATGGGGGATGGAAGACAAGACGG	9	
	F	CAAAGCGACGTGAAGAGGCGATAAAACTG	10	902
	R	GGGATGGAAGACAAGACGGCTCAAGTGC	7	
				
AAEL005026	R	CGACGATGTTTTATCTTCCAACAGTGTCTGAGG	10	538
	F	TAGGGTTACATCAGTTGCGAAGCCGTCTG	2	
	F	CGGTTGAATCAAAATCTGAAAACGCCAC	10	814
	R	CATCCGATAAAACGACGATGTTTTATCTTCC	10	
				
AAEL013438	R	GCAGTTGAAGCACTGACTGTGGTAGTTGTTG	10	408
	F	GACATCGGATTCGTAGAGGAGAAGGGTG	6	
	F	CGTAGAGGAGAAGGGTGATCTGTACTGCGAG	10	386
	F	CACTGACTGTGGTAGTTGTTGTTCAGGGCCTC	10	

				
AAEL011704 & AAEL014843	F	CTCTCTGCTCTGGGCTCTGGAACAGTCTC	7	296
	R	CACGTTGCTTGCCTCGATCTAACTCGC	4	
	F	GCCAATCCCACTTCTCCTCCACGAAAGAG	19	489
	R	GCACTTGCAGAATATTTCACGTTGCTTGCCTCG	26	
AAEL011708 & AAEL014845	R	CCTGCAATAATCCATCCTCCAAATAAATCC	51	493
	F	GTCCTTCCGGATCACTAGAAACACAACAC	44	
	R	GCAATAATCCATCCTCCAAATAAATCCAC	39	499
	F	GTCCTTCCGGATCACTAGAAACACAACAC	44	
				

Tables 3.8 – 3.12 Chromatograms for bases in AAEL000817, AAEL005026, AAEL013438 and AAEL011704 SNPs by phenotype. AAEL000817 sequence data shows inheritance patterns consistent with those expected if SNPs segregate with RNAi SNPs, where AAEL005026, AAEL013438 and AAEL011704 sequence data shows inheritance patterns consistent with that expected if SNPs do not segregate with siRNA phenotypes. Transcript names are indicated at the top of each table. For reference are the orientation of the VectorBase sequence, the ten bases following each SNP in the VectorBase sequence and the VectorBase and consensus base at each SNP. Shown in each grid are chromatograms of sequence data in duplicate for each parent and offspring pool each for families 15, 16 and 20. Families are identified by their mate pair cross numbers.

AAEL000817

VectorBase <<	SNP before TCTGTTTAAT			SNP before ATTTGAAAAA		
Consensus	W			K		
VectorBase	T			G		
Family	15	16	20	15	16	20
F_0 Bright	A	A	A	G	G	G
F_0 Dim	T	T	T	K	K	K
F_2 Bright	A	A	A	G	G	G
F_2 Dim	T	T	T	K	K	K
5x enriched Bright		A			G	
5x enriched Dim		T			K	

AAEL005026

VectorBase >>	SNP before CCTACTCGGA			SNP before CAACTGATTT			SNP before TAACGGTGGC		
Consensus		R			Y			M	
VectorBase		G			C			C	
Family	15	16	20	15	16	20	15	16	20
F_0 Bright	R	A	R	Y	C	Y	M	A	M
F_0 Dim	G	G	G	T	T	T	C	C	C
F_2 Bright	R	R	R	Y	Y	Y	M	M	M
F_2 Dim	R	R	R	Y	Y	Y	M	M	M
5x enriched Bright		A			C			A	
5x enriched Dim		G			T			C	

AAEL013438 SNPs 1 & 2

VectorBase <<	SNP before AAATTTCAGA			SNP before GGGGCCAAGA		
Consensus	W			R		
VectorBase	A			T		
Family	15	16	20	15	16	20
F_0 Bright	T	K	K	G	D	D
F_0 Dim	A	R	R	A	A	W
F_2 Bright	W	W	W	R	R	R
F_2 Dim	W	W	D	R	R	R
5x enriched Bright		T			G	
5x enriched Dim		A			A	

AAEL013438 SNPs 3 & 4

VectorBase <<	SNP before AATGGAAATT			SNP before CAGTACAGAT		
Consensus		M			R	
VectorBase		C			C	
Family	15	16	20	15	16	20
F_0 Bright	A	V	V	G	S	S
F_0 Dim	C	S	S	A	A	A
F_2 Bright	M	M	M	R	R	R
F_2 Dim	M	M	M	R	R	R
5x enriched Bright		A			G	
5x enriched Dim		C			A	

AAEL011704

VectorBase <<	SNP before CCGCTAGAAA			SNP before GCTATTGTGG			SNP before AGCGAGAGAC		
Consensus	C			Y			Y		
VectorBase	C			T			C		
Family	15	16	20	15	16	20	15	16	20
F_0 Bright	A	A	A	T	T	T	T	T	T
F_0 Dim	C	C	C	C	C	C	C	C	C
F_2 Bright	M	M	M	Y	Y	Y	Y	Y	B
F_2 Dim	M	C	C	Y	Y	Y	Y	Y	Y
5x enriched Bright		A			T			T	
5x enriched Dim		C			C			C	

Chapter 4

Summary

The research herein describes how putative contributors to RNAi were elucidated in *Aedes aegypti*, an important vector of viral diseases. Figure 4.1 displays a schematic summarizing the experimental pipeline.

First, a mosquito line homozygous for a transgenic cassette that reflects the function of siRNA knockdown was developed. The construct is comprised of genes for DsRED, EGFP and an inverted repeat of EGFP all driven by an eye-specific promoter and are easily detected in mosquito eyes with fluorescence microscopy. The DsRED gene gives red fluorescence and serves here an easily detectable gene for confirming that a mosquito has at least one copy of the transgenic construct; EGFP gives green fluorescence and serves here as a selectable marker for siRNA knockdown capabilities. The construct was designed and implemented previously by the Adelman laboratory in another *Aedes aegypti* strain and shown to represent a mosquito's viral siRNA knockdown capabilities through EGFP expression (Adelman et al., 2008; Adelman et al., 2013). Here, a previously developed white-eyed strain with a Liverpool genetic background was transformed with the transgenic cassette in order for following studies to align with the published *Aedes aegypti* genome (Aryan et al., 2013). The construct reflects an individual mosquito's siRNA pathway capabilities by high or low expression of the selectable marker gene EGFP. When transcribed, the inverted repeat of this selectable marker forms a transcript that binds to itself, by design of the inversion of the EGFP sequences, creating dsRNA. The dsRNA triggers the siRNA silencing pathway, resulting in silencing of the EGFP transcripts; therefore, dim green fluorescence indicates efficient viral RNA knockdown and bright green fluorescence represents poor viral RNA knockdown. Flanking the transgenic cassette with transposon sequences from

another organism allowed for random insertion of the cassette into the genome. This random insertion allowed for a range of insertion sites with differing levels of experimental advantages. The microinjected embryos were then reared and segregated into pools and outcrossed to the parent strain. The pools that yielded progeny positive for the transgenes were outcrossed to the parent strain for two additional generations to restore health. Next, DsRED and EGFP expression profiles showed variable fluorescence expression among the pools. Mosquitoes positive for the transgene were intercrossed and goodness-of-fit tests on the progeny were conducted to estimate the survivability of the mosquitoes homozygous for the transgenes. Based on the estimated homozygote survivability results and the EGFP profiles, mosquitoes positive for the cassette from two favored pools were intercrossed an additional three times. Homozygosity was enriched for and verified through self-crosses and an outcross challenge. The outcross challenge verifies homozygosity by outcrossing females from the putative homozygous line with males from the parent strain; all offspring should be heterozygous for the transgene if the putative strains are indeed homozygous for the transgene. Next one additional intercross and screening for 100% transgene positive progeny of the homozygous line. An EGFP profile of the finalized, homozygous strain was obtained and shows continuous EGFP variability despite controlling for environmental factors.

Next, the two extreme phenotypes of efficient and poor siRNA knockdown, represented by dim and bright green fluorescing eyes, were enriched for in compounding crosses. The phenotype enrichments began in selecting the dimmest and brightest eyed mosquitoes from the recently developed line of mosquitoes homozygous for the sensor construct. The mosquitoes with the highest, or brightest EGFP eye expression were placed in a cage to intercross while the mosquitoes with the lowest, or dimmest, EGFP expression were placed in a separate cage to

intercross. Then the embryos from the brightest intercross were hatched, reared and the approximately top ten percent brightest mosquitoes of this batch were intercrossed. The same was done for the dimmest intercross progeny. Each phenotype was enriched by these sequential self-crosses five times.

Southern analysis validated that the enriched dim and bright lines have equivalent copies of the transgene. The objective of the enrichments was to segregate and enrich genomic characteristics of extreme siRNA knockdown capabilities. Theoretically, two copies of the transgene could manifest in brighter eye fluorescence; more copies of EGFP should yield more EGFP protein which should equate to brighter green fluorescence. Therefore, it was necessary to determine the transgene copy number in each enriched phenotype line to validate that the enrichments represent extreme siRNA knockdown capabilities and not a difference in copy number. Equivalent band numbers and sizes were seen in both enriched lines with multiple restriction enzymes, which is expected with equivalent transgene copies that are roughly in analogous genome location between both enriched cohorts.

Next, transcriptome comparisons of the enriched phenotypes revealed 11,562 quality-filtered transcripts that mapped uniquely to the published transcriptome. Of these significantly different transcripts, 590 were significantly upregulated and 530 downregulated in the bright cohort relative to the dim cohort. EGFP expression was 2.86 log2 fold-change higher in the bright substrain (FDR = 1.15e-84), demonstrating quantitatively the disparity in EGFP expression in the two, opposite extreme phenotypes. The majority of the significantly different genes have yet to be characterized. A 6.14 log2 fold-change deficiency AAEL011708 expression was detected in the cohort representing poor RNAi silencing abilities. The AAEL011708 ortholog in *Drosophila* plays an integral role in the siRNA pathway, in creating a conformational

change in another protein, Agonaute-2, making it able to bind the siRNA as well as in unwinding the siRNA. AAEL011708 has three paralogs in *Aedes aegypti* and it is possible that this redundancy resulted from duplication events.

Quantitative trait loci crosses of the extreme, enriched lines were performed to detect potential SNPs in select genes in three families. By cross mating the two, opposite, enriched cohorts and self-crossing their F_1 offspring, the resulting F_2 progeny should theoretically give two sets of opposite, genetically distinct cohorts homozygous in genomic regions contributing to the siRNA pathway and heterozygous at genomic regions not contributing to the siRNA pathway. An upstream region upstream of an AAEL011708 paralog was sequenced; no SNPs showing linkage to RNAi silencing phenotypes was detected. Intronic regions of genes significantly different in the RNA-seq data but not known to play a role in the siRNA pathway were also sequenced; one such gene, a rhomboid, showed two SNPs with inheritance patterns consistent with siRNA silencing phenotype segregation.

Hundreds of genes potentially influencing RNAi silencing abilities in Aedes aegypti have been identified. Future studies can elucidate the relevance these genes as the effort to reduce human suffering by diseases vectored by *Aedes aegypti* continues.

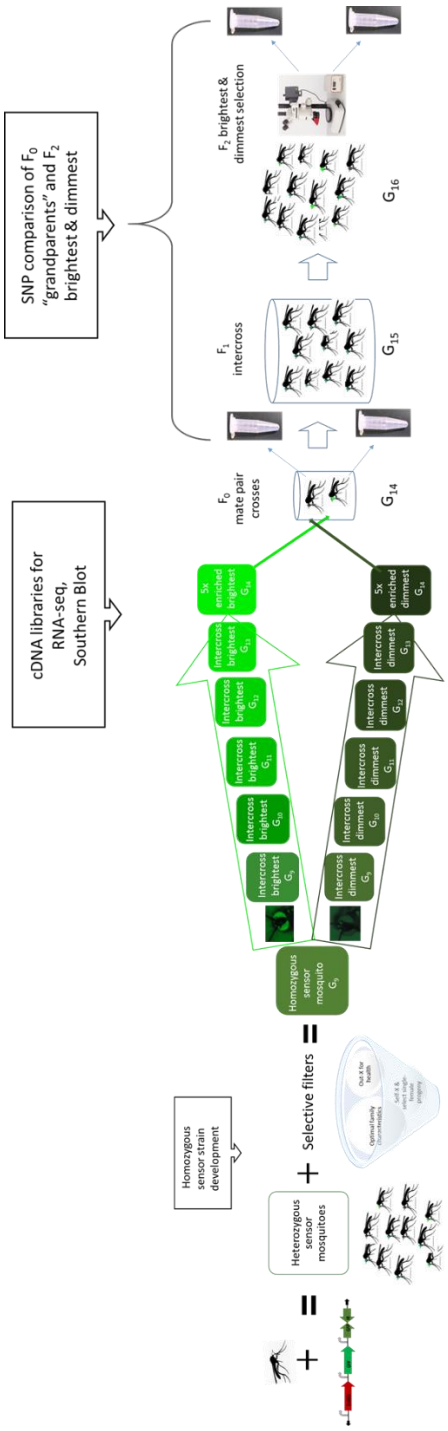


Figure 4.1 Project overview. Mosquito silhouette from www.craftmanspace.com, used under fair use, 2015. Leica MZ-16FL stereofluorescent dissecting microscope image used with permission from Anja Schue.

References

2012. "Dengue and the *Aedes aegypti* mosquito." http://www. cdc. gov/dengue/resources/30jan2012/aegyptifactsheet. pdf.

2012. "WHO report: Global Strategy for dengue prevention and control, 2012–2020." http://www.who.int/denguecontrol/9789241504034/en/

2014. "Chikungunya." http://www.who.int/mediacentre/factsheets/fs327/en/

Adelman, Z. N., M. A. Anderson, E. M. Morazzani and K. M. Myles (2008). "A transgenic sensor strain for monitoring the RNAi pathway in the yellow fever mosquito, *Aedes aegypti*." Insect Biochem Mol Biol **38**(7): 705-713.

Adelman, Z. N., M. A. Anderson, M. R. Wiley, M. G. Murreddu, G. H. Samuel, E. M. Morazzani and K. M. Myles (2013). "Cooler temperatures destabilize RNA interference and increase susceptibility of disease vector mosquitoes to viral infection." PLoS Negl Trop Dis **7**(5): e2239.

Aitken, T. H. G., R. B. Tesh, B. J. Beaty and L. Rosen (1979). "Transovarial Transmission of Yellow-Fever Virus by Mosquitos (*Aedes-Aegypti*)." American Journal of Tropical Medicine and Hygiene **28**(1): 119-121.

Alphey, L. (2014). "Genetic control of mosquitoes." Annu Rev Entomol **59**: 205-224.

Andrews, S. (2010). "FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc."

Aryan, A., M. A. Anderson, K. M. Myles and Z. N. Adelman (2013). "TALEN-based gene disruption in the dengue vector *Aedes aegypti*." PLoS One **8**(3): e60082.

Barnett, E. D. (2007). "Yellow fever: epidemiology and prevention." Clin Infect Dis **44**(6): 850-856.

Bennett, K. E., K. E. Olson, L. Munoz Mde, I. Fernandez-Salas, J. A. Farfan-Ale, S. Higgs, W. C. t. Black and B. J. Beaty (2002). "Variation in vector competence for dengue 2 virus among 24 collections of *Aedes aegypti* from Mexico and the United States." Am J Trop Med Hyg **67**(1): 85-92.

Black, W. C. t., K. E. Bennett, N. Gorrochotegui-Escalante, C. V. Barillas-Mury, I. Fernandez-Salas, M. de Lourdes Munoz, J. A. Farfan-Ale, K. E. Olson and B. J. Beaty (2002). "Flavivirus susceptibility in *Aedes aegypti*." Arch Med Res **33**(4): 379-388.

Brown, J. E., C. S. McBride, P. Johnson, S. Ritchie, C. Paupy, H. Bossin, J. Lutomiah, I. Fernandez-Salas, A. Ponlawat, A. J. Cornel, W. C. t. Black, N. Gorrochotegui-

Escalante, L. Urdaneta-Marquez, M. Sylla, M. Slotman, K. O. Murray, C. Walker and J. R. Powell (2011). "Worldwide patterns of genetic differentiation imply multiple 'domestications' of *Aedes aegypti*, a major vector of human diseases." Proc Biol Sci **278**(1717): 2446-2454.

Brown, J. E., V. Obas, V. Morley and J. R. Powell (2013). "Phylogeography and spatio-temporal genetic variation of *Aedes aegypti* (Diptera: Culicidae) populations in the Florida Keys." J Med Entomol **50**(2): 294-299.

Bush, W. S. and J. H. Moore (2012). "Chapter 11: Genome-Wide Association Studies." Plos Computational Biology **8**(12).

Campbell, C. L., K. M. Keene, D. E. Brackney, K. E. Olson, C. D. Blair, J. Wilusz and B. D. Foy (2008). "*Aedes aegypti* uses RNA interference in defense against Sindbis virus infection." BMC Microbiol **8**: 47.

Carpenetti, T. L., A. Aryan, K. M. Myles and Z. N. Adelman (2012). "Robust heat-inducible gene expression by two endogenous hsp70-derived promoters in transgenic *Aedes aegypti*." Insect Mol Biol **21**(1): 97-106.

Chepkorir, E., J. Lutomiah, J. Mutisya, F. Mulwa, K. Limbaso, B. Orindi, Z. Ng'ang'a and R. Sang (2014). "Vector competence of *Aedes aegypti* populations from Kilifi and Nairobi for dengue 2 virus and the influence of temperature." Parasit Vectors **7**: 435.

Christophers, S. R. (1960). *Aedes aegypti*: the yellow fever mosquito. U.K., Cambridge University Press.

Clemons, A., M. Haugen, E. Flannery, M. Tomchaney, K. Kast, C. Jacowski, C. Le, A. Mori, W. Simanton Holland, J. Sarro, D. W. Severson and M. Duman-Scheel (2010). "*Aedes aegypti*: an emerging model for vector mosquito development." Cold Spring Harb Protoc **2010**(10): pdb emo141.

Coates, C. J., N. Jasinskiene, L. Miyashiro and A. A. James (1998). "Mariner transposition and transformation of the yellow fever mosquito, *Aedes aegypti*." Proc Natl Acad Sci U S A **95**(7): 3748-3751.

Cook, W. J., S. E. Ealick, P. Reichert, G. S. Hammond, H. V. Le, T. L. Nagabhushan, P. P. Trotta and C. E. Bugg (1991). "Crystallization and Preliminary-X-Ray Investigation of Recombinant Human Interleukin-4." Journal of Molecular Biology **218**(4): 675-678.

Cooper, G. M., J. A. Johnson, T. Y. Langaee, H. Feng, I. B. Stanaway, U. I. Schwarz, M. D. Ritchie, C. M. Stein, D. M. Roden, J. D. Smith, D. L. Veenstra, A. E. Rettie and M. J. Rieder (2008). "A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose." Blood **112**(4): 1022-1027.

Cornel, A. J., M. Q. Benedict, C. S. Rafferty, A. J. Howells and F. H. Collins (1997).

"Transient expression of the *Drosophila melanogaster* cinnabar gene rescues eye color in the white eye (WE) strain of Aedes aegypti." Insect Biochem Mol Biol **27**(12): 993-997.

Davis, F. A. (2009). Taber's Cyclopedic Medical Dictionary, F. A Davis Company.

de Andrade, D. C., S. Jean, P. Clavelou, R. Dallel and D. Bouhassira (2010). "Chronic pain associated with the Chikungunya Fever: long lasting burden of an acute illness." BMC Infect Dis **10**: 31.

Doudna, J. A. and E. Charpentier (2014). "Genome editing. The new frontier of genome engineering with CRISPR-Cas9." Science **346**(6213): 1258096.

Fodor, S. P., J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu and D. Solas (1991). "Light-directed, spatially addressable parallel chemical synthesis." Science **251**(4995): 767-773.

Fontenille, D., M. Diallo, M. Mondo, M. Ndiaye and J. Thonnon (1997). "First evidence of natural vertical transmission of yellow fever virus in *Aedes aegypti*, its epidemic vector." Transactions of the Royal Society of Tropical Medicine and Hygiene **91**(5): 533-535.

Forrester, N. L., Coffey, L. L. and Weaver, S. C. (2014). "Arboviral bottlenecks and challenges to maintaining diversity and fitness during mosquito transmission." Viruses, 6(10), 3991-4004. doi: 10.3390/v6103991

Franz, A. W., N. Jasinskiene, I. Sanchez-Vargas, A. T. Isaacs, M. R. Smith, C. C. Khoo, M. S. Heersink, A. A. James and K. E. Olson (2011). "Comparison of transgene expression in *Aedes aegypti* generated by mariner Mos1 transposition and PhiC31 site-directed recombination." Insect Mol Biol **20**(5): 587-598.

Frierson, J. G. (2010). "The yellow fever vaccine: a history." Yale J Biol Med **83**(2): 77-85.

Gao, D., J. Kim, H. Kim, T. L. Phang, H. Selby, A. C. Tan and T. Tong (2010). "A survey of statistical software for analysing RNA-seq data." Hum Genomics **5**(1): 56-60.

Garver, L. S., Z. Y. Xi and G. Dimopoulos (2008). "Immunoglobulin superfamily members play an important role in the mosquito immune system." Developmental and Comparative Immunology **32**(5): 519-531.

Gilman, J. G. and T. H. Huisman (1985). "A mutation associated with elevated G gamma chain in sickle cell anemia and hereditary persistence of fetal hemoglobin." Prog Clin Biol Res **191**: 141-149.

Goll, M. G. and T. H. Bestor (2005). "Eukaryotic cytosine methyltransferases." Annu Rev Biochem **74**: 481-514.

Gordon, R. M. a. L. W. H. R. (1939). "A study of the behaviour of the mouthparts of mosquitoes when taking up blood from living tissue; together with some observations on

the ingestion of microfilariae." Ann. Trop. Med. Parasitol. **33**: 259-278.

Gross, L. A., G. S. Baird, R. C. Hoffman, K. K. Baldridge and R. Y. Tsien (2000). "The structure of the chromophore within DsRed, a red fluorescent protein from coral." Proc Natl Acad Sci U S A **97**(22): 11990-11995.

Gubler, D. J., S. Nalim, R. Tan, H. Saipan and J. Sulianti Saroso (1979). "Variation in susceptibility to oral infection with dengue viruses among geographic strains of *Aedes aegypti*." Am J Trop Med Hyg **28**(6): 1045-1052.

Guzman, M. G. and E. Harris (2014). "Dengue." Lancet.

Haac, M. E., M. A. Anderson, H. Eggleston, K. M. Myles and Z. N. Adelman (2015). "The hub protein loquacious connects the microRNA and short interfering RNA pathways in mosquitoes." Nucleic Acids Res **43**(7): 3688-3700.

Halstead, S. B. (2002). "Dengue." Curr Opin Infect Dis **15**(5): 471-476.

Hemungkorn, M., U. Thisyakorn and C. Thisyakorn (2007). "Dengue infection: a growing global health threat." Biosci Trends **1**(2): 90-96.

Hull, B., E. Tikasingh, M. de Souza and R. Martinez (1984). "Natural transovarial transmission of dengue 4 virus in *Aedes aegypti* in Trinidad." Am J Trop Med Hyg **33**(6): 1248-1250.

Hurd, P. J. and C. J. Nelson (2009). "Advantages of next-generation sequencing versus the microarray in epigenetic research." Brief Funct Genomic Proteomic **8**(3): 174-183.

Ivics, Z. and Z. Izsvak (2010). "The expanding universe of transposon technologies for gene and cell engineering." Mob DNA **1**(1): 25.

Iwasaki, S., H. M. Sasaki, Y. Sakaguchi, T. Suzuki, H. Tadakuma and Y. Tomari (2015). "Defining fundamental steps in the assembly of the *Drosophila* RNAi enzyme complex." Nature.

James, A. A. (2003). "Blocking malaria parasite invasion of mosquito salivary glands." J Exp Biol **206**(Pt 21): 3817-3821.

Joshi, V., M. Singhi and R. C. Chaudhary (1996). "Transovarial transmission of dengue 3 virus by *Aedes aegypti*." Trans R Soc Trop Med Hyg **90**(6): 643-644.

Khin, M. M. and K. A. Than (1983). "Transovarial transmission of dengue 2 virus by *Aedes aegypti* in nature." Am J Trop Med Hyg **32**(3): 590-594.

Knox, T. B., B. H. Kay, R. A. Hall and P. A. Ryan (2003). "Enhanced vector competence of Aedes aegypti (Diptera: Culicidae) from the Torres Strait compared with mainland

Australia for dengue 2 and 4 viruses." J Med Entomol **40**(6): 950-956.

Kuno, G. (2010). "Early History of Laboratory Breeding of *Aedes aegypti* (Diptera: Culicidae) Focusing on the Origins and Use of Selected Strains." Journal of Medical Entomology **47**(6): 957-971.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nat Methods **9**(4): 357-359.

Lee, H. and H. Tang (2012). "Next-generation sequencing technologies and fragment assembly algorithms." Methods Mol Biol **855**: 155-174.

Marchoux, E., Simmond, P.L. (1905). "La transmission hereditaire du virus de la fievre jaune chez le Stegomyia fasciata." C R Soc Biol **59**: 259-260.

Mavale, M., D. Parashar, A. Sudeep, M. Gokhale, Y. Ghodke, G. Geevarghese, V. Arankalle and A. C. Mishra (2010). "Venereal transmission of chikungunya virus by *Aedes aegypti* mosquitoes (Diptera: Culicidae)." Am J Trop Med Hyg **83**(6): 1242-1244.

Meister, G. and T. Tuschl (2004). "Mechanisms of gene silencing by double-stranded RNA." Nature **431**(7006): 343-349.

Michelmore, R. W., I. Paran and R. V. Kesseli (1991). "Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations." Proc Natl Acad Sci U S A **88**(21): 9828-9832.

Miles, C. M., Wayne, M. (2008). "Quantitative trait locus (QTL) analysis." Nature Education **1**(1).

Miyoshi, T., A. Takeuchi, H. Siomi and M. C. Siomi (2010). "A direct role for Hsp90 in pre-RISC formation in Drosophila." Nat Struct Mol Biol **17**(8): 1024-1026.

Mourya, D. T. (1987). "Absence of transovarial transmission of Chikungunya virus in *Aedes aegypti* & *Ae. albopictus* mosquitoes." Indian J Med Res **85**: 593-595.
Mourya, D. T., Gokhale, A. Basu, P. V. Barde, G. N. Sapkal, V. S. Padbidri and M. M. Gore (2001). "Horizontal and vertical transmission of dengue virus type 2 in highly and lowly susceptible strains of *Aedes aegypti* mosquitoes." Acta Virol **45**(2): 67-71.

Myles, K. M., M. R. Wiley, E. M. Morazzani and Z. N. Adelman (2008). "Alphavirus-derived small RNAs modulate pathogenesis in disease vector mosquitoes." Proc Natl Acad Sci U S A **105**(50): 19938-19943.

Napoli, C., C. Lemieux and R. Jorgensen (1990). "Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans." Plant Cell **2**(4): 279-289.

Nene, V., J. R. Wortman, D. Lawson, B. Haas, C. Kodira, Z. J. Tu, B. Loftus, Z. Xi, K. Megy, M. Grabherr, Q. Ren, E. M. Zdobnov, N. F. Lobo, K. S. Campbell, S. E. Brown, M. F. Bonaldo, J. Zhu, S. P. Sinkins, D. G. Hogenkamp, P. Amedeo, P. Arensburger, P. W. Atkinson, S. Bidwell, J. Biedler, E. Birney, R. V. Bruggner, J. Costas, M. R. Coy, J. Crabtree, M. Crawford, B. Debruyn, D. Decaprio, K. Eiglmeier, E. Eisenstadt, H. El-Dorry, W. M. Gelbart, S. L. Gomes, M. Hammond, L. I. Hannick, J. R. Hogan, M. H. Holmes, D. Jaffe, J. S. Johnston, R. C. Kennedy, H. Koo, S. Kravitz, E. V. Kriventseva, D. Kulp, K. Labutti, E. Lee, S. Li, D. D. Lovin, C. Mao, E. Mauceli, C. F. Menck, J. R. Miller, P. Montgomery, A. Mori, A. L. Nascimento, H. F. Naveira, C. Nusbaum, S. O'Leary, J. Orvis, M. Pertea, H. Quesneville, K. R. Reidenbach, Y. H. Rogers, C. W. Roth, J. R. Schneider, M. Schatz, M. Shumway, M. Stanke, E. O. Stinson, J. M. Tubio, J. P. Vanzee, S. Verjovski-Almeida, D. Werner, O. White, S. Wyder, Q. Zeng, Q. Zhao, Y. Zhao, C. A. Hill, A. S. Raikhel, M. B. Soares, D. L. Knudson, N. H. Lee, J. Galagan, S. L. Salzberg, I. T. Paulsen, G. Dimopoulos, F. H. Collins, B. Birren, C. M. Fraser-Liggett and D. W. Severson (2007). "Genome sequence of Aedes aegypti, a major arbovirus vector." Science **316**(5832): 1718-1723.

Nikolayeva, O. and M. D. Robinson (2014). "edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology." Methods Mol Biol **1150**: 45-79.

Pascini, T. V., M. Ramalho-Ortigao and G. F. Martins (2012). "Morphological and morphometrical assessment of spermathecae of Aedes aegypti females." Mem Inst Oswaldo Cruz **107**(6): 705-712.

Perrotta, A. T. and M. D. Been (1991). "A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA." Nature **350**(6317): 434-436.

Powell, J. R. and W. J. Tabachnick (2013). "History of domestication and spread of Aedes aegypti--a review." Mem Inst Oswaldo Cruz **108 Suppl 1**: 11-17.

Prendergast, F. G. and K. G. Mann (1978). "Chemical and physical properties of aequorin and the green fluorescent protein isolated from *Aequorea forskalea*." Biochemistry **17**(17): 3448-3453.

Reed, W. and J. Carroll (1901). "The Prevention of Yellow Fever." Public Health Pap Rep **27**: 113-129.

Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." Bioinformatics **26**(1): 139-140.

Rosen, L., D. A. Shroyer, R. B. Tesh, J. E. Freier and J. C. Lien (1983). "Transovarial transmission of dengue viruses by mosquitoes: *Aedes albopictus* and *Aedes aegypti*." Am J Trop Med Hyg **32**(5): 1108-1119.

Ruohola-Baker, H., E. Grell, T. B. Chou, D. Baker, L. Y. Jan and Y. N. Jan (1993). "Spatially localized rhomboid is required for establishment of the dorsal-ventral axis in *Drosophila* oogenesis." Cell **73**(5): 953-965.

Sanchez-Vargas, I., J. C. Scott, B. K. Poole-Smith, A. W. Franz, V. Barbosa-Solomieu, J. Wilusz, K. E. Olson and C. D. Blair (2009). "Dengue virus type 2 infections of *Aedes aegypti* are modulated by the mosquito's RNA interference pathway." PLoS Pathog **5**(2): e1000299.

Sanchez-Vargas, I., E. A. Travanty, K. M. Keene, A. W. Franz, B. J. Beaty, C. D. Blair and K. E. Olson (2004). "RNA interference, arthropod-borne viruses, and mosquitoes." Virus Res **102**(1): 65-74.

Scott, T. W., G. G. Clark, L. H. Lorenz, P. H. Amerasinghe, P. Reiter and J. D. Edman (1993). "Detection of multiple blood feeding in *Aedes aegypti* (Diptera: Culicidae) during a single gonotrophic cycle using a histologic technique." J Med Entomol **30**(1): 94-99.

Scott, T. W., A. Naksathit, J. F. Day, P. Kittayapong and J. D. Edman (1997). "A fitness advantage for *Aedes aegypti* and the viruses it transmits when females feed only on human blood." Am J Trop Med Hyg **57**(2): 235-239.

Scott, T. W. and W. Takken (2012). "Feeding strategies of anthropophilic mosquitoes result in increased risk of pathogen transmission." Trends Parasitol **28**(3): 114-121.

Sethuraman, N., M. J. Fraser, Jr., P. Eggleston and D. A. O'Brochta (2007). "Post-integration stability of piggyBac in *Aedes aegypti*." Insect Biochem Mol Biol **37**(9): 941-951.

Severson, D. W., D. L. Knudson, M. B. Soares and B. J. Loftus (2004). "*Aedes aegypti* genomics." Insect Biochem Mol Biol **34**(7): 715-721.

Sheng, G., E. Thouvenot, D. Schmucker, D. S. Wilson and C. Desplan (1997). "Direct regulation of rhodopsin 1 by Pax-6/eyeless in *Drosophila*: evidence for a conserved function in photoreceptors." Genes Dev **11**(9): 1122-1131.

Shepard, D. S., L. Coudeville, Y. A. Halasa, B. Zambrano and G. H. Dayan (2011). "Economic impact of dengue illness in the Americas." Am J Trop Med Hyg **84**(2): 200-207.

Slosek, J. (1986). "*Aedes aegypti* mosquitoes in the Americas: a review of their interactions with the human population." Soc Sci Med **23**(3): 249-257.

Southern, E. (2006). "Southern blotting." Nat Protoc **1**(2): 518-525.

Sturtevant, M. A., M. Roark and E. Bier (1993). "The *Drosophila* rhomboid gene mediates the

localized formation of wing veins and interacts genetically with components of the EGF-R signaling pathway." Genes Dev **7**(6): 961-973.

Thavara, U., A. Tawatsin, T. Pengsakul, P. Bhakdeenuan, S. Chanama, S. Anantapreecha, C. Molito, J. Chompoosri, S. Thammapalo, P. Sawanpanyalert and P. Siriyasatien (2009). "Outbreak of chikungunya fever in Thailand and virus detection in field population of vector mosquitoes, *Aedes aegypti* (L.) and *Aedes albopictus* Skuse (Diptera: Culicidae)." Southeast Asian J Trop Med Public Health **40**(5): 951-962.

Torres, J. R. and J. Castro (2007). "The health and economic impact of dengue in Latin America." Cad Saude Publica **23 Suppl 1**: S23-31.

Vijayendran, D., P. M. Airs, K. Dolezal and B. C. Bonning (2013). "Arthropod viruses and small RNAs." J Invertebr Pathol **114**(2): 186-195.

Wang, K., M. Li and H. Hakonarson (2010). "Analysing biological pathways in genome-wide association studies." Nat Rev Genet **11**(12): 843-854.

Appendix A

Supplementary File of Complete RNA-seq and BioMart Data

Excel file containing filtered RNA-seq data and gene descriptions, locations and other information from BioMart.

Email Dr. Zach N. Adelman (zachadel@vt.edu) or Angie Saadat (saadat13@vt.edu) for a copy of this supplemental file.