UNIVERSITY OF CALIFORNIA
RIVERSIDE

Transposable Elements and Their Application as Genetic Tools in Arthropods

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

Doctor of Philosophy

in

Cell, Molecular and Developmental Biology

by

Anna-Louise Anita Doss

September 2019

Dissertation Committee:
Dr. Peter Atkinson, Chairperson
Dr. Patricia Springer
Dr. Naoki Yamanaka

The Dissertation of Anna-Louise Anita Doss is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Peter Atkinson, for the opportunity he gave me to join his laboratory. Working with Dr. Atkinson has made me a better writer, presenter, and scientist. I am grateful for his guidance, humor, and patience, which persisted even through very difficult situations.

I would like to acknowledge the support of the W.M. Keck Foundation, which funded the research presented in this dissertation.

Thank you to our laboratory's former Senior Research Associate, the wise Robert Hice, who likes to say that he taught me everything I know.

I am ever appreciative for the support, laughter and companionship of my lab members, Michael Han, Sally Ireri, Simran Sandhu, and Baoju Wang.

I would like to thank my Committee members, Dr. Patty Springer and Dr. Naoki Yamanaka for their advice, attention, and thoughtful questions over the years.

Thank you to my life partner and best friend, Samuel Aguayo, who has supported me in working towards my goals, even when he had to give up his season tickets.

Thank you especially to my Mom.

ABSTRACT OF THE DISSERTATION

Transposable Elements and Their Application
as Genetic Tools in Arthropods

by

Anna-Louise Anita Doss

Doctor of Philosophy, Graduate Program in Cell, Molecular and Developmental Biology
University of California, Riverside, September 2019
Dr. Peter Atkinson, Chairperson

The aim of the research presented in this thesis is to examine the transposition mechanism of class 2 transposons and to assess their employment as genetic tools in arthropod genomes. The first chapter reviews transposable genetic elements and their intimate relationship with eukaryotic genome diversity, and methods of host genome control for maintaining genome integrity, including the piRNA system. Chapter Two introduces a *hAT* transposable element, *Hermes*, that originates from the genome of the housefly, Musca *domestica.* The *Hermes* element is an active transposon that has been shown to transform several non-homologous arthropod species. Our laboratory has participated in a collaborative research project to understand the molecular details of *Hermes* transposition through crystal structure and biochemical properties. The research outlined in Chapter Two provides a rationale for the occurrence of *Hermes* as an octameric protein with multiple DNA binding domains. *Hermes* is the only *hAT* transposable element for which a co-crystal structure has been successfully produced.

Chapter Three introduces a newly described *MULE* element from the genome of the mosquito *Aedes aegypti*, a major global arboviral vector. The research outlined demonstrates that the *Ae. aegypti Muta1* element is active in its host, in human cell culture and can transform the model dipteran, *Drosophila melanogaster*, and remobilize in its germline. Chapter Four examines the potential of the *Muta1* element as the basis for a novel forward genetics-based tool in its host, *Aedes aegypti*, for which no enhancer trap system has been effectively developed. This chapter also outlines potential challenges to implementing a transposon-based tool in this species due to its evolutionarily expanded piRNA pathway. Chapter Five explains the ability of the *Muta1* element to transform and remobilize in two *Anopheles* mosquito species and explores the potential benefit of having a novel enhancer trap system to deploy in these mosquito species.

TABLE OF CONTENTS

List of tables

List of figures

Chapter 1: An introduction to transposable elements

1.1 Transposable elements

Transposable elements were first described by Barbara McClintock in 1931. McClintock's experiments in *Zea mays* led to her observation of genetic "controlling elements" (McClintock 1931) and her contribution to the field of genetics eventually led to her being awarded the Nobel Prize in Physiology in 1983 (a review by Nathaniel C. Comfort discusses the delay in the recognition of Dr. McClintock's achievements as a result of gender bias in the field of science). Transposable elements were once thought of as "selfish" genetic elements that could propagate at the expense of the host genome, or "junk" DNA, primarily neutral elements that added to genome size (Kidwell and Lisch, 2000). These early observations have changed overtime, however, as more research has shown an intimate evolutionary relationship between transposable elements and their hosts. Indeed, it was eventually discovered that a transposable element insertion into a metabolic gene was responsible for the "wrinkled" pea phenotype that facilitated Gregor Mendel's studies of inheritance (Bhattacharyya *et al.,* 1990). It is now known that transposable elements are ubiquitous in the genomes of all organisms, from archaea to eukaryotes (Friedli and Trono, 2015), and they are drivers of genome evolution and genetic diversity (Chuong *et al.*, 2017).

Transposable elements can broadly be placed into two classes depending on their replication intermediate and can be classified as autonomous or non-autonomous. Autonomous elements comprise one or more genes for producing the

proteins required for mobilization positioned between 5' (left) and 3' (right) flanking sequences that are recognized and bound by transposon enzymes. Non-autonomous transposons have lost coding capacity but retain flanking sequences that can be recognized by complementary transposition enzymes that have been produced in *trans* by full length elements.

Class 1 transposable elements, also called retrotransposons or copy-and-paste transposons, use an RNA intermediate and rely on reverse transcription for propagation. Retrotransposons are further classified based on their encoded proteins and the sequence repeats in the 5' and 3' ends of the element. LTR (long terminal repeat) retrotransposons are very similar to retroviruses in their replication method. They comprise 5' end and 3' end long terminal repeats flanking several genes, including the *gag* core viral proteins, and *pol*, which codes for reverse transcriptase, integrase, and RNase-H. Retroviruses, however, unlike retrotransposons, are capable of budding from the host cell (Finnegan, 2012). Non-LTR retrotransposons include autonomous LINEs (long interspersed elements) and non-autonomous SINEs (short interspersed elements). While non-LTR retrotransposons also move through an RNA intermediate, the mechanisms of replication are different from LTR elements. LINEs characteristically comprise two open reading frames (in most cases), *orf1* and *orf2*, that are transcribed into mRNA via RNA polymerase II from an internal promoter. Once in the cytoplasm, the mRNA is recognized and bound by its proteins ORF1 and ORF2, and the resulting ribonucleoprotein is transported into the nucleus and the element is integrated by target-primed reverse transcription (TPRT), in which the target DNA is cleaved on the 3' strand

and the nicked DNA acts as a primer for reverse transcription (Han, 2010). Non-autonomous SINEs, which lack coding capabilities, contain an internal RNA polymerase III promoter, but must rely on reverse transcription and integration by proteins produced in trans by other active elements (Deininger, 2011).

Class 2 transposable elements, or cut-and-paste transposons, use a single or double-stranded DNA intermediate and rely on a transposase enzyme for proliferation. There is an exception for both *Helitrons* and *Polintons* (formerly called *Maverick*), which are described below. The largest group of eukaryotic DNA elements encode a transposase flanked by 5' and 3' DNA arms that end in terminal inverted repeats. These cut-and-paste transposons can be divided into almost twenty different major superfamilies based on their encoded transposase (and occasionally an accessory protein) and the sequence repeats in the 5' and 3' ends of the element. Because of their copy-and-paste mechanism of propagation, DNA elements do not usually make up a large portion of the DNA content of genomes as compared to retroelements, although they are certainly ubiquitous across eukaryotic genomes. For example, the human genome contains ~45% transposable element derived sequences, but only 3% of that originates from DNA transposons, none of which show any evidence of current activity (Lander *et al.,* 2001). Contrariwise, the genome of the protozoan parasite *Trichomonas vaginalis* contains almost ~65% sequence derived from DNA transposons, although this trait is somewhat unique amongst eukaryotes (Carlton *et al.,* 2007). DNA transposons can also be grouped according to structurally distinct catalytic domains in their enzymes. The largest group of eukaryotic DNA transposons is categorized by the

presence of an RNase H-like fold containing an acidic DDE/D catalytic motif that makes up the nuclease domain. This nuclease domain is also found in prokaryotic transposases, like *Tn5* and many bacterial insertion sequences (Hickman and Dyda, 2015). A unique subclass of DNA transposons, called *Helitrons*, use a type of rolling circle replication via a single stranded DNA intermediate and encode a HUH nuclease (containing two histidines separated by a hydrophobic residue) with a DNA helicase domain, and they contain a short palindromic 3' repeat rather than terminal inverted repeats (Grabundzija *et al.,* 2016). *Helitrons* have been identified *in silico* in a wide range of eukaryotic hosts, including plants, insects, and small bats, however, there is still information to be discerned about their transposition and propagation mechanisms (Thomas and Pritham, 2015). Another subclass of DNA transposons called *Polintons* (formerly *Maverick*) are much larger than common DDE/D transposases, with lengths of ~15-25 kilobases, and self-replicate with their own DNA polymerase and integrase, among other proteins (Haapa-Paananen *et al.,* 2014). *Polintons* have mysterious origins and are thought to be closely related to virophages, or perhaps are derived from an ancient integrated virophage (Koonin and Krupovic, 2017).

Transposable elements rely on vertical transmission to maintain their presence within a species. Transposable elements can also be spread between species, even between different kingdoms. Horizontal transfer of transposons occurs frequently in prokaryotes, which commonly pick up and exchange genetic information via transformation and conjugation. The first report on horizontal transmission of a transposable element examined the early invasion of the *P* element into *D. melanogaster*

from the closely related *D. willistoni* genome (Daniels *et al.,* 1990). The recent plethora of genome sequencing data has made it possible to survey genomes for transposable element families using bioinformatic programs. Evidence for horizontal transmission of transposons has been found between a broad range of eukaryotes, from prokaryotes to eukaryotes, involving almost all transposable element families (Gilbert and Cordaux, 2013). One must assume that for horizontal transfer to occur, two species must have close contact with each other, but the exact mode of genetic transmission is somewhat nebulous. There is an increased amount of horizontal transfer events observed between parasite and host species and increased transfer between species occupying the same biogeographical areas (Gilbert and Feschotte, 2018). Although retroelements generally make up a larger portion of genomic repeat elements, horizontal transfer is more likely to be observed for DNA transposons than retroelements, possibly due to the overall stability of the DNA transpososome and the lack of cofactors required for mobilization (Reiss *et al.,* 2019).

1.1.2 Genome evolution and transposons

Evidence of transposable element proliferation and dissemination is observable across genera in the form of active, inactive, and truncated elements, and the cryptic footprints left in the host genome during duplication. Sequences derived from transposable elements and their remnants can make up a large portion of genome content and there is a known overall positive correlation between genome size and transposon load in eukaryotes (Elliott and Gregory, 2015). For example, nearly half of the ~3.1 Gb

human genome is transposable element sequence (Jain *et al.,* 2018, Lander *et al.,* 2001), and almost 90% of the ~5.5 Gb *Hordeum vulgare* (barley) genome is transposon derived (Wicker *et al.,* 2009). It is now established that transposable elements provide material for genetic evolution and are the source the extreme genome size variation observed in higher eukaryotes, which is called the C-value paradox (Joseph Gall, 1981). The C-value paradox is most apparent in angiosperm genomes, where the haploid genome content sees the most extreme variation among related species, which may be due, in part, to the evolutionarily disparate epigenetic transposon control mechanisms possessed by flowering plants (Federoff, 2012). The Liliaceae family provides a good example of the C-value paradox, where the genome sizes (1C) have been observed to range from ~0.3 Gb for *Prosartes smithii* (Largeflower fairy bells) to ~87.4 Gb for *Fritillaria uva-vulpis* (fox-grape flower) (Leitch *et al.,* 2007). Improved sequencing methods have provided evidence that the massive genome sizes observed in some Liliaceae species are not due to the presence of one or several high-copy-number elements but are a result of inefficient DNA removal in the species (Kelly *et al.,* 2015). In addition to epigenetic silencing mechanisms, plants have the ability to remove genomic DNA by homologous and illegitimate recombination (Bennetzen and Wang, 2014).

There are numerous examples of exaptation, or molecular domestication, of transposon sequences and genes by host genomes, as sequences can become transcription factors and enhancers, or new genes and pseudogenes. In vertebrates, the *ZBED* gene family, with members that serve as important transcription factors during growth and development, is derived from domesticated *hAT* transposons (Hayward *et al.,*

2013). Fundamental components of mammalian innate immunity have also evolved as a result of the exaptation of transposons. Specifically, the process of antibody production via V(D)J recombination is known to have evolved from an exapted *Transib* DNA transposon (Sakano *et al.,* 1979, Kapitonov and Jurka, 2005). One of the core proteins involved in V(D)J recombination, recombination activating protein 1 (RAG1), is known to be responsible for recognizing discrete DNA repeats, called recombination signal sequences, within the *Tcbr* locus (Schatz and Swanson, 2011). The shuffling and joining of recombination signal sequences within the *Tcbr* coding sequence is what generates millions of different antigen receptors during T and B lymphocyte development (Huang *et al.,* 2016). This ancient exaptation of a *Transib* element is considered to be the fundamental event responsible for the development of the adaptive immune system in jawed vertebrates (Feschotte and Pritham, 2007).

It is also well recognized that transposable elements have contributed to genomic regulatory networks over the course of evolution. In addition to providing coding sequences that may be co-opted by host genomes and repurposed for host cellular functions, transposable elements are also a rich source of non-coding DNA, like enhancers and regulatory sequences. Many transposons contain their own internal promoters in order to initiate replication independent of host cell factors. For example, the human LINE-1 retrotransposon has its own internal RNA polymerase II promoter sequence within the element 5'UTR (Richardson *et al.,* 2015). LTR retrotransposons, including endogenous retroviruses (ERVs), have enhancer and transcription factor binding motifs in both the 5' and 3' LTR sequences, such as OCT4, SOX2, and Nanog

binding sites, which are crucial factors for mammalian development (Rebollo *et al.,* 2012). From the perspective of the LTR element, possessing transcription factor binding sites that are active in rapidly dividing cells is a trait that will be selected for over time, as it increases the chance of heritable transposition events. However, there is also a plethora evidence for selection on the part of the host genome. Genome wide analysis methods in mammals, including ChIP-seq (chromatin immunoprecipitation followed by sequencing) and CAGE-seq (cap analysis of gene expression followed by sequencing), have revealed that transposable elements make up a substantial fraction of transcription factor binding sites and RNA polymerase II initiation sites (Chuong *et al., 2017*).

Transposable elements themselves also experience evolutionary selection for properties that make them less likely to be targeted by genome defenses, and contrariwise, for properties that positively impact vertical transmission. For instance, a comprehensive analysis of over 40,000 *Mu* insertion sites within the maize genome show a statistically significant preference for euchromatic, actively transcribed genomic regions and 5' gene-adjacent sites (Liu *et al.,* 2009). Conversely, some transposable elements show an insertion preference for heterochromatin and pericentromeric regions that show low expression, perhaps as a mechanism to avoid deleterious mutations in the host or to evade host silencing mechanisms (Vini Pereira, 2004, Rebollo *et al.,* 2012). Adaptive selection on the side of transposable elements has not seen the same breadth of analysis as the transposable element impact on host genomes. However, the co-evolution of both transposable elements and their host genomes are clearly eternally linked to one another.

1.1.3 Controlling transposable element proliferation

The ability of transposons to self-propagate and modify host genomes have led to transposon genes being the most abundant genes in nature (Aziz *et al.,* 2010). Nevertheless, transposon mobility causes mutations that can be deleterious to the host, and consequently, organisms have developed methods of controlling or silencing active transposable elements.

Plants have numerous mechanisms for controlling transposable element proliferation, including various epigenetic silencing mechanisms and DNA removal. LTR retrotransposons are the largest contributor to the significant genome size variation observed in angiosperms, not only because of their ability to proliferate, but also because there are observed disparities in the occurrence of DNA removal between plant species (Feschotte *et al.,* 2002, Tenaillon *et al.,* 2010). DNA removal is accomplished by two mechanisms; unequal intra-strand homologous recombination (UR) and illegitimate recombination (IR). Unequal intra-strand homologous recombination can occur between the LTRs of a single element, or between the LTRs of similar elements inserted at varying distances, which can remove large amounts of DNA at once and typically leaves behind solo LTRs in the host genome (Ma *et al.,* 2004, Bennetzen and Wang, 2014). Illegitimate recombination, which is not homology driven, usually occurs as the result of a double-stranded breaks and is evidenced by elements with significant truncations or deletions (Devos *et al.,* 2002). There are numerous studies detailing various propensities among plant species for DNA loss through unequal and illegitimate recombination, and these data imply that "massive" plant genomes see less

DNA removal (Vicient *et al.,* 1999, Devos *et al.,* 2002, Ma *et al.,* 2004, Cossu *et al.,* 2017). Interestingly, despite the fact that LTR retrotransposons should experience the same amount of removal by unequal recombination, different LTR families appear to be able to maintain different copy numbers in genomes due to differences in insertion site preference (Baucom *et al.,* 2009). In the foxtail millet genome, *Gypsy* retroelements maintain a greater genomic copy number by preferential insertion into heterochromatic regions where unequal homologous recombination is less likely to occur, while *Copia* elements, which have been described as having a preference for insertion into euchromatin, have fewer retained copies in foxtail millet due to recombinational loss (Bennetzen *et al.,* 2012, Vini Pereira, 2004).

Plants also have a diverse array of epigenetic silencing mechanisms to control transposable element proliferation, and these methods of control are both heritable and reversible. While animals typically undergo an epigenetic "reset" during development, plants do not maintain a separate germline tissue and it is easier for somatic *de novo* methylation to be carried on into the next generation (Lisch, 2009). Epigenetic silencing in plants involves small interfering RNA (siRNA) pathways that are sequence-specific, and both target transcripts for degradation and induce chromatin silencing by RNA-dependent DNA methylation (RdDM) (Law and Jacobsen, 2010). siRNAs are single stranded RNAs that are approximately 21-24 nucleotides long and their biogenesis starts with the processing of longer double-stranded RNA templates that originate from genomic transposon copies and repeats, as well as virally derived RNAs (Matzke *et al.,* 2015). There are several plant RNA polymerases involved in siRNA biogenesis before

downstream processing by *Dicer*-like endoribonucleases and members of the AGO clade of Argonaute proteins before silencing occurs (Freidli and Trono, 2015). Post-transcriptional silencing occurs when processed siRNAs target transposon transcripts for cleavage through homology and the assembly of Argonaute-containing RNA induced silencing complexes (RISCs) (Cui and Cao, 2014). Chromatin silencing occurs when processes siRNAs associate with RNAi machinery in the nucleus to direct the recruitment of histone modifying enzymes and methyltransferases to genomic transposable element insertions for either the maintenance or establishment of silencing (Fultz *et al.,* 2015). Metazoans also utilize a small RNA-based pathway for transposable element control. The PIWI-interacting RNA (piRNA) pathway, which maintains germline stability by preventing transposition, has been most comprehensively characterized in the model Dipteran, *Drosophila melanogaster* (Aravin *et al.,* 2001, Zamore and Haley, 2005, Brennecke *et al.,* 2007). The piRNA pathway consists of a class of small RNAs that associate with the PIWI clade of Argonaute proteins, including Piwi, Aubergine (Aub), and AGO3. The expression of PIWI proteins is primarily confined to the *D. melanogaster* germline and somatic germline support cells, which, together with their associated piRNAs, are known to have a vital role in repressing the expression and mobilization of transposons (Brennecke *et al.,* 2007, Malone *et al.,* 2009). *In D. melanogaster*, precursor piRNAs are transcribed from distinct genomic loci called piRNA clusters, which exist as dualstrand (bidirectional) or unistrand (unidirectional), depending on the direction of transcription (Brennecke *et al.,* 2007). piRNA precursors are subsequently processed into mature primary piRNAs within cytoplasmic granules

referred to as nuage in germ cells (Lim and Kai, 2007) and Yb bodies in somatic support cells (Saito *et al.,* 2010). These mature primary piRNAs are generally 26 – 30 nucleotides in length and are loaded into the protein Piwi and transported back into the nucleus where the complementary transposable elements are silenced through addition of repressive histone marks (Sienski *et al.* 2012). Two other PIWI family proteins, Aub and AGO3, generate secondary piRNAs within the cytoplasm by cleaving sense and antisense transposable element transcripts, the continued generation of which produces additional secondary piRNAs (Brennecke *et al.* 2007). This positive feedback loop of secondary piRNA production is referred to as the ping-pong loop (Brennecke et al. 2007).

1.1.4 Transposons provide indispensable genetic tools

Insertional mutagenesis has long been used as a forward genetics approach for understanding gene function and regulation. The first use of insertional mutagenesis to identify an "enhancer element" was described by Weber *et al.* (1984) after the discovery of the SV40 regulatory sequence from simian vacuolating virus 40. The first successful transposon-based enhancer trap experiments in eukaryotes were achieved using the *P* element in *D. melanogaster* (O'Kane and Gehring, 1987). In these first experiments, a *P* element cassette containing a *P* element promoter and bacterial *lacZ* fusion was used as a reporter to visualize insertion events near enhancer sequences. This experiment proved to be quite successful, showing different developmental expression patterns in developing embryos. The success of this system is due, in part, to the non-random insertion site preference of the *P* element, which shows an insertion site preference for regions 5'

(upstream) of actively translated genes (Liao *et al.,* 2000). Transposable elements have since made it possible to do forward genetic screens, like enhancer, promoter, and gene trapping, in a wide variety of eukaryotes, including plants, insects, zebrafish, mice, and others (Sundaresan *et al.,* 1995, O'Kane and Gehring, 1987, Kawakami *et al.,* 2004, Gossler *et al.,* 1989).

The concept of gene therapy has existed for decades. It was originally imagined to be of therapeutic benefit in treating inherited simple genetic disorders, such as hemophilia or severe combined immunodeficiency, in which the function of the affected gene could be restored by introducing the wild-type gene back into a patient stem cell population. The challenge to gene therapy has often been the system of delivery. The first vectors for gene therapy were derived from recombinant viruses, either retroviral vectors, such as lentiviruses and γ-retroviruses, or adeno-associated viral vectors. These systems have the ability to integrate therapeutic expression cassettes of approximately 5kb to 10kb in size into the host cell genome (Dunbar, *et al.,* 2018). However, the viral vector delivery system has shown a variety of immunogenic effects in trial patients. Viral particles, like capsid or envelope proteins can cause a strong immune system response and cytotoxicity. For example, viral vectors have previously and successfully been used to treat some immunological diseases, including SCID-X1 (X-linked severe combined immunodeficiency) and WAS (Wiskott-Aldrich syndrome), but have led to the later development of leukemia in clinical trial patients due to vector insertion in or near protooncogenes. (Kebriaei *et al.,* 2018).

Active class 2 transposons offer an alternative to viral vectors and the *Sleeping Beauty* transposon has shown the most potential as a therapeutic non-viral vector for human gene therapy (Kebriaei *et al.,* 2018). The *Sleepy Beauty* transposon is a resurrected mammalian class 2 DNA element and was the first transposon shown to be capable of efficient transposition and gene delivery in vertebrate cells (Ivics and Izsvák, 2010). The active *Sleeping Beauty* transposase was constructed by surveying a number of fish genomes, examining the coding sequences of extant *Mariner*-like DNA elements, and correcting deleterious mutations based on data from non-mammalian active *Mariner* family elements, such as *Minos* from *Drosophila hydei* (Ivics *et al.,* 1997, Loukeris *et al.,* 1995). A hyperactive version of a *Sleeping Beauty* vector was shown to be highly efficient at mediating gene transfer of therapeutic cassettes in mouse models and in porcine zygotes (Mátés *et al.,* 2009, Garrels *et al.,* 2011). In a study comparing transgene insertion site preferences in human progenitor CD4+ T cells, *Sleeping Beauty* was found to have a nearly random insertion site preference compared to both *piggyBac* transposon vectors and an MLV (murine leukemia virus) vectors, which showed a preference for insertion at transcription start sites and BRD4-associated loci (a known factor in the development of several cancers) (Gogol-Döring *et al.,* 2016). Phase I clinical trials are currently being carried out using chimeric antigen receptor T cells (CAR T cells) engineered by a *Sleeping Beauty* vector; Donor or patient derived T cells have been engineered to express receptors for CD19+ B-lymphocytes in patients with non-Hodgkin lymphoma (NHL) and acute lymphoblastic leukemia (ALL) to combat minimal residual disease and relapse (Kebriaei *et al.,* 2016).

1.5 References

Aravin, A. A., N. M. Naumova, A. V. Tulin, V. V. Vagin, Y. M. Rozovsky, and V. A. Gvozdev. 2001. "Double-Stranded RNA-Mediated Silencing of Genomic Tandem Repeats and Transposable Elements in the D. Melanogaster Germline." *Current Biology: CB* 11 (13): 1017–27.

Aziz, Ramy K., Mya Breitbart, and Robert A. Edwards. 2010. "Transposases Are the Most Abundant, Most Ubiquitous Genes in Nature." *Nucleic Acids Research* 38 (13): 4207–17.

Bhattacharyya, Madan K., Alison M. Smith, T. H. Noel Ellis, Cliff Hedley, and Cathie Martin. 1990. "The Wrinkled-Seed Character of Pea Described by Mendel Is Caused by a Transposon-like Insertion in a Gene Encoding Starch-Branching Enzyme." Cell 60 (1): 115–22.

Baucom, Regina S., James C. Estill, Cristian Chaparro, Naadira Upshaw, Ansuya Jogi, Jean-Marc Deragon, Richard P. Westerman, Phillip J. Sanmiguel, and Jeffrey L. Bennetzen. 2009. "Exceptional Diversity, Non-Random Distribution, and Rapid Evolution of Retroelements in the B73 Maize Genome." *PLoS Genetics* 5 (11): e1000732.

Bennetzen, Jeffrey L., Jeremy Schmutz, Hao Wang, Ryan Percifield, Jennifer Hawkins, Ana C. Pontaroli, Matt Estep, et al. 2012. "Reference Genome Sequence of the Model Plant Setaria." *Nature Biotechnology* 30 (6): 555–61.

Brennecke, Julius, Alexei A. Aravin, Alexander Stark, Monica Dus, Manolis Kellis, Ravi Sachidanandam, and Gregory J. Hannon. 2007. "Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila." *Cell* 128 (6): 1089–1103.

Carlton, Jane M., Robert P. Hirt, Joana C. Silva, Arthur L. Delcher, Michael Schatz, Qi Zhao, Jennifer R. Wortman, et al. 2007. "Draft Genome Sequence of the Sexually Transmitted Pathogen Trichomonas Vaginalis." *Science* 315 (5809): 207–12.

Chuong, Edward B., Nels C. Elde, and Cédric Feschotte. 2016. "Regulatory Activities of Transposable Elements: From Conflicts to Benefits." *Nature Reviews. Genetics* 18 (November): 71.

Comfort, N. C. 2001. "From Controlling Elements to Transposons: Barbara McClintock and the Nobel Prize." *Endeavour* 25 (3): 127–30.

Cossu, Rosa Maria, Claudio Casola, Stefania Giacomello, Amaryllis Vidalis, Douglas G. Scofield, and Andrea Zuccolo. 2017. “LTR Retrotransposons Show Low Levels of Unequal Recombination and High Rates of Intraelement Gene Conversion in Large Plant Genomes.” *Genome Biology and Evolution* 9 (12): 3449–62.

Cui, Xiekui, and Xiaofeng Cao. 2014. “Epigenetic Regulation and Functional Exaptation of Transposable Elements in Higher Plants.” *Current Opinion in Plant Biology*. https://doi.org/10.1016/j.pbi.2014.07.001.

Daniels, S. B., K. R. Peterson, L. D. Strausbaugh, M. G. Kidwell, and A. Chovnick. 1990. “Evidence for Horizontal Transmission of the P Transposable Element between Drosophila Species.” *Genetics* 124 (2): 339–55.

Deininger, Prescott. 2011. “Alu Elements: Know the SINEs.” *Genome Biology* 12 (12): 236.

Devos, Katrien M., James K. M. Brown, and Jeffrey L. Bennetzen. 2002. “Genome Size Reduction through Illegitimate Recombination Counteracts Genome Expansion in Arabidopsis.” *Genome Research* 12 (7): 1075–79.

Dunbar, Cynthia E., Katherine A. High, J. Keith Joung, Donald B. Kohn, Keiya Ozawa, and Michel Sadelain. 2018. “Gene Therapy Comes of Age.” *Science* 359 (6372): eaan4672.

Federoff, Nina V. 1988. Transposable elements and process for using same. USPTO 4732856. *US Patent*, filed April 3, 1984, and issued March 22, 1988. https://patentimages.storage.googleapis.com/6c/58/e8/c867d3a33bce32/US4732856.pdf.

Feschotte, Cédric, Ning Jiang, and Susan R. Wessler. 2002. “Plant Transposable Elements: Where Genetics Meets Genomics.” *Nature Reviews. Genetics* 3 (5): 329–41.

Feschotte, Cédric, and Ellen J. Pritham. 2007. “DNA Transposons and the Evolution of Eukaryotic Genomes.” *Annual Review of Genetics* 41 (1): 331–68.

Finnegan, David J. 2012. “Retrotransposons.” *Current Biology: CB* 22 (11): R432–37.

Friedli, Marc, and Didier Trono. 2015. “The Developmental Control of Transposable Elements and the Evolution of Higher Species.” *Annual Review of Cell and Developmental Biology* 31 (1): 429–51.

Fultz, Dalen, Sarah G. Choudury, and R. Keith Slotkin. 2015. “Silencing of Active Transposable Elements in Plants.” *Current Opinion in Plant Biology* 27 (October): 67–76.

Gall, J. G. 1981. “Chromosome Structure and the C-Value Paradox.” *The Journal of Cell Biology* 91 (3 Pt 2): 3s – 14s.

Garrels, Wiebke, Lajos Mátés, Stephanie Holler, Anna Dalda, Ulrike Taylor, Björn Petersen, Heiner Niemann, Zsuzsanna Izsvák, Zoltán Ivics, and Wilfried A. Kues. 2011. “Germline Transgenic Pigs by Sleeping Beauty Transposition in Porcine Zygotes and Targeted Integration in the Pig Genome.” *PloS One* 6 (8): e23573.

Gilbert, Clément, and Richard Cordaux. 2013. “Horizontal Transfer and Evolution of Prokaryote Transposable Elements in Eukaryotes.” *Genome Biology and Evolution* 5 (5): 822–32.

Gilbert, Clément, and Cédric Feschotte. 2018. “Horizontal Acquisition of Transposable Elements and Viral Sequences: Patterns and Consequences.” *Current Opinion in Genetics & Development* 49 (April): 15–24.

Gogol-Döring, Andreas, Ismahen Ammar, Saumyashree Gupta, Mario Bunse, Csaba Miskey, Wei Chen, Wolfgang Uckert, Thomas F. Schulz, Zsuzsanna Izsvák, and Zoltán Ivics. 2016. “Genome-Wide Profiling Reveals Remarkable Parallels Between Insertion Site Selection Properties of the MLV Retrovirus and the piggyBac Transposon in Primary Human CD4+ T Cells.” *Molecular Therapy: The Journal of the American Society of Gene Therapy* 24 (3): 592–606.

Gossler, A., A. L. Joyner, J. Rossant, and W. C. Skarnes. 1989. “Mouse Embryonic Stem Cells and Reporter Constructs to Detect Developmentally Regulated Genes.” *Science* 244 (4903): 463–65.

Grabundzija, Ivana, Simon A. Messing, Jainy Thomas, Rachel L. Cosby, Ilija Bilic, Csaba Miskey, Andreas Gogol-Döring, et al. 2016. “A Helitron Transposon Reconstructed from Bats Reveals a Novel Mechanism of Genome Shuffling in Eukaryotes.” *Nature Communications* 7 (March): 10716.

Haapa-Paananen, Saija, Niklas Wahlberg, and Harri Savilahti. 2014. “Phylogenetic Analysis of Maverick/Polinton Giant Transposons across Organisms.” *Molecular Phylogenetics and Evolution* 78 (September): 271–74.

Han, Jeffrey S. 2010. “Non-Long Terminal Repeat (non-LTR) Retrotransposons: Mechanisms, Recent Developments, and Unanswered Questions.” *Mobile DNA* 1 (1): 15.

Hayward, Alexander, Awaisa Ghazal, Göran Andersson, Leif Andersson, and Patric Jern. 2013. “ZBED Evolution: Repeated Utilization of DNA Transposons as Regulators of Diverse Host Functions.” *PloS One* 8 (3): e59940–e59940.

Hickman, Alison B., and Fred Dyda. 2015. “Mechanisms of DNA Transposition.” *Microbiology Spectrum* 3 (2): MDNA3–0034 – 2014.

Huang, Shengfeng, Xin Tao, Shaochun Yuan, Yuhang Zhang, Peiyi Li, Helen A. Beilinson, Ya Zhang, et al. 2016. “Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination.” *Cell* 166 (1): 102–14.

Ivics, Zoltán, Perry B. Hackett, Ronald H. Plasterk, and Zsuzsanna Izsvák. 1997. “Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells.” *Cell* 91 (4): 501–10.

Ivics, Zoltán, and Zsuzsanna Izsvák. 2010. “The Expanding Universe of Transposon Technologies for Gene and Cell Engineering.” *Mobile DNA* 1 (1): 25.

Jain, Miten, Sergey Koren, Karen H. Miga, Josh Quick, Arthur C. Rand, Thomas A. Sasani, John R. Tyson, et al. 2018. “Nanopore Sequencing and Assembly of a Human Genome with Ultra-Long Reads.” *Nature Biotechnology* 36 (January): 338.

Kapitonov, Vladimir V., and Jerzy Jurka. 2005. “RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons.” *PLoS Biology* 3 (6): e181.

Kawakami, Koichi, Hisashi Takeda, Noriko Kawakami, Makoto Kobayashi, Naoto Matsuda, and Masayoshi Mishina. 2004. “A Transposon-Mediated Gene Trap Approach Identifies Developmentally Regulated Genes in Zebrafish.” *Developmental Cell* 7 (1): 133–44.

Kebriaei, Partow, Zsuzsanna Izsvák, Suneel A. Narayanavari, Harjeet Singh, and Zoltán Ivics. 2017. “Gene Therapy with the Sleeping Beauty Transposon System.” *Trends in Genetics: TIG* 33 (11): 852–70.

Kebriaei, Partow, Harjeet Singh, M. Helen Huls, Matthew J. Figliola, Roland Bassett, Simon Olivares, Bipulendu Jena, et al. 2016. “Phase I Trials Using Sleeping Beauty to Generate CD19-Specific CAR T Cells.” *The Journal of Clinical Investigation* 126 (9): 3363–76.

Kelly, Laura J., Simon Renny-Byfield, Jaume Pellicer, Jiří Macas, Petr Novák, Pavel Neumann, Martin A. Lysak, et al. 2015. “Analysis of the Giant Genomes of Fritillaria (Liliaceae) Indicates That a Lack of DNA Removal Characterizes Extreme Expansions in Genome Size.” *The New Phytologist* 208 (2): 596–607.

Kidwell, M. G., and D. R. Lisch. 2000. “Transposable Elements and Host Genome Evolution.” *Trends in Ecology & Evolution* 15 (3): 95–99.

Koonin, Eugene V., and Mart Krupovic. 2017. “Polintons, Virophages and Transpovirons: A Tangled Web Linking Viruses, Transposons and Immunity.” *Current Opinion in Virology* 25 (August): 7–15.

Lander, Eric S., Lauren M. Linton, Bruce Birren, Chad Nusbaum, Michael C. Zody, Jennifer Baldwin, Keri Devon, et al. 2001. “Initial Sequencing and Analysis of the Human Genome.” *Nature* 409 (6822): 860–921.

Law, Julie A., and Steven E. Jacobsen. 2010. “Establishing, Maintaining and Modifying DNA Methylation Patterns in Plants and Animals.” *Nature Reviews. Genetics* 11 (February): 204.

Leitch, I. J. 2007. “Genome Sizes through the Ages.” *Heredity* 99 (2): 121–22.

Liao, G. C., E. J. Rehm, and G. M. Rubin. 2000. “Insertion Site Preferences of the P Transposable Element in Drosophila Melanogaster.” *Proceedings of the National Academy of Sciences of the United States of America* 97 (7): 3347–51.

Lim, Ai Khim, and Toshie Kai. 2007. “Unique Germ-Line Organelle, Nuage, Functions to Repress Selfish Genetic Elements in Drosophila Melanogaster.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (16): 6714–19.

Lisch, Damon. 2009. “Epigenetic Regulation of Transposable Elements in Plants.” *Annual Review of Plant Biology* 60: 43–66.

Liu, Sanzhen, Cheng-Ting Yeh, Tieming Ji, Kai Ying, Haiyan Wu, Ho Man Tang, Yan Fu, Dan Nettleton, and Patrick S. Schnable. 2009. “Mu Transposon Insertion Sites and Meiotic Recombination Events Co-Localize with Epigenetic Marks for Open Chromatin across the Maize Genome.” *PLoS Genetics* 5 (11): e1000733.

Loukeris, Thanasis G., Ioannis Livadaras, Bruno Arcà, Sophia Zabalou, and Charalambos Savakis. 1995. “Gene Transfer into the Medfly, Ceratitis Capitata, with a Drosophila Hydei Transposable Element.” *Science* 270 (5244): 2002.

Ma, Jianxin, Katrien M. Devos, and Jeffrey L. Bennetzen. 2004. “Analyses of LTR-Retrotransposon Structures Reveal Recent and Rapid Genomic DNA Loss in Rice.” *Genome Research* 14 (5): 860–69.

Malone, Colin D., Julius Brennecke, Monica Dus, Alexander Stark, W. Richard McCombie, Ravi Sachidanandam, and Gregory J. Hannon. 2009. “Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary.” *Cell* 137 (3): 522–35.

Mátés, Lajos, Marinee K. L. Chuah, Eyayu Belay, Boris Jerchow, Namitha Manoj, Abel

Acosta-Sanchez, Dawid P. Grzela, et al. 2009. “Molecular Evolution of a Novel Hyperactive Sleeping Beauty Transposase Enables Robust Stable Gene Transfer in Vertebrates.” *Nature Genetics* 41 (May): 753.

Matzke, Marjori A., Tatsuo Kanno, and Antonius J. M. Matzke. 2015. “RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants.” *Annual Review of Plant Biology* 66: 243–67.

McClintock, Barbara. 1931. *Cytological Observations of Deficiencies Involving Known Genes, Translocations and an Inversion in Zea Mays*. University of Missouri, College of Agriculture, Agricultural Experiment Station.

O’Kane, C. J., and W. J. Gehring. 1987. “Detection in Situ of Genomic Regulatory Elements in Drosophila.” *Proceedings of the National Academy of Sciences of the United States of America* 84 (24): 9123.

Pereira, Vini. 2004. “Insertion Bias and Purifying Selection of Retrotransposons in the Arabidopsis Thaliana Genome.” *Genome Biology* 5 (10): R79.

Reiss, Daphné, Gladys Mialdea, Vincent Miele, Damien M. de Vienne, Jean Peccoud, Clément Gilbert, Laurent Duret, and Sylvain Charlat. 2019. “Global Survey of Mobile DNA Horizontal Transfer in Arthropods Reveals Lepidoptera as a Prime Hotspot.” *PLoS Genetics* 15 (2): e1007965.

Richardson, Sandra R., Aurélien J. Doucet, Huira C. Kopera, John B. Moldovan, José Luis Garcia-Perez, and John V. Moran. 2015. “The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes.” *Microbiology Spectrum* 3 (2): MDNA3–2014.

Saito, Kuniaki, Hirotsugu Ishizu, Miharu Komai, Hazuki Kotani, Yoshinori Kawamura, Kazumichi M. Nishida, Haruhiko Siomi, and Mikiko C. Siomi. 2010. “Roles for the Yb Body Components Armitage and Yb in Primary piRNA Biogenesis in Drosophila.” *Genes & Development* 24 (22): 2493–98.

Sakano, Hitoshi, Konrad Hüppi, Günther Heinrich, and Susumu Tonegawa. 1979. “Sequences at the Somatic Recombination Sites of Immunoglobulin Light-Chain Genes.” *Nature* 280 (5720): 288–94.

Sienski, Grzegorz, Derya Dönertas, and Julius Brennecke. 2012. “Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression.” *Cell* 151 (5): 964–80.

Sundaresan, V., P. Springer, T. Volpe, S. Haward, J. D. Jones, C. Dean, H. Ma, and R. Martienssen. 1995. “Patterns of Gene Action in Plant Development Revealed by

Enhancer Trap and Gene Trap Transposable Elements." *Genes & Development* 9 (14): 1797–1810.

Tenaillon, Maud I., Jesse D. Hollister, and Brandon S. Gaut. 2010. "A Triptych of the Evolution of Plant Transposable Elements." *Trends in Plant Science* 15 (8): 471–78.

Thomas, Jainy, and Ellen J. Pritham. 2015. "Helitrons, the Eukaryotic Rolling-Circle Transposable Elements." *Microbiology Spectrum*. https://doi.org/10.1128/microbiolspec.mdna3-0049-2014.

Vicient, C. M., A. Suoniemi, K. Anamthawat-Jónsson, J. Tanskanen, A. Beharav, E. Nevo, and A. H. Schulman. 1999. "Retrotransposon BARE-1 and Its Role in Genome Evolution in the Genus Hordeum." *The Plant Cell* 11 (9): 1769–84.

Weber, Frank, Jean de Villiers, and Walter Schaffner. 1984. "An SV40 'enhancer Trap' Incorporates Exogenous Enhancers or Generates Enhancers from Its Own Sequences." *Cell* 36 (4): 983–92.

Wicker, T., S. Taudien, A. Houben, and B. Keller. 2009. "… -genome Snapshot of 454 Sequences Exposes the Composition of the Barley Genome and Provides Evidence for Parallel Evolution of Genome Size in Wheat and Barley." *The Plant*. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-313X.2009.03911.x.

Zamore, Phillip D., and Benjamin Haley. 2005. "Ribo-Gnome: The Big World of Small RNAs." *Science* 309 (5740): 1519–24.

Chapter 2 – Transposon end sequence requirements for *Hermes* transposition

2.1 Introduction

2.1.1 The *hAT* superfamily of Class 2 transposable elements

There are almost 20 different superfamilies of cut and paste DNA elements, which are mainly categorized based on their transposase amino acid sequence. All of the eukaryotic cut and paste superfamilies contain an acidic amino acid catalytic domain, known as the DDE/D motif, and share a signature string of conserved amino acids within the DDE/D catalytic domain (Yuan and Wessler, 2011). All *hAT* elements comprise a single open reading frame (ORF), which encodes the transposase enzyme, several hundred base pairs of flanking 5' and 3' sequence, the presence of terminal inverted repeats on the extreme 5' and 3' transposon ends, the presence of numerous subterminal repeat, or palindromic, sequences within the transposon ends, and the formation of 8-base pair target site duplications, which are formed by repaired staggered breaks in the host DNA (reviewed by Atkinson, 2015).

The *hAT* superfamily is one of the largest families of cut and paste elements (Arensburger *et al.,* 2011). This superfamily is named after several of the DNA elements initially described; *hobo* from *Drosophila melanogaster*, Barbara McClintock's *Ac* from maize, and *Tam3* from snapdragon (Strek *et al.*, 1986, McClintock, 1950, Calvi *et al.,* 1991). The *hAT* superfamily includes multiple active and well-studied transposons, including *Ac*, the *Tol2* element from Medaka fish (Kawakami and Shima, 1999), and *Hermes*, from the common housefly *Musca domestica* (Warren *et al.,* 1994). Phylogenetic analysis of transposon amino acid sequences has further divided the *hAT*

superfamily into three subfamilies. The first two described include *Ac*, which appears to be restricted to plants and fungi, and *Buster,* which has a wide distribution among animal genomes, including bats, insects, and tunicates (Arensburger *et al.,* 2011). Both *Ac* and *Buster* are the larger of the *hAT* sub-families. A smaller third family, comprising the *Tip* elements, are named after *Tip100* from morning glory (Habu *et al.,* 1998) and show a sporadic distribution in plants and insects (Atkinson, 2015).

Several active *hAT* elements have exhibited activity in a variety of organisms and have served as valuable biotechnology tools in the field of genetics. Barbara McClintock's *Ac/Ds* system from maize has been deployed over decades in a wide variety of plant species for insertional mutagenesis and trapping experiments, as well as various species of yeast (Fladung and Polak, 2012, Mielich *et al.,* 2018). The activity of the *Tol2* element in vertebrate systems has made it an important tool for developmental biologists; the element has been used for forward genetic screens in the model organism, *Danio rerio,* the zebrafish (Kawakami 2007, Holtzman *et al.,* 2016). *Tol2* has also been used in other vertebrate systems, including mouse models (Abe *et al.,* 2011) and human cell culture (Grabundzija *et al.,* 2010). The elements *Hermes* and *hobo*, both isolated from diptera, have been used for transgene delivery in various insects (Atkinson *et al.,* 2007), with *Hermes* being particularly successful. The *Hermes* transposon, which is the focus of this chapter, has been shown to stably transform diverse insect species, including the yellow fever mosquito, *Aedes aegypti*, the Mediterranean fruit fly, *Ceratitis capitata*, and the African satyrid butterfly (Jasinskiene *et al.,* 1998, Michel *et al.,* 2001, Marcus *et al.,* 2004).

2.1.2 Experimental methods for determining mechanism of transposition

The biotechnological utility of transposable elements is part of the necessity for understanding how these mobile genetic elements function. While all eukaryotic DNA transposons share the same conserved catalytic motif, the mechanism of transposition and transpososome assembly appears to vary between the superfamilies. Elucidation of the mechanism of transposition can be carried out by a variety of experiments. The relative frequency of transposition by active elements can be observed *ex vivo* in cell culture and *en vivo* in embryo transposition assays. These assays also provide evidence for insertion and target site preference. The use of *in vitro* binding assays, including plasmid cleavage assays and electrophoretic mobility shift assays (EMSA), can help determine the affinity of transposase binding to terminal inverted repeats (TIRS) and sub-terminal repeats. These same techniques can be combined to determine critical catalytic amino acids and minimal DNA recognition sequences required for various steps of transposition, like target recognition and binding, strand cleavage and insertion. The caveat to using these techniques is that results can vary between experimental systems. Different cell types and/or species will have a different repertoire of host factors available that may be required for transposition and these varying conditions can impact transposition efficiency (Miskey *et al.,* 2005).

The type of experimental data described above is even more informative when combined with protein crystal structure data. There are difficulties in obtaining protein crystals, however, and sometimes numerous crystallization conditions need to be assayed to find the right combination of conditions for crystallization (Kurpiewska and Lewinski,

2010). Partial crystal and co-crystal structures have been obtained for the *Mariner* superfamily DNA element, *Mos1*, from *D. melanogaster*, and the *Mu* bacteriophage transposon (Richardson *et al.*, 2009, Montaño *et al.,* 2012). Crystal structure has also been solved for the prokaryotic *Tn5* transposon, which forms hair-pin DNA intermediates, similar to eukaryotic *hAT* elements (Davies *et al.,* 2000).

2.1.3 Mechanisms of *hAT* transposition

The isolation of active transposase proteins is necessary for both *in vitro* biochemical studies and protein crystallography. Despite the early identification of *Ac* from maize and the plethora of other identified *hAT* elements, only the *Herme*s transposase has been studied in great detail, due to the ability to produce and purify the protein (Hickman *et al.,* 2005). Early research on *Hermes* revealed a partial-crystal structure containing an RNase-H like domain interrupted by an alpha-helix "insertion" domain, and an N-terminal dimerization domain (Hickman *et al.,* 2005, Zhou *et al.,* 2004). A truncated version of *Hermes* has been crystallized while the active site is bound to a 16-base-pair oligomer of the left end terminal inverted repeat (Hickman *et al.,* 2014) and this has provided insight into how transposases recognize their targets. The *Hermes* element has catalytic activity *in vitro* as a dimeric protein in low ionic (non-physiological) conditions and it was observed to form a multimeric structure when expressed in insect cell culture (Hickman *et al.,* 2005). The 2014 report, however, showed that the *Hermes* element is active *in vivo* as an octamer, or a tetramer of dimers, which makes it unique among studied transposons (Hickman *et al.,* 2014).

Indeed, it is the *Hermes* transposase that has given us the greatest insight into the biochemical mechanisms of *hAT* element transposition. The mechanism of *Hermes* element transposition includes the generation of a double-stranded DNA break and the formation of a DNA hair-loop. Hairpin loops are also formed during transposase mediated DNA cleavage by *piggyBac*, *Transib*, and *Mutator* elements, and the prokaryotic *Tn5* element (Liu and Wessler, 2017, Davies *et al.,* 2000), however, *hAT* elements are unique in that the hairpin loop is formed on the flanking DNA (Hickman *et al.,* 2014). The chemical reaction for excision starts with a single-strand nick on the 5' end of the element, after which a 3'OH group is freed on each end, which subsequently performs a nucleophilic attack on the adjacent strand, ending in the formation of a hairpin loop on the adjacent DNA, and producing another 3'OH on the transposon side of the DNA , which will later attack the target DNA strands during integration (Zhou *et al.,* 2004). Several crucial domains within *Hermes* are recognized as being necessary to stabilize and facilitate transposition. While the DDE/D catalytic motif is known to be necessary for excision, several residues have been shown to facilitate the stabilization of the transferred DNA strands and conformational change within the *Hermes* transpososome. For example, a hydrophobic tryptophan, W319, which resides inside of the *Hermes* catalytic pocket, is required for stable hairpin formation (Hickman *et al.,* 2005). Histidine H268 is part of a conserved CxxH/CxxC motif that is conserved in *hAT* elements, as well as in the *P*, *MULE*, and *Kolobok* DNA elements (Yuan and Wessler, 2011). The H268 residue of the conserved CxxH motif is predicted to stabilize the DNA within the hydrophobic catalytic pocket (Hickman *et al.,* 2014).

2.1.4 The role of *hAT* element subterminal repeats

The subterminal repeats found within *hAT* element ends are a discrete feature that is known to be necessary for accurate transposition (Kunze, *et al.,* 1989, Urasaki *et al.,* 2006). During transposition, the transposon terminal inverted repeats (TIRs) are bound more weakly than the sub-terminal repeats, which are presumed to be bound by the C_2H_2 Zn^{2+} finger binding domain, called the BED domain, present in *hAT* element N-terminal ends (Kahlon *et al.,* 2011, Hickman *et al.,* 2014). The 5' and 3' ends of the *Hermes* element are not completely symmetric; the 17-base-pair TIRs on each end are imperfect repeats, with the left TIR showing a stronger affinity for active site binding (Zhou *et al.,* 2004). Additionally, the subterminal repeat of 5'-**GTGGC**-3' is present three times in the left end, at positions 13-25, 36-48, 69-81, and 281-293 bases in from the TIR, while the right end repeats appear at 13-25, 30-42, 186-197, with a cryptic motif of 5'-**GTG**TT-3' at 49-61 bases in from the TIR (Smith and Atkinson, 2011, Hickman *et al.,* 2014).

The *Hermes* co-crystal structure does not include the N-terminal BED domain – it was found that the full-length transposase with an N-terminal Histidine tag would form aggregates (Hickman, *et al.,*2005). The *Hermes* co-crystal structure successfully used for crystallization, Hermes79-612, was truncated at the N-terminal end and the BED domain was absent (Hickman *et al.,* 2014). This protein was catalytically active *in vitro* using 30-base-pair oligomers of the left terminal inverted repeat. However, the crystal structure revealed that the transposase is an octameric ring, containing a tetramer of dimers, with a single dimer binding the terminal inverted repeats and additional modeling suggested that the eight absent BED domains would reside arrayed along the inside of the octameric ring

(Hickman *et al.,* 2014). This implies that the eight BED domains may be in a position to recognize and bind to the numerous discreet subterminal repeats of *Hermes.* Previous experiments using *Herves*, an active *hAT* element from the malaria mosquito, *Anopheles gambiae*, suggested that there is cooperative binding in at least two locations in each transposon ends that correspond to the locations of the subterminal repeats in *Herves* (Kahlon *et al.,* 2011).

2.1.5 Chapter aims

The goal of the research described in this chapter was to understand the importance of *Hermes* transposon ends in transposition frequency and accuracy. Previous experiments have indicated that the discrete subterminal repeats found in *hAT* elements may be bound by the BED domain and necessary for recognition and transposition *in vivo* and that there is some asymmetry between the right and left transposons ends (Hickman *et al.,* 2005, Kahlon *et al.,* 2011, Kim *et al.,* 2011). The following describes plasmid-based transposition assays using length variations of *Hermes* transposon ends for *ex vivo* experiments. Additionally, sequence analysis revealed a natural polymorphism in the *Hermes* transposon in one of the subterminal repeats of the *Hermes* left end. The first subterminal repeat of the *Hermes* right end, bases 13-25, show a polymorphism of 5'-**GTGAC**-3' as opposed to the expected 5'-**GTGGC**-3' repeat. Additional plasmid-based transposition assays have been performed to determine if the presence of this natural single-base polymorphism would impact recognition of the ends and change transposition frequency or impact target-site preference.

2.2 Materials and methods

2.2.1 Plasmid construction

Construction methods for *Hermes* donor plasmid pHDG1 are described by Wright *et al.,* 2013.

Construction of donor pHDG8: Short *Hermes* ends (30 bp LE and RE) along with vector sequences (ampicillin resistance and a replication origin) and homology sequences were prepared by PCR using a 4kb fragment of clone pHDG1 digested with *Xho*I and *Bgl*II as template.

Primer Hermes 1-30 For:

5'-*GACACTATTCAACTAC*GTTTGCCTGTGACTTGTTGAAGT-3'

Underlined sequence has homology to the *Hermes* RE TIR, while the italicized sequence provides homology to the stuffer fragment.

Primer *Hermes* 1-30 Rev:

5'-CATGCCCTTGGCTA*G*TCAAAATAAGCCACTTGTTGTTGTTCTCTG-3'

Underlined sequence has homology to the Hermes LE TIR, while the italicized sequence provides homology to the pGoE plasmid.

The stuffer fragment was amplified from the *piggyBac* transposon plasmid pBac3xP3dsRed/AgB2tEGFP, and was included in the construct as "spacer" DNA to make the Hermes elements from pHDG7 and pHDG8 the same size. This PCR fragment also has homology to the *Hermes* PCR fragment on one end and the pGoE plasmid on the other for assembly purposes.

Primer Stuffer Fragment For:

5’-CTTGCTTTAAAGCTAGGTCAGTCAGAAACAACTTTGGC-3’

Underlined region has homology to the piggyBac element, while the italicized region has homology to the pGoE vector for assembly purposes.

Primer Stuffer Fragment Rev:

5’-TAGTTGAATAGTGTCGGTATACTTATTATCATCTTGTGATGAGGA-3’

Underlined region is homologous to the *piggyBac* element, while the italicized region has homology to the *Hermes* vector fragment for assembly.

Plasmid pGoE, which has a gentamycin resistance gene, a replication origin, and EGFP fused in-frame with the lacZ-alpha fragment was digested with *Nhe*I, run on an agarose gel, and purified. The *Hermes* vector fragment, the stuffer fragment, and the pGoE vector were assembled using the GeneArt Seamless Cloning and Assembly Kit (Life Technologies), transferred to TOP10 cells, and plated onto LB agar containing ampicillin and gentamycin.

Construction of pHDG7: The Hermes clone with long ends (711 bp LE, 520 bp RE) was constructed from multiple fragments. Clone pHDG1 was used as the starting vector, and was digested with MfeI and BglII, and the 3kb fragment containing the appropriate target site duplications and partial ends of *Hermes* was purified on an agarose gel. Fragments of the remainder of the longer ends were constructed by amplifying separately the *Hermes* LE and RE from the MfeI sites to base 711 (left) or 520 (right) using the 1kb fragment of vector pBSHermes digested with MfeI as template.

Primer Hermes LE For:

5’-TGCTACTTATGAGTACAATTGTGCTTTGCCACTTGAAC-3’

Underlined region is homologous to the pBSHermes MfeI fragment, while the italicized portion provides homology to the partial left end of the pHDGI fragment.

Primer Hermes LE Rev:

5’-CATGCCCTTGGCTAGCTGAAACAGTTTTTAATTCTCGGGATT-3’

Underlined region is homologous to the left end of pBSHermes, while the italicized portion provides homology to the pGentOriAlpha vector.

Primer Hermes RE For:

5’-GAATGGCGATAAGCTAGTCGACAGCTTGTTATTTTTAAATTCC-3’

Underlined region has homology to the Hermes right end from pBSHermes, while the italicized region provides homology to the pGentOriAlpha vector.

Primer Hermes RE Rev:

5’-GTGCGATTTGTCAATTGGCAAATTATACTCACTTCTTGTTG-3’

Underlined region has homology to the Hermes right end MfeI region, while the italicized portion provides homology to the partial right end of the pHDGI fragment.

pGentOriAlpha, containing a gentamycin resistance gene, the LacZ-Alpha fragment, and a replication origin was digested with NheI and gel purified from an agarose gel. The four fragments pHDGI MfeI, the partial right and left end PRC products, and the pGentOriAlpha vector were assembled and plated as above.

Construction of pHDG-EGFP: pHDG1 was used as template for Taq polymerase to amplify 305 bases of the *Hermes* LE flanked by SacI and XbaI sites. The PCR product

was purified, digested with SacI and XbaI, and cloned into pBluescipt SK+ digested with the same enzymes to give clone pHermesL305. A *Hermes* RE of 307 bases was amplified by PCR also using pHDG1 as template, digested with PstI and EcoRV, and also cloned into pBluescipt SK+ digested with the same enzymes. The RE of *Hermes* was digested with the above enzymes, purified on an agarose gel, and ligated to the left end clone also digested with PstI and EcoRV to give clone pHermesL305R307. This clone was digested with XbaI and ligated to pGoE vector which had been digested with NheI to give clone pHDG-EGFP.

HL8bpFSac:

5’ - TGAGAGCTCGTCTGTATCAGAGAACAACAACAAGTGGCTTATTTTG – 3’

Underlined region is homologous to the *Hermes* LE.

Hermes L305 Xba Rev:

5’-GATTCTAGACACACTCAAGTGCATAAGCCACTTGTTAGC-3’

Underlined region is homologous to the *Hermes* LE.

Hermes R307 Pst For:

5’-CATCTGCAGCAGAATCATATGCAATACTACAAACAATAGCACACAC-3’.

Underlined region is homologous to the *Hermes* RE.

Hermes R TSD RV Rev:

5’-GATGATATCATACAGACCAGAGAACTTCAACAAGTCACAGGC-3’

Underlined region is homologous to the *Hermes* RE.

Construction of pHDGLL: A second LE of 305 bases flanked by XhoI and ClaI sites was amplified by PCR and cloned into the same sites in the pBSHL305G clone to give pHDGLL.

Hermes L2 Xho For:

5'-GATCTCGAGATACAGACCAGAGAACAACAACAAGTGGCTTATTTTGATAC-3'

The underlined region is homologous to the *Hermes* LE.

Hermes L2 ClaI seq Rev:

5'-

GATGATATCGATGTTTTGGGAAATCATCCACACTCAAGTGCATAAGCCACTTGTTAGC-3'

The underlined region is homologous to the *Hermes* LE; the rest is "spacer" DNA.

2.2.2 Interplasmid transposition assays

Assays in *Drosophila* S2 DEV8 cells: S2 DEV8 (stably expressing *Hermes* transposase, Michel *et al.*, 2003) cells were seeded in treated 6-well plates at a density of $2x10^6$ cells per well. Cells were transfected the next day following the X-tremeGENE HP DNA Transfection Reagent protocol (Roche). For three plasmid comparison assays, donor plasmids and their corresponding positive control plasmid (1 μg each) were added along with pGDV1 (Bron *et al.,* 1991) target plasmid (2 μg) to transfection reagent. For two plasmids assays, 2 μg each of donor and target were added. Plasmid Actin5C-EGFP was used as a positive control for transfection. DEV8 cells were induced to express

Hermes the day following transfection by the addition of $CuSO_4$ to a concentration 250 μM to each well (excluding transfection control). Cells were harvested two days after transfection, washed with PBS, and frozen for plasmid preparations. Cell pellets were resuspended with 500 μl of grinding buffer (0.5% SDS, 0.08 M NaCl), 54.7 mg/ml sucrose, 0.06 M EDTA, 120 mM Tris pH9.0) before being incubated at 65°C for 30 min. Potassium acetate was added to a final concentration of 1 M prior to incubating on ice for 30 min. Samples were centrifuged, and the DNA in the supernatant was precipitated with ethanol. Reactions were resuspended in nuclease free water and electroporated into bacteria and plated. Colonies were selected by dual antibiotic resistance as a transposition events into the target plasmid would confer both gentamicin and chloramphenicol resistance. Selected colonies mini-prepped using 5PRIME Fast Plasmid Mini-Prep kit and assayed by restriction enzyme digest. Colonies that passed restriction enzyme digestion were sequenced using Sanger sequencing (Sanger *et al.,* 1977) from both the right and left *Hermes* transposon ends using the following primers:

Hermes 2529:

5’ – AATTTGCCAATTGACAAATCGCACACGTCC – 3’

GentOut:

5’ - GTTGTTCGGTAAATTGTCACAAC – 3’

Target site duplications and integration locations into target plasmid pGDV1 were recorded. Transposition frequencies were calculated using only confirmed transposition events after sequencing.

2.3 Results

2.3.1 Variation of *Hermes* end lengths

Variation in *Hermes* transposon end length affected transposition rates *ex vivo* in interplasmid transposition assays. The plasmid pHDG1, with moderate length ends, includes 387 bases of the *Hermes* right end and 444 bases of the left end, while the plasmid pHDG7 has full length right and left *Hermes* ends (520 bases, and 711 bases, respectively). Transposition frequency *ex vivo* was ten-fold higher for the longer *Hermes* donor, pHDG7, versus the moderate length donor, pHDG1. Notably, transposition events were never recovered for plasmid pHDG8, which includes the outermost 30 bases, including the TIRs and the first subterminal repeat motif of each *Hermes* end (Table 2.1).

2.3.2 *Hermes* end symmetry

Variation in *Hermes* end symmetry effected transposition rates *ex vivo* in interplasmid transposition assays. Plasmid pHDG-EGFP contains 305 bases of the *Hermes* right and left ends, and was compared to either plasmid pHDGLL, which has two 305 base pair left ends (Figure 2.2), or plasmid pHDGLRL, which has one 305 base pair left end and a 307 base pair right end immediately followed by another 305 base pair left end (Figure 2.1). No transposition events were recovered for the plasmids with asymmetric ends (pHDGLL, pHDGLRL) after the conclusion of ten total interplasmid transposition assays (Table 2.2). Only *Hermes* donors retaining moderate to full length right and left ends in the natural orientation were capable of transposition.

2.3.3 A natural polymorphism in the *Hermes* subterminal repeat

Sequencing of *Hermes* plasmids (derived from genomic sequence of *Musca domestica*) revealed a natural single base pair polymorphism in the first subterminal repeat, bases 13-25, of the *Hermes* right end, which is 5'-**GTGAC**-3' as opposed to the expected 5'-**GTGGC**-3' repeat. Eight total interplasmid transposition assay experiments were carried out to assess if the rate of *Hermes* transposition is affected by the natural polymorphism in the outermost right end subterminal repeat (Table 2.3). There was no observable impact on transposition frequency. The target site duplications produced by all transposition events in these experiments were collected and compared to evaluate if the single base pair polymorphism impacted the accuracy of transposition compared to the canonical subterminal repeat sequence (Table 2.4). There were no observable differences in transposition accuracy and target site preferences that could be attributed to the 5'-**GTGAC**-3' polymorphism.

2.3.4 Summary of results

Longer *Hermes* transposon ends showed a higher transposition frequency compared to moderate length ends. While short *Hermes* oligomers are active *in vitro*, no transposition events were recovered using *Hermes* 30-mer TIR ends in *ex vivo* experiments. The end symmetry of *Hermes* also impacts transposition, as constructs containing either two *Hermes* left ends, or a right end immediately joined to an additional left end had no activity in *ex vivo* experiments. The presence of a natural single base pair

polymorphism in the outermost right end subterminal repeat did not have an impact on *Hermes* transposition frequency or accuracy.

2.4 Discussion

The *hAT* superfamily of transposable elements is ancient and widespread amongst eukaryotic genomes. Due in part to their ubiquity, identified active *hAT* elements have been successfully deployed as biotechnology tools for genetic research in plants, insects, and animal systems. Additionally, the occurrence of transposon exaptation by host genomes has been observed across eukaryotic species, including the human genome (Huda *et al.,* 2010). Humans have an entire gene family, called ZBED, that has arisen from exaptation of the *Buster* subfamily of *hAT* transposons (Hayward *et al.,* 2013).

Previously, many biochemical studies of DNA transposition have relied on prokaryotic transposases, such as *Tn5* (reviewed by Hickman *et al.,* 2010), which are not easily comparable to all eukaryotic transposons, including the *hAT* superfamily. A previous study using a partial protein, $Hermes_{79\text{-}612}$, obtained crystal structure data suggesting that the *Hermes* element forms a ring-shaped protein composed of a trimer of dimers (Hickman *et al.,* 2005). This data was refined after subsequent crystallography data revealed that *Hermes* forms an octameric ring as a tetramer of dimers (Hickman *et al.,* 2014). This *Hermes* co-crystal structure was achieved using a truncated version of the protein, $Hermes_{79\text{-}612}$, due to the inability to crystalize the complete protein (Hickman *et al.,* 2005). The first 78 amino acids of *Hermes* comprise the N-terminal BED domain, from residues 27-78. The BED domain contains zinc-chelating residues and is a

conserved eukaryotic DNA-binding domain (Aravind, 2000). This co-crystal structure also used short, 16-base pair left end TIRs, which is the outermost portion of the transposon end that will be bound is the active site. Indeed, the co-crystal structure showed an array of residues within protein interacting with the TIR DNA from the first nucleotide up to the eleventh nucleotide. Modeling and experimental data have been used to predict the complete octameric structure and suggest that the eight N-terminal BED domains would reside within the center of the octamer in an ordered array (Hickman *et al.,* 2014). These BED domains would be positioned in such a way that multiple subterminal repeats could be bound it the same time, while the transposon TIRs would be bound in a single active site out of the four available within the octamer (Hickman *et al.,* 2014). Indeed, while a dimeric version of *Hermes* retains catalytic activity *in vitro* with 30-mer TIRs, the 30-mer *Hermes* donors, which contain only the first subterminal repeat motif, are not active in *ex vivo*, in the transposition assays outlined in this chapter, confirming the importance of the subterminal repeats for binding.

The asymmetric *Hermes* donors, pHDGLL and pHDGLRL, used in the *ex vivo* experiments described here showed no activity. This indicates that both the left and right ends of *Hermes* are required for transposition. This is distinct from the *Mos1* and *piggyBac* transposases, which both exhibit similar or increased transposition frequency with completely symmetric ends (Zhang *et al.,* 2001, Augé-Gouillou *et al.,* 2001, Elick *et al.,* 1997). While both *piggyBac* and *Mos1* belong to the diverse group of DDE/D class 2 transposons, the *piggyBac* and *Mariner* superfamilies are more distantly related to the *hAT* superfamily (Yuan and Wessler, 2011), and both transposases are structurally

distinct from *Hermes*. While there is little structural information known about *piggyBac*, recent studies show that the C-terminal Cysteine-Rich Domain (CRD) is responsible for binding a specific motif within the TIRs (Morellet *et al.,* 2018). The *Mos1* transposon is known to function as a homodimer that binds short 28-base-pair TIRs (Richardson *et al.,* 2009). Similar to Hermes, the *hAT* elements *Ac*, and *Tol2*, are also inactive as donors possessing right and left ends arranged out of natural order (Coupland *et al.,* 1989, Urasaki *et al.,* 2006). Additionally, the *P* element, the namesake of the closely-related *P* family, is also inactive as a donor when the transposon ends are arranged out of natural order (Yuan and Wessler, 2001, Mullins *et al.,* 1989). This suggests that other *hAT* elements, and potentially closely related *P* family elements, may be structurally similar to *Hermes,* and possess similar mechanisms for recognizing their target DNA amongst host DNA.

2.5 References

Abe, Gembu, Maximilliano L. Suster, and Koichi Kawakami. 2011. "Chapter 2 - Tol2-Mediated Transgenesis, Gene Trapping, Enhancer Trapping, and the Gal4-UAS System." In , edited by H. William Detrich, Monte Westerfield, and Leonard I. B. T-Methods in Cell Biology Zon, 104:23–49. Academic Press.

Aravind, L. 2000. "The BED Finger, a Novel DNA-Binding Domain in Chromatin-Boundary-Element-Binding Proteins and Transposases." *Trends in Biochemical Sciences* 25 (9): 421–23.

Arensburger, Peter, Robert H. Hice, Liqin Zhou, Ryan C. Smith, Ariane C. Tom, Jennifer A. Wright, Joshua Knapp, David A. O'Brochta, Nancy L. Craig, and Peter W. Atkinson. 2011. "Phylogenetic and Functional Characterization of the hAT Transposon Superfamily." *Genetics* 188 (1): 45 LP – 57.

Atkinson, P. W., D. A. O'brochta, and N. L. Craig. 2007. "The Hobo, Hermes and Herves Transposable Elements of Insects." In *Area-Wide Control of Insect Pests*, 61–71. Springer Netherlands.

Atkinson, Peter W. 2015. "hAT Transposable Elements." In , 775–802. American Society of Microbiology.

Augé-Gouillou, C., M-H Hamelin, M-V Demattei, M. Periquet, and Y. Bigot. 2001. "The Wild-Type Conformation of the Mos-1 Inverted Terminal Repeats Is Suboptimal for Transposition in Bacteria." *Molecular Genetics and Genomics: MGG* 265 (1): 51–57.

Bhattacharyya, Madan K., Alison M. Smith, T. H. Noel Ellis, Cliff Hedley, and Cathie Martin. 1990. "The Wrinkled-Seed Character of Pea Described by Mendel Is Caused by a Transposon-like Insertion in a Gene Encoding Starch-Branching Enzyme." *Cell* 60 (1): 115–22.

Calvi, Brian R., Timothy J. Hong, Seth D. Findley, and William M. Gelbart. 1991. "Evidence for a Common Evolutionary Origin of Inverted Repeat Transposons in Drosophila and Plants: Hobo, Activator, and Tam3." *Cell* 66 (3): 465–71.

Coupland, G., C. Plum, S. Chatterjee, A. Post, and P. Starlinger. 1989. "Sequences near the Termini Are Required for Transposition of the Maize Transposon Ac in Transgenic Tobacco Plants." *Proceedings of the National Academy of Sciences of the United States of America* 86 (23): 9385.

Davies, Douglas R., Igor Y. Goryshin, William S. Reznikoff, and Ivan Rayment. 2000. "Three-Dimensional Structure of the Tn5 Synaptic Complex Transposition Intermediate." *Science* 289 (5476): 77 LP – 85.

Dijl, Jan Maarten van, Anne de Jong, Hilde Smith, Sierd Bron, and Gerard Venema. 1991. "Non-Functional Expression of Escherichia Coli Signal Peptidase I in Bacillus Subtilis." *Microbiology* 137 (9): 2073–83.

Elick, T. A., N. Lobo, and M. J. Fraser Jr. 1997. "Analysis of the Cis-Acting DNA Elements Required for piggyBac Transposable Element Excision." *Molecular & General Genetics: MGG* 255 (6): 605–10.

Fladung, Matthias, and Olaf Polak. 2012. "Ac/Ds-Transposon Activation Tagging in Poplar: A Powerful Tool for Gene Discovery." *BMC Genomics* 13 (1): 61–61.

Grabundzija, Ivana, Zsuzsanna Izsvák, and Zoltán Ivics. 2011. "Insertional Engineering of Chromosomes with Sleeping Beauty Transposition: An Overview." *Methods in Molecular Biology* 738: 69–85.

Habu, Yoshiki, Yasuyo Hisatomi, and Shigeru Iida. 1998. "Molecular Characterization of the Mutable Flaked Allele for Flower Variegation in the Common Morning Glory." *The Plant Journal: For Cell and Molecular Biology* 16 (3): 371–76.

Hayward, Alexander, Awaisa Ghazal, Göran Andersson, Leif Andersson, and Patric Jern. 2013. "ZBED Evolution: Repeated Utilization of DNA Transposons as Regulators of Diverse Host Functions." *PloS One* 8 (3): e59940–e59940.

Hickman, Alison B., Hosam E. Ewis, Xianghong Li, Joshua A. Knapp, Thomas Laver, Anna-Louise Doss, Gökhan Tolun, et al. 2014. "Structural Basis of hAT Transposon End Recognition by Hermes, an Octameric DNA Transposase from Musca Domestica." *Cell* 158 (2): 353–67.

Hickman, Alison B., Zhanita N. Perez, Liqin Zhou, Primrose Musingarimi, Rodolfo Ghirlando, Jenny E. Hinshaw, Nancy L. Craig, and Fred Dyda. 2005. "Molecular Architecture of a Eukaryotic DNA Transposase." *Nature Structural & Molecular Biology* 12 (8): 715–21.

Hickman, Alison Burgess, Michael Chandler, and Fred Dyda. 2010. "Integrating Prokaryotes and Eukaryotes: DNA Transposases in Light of Structure." *Critical Reviews in Biochemistry and Molecular Biology* 45 (1): 50–69.

Holtzman, Nathalia G., M. Kathryn Iovine, Jennifer O. Liang, and Jacqueline Morris. 2016. "Learning to Fish with Genetics: A Primer on the Vertebrate Model Danio Rerio." *Genetics* 203 (3): 1069–89.

Huang, Shengfeng, Xin Tao, Shaochun Yuan, Yuhang Zhang, Peiyi Li, Helen A. Beilinson, Ya Zhang, et al. 2016. “Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination.” *Cell* 166 (1): 102–14.

Huda, Ahsan, Leonardo Mariño-Ramírez, and I. King Jordan. 2010. “Epigenetic Histone Modifications of Human Transposable Elements: Genome Defense versus Exaptation.” *Mobile DNA* 1 (1): 2–2.

Jasinskiene, Nijole, Craig J. Coates, Mark Q. Benedict, Anthony J. Cornel, Cristina Salazar Rafferty, Anthony A. James, and Frank H. Collins. 1998. “Stable Transformation of the Yellow Fever Mosquito, Aedes Aegypti, with the Hermes Element from the Housefly.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (7): 3743.

Kahlon, Amandeep S., Robert H. Hice, David A. O’Brochta, and Peter W. Atkinson. 2011. “DNA Binding Activities of the Herves Transposase from the Mosquito Anopheles Gambiae.” *Mobile DNA* 2 (1): 9–9.

Kawakami, K., and A. Shima. 1999. “Identification of the Tol2 Transposase of the Medaka Fish Oryzias Latipes That Catalyzes Excision of a Nonautonomous Tol2 Element in Zebrafish Danio Rerio.” *Gene* 240 (1): 239–44.

Kawakami, Koichi. 2007. “Tol2: A Versatile Gene Transfer Vector in Vertebrates.” *Genome Biology* 8 (1): S7–S7.

Kim, Yu Jung, Robert H. Hice, David A. O’Brochta, and Peter W. Atkinson. 2011. “DNA Sequence Requirements for Hobo Transposable Element Transposition in Drosophila Melanogaster.” *Genetica* 139 (8): 985–985.

Kunze, R., and P. Starlinger. 1989. “The Putative Transposase of Transposable Element Ac from Zea Mays L. Interacts with Subterminal Sequences of Ac.” *The EMBO Journal* 8 (11): 3177–85.

Kurpiewska, Katarzyna, and Krzysztof Lewiński. 2010. “High Pressure Macromolecular Crystallography for Structural Biology: A Review.” *Central European Journal of Biology* 5 (5): 531–42.

Lazarow, Katina, My-Linh Du, Ruth Weimer, and Reinhard Kunze. 2012. “A Hyperactive Transposase of the Maize Transposable Element Activator (Ac).” *Genetics* 191 (3): 747.

Liu, Dong, Alyson Mack, Rongchen Wang, Mary Galli, Jason Belk, Nan I. Ketpura, and Nigel M. Crawford. 2001. “Functional Dissection of the Cis-Acting Sequences of the Arabidopsis Transposable Element Tag1 Reveals Dissimilar Subterminal Sequence and Minimal Spacing Requirements for Transposition.” *Genetics* 157 (2): 817.

Liu, Kun, and Susan R. Wessler. 2017. “Transposition of Mutator–like Transposable Elements (MULEs) Resembles hAT and Transib Elements and V(D)J Recombination.” *Nucleic Acids Research* 45 (11): 6644–55.

Marcus, Jeffrey M., Diane M. Ramos, and Antónia Monteiro. 2004. “Germline Transformation of the Butterfly Bicyclus Anynana.” *Proceedings. Biological Sciences / The Royal Society* 271 Suppl 5 (August): S263–65.

McClintock, Barbara. 1950. “The Origin and Behavior of Mutable Loci in Maize.” *Proceedings of the National Academy of Sciences* 36 (6): 344 LP – 355.

Michel, Agnès H., Riko Hatakeyama, Philipp Kimmig, Meret Arter, Matthias Peter, Joao Matos, Claudio De Virgilio, and Benoît Kornmann. 2017. “Functional Mapping of Yeast Genomes by Saturated Transposition.” *eLife* 6: e23570–e23570.

Michel, K., At Stamenova, A. C. Pinkerton, G. Franz, A. S. Robinson, A. Gariou-Papalexiou, A. Zacharopoulou, D. A. O’Brochta, and P. W. Atkinson. 2001. “Hermes-Mediated Germ-Line Transformation of the Mediterranean Fruit Fly Ceratitis Capitata.” *Insect Molecular Biology* 10 (2): 155–62.

Michel, K., D. A. O’Brochta, and P. W. Atkinson. 2003. “The C-Terminus of the Hermes Transposase Contains a Protein Multimerization Domain.” *Insect Biochemistry and Molecular Biology* 33 (10): 959–70.

Mielich, Kevin, Ella Shtifman-Segal, Julia C. Golz, Guisheng Zeng, Yue Wang, Judith Berman, and Reinhard Kunze. 2018. “Maize Transposable Elements Ac/Ds as Insertion Mutagenesis Tools in Candida Albicans.” *G3: Genes|Genomes|Genetics* 8 (4): 1139 LP – 1145.

Miskey, C., Z. Izsvák, K. Kawakami, and Z. Ivics. 2005. “DNA Transposons in Vertebrate Functional Genomics.” *Cellular and Molecular Life Sciences: CMLS* 62 (6): 629–629.

Montaño, Sherwin P., Ying Z. Pigli, and Phoebe A. Rice. 2012. “The Mu Transpososome Structure Sheds Light on DDE Recombinase Evolution.” *Nature* 491 (November): 413.

Mullins, M. C., D. C. Rio, and G. M. Rubin. 1989. “Cis-Acting DNA Sequence Requirements for P-Element Transposition.” *Genes & Development* 3 (5): 729–38.

Richardson, Julia M., Sean D. Colloms, David J. Finnegan, and Malcolm D. Walkinshaw. 2009. “Molecular Architecture of the Mos1 Paired-End Complex: The Structural Basis of DNA Transposition in a Eukaryote.” *Cell* 138 (6): 1096–1108.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. “DNA Sequencing with Chain-Terminating Inhibitors.” *Proceedings of the National Academy of Sciences* 74 (12): 5463 LP – 5467.

Smith, Ryan C., and Peter W. Atkinson. 2011. “Mobility Properties of the Hermes Transposable Element in Transgenic Lines of Aedes Aegypti.” *Genetica* 139 (1): 7–22.

Streck, R. D., J. E. MacGaffey, and S. K. Beckendorf. 1986. “The Structure of Hobo Transposable Elements and Their Insertion Sites.” *The EMBO Journal* 5 (13): 3615–23.

Urasaki, Akihiro, Ghislaine Morvan, and Koichi Kawakami. 2006a. “Functional Dissection of the Tol2 Transposable Element Identified the Minimal Cis-Sequence and a Highly Repetitive Sequence in the Subterminal Region Essential for Transposition.” *Genetics* 174 (2): 639–49.

Warren, William D., Peter W. Atkinson, and David A. O’Brochta. 1994. “The Hermes Transposable Element from the House Fly, Musca Domestica, Is a Short Inverted Repeat-Type Element of the Hobo, Ac, and Tam3 (hAT) Element Family.” *Genetical Research* 64 (2): 87–97.

Wright, Jennifer A., Ryan C. Smith, Xianghong Li, Nancy L. Craig, and Peter W. Atkinson. 2013. “IPB7 Transposase Behavior in Drosophila Melanogaster and Aedes Aegypti.” *Insect Biochemistry and Molecular Biology* 43 (10): 899–906.

Yuan, Yao-Wu, and Susan R. Wessler. 2011. “The Catalytic Domain of All Eukaryotic Cut-and-Paste Transposase Superfamilies.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (19): 7884–89.

Zhang, Lei, Angela Dawson, and David J. Finnegan. 2001. “DNA-Binding Activity and Subunit Interaction of the Mariner Transposase.” *Nucleic Acids Research* 29 (17): 3566–75.

Zhou, Liqin, Rupak Mitra, Peter W. Atkinson, Alison Burgess Hickman, Fred Dyda, and Nancy L. Craig. 2004. “Transposition of hAT Elements Links Transposable Elements and V(D)J Recombination.” *Nature* 432 (7020): 995–1001.

2.6 Figures and Tables

Donor plasmid	*Hermes* ends	Donors screened	True events	Frequency per donor
pHDG8	30-mer ends	341,575	0	0
pHDG7	long (R=520bp, L=711bp)	386,150	30	7.77E-05
pHDG8	30-mer ends	657,833	0	0
pHDG8	30-mer ends	810,875	0	0
pHDG1	medium (R=387bp, L=444bp)	699,500	5	7.15E-06

Table 2.1 – Transposition frequencies observed in *ex vivo* inter-plasmid transposition assays using *Hermes* donors with varying transposon end lengths.

Donor plasmid	*Hermes* ends	Donors screened	True events	Frequency per donor
pHDG-EGFP	medium (305bp each)	1,154,125	3	2.60E-06
pHDGLL	two left ends (305bp each)	1,560,125	0	0
pHDG-EGFP	medium (305bp each)	1,082,123	4	3.70E-06
pHDGLRL	one left end (305bp), one right end (307bp) followed by one left end (305bp)	770,251	0	0

Table 2.2 – Transposition frequencies observed in *ex vivo* inter-plasmid transposition assays using *Hermes* donors with asymmetric left and right ends.

Donor plasmid	*Hermes* ends	Donors screened	True events	Frequency per donor
pHDG-gtggc-EGFP	medium (R=387bp, L=444bp)	2,525,055	28	1.11E-05
pHDG1	medium (R=387bp, L=444bp)	2,931,137	29	9.89E-06
pHDG-gtggc	medium (R=387bp, L=444bp)	1,046,067	12	1.15E-05
pHDG1-EGFP	medium (R=387bp, L=444bp)	1,737,917	26	1.50E-05

Table 2.3 – Transposition frequencies observed in *ex vivo* inter-plasmid transposition assays using *Hermes* donors with the canonical first right-end subterminal repeat, 5'-**GTGGC**-3', versus the natural polymorphism, 5'-**GTGAC**-3'.

Colony ID	Hermes donor	Right end motif	Right end TSD	Left end TSD	pGDV1 integration site	Comments
36	pHDG1	gtgac	ACTCATAA	NA	200	no left end sequence
35	pHDG1	gtgac	ACTCATAA	ACTCATAA	*312*	perfect integration
8	pHDG1	gtgac	GTCGTAAT	GTCGTAAT	624	perfect integration
33	pHDG1	gtgac	GTTCAGAC	GTTCAGAC	*744*	perfect integration
27	pHDG1	gtgac	GTGTAAAT	GTGTAAAT	809	perfect integration
152	pHDG1	gtgac	ATAGCAAC	ATAGCAAC	1904	perfect integration
29	pHDG1	gtgac	GTGCATAC	GTGCATAC	*2154*	perfect integration
142	pHDG1	gtgac	GTGCATAC	GTGCATAC	2154	perfect integration
88	pHDG1	gtgac	GTGCATAC	GTGCATAC	*2154*	perfect integration
107	pHDG1	gtgac	GTCGGAAC	GTCGGAAC	*2303*	perfect integration
146	pHDG1	gtgac	ATTCAGAG	ATTCAGAG	*2358*	perfect integration
106	pHDG1	gtgac	AACGACAT	AACGACAT	2549	perfect integration
32	pHDG-EGFP	gtgac	GTACAGAG	GTACAGAG	318	perfect integration
31	pHDG-EGFP	gtgac	GTCACGAA	GTCACGAA	*668*	perfect integration
50	pHDG-EGFP	gtgac	ATAGCAAC	ATAGCAAC	1904	perfect integration
53	pHDG-EGFP	gtgac	ATAGCAAC	ATAGCAAC	1904	perfect integration
30	pHDG-EGFP	gtgac	GTATGGGA	GTATGGGA	2070	perfect integration
51	pHDG-EGFP	gtgac	GTGCATAC	GTGCATAC	*2151*	perfect integration
4	pHDG-EGFP	gtgac	GTGCATAC	GTGCATAC	*2154*	perfect integration
41	pHDG-EGFP	gtgac	ATAGAGGT	ATAGAGGT	*2156*	perfect integration
9	pHDG-EGFP	gtgac	ATGCATTT	ATGCATTT	2171	perfect integration
3	pHDG-EGFP	gtgac	GTTCCGAC	GTTCCGAC	2303	perfect integration
22	PHDG-gtggc	**gtggc**	GTCACGAA	GTCACGAA	*668*	perfect integration
17	PHDG-gtggc	**gtggc**	GTTCAGAC	GTTCAGAC	*744*	perfect integration
15	PHDG-gtggc	**gtggc**	ATAGCAAC	ATAGCAAC	1904	perfect integration
1	PHDG-gtggc	**gtggc**	CTCTAGAG	CTCTAGAG	1993	perfect integration
2	PHDG-gtggc	**gtggc**	GTGCATAC	NA	*2154*	no left end sequence
16	PHDG-gtggc	**gtggc**	GTATGCAC	GTATGCAC	2154	perfect integration
24	PHDG-gtggc	**gtggc**	GTGCATAC	GTGCATAC	*2154*	perfect integration
25	PHDG-gtggc	**gtggc**	GTATGCAC	GTATGCAC	2154	perfect integration
18	PHDG-gtggc	**gtggc**	GTTGGTAC	GTTGGTAC	*2271*	perfect integration
20	pHDG-gtggc-EGFP	**gtggc**	ATATACAT	ATATACAT	56	perfect integration
117	pHDG-gtggc-EGFP	**gtggc**	ATTTGAAC	ATTTGAAC	276	perfect integration
116	pHDG-gtggc-EGFP	**gtggc**	ATTACGAC	ATTACGAC	*631*	perfect integration
132	pHDG-gtggc-EGFP	**gtggc**	GTCTGAAC	GTCTGAAC	736	perfect integration
21	pHDG-gtggc-EGFP	**gtggc**	ATAGCAAC	ATAGCAAC	1904	perfect integration
62	pHDG-gtggc-EGFP	**gtggc**	ATCCCCGG	ATCCCCGG	*2002*	perfect integration
63	pHDG-gtggc-EGFP	**gtggc**	CTCGGTCA	CTCGGTCA	*2010*	perfect integration
7	pHDG-gtggc-EGFP	**gtggc**	GTATGCAC	GTATGCAC	2154	perfect integration
26	pHDG-gtggc-EGFP	**gtggc**	GTATGCAC	GTATGCAC	2154	perfect integration
131	pHDG-gtggc-EGFP	**gtggc**	GTGCATAC	GTGCATAC	*2154*	perfect integration
115	pHDG-gtggc-EGFP	**gtggc**	ATGTGTTA	ATGTGTTA	2211	perfect integration
93	pHDG-gtggc-EGFP	**gtggc**	GTTCCGAC	GTTCCGAC	2303	perfect integration
119	pHDG-gtggc-EGFP	**gtggc**	ATACTTAC	NA	*2309*	no left end sequence

Table 2.4 – Target plasmid integration location and target site duplication sequences observed in *ex vivo* inter-plasmid transposition assays using *Hermes* donors with the canonical first right-end subterminal repeat, 5'-**GTGGC**-3', versus the natural polymorphism, 5'-**GTG**A**C**-3'. The pGDV1 integration location is italicized to denote right end integration into the negative strand.

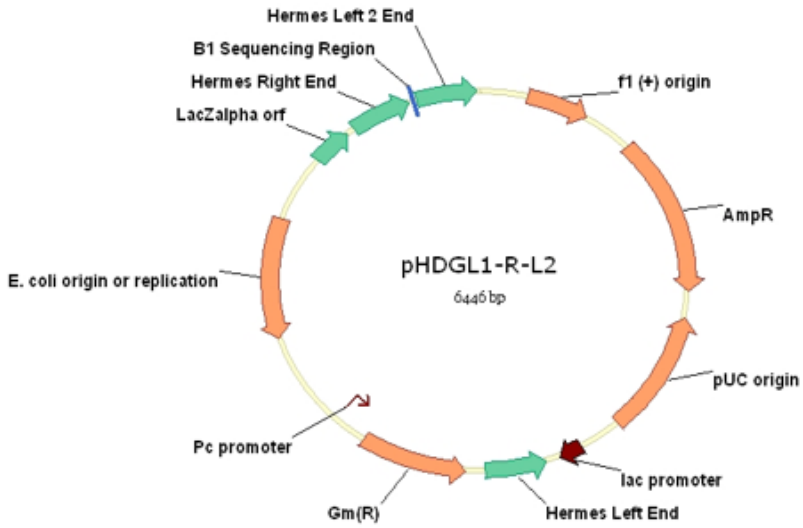

Figure 2.1 – The *Hermes* donor pHDGLRL containing one left end and one right end immediately proceeded by another left end.

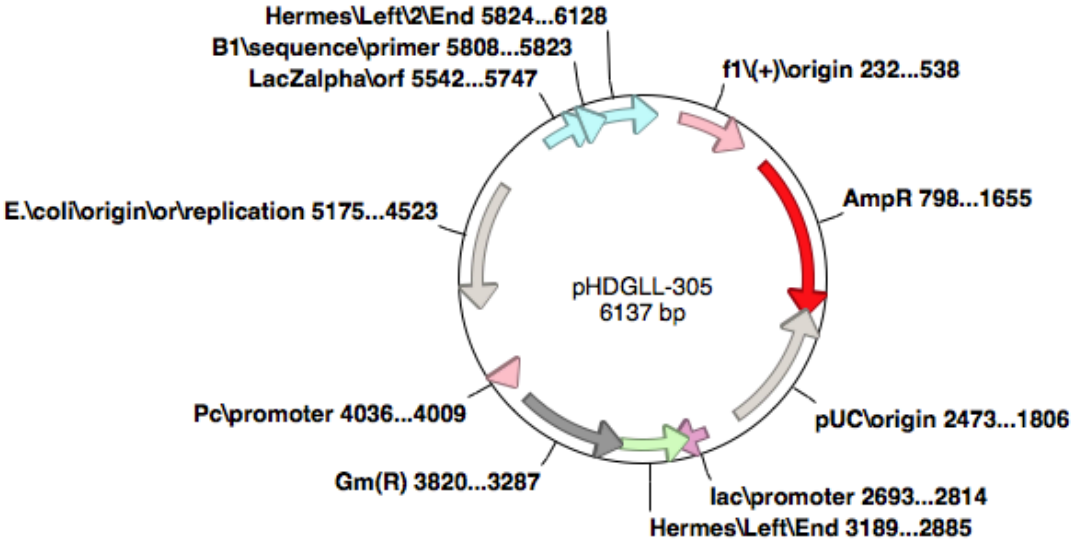

Figure 2.2 – The *Hermes* donor pHDGLL containing two left end sequences.

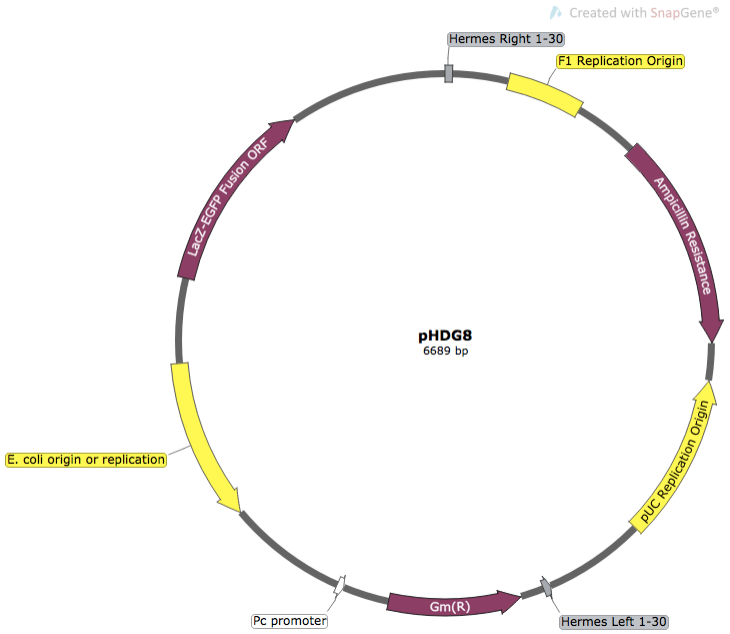

Figure 2.3 – The *Hermes* donor with 30-mer transposon ends.

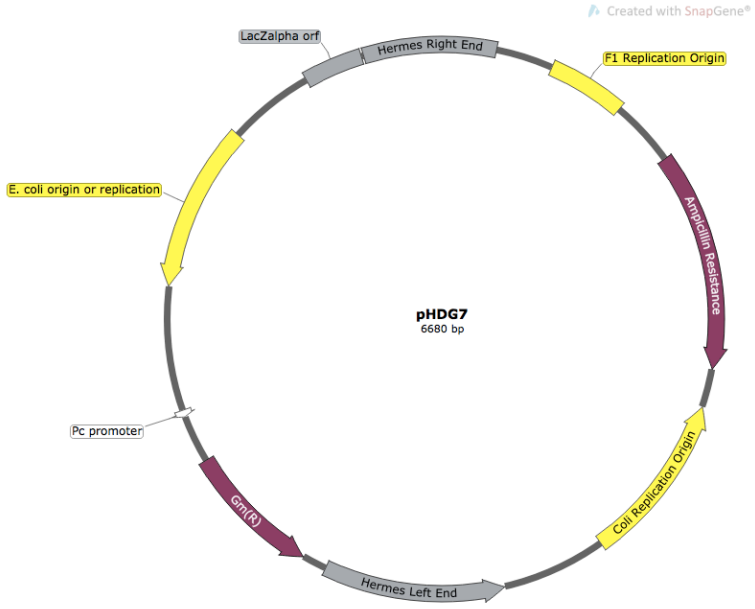

Figure 2.4 – The *Hermes* donor with long transposon ends.

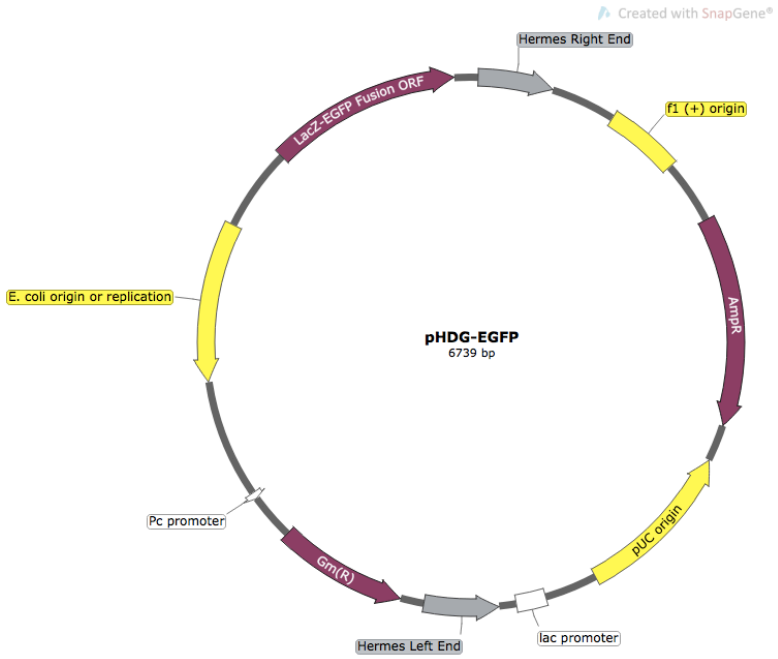

Figure 2.5 – The *Hermes* donor with intermediate length transposon ends.

Figure 2.6 – Target plasmid for interplasmid transposition assays.

Chapter 3 – The *Muta1* transposable element is active *ex vivo* in cell culture and *in vivo* in *Drosophila melanogaster*

3.1 Introduction

3.1.1 The *MULE* superfamily of class 2 transposable elements

The *MULE* (*Mutator*-like elements) superfamily of class 2 transposons is widespread throughout eukaryotic genomes, particularly in plants, in addition to fungi, insects, and even diatoms (Feschotte and Pritham, 2007). The presence of *MULE* elements in vertebrate genomes is currently under examination, but their presence has not been substantiated as of yet (Dupeyron *et al.,* 2019). Bioinformatic and phylogenetic analysis indicate that *MULE* elements are most closely related to the *P* and *hAT* element superfamilies. (Yuan and Wessler, 2011). The *MULE* superfamily of transposons, like most eukaryotic superfamilies, contain the catalytic DDE/D acidic amino acid motif that forms the transposase catalytic pocket (Yuan and Wessler, 2011), and also possess one or more zinc finger binding domains. *MULEs* also contain the C/DxxH and RW amino acid motifs that are present in *hAT*, *Kolobok*, and *P* elements (Yuan and Wessler, 2011), and these conserved motifs are believed to help stabilize cleaved DNA within the active site (Hickman *et al.,* 2018). One of the defining features of *MULEs*, although absent from a small subset, are the atypically long and repetitive terminal inverted repeats, which can vary from 14 – 500 base pairs, and longer target site duplications of 8 to 9 base pairs (Dupeyron *et al.,* 2019, Lisch, 2002). There are cases of *MULE* elements containing additional ORFs, some that may increase the efficiency of one or more of the steps of

transposition and some that may encode anti-silencing factors (Lisch 2002, Fu *et al.,* 2013). It has been suggested that these additional ORFs are likely obtained from the host genome, as *MULE* elements exhibit a propensity for picking up gene fragments, which are called Pack-MULEs and described in greater detail below (Feschotte and Pritham, 2007). *MULEs* are also typically known to have many non-autonomous copies, called MITES, which have lost transposase coding capacity but still have transposon ends and terminal inverted repeats that can be recognized and mobilized by transposases expressed by autonomous elements. While several active *Mutator* elements have been discovered (described below), biochemical experiments to assess the *MULE* mechanisms of transposition have been somewhat limited. This is due in part to the fact that *MULE* transposases have not been amenable to cloning and purification from *E. coli* systems (reviewed by Lisch, 2015), thus, biochemical assays await a better method for expressing and purifying active *MULE* transposases.

3.1.2 *MULE* elements in plants

The first *MULE* element was initially called *Mutator Don Robertson*, or *MuDR*, due to obvious phenotypic mutagenic effects, and was discovered in maize (Donald Robertson, 1978). *Mutator*-like elements have since been most widely studied in maize and other plant genomes. Because active *MULE* elements are highly mutagenic and have a measurable propensity for insertion in or near genes (Cresse *et al.,* 1995), *MULEs* have been employed as a tool in maize for forward genetic screens for many years (McCarty and Meeley, 2009). Indeed, a study published in 2009 found that the maize element *Mu*,

and its non-autonomous derivatives had the greatest insertion site preferences for both the 5' regions of genes and euchromatic regions containing open epigenetic chromatin marks (Liu *et al.,* 2009). This is a feature that makes *MULEs* a good candidate for forward genetics tools like promoter and gene trapping experiments.

In addition to the active *MuDR* element, several other active *MULEs* have been identified in maize, including *TED* and *Jittery*, although *Jittery* has a unique observed behavior of excision without integration (Li *et al.,* 2013, Xu *et al.,* 2004). Active *MULEs* have also been reported in rice, *Arabidopsis thaliana*, and the fungus *Fusarium oxysporum* (Zhao *et al.,* 2015, Chalvet *et al.,* 2003, Singer *et al.,* 2001, Fu *et al.,* 2013). However, there have only been two examples of an active *MULE* element that has been shown to transform and remobilize in a heterologous host. In one instance, the rice *Mutator*-like element, *Os3378*, was shown to be able to catalyze excision and reinsertion in the *Saccharomyces cerevisiae* genome (Zhao *et al.,* 2015). The other example is the recently described *MULE* element from an arthropod, the yellow fever mosquito, called *Muta1,* which will be discussed in greater detail in this chapter (Liu and Wessler, 2017).

The *MULE* superfamily is most well studied within its plant hosts and is known particularly for the observation of Pack-MULEs, non-autonomous MULE elements that have lost their transposases coding sequences and have picked up one or more fragments of host genes. The *Mu* transposons in maize have been shown to have their own promoter sequences within each of the terminal inverted repeats (Raizada *et al.,* 2001). It has also been observed that *MULEs* have a preference for inserting in or near genes, particularly 5' to genes (Jiang *et al.,* 2011). These two features of *MULE* elements and

the large-scale sequence analysis of certain plant genomes have led researchers to conclude that Pack-MULEs provide novel material for genetic evolution and provide a rationale for the retained heavy Pack-MULE load in some plant genomes (Ferguson and Jiang, 2011).

3.1.3 The *Muta1* element from *Aedes aegypti*

The *Muta1* element from *Ae. aegypti* is a newly described active *MULE* element. *MULE* elements have not been well studied outside of plants and the identification of *Muta1* marked the first time an active *MULE* has been observed in invertebrates (Liu and Wessler, 2017). According to recent phylogenetic analysis of available *MULE* sequences, the *Ae. aegypti Muta1* element belongs to a subfamily of *MULEs*, called *Phantom*, which occur in primarily in arthropods with a few observations in tunicate, helminth, and a copepod genome (Dupeyron *et al.,* 2019). There is a very close phylogenetic relationship between different *MULE* elements of the plant kingdom; outside of plants, the broad range of host species and rather disjointed phylogenies strongly suggest that *MULE* elements are amenable to horizontal transfer and have seen multiple horizontal transfer events over evolutionary time (Dupeyron *et al.,* 2019). Why *MULE* elements have seen many successful horizonal transfer events, particularly into arthropods and stramenopiles is an open question (Dupeyron *et al.,* 2019), but recent studies have identified *MULE Phantom* subfamily elements in two different insect viruses, suggesting a unique means of horizontal transfer into arthropod species (Marquez and Pritham, 2010). A recent study has identified two new *MULE* clades in arthropods

(Dupeyron *et al.,* 2019). The *Ghost* clade has been identified in various aphid and arachnid species, as well as in several cnidarian genomes, but none of the *Ghost* elements are complete, full length copies. The *Spectre* clade was identified only in arthropods, including aphid, arachnid, beetle, and true bug species, including the major crop pest, *Bemesia tabaci*, and some of these elements are full-length copies (Dupeyron *et al.,* 2019).

The *Aedes aegypti* mosquito, the host of the active *Muta1* transposon, is a unique pest with unique genome features, outlined in this section. The *Ae. aegypti* mosquito is a major global arboviral vector, and as such, is an intensely studied vector species (Bhatt *et al.,* 2013). *Ae. aegypti* contains a large genome of 1.38 Gb in size with a transposon load of about 47% of the genome (Nene *et al.,* 2007). For perspective, the model dipteran *Drosophila melanogaster* has a 144 Mb sized genome with only a 16% transposon genome load, and the mean dipteran genome size is about 700 Mb (Arensburger *et al.,* 2011, Hanrahan and Johnston, 2011). The genome size ranges observed in arthropods see great variation, although not to the same extremes observed in plants (Leitch *et al.,* 2007). The smallest arthropod genome observed belongs to the two-spotted spider mite and is 91 Mb, while the largest genome observed belongs to the broad-tipped conehead, at a size of 7.7 Gb (in females), more than twice the size of the human genome (Lander *et al.,* 2001, Hanrahan and Johnston, 2011).

In the first study of *Muta1* by Liu and Wessler (2017), a total of 14 *MUTA* families were identified in the *Ae. aegypti* genome, with the *Muta1* element suspected of having catalytic activity. This was due to the fact that the genome contains eight full

length *Muta1* copies with intact terminal inverted repeats and coding sequence, as well as hundreds of copies of non-autonomous *Muta1* MITEs. It was also observed that a subset of non-autonomous *Ae. aegypti MULEs* contained captured host gene fragments, just as observed in plant Pack-*MULEs* (Liu and Wessler, 2017). This initial study presented promising data: *Muta1* constructs were capable of efficiently catalyzing excision and insertion in a heterologous host (yeast), without an obvious target site preference and a non-random preference for inserting in or near genes, as has been reported for other active plant *MULE* elements (Liu and Wessler, 2017). Additionally, the *Muta1* element showed ~90% precise excision rate in yeast assays (Liu and Wessler, 2010). Following transposon excision, the DNA double-strand break is typically repaired by host mechanisms, like non-homologous end-joining (NHEJ), and "foot prints" are left at the excision site (Zhou *et al.,* 2004). Several elements, however, show a majority of precise excisions in which one of the two target site duplications (TSD) is excised along with the element. An active *MULE* element from rice, *Os3378,* and the *piggyBac* transposon demonstrate precise excision, a feature that is preferred in gene therapy vectors (Zhao *et al.,* 2015, Fraser *et al.,* 1996, Wilson *et al.,* 2007). These features of the *Muta1* element make it an attractive candidate for promoter and gene trap experiments, which will be discussed in greater detail in Chapter 4.

3.1.4 Chapter aims

The goal of the research described in this chapter was to create and test the transformation ability and mobility of a *Muta1* construct for potential use in insect systems. The study by Liu and Wessler (2017) suggests that *Muta1* may have the capacity to be used as a robust forward genetics tool, perhaps as efficacious as the *MuDR* system in maize. I first assessed the ability of a *Muta1*-based construct to remobilize in cell culture using both *Drosophila* S2 cell culture, and in HeLa cell culture. I examined potential patterns of target site preference in HeLa cell culture. I tested the ability of *Muta1* to transform the germline *Drosophila melanogaster.* I made transgenic enhancer trap lines in *D. melanogaster* to examine the remobilization potential of a *Muta1* enhancer trap system for potential expansion into other insect species.

3.2 Materials and methods

3.2.1 Plasmid construction

Construction of helper pKH70Muta1 (Figure 3.4): The *Muta1* open reading frame was amplified from *Aedes aegypti* Liverpool strain cDNA using the primers below. The PCR fragment was digested by *SpeI* and *XhoI* and ligated into plasmid pKhsp70new (Arensburger *et al.,* 2005).

Muta1 ORF For:

5'- GCCACTAGTATGGACTCGGACAGCGATAGC-3'

Muta1 ORF Rev:

5'- GTACTCGAGTCTTATTTTGATTTTGATCCTAAGTGAGCTG-3'

Construction of pMutaENT3 (Figure 3.1): An *ex vivo* transposition event (pGDV1event33) was recovered from a *Muta1* donor plasmid (unpublished) into *Bacillus* plasmid pGDV1 (described in chapter 2, Firgure 2.) and was digested with *PvuII* and self-ligated to recover plasmid containing the *Muta1* left end and chloramphenicol resistance coding sequence. The resulting plasmid was transformed into electrocompetent cells and miniprepped with 5PRIME Fast Plasmid Mini-Prep kit. This plasmid was then cut with *SphI* and *NsiI* and a linear 5,389 base pair fragment was gel purified using Qiagen QIAquick Gel Purification Kit (Cat. No. 28704), and the purified linear product was blunted and ligated using Thermo Fisher Scientific CloneJET PCR Cloning Kit (Cat. No. K1231) blunting and ligation reaction components, following manufacturer protocol. The resulting plasmid was transformed into DBH10 electrocompetent cells and miniprepped with 5PRIME Fast Plasmid Mini-Prep kit. The

resulting plasmid was cut with *XhoI* and BamHI and purified. A fragment from pGoE vector was cut by *NheI* and treated with Thermo Fisher Scientific FastAP (Cat. No. EF0652). An attP-loxP DNA fragment was synthesized by IDT and cloned into vector pUC57. The following primers were using to amplify and purify an att-P PCR fragment.

Att-P For:

5' – GTATCTTATACTGACGGACACAC -3'

Att-P Rev:

5' – GCTGGCAGTCAGCGCGCTCGCGCTT – 3'

The three fragments were assembled as pVectorStep3 using the Thermo Fisher Scientific GeneArt Seamless Cloning and Assembly kit following the kit protocol (Cat. No. A13288).

The following primers were used to create PCR fragments from plasmid pMos[3xP3-DsRed] (Smith *et al.,* 2007).

SV40-polyA For:

5' – TAAGATACATTGATGAGTTTGGACAAACC – 3'

SV40-polyA Rev:

5' – CTAGATCTCAGCGCCGGCGAT – 3'

Pax3dsRed For:

5' – CAGGAACAGGTGGTGGCG – 3'

Pax3dsRed Rev:

5' – CAGAGATTAACTTAATCTAGGG – 3'

The plasmid pVectorStep3 was digested with *FspI* and *PvuII* and purified. This was assembled with the three above PCR fragments using the Thermo Fisher Scientific GeneArt Seamless Cloning and Assembly kit (Cat. No. A13288).

Construction of pBac3ChspMuta1 (Figure 3.2): The plasmid pBac[3xP3afm] (Kokoza *et al.,* 2001) was digested with *AvrII* and treated with Thermo Scientific FastAP (Cat. No. EF0652). A fragment containing hsp70-Muta1 transposase was amplified from the plasmid pKH70Muta1 (Figure 3.4) using the following primers.

hsp70 Nhe F:

5'-AATGCTAGCGAATTCGTGCTCTCGTTGGTTC-3'

Hsp 3' Poly-A AvrII Rev:

5'-CCTCCTAGGAAGCTTGGATCTAAACGAGTTTTTAAGC-3'

The resulting PCR fragment was digested with *NheI* and ligated to the *AvrII* fragment from pBac[3xP3afm]. The resulting plasmid, pBacEHspMuta, was digested with NotI and *NcoI* and an 8,484 base pair vector fragment was gel purified following Qiagen QIAquick Gel Purification kit protocol. The plasmid pMos[3PX3-ECFP] (Smith *et al.,* 2011) was digested with *NotI* and *NcoI* and a 716 base pair fragment containing the ECFP coding sequence was gel purified using Qiagen QIAquick Gel Purification Kit (Cat. No. 28704). The two *NcoI/NotI* fragments were ligated together using Thermo Scientific T4 DNA Ligase (Cat. No. EL0014).

Construction of pCMV-Muta1: The plasmid pCMV-HSB16 (Woodard *et al.,* 2012) was digested with *KpnI* and *XhoI*. A PCR fragment for the *Muta1* coding sequence was amplified using the below primers and digested with *KpnI* and *XhoI*.

Muta1 Orf F:

5’-GATGGTACCACTAGTATGGACTCGGACAGCGATAG-3’

Muta1 Orf R:

5’-GATCTCGAGTTATTTTGATTTTGATCCTAAGTGAGCTGC-3’

The two fragments were ligated using Thermo Scientific T4 DNA Ligase (Cat. No. EL0014) following manufacturer protocol.

3.2.2 Interplasmid transposition assays

S2 *D. melanogaster* cell lines were transfected with plasmids pKH70Muta1 and pMutaENT3 using the following transfection protocol. S2 cells were seeded in treated 6-well plates at a density of $2x10^6$ cells per well. Cells were transfected the next day following the X-tremeGENE HP DNA Transfection Reagent protocol (Roche), using the donor plasmid, pMutaENT3 (1 μg), helper plasmid pKH70Muta1 (1 ug) and the pGDV1 (Bron *et al.,* 1991) target plasmid (2 μg).

Plasmid Actin5C-EGFP was used as a positive control for transfection.

The S2 cell plates were heat shocked at 37°C for 2 hours one day following transfection. Cells were harvested two days after transfection, washed with PBS, and frozen for plasmid preparations. Cell pellets were resuspended with 500 μl of grinding buffer (0.5% SDS, 0.08 M NaCl), 54.7 mg/ml sucrose, 0.06 M EDTA, 120 mM Tris pH9.0) before being incubated at 65°C for 30 min. Potassium acetate was added to a final concentration of 1 M prior to incubating on ice for 30 min. Samples were centrifuged, and the DNA in the supernatant was precipitated with ethanol. Reactions were

resuspended in nuclease free water and electroporated into bacteria and plated. Colonies were selected by dual antibiotic resistance. Selected colonies mini-prepped using 5PRIME Fast Plasmid Mini-Prep kit and assayed by restriction enzyme digest. Colonies that passed restriction enzyme digestion were sequenced using Sanger sequencing (Sanger *et al.,* 1977).

HeLa cell transformations:

HeLa cells were maintained in sterile filtered DMEM with 5% FBS, and 1x antibiotic-antimycotic (Streptomycin, Amphotericin B, Penicillin) grown in 37°C incubator with 5% CO_2.

For interplasmid transpositions assays, HeLa cells were trypsinized and seeded at $1.5x10^5$ cells/mL in Falcon tissue culture treated 6-well plates (#353046) to obtain 50-80% confluent cells the following day.

Transfections were performed the following day using FuGENE6 Transfection Reagent (Promega) at a 3:1 reagent to DNA ratio. A control well was transfected using 1ug plasmid pCMV-EGFP (Robert Hice) for assessing transfection reagent efficiency. Experimental wells were transfected with 500ng of target *Bacillus subtilis* plasmid, pGDV1 (Sarkar *et al.* 1997), 250ng of *Muta1* helper plasmid, pCMV-Muta1 (Robert Hice), and 250ng *Muta1* donor plasmid, pMutaETS (Anna-Louise Doss).

Two days following the transfection, cells from the transfection control well were trypsinized and counted on a hemocytometer under a fluorescent microscope to determine the percentage of EGFP positive cells.

The cells from experimental wells were collected for DNA isolation using a Wizard Genomic DNA Purification kit (Promega). Purified HeLa cell DNA was electroporated into DBH10 competent *E. coli* cells. Transposition events were recovered by selecting for *E. coli* colonies with dual gentamicin and chloramphenicol resistance. Resistant colonies were picked and grown overnight for plasmid preparations using a FastPlasmid Mini Kit (5 PRIME).

Putative transposition events were first characterized by *BamHI* plasmid digest (Fast Digest enzymes, ThermoFisher). Plasmids with accurate digest patterns were sequenced using the following primers to obtain integration location in target plasmid and to confirm target site integration sequences:

Mule LE RP:

5' - GATCTCGAGCATTTCCAGCTTCGTAGTACAAATATC – 3'

Mule RE FP:

5' – GCGTCTAGAGATTATCTTGAGGCAATTGCAG – 3'

The online tool WebLogo at weblogo.berkeley.edu was used to create web logos for *Muta1* 9 base pair target site integrations (Crooks *et al.,* 2004).

3.2.3 *Drosophila melanogaster* transformation

The Canton S White strain of *D. melanogaster* was used for the work described in this chapter.

For germline transformation, embryo injections using pre-blastoderm embryos were performed as described in Sarkar *et al.,* 1997. A four-plasmid injection mix was

prepared as described (Sarkar *et al.,* 1997) using 500ng/ul of pMutaENT3 (Figure 3.1), 250ng/ul pBac3ChspMuta1 (Figure 3.2), and pKH70Muta1 (Figure 3.4). Surviving G_0 injected flies were setup in fertile crosses and the G_1 progeny were screened as newly emerged adults and scored for the presence of eye fluorescence. Individually isolated transgenic lines were backcrossed and reared under standard conditions.

3.2.4 *D. melanogaster* enhancer trap crosses

Crosses were established using enhancer trap line 1126N (pMutaENT3) and helper line 1109A (pBac3ChspMuta1). Twenty males from enhancer trap line were crossed with twenty virgin females from the helper line. The reciprocal crosses were also established. Progeny with all three eye markers are selected and self-crossed after eclosion. Twenty virgin females and twenty males are self-crossed every generation. Self-crossed adults are heat-shocked daily for one hour at 37ºC and progeny are screened every generation for new phenotypes.

3.2.5 Molecular verification of transgenic *D. melanogaster* lines

The genomic DNA of individual flies was prepared using a Qiagen Insect Protocol with the Qiagen DNeasy Blood and Tissue Kit (Cat. No. 69506).

Transgenic flies from the enhancer trap line (plasmid pMutaENT3) were characterized by plasmid rescue. Genomic DNA was digested with *XbaI*, *NheI*, and *AvrII* and digest was column purified (Qiaquick PCR Purification Kit). Purified products were

self-ligated with T4 DNA Ligase (Thermo Fisher Scientific) and chloroform extracted following Thermo Fisher kit protocol. Self-ligated DNA was electroporated into DHB10 *E. coli*. Gentamicin resistant colonies were selected, and plasmids were purified using FastPlasmid Mini Kit (5 PRIME). Plasmids were sequenced with the primers below.

Mule LE RP - To obtain *Muta1* left end integration:

5'-GATCTCGAGCATTTCCAGCTTCGTAGTACAAATATC-3'

Mule RE FP - To obtain *Muta1* right end integration:

5'- GCGTCTAGAGATTATCTTGAGGCAATTGCAG-3'

Transgenic flies from the helper line (plasmid pBac3ChspMuta1) were characterized by inverse PCR (Martin and Mohn, 1999). Genomic DNA was purified as above and digested with *BamHI* and *BglII,* self-ligated wth T4 DNA Ligase and chloroform extracted. This was used as template for two rounds of nested PCR in AccuPower PCR Premix tubes with the following primers.

PB RE RP1:

5'-CAACATGACTGTTTTTAAAGTACAAA-3'

PB RE FP1:

5'-GTCAGAAACAACTTTGGCACATATC-3'

PB RE RP2:

5'-CCTCGATATACAGACCGATAAAAC-3'

PB RE FP2:

5'-TGCATTTGCCTTTCGCCTTAT-3'

PCR products were purified as above and cloned into pJET1.2/BLUNT (CloneJET PCR Cloning kit, ThermoFisher). Ampicillin resistant clones were selected and colony PCR was performed using the pJET1.2 colony PCR primers provided in the kit. PCR products were purified as above and sequenced with the forward pJET1.2 kit primer to obtain the *piggyBac* left genomic integration location. The *piggyBac* right end sequences and TSDs were confirmed using genomic DNA as template for PCR with primers PB RE FP1 (see above), and the following:

Slob to RE For:

5' – GTAACCTCCGTCAATAAAGCG – 3'

ch.2R RE:

5' – GCTTAAACTTCCTGATACCCTG – 3'

ch.X RE

5' – CGTGTTCCATAATCATGTTCGCC – 3'

AccuPower PCR Premix tubes (Bioneer) were used for PCR amplification following Bioneer protocol. PCR products were purified as above and sequenced with primer PB RE FP1 (see above).

The transgenic flies of the helper line (pBac3ChspMuta1) were evaluated for the levels of *Muta1* transposase expression using qRT-PCR. Three biological replicates were used for each sample, and each biological replicate included three females, 2 to 5 days old. RNA from each biological replicate was extracted using ThermoFisher TRIzol reagent following manufacturer protocol for RNA isolation from 50-100mg tissue

samples. The RNA samples were treated with the ThermoFisher TURBO DNA-free Kit according to manufacturer protocol, with a longer incubation period of 1 hour. All RNA samples were diluted to 50ng/uL stocks before cDNA synthesis using the New England Biolabs ProtoscriptII First Strand cDNA Synthesis Kit, using half-reactions.

All cDNA for each biological replicate was diluted to 100ng/ul stocks for subsequent qRT-PCR reactions.

Quantitative RT-PCR was performed using Bio-Rad iQ SYBR Green Supermix following protocol for half reactions. Thermocycling reactions were performed on a Bio-Rad MyiQ Detection system.

For the Actin housekeeping gene, a standard curve was generated using 300nM primer concentrations and an annealing temperature of 57ºC. The standard curve for this primer set had an E value of 97.4 and an R^2 value of 0.993.

Actin DM For:

5' – CGCTCGGTCAATTCAATCTT – 3'

Actin DM Rev:

5' – AAGCTGCAACCTCTTCGTCA – 3'

For the *Muta1* transgene, a standard curve was generated using 300nM primer concentrations and an annealing temperature of 57ºC. The standard curve for this primer set had an E value of 110.6 and an R^2 value of 0.998.

Muta1 2015 For:

5' – GCGTATGGTAACGTTCAAGGC – 3'

Muta1 2015 Rev:

5’ – GTACTATTTTCGCTGGCGTTG – 3’

All quantitative RT-PCR data was analyzed using the Pfaffl equation (Pfaffl, 2001).

3.3 Results

3.3.1 The *Muta1* element remobilizes in S2 and HeLa cell culture without strict target site preference

The plasmid pMutaETS (Figure 3.3) was used for S2 and HeLa cell transformations for interplasmid transposition assays. Several remobilization events were recovered in S2 cell culture (data not shown, small sample size) and 24 remobilization events were recovered from HeLa cell culture (Table 3.1). In HeLa cell culture the *Muta1* element showed a higher remobilization frequency than the *piggyBac* positive control (Table 3.1). Target site duplications were recovered from all 24 HeLa cell remobilization events. An additional 17 HeLa cell remobilization events were also recovered by Robert Hice, the Senior Research Associate in the laboratory. All 41 of the 9-base-pair target site duplications were analyzed together using WebLogo software to identify a consensus target site preference sequence. The consensus sequence generated showed only a slight preference for adenine at bases 6-8 of the target site duplication (Figure 3.5). The donor plasmid, pMutaETS (Figure 3.3), in addition to all *Muta1* donors described in this thesis, are flanked by 9-base-pair target site duplications.

3.3.2 The *Muta1* element transforms *D. melanogaster*

The plasmid pMutaENT3 (Figure 3.1) was used for germline transformation in *D. melanogaster* Canton S White. I was able to obtain 3 *Muta1* transgenic lines at a transformation frequency of 4.9% (Table 3.2, Figure 3.6). A total of 9 *piggyBac* helper lines were generated at a frequency of 14.8% (Table 3.2, Figure 3.6). The integration

locations of several transformed lines were analyzed (Table 3.4). The *D. melanogaster* helper line (pBac3ChspMuta1) was evaluated for *Muta1* transposase expression using qRT-PCR. The results show that the HSP70 promoter driving *Muta1* transposase expression is leaky, as all non-heat-shocked samples showed detectable *Muta1* expression. However, heat-shock treatment increased *Muta1* expression levels up to 90-fold higher in the transgenic strain (Table 3.3).

3.3.3 The *Muta1* element remobilizes at a low frequency in *D. melanogaster*

Transgenic fly crosses were screened every generation for new fluorescent phenotypes that could indicate a *Muta1* remobilization event. One transgenic cross, 1B, showed an unexpected phenotype of the loss of the EGFP eye marker in a subset of flies, indicating a germline remobilization event in a parent (Figure 3.7). Subsequent molecular analysis showed that the *Muta1* construct had indeed remobilized (Table 3.5). The parental integration was not recovered in this line. The new integration location was 272 bases upstream of the parental integration, a phenomenon called local hopping (Guimond *et al.,* 2003). The loss of EGFP in the transgenic cross, 1B, was due to imperfect integration, as the entire left end of the construct was missing. All crosses were evaluated for the potential presence of jumps that lacked new phenotypes and five other remobilization events were recovered, though there is the potential that some or all of these jumps represent somatic remobilization (Table 3.5). The loss of EGFP phenotype was observed in F_{15} of line 1B. This gives a remobilization frequency of 1 remobilization event / (40 adults*15 generations) = 0.167%. This remobilization

frequency is too low to be an effective enhancer trap system in the case of *D. melanogaster,* but this species is not lacking in forward genetics tools.

3.4 Discussion

In interplasmid transposition assays in HeLa cell culture, the *Muta1* element had approximately a 2.5-fold greater integration frequency compared to the *piggyBac* element (Table 3.1). Currently there is no literature reporting the remobilization of a *MULE* element in human cell culture, and as such, this appears to be the first observation recorded. DNA transposons that show activity and stability in human cells, like the *Sleeping Beauty* transposon, have been investigated as alternatives to viral-derived vectors for gene therapy (Mátés *et al.,* 2009, Garrels *et al.,* 2011). *Muta1* has also been reported to remobilize in *Saccharomyces cerevisiae* (Liu and Wessler, 2017), indicating that *Muta1* has a wide host range for activity.

Previous experiments performed in our laboratory have analyzed *Muta1* element excision. In *D. melanogaster* embryo interplasmid transposition assays, the *Muta1* element was capable of perfect excision without leaving a footprint in the majority of excision events recovered (Shah, 2015). This precise excision has also been observed in *Saccharomyces cerevisiae* experiments (Liu and Wessler, 2017). The *MULE* element from rice, *Os3378,* and the *piggyBac* element are the only other class 2 transposons that have shown precise excision events in which the second TSD is removed from the excision site (Zhao *et al.,* 2015, Fraser *et al.,* 1996). The relatively high transposition frequency of *Muta1* in HeLa cell culture (Table 3.1), coupled with the observed ability to

excise without leaving a footprint within the excision site, makes *Muta1* a potential candidate for a human gene therapy vector. The resurrected *Sleeping Beauty* element, which is currently being used in gene therapy clinical trials, has been shown to leave various length footprints within excision sites, depending on the type of human cells used (Kebriaei *et al.,* 2016, Liu *et al.,* 2004).

A total of twenty-four *Muta1* 9-base-pair target-site duplications were recovered and analyzed (Table 3.1). The target-site duplication consensus sequence showed only a slight preference for adenine at bases 6-8 of the duplication (Figure 3.5). Liu and Wessler (2017) analyzed 8- and 9-base target sites from both the *Ae. aegypti* genome and from *S. cerevisiae* transposition assays and found that the 8-base-pair target-site duplications do not occur as frequently as 9-base-pair target-site duplications and neither appear to have a stringent target site preference, although a slight preponderance of adenine is observed around bases 6-8 (Liu and Wessler, 2017). The length and presence of the target-site duplication (TSD) in the *Muta1* donor have an effect on transposition. In *D. melanogaster* S2 cell culture interplasmid transposition assays, a *Muta1* donor plasmid with a 9-base-pair TSD showed a greater transposition frequency and more precise transposition events than compared to a *Muta1* donor plasmid with an 8-base-pair TSD (Shah, 2015). In yeast cell transposition assays, a *Muta1* donor lacking flanking TSDs had a lower excision frequency and various footprint lengths compared to *Muta1* donors with 8 or 9-base-pair TSDs (Liu and Wessler, 2017). This demonstrates that the flanking TSDs of a *Muta1* donor can affect transposition. The *Muta1* donors used in the

experiments described in this thesis have the same 9-base-pair target site duplication (Figure 3.1, Figure 3.3).

The *Muta1* element was capable of transforming *Drosophila melanogaster*, a model genetic organism. The transformation frequency I obtained for *Muta1* was 4.9%, about 3-fold lower than *piggyBac* (Table 3.2). The *Muta1* element has previously been used to transform *D. melanogaster* in our laboratory and a varied germline transformation frequency of 4.9% - 14.8% was reported (Shah, 2015). *D. melanogaster* has also been transformed by other DNA transposons with varying transformation frequencies. The *piggyBac* element is reported to have a 3% transformation frequency, (Handler and Harrell, 2001), and the *Herves* element has a 30.1% transformation frequency (Arensburger *et al.,* 2005).

Production of *Muta1* transposase was assessed for the helper line, 1109A, which was used for remobilization experiments, as well as the crossed enhancer trap lines (Table 3.3, Table 3.4). *Muta1* transposase was expressed in both the parental line, 1109A, and the crossed line, even in the absence of heat shock protocol. However, the *Drosophila* HSP70 promoter has been observed to be "leaky" at about room temperature or 25º Celsius (D'Souza *et al.,* 1999). The parental line, 1109A, showed about a 90-fold increase of *Muta1* expression after experimental heat shock protocol. The crossed line sampled consisted of F_1 progeny that were heterozygous for each transgene insertion from pMutaENT3 and pBac3ChspMuta1. The F_1 flies showed about a 28-fold increase in *Muta1* transposase production after heat shock treatment (Table 3.3).

Previous to these experiments in *D. melanogaster*, *Muta1* remobilization had only been assessed *ex vivo* in yeast culture (Liu and Wessler, 2017) and in the HeLa cell experiments described above. The enhancer trap lines made in *D. melanogaster* were screened every generation for the novel fluorescent phenotypes as evidence of remobilization and insertion of the enhancer trap element. The *Muta1* element remobilizes in *D. melanogaster*, but at a low frequency. For example, the first novel fluorescent phenotype observed, the loss of the EGFP eye marker, occurred in G_{15} (Table 3.5, Figure 3.7). This phenotype was attributed to loss of the left end of the pMutaENT3 construct following local hopping of the parental integration. This was the only obvious novel fluorescent phenotype observed. One limitation, however, is the flies are screened only as newly eclosed adults. The experiments were not designed for screening larvae, pupae, or dissected material, which may lead to overlooking temporally driven fluorescent expression or fluorescent expression in internal tissues. Flies without obvious changes in phenotype were also sampled for possible jumps (Table 3.5), however none of the jumps have been ruled out as being somatic. The five additional pMutaENT3 remobilization events recovered, and as with the parental integrations, all insertions were in genic regions, either exonic or intronic. While this sample size is quite small, it is still in agreement with previous reports on *MULE* integration preferences for genic regions (Jiang *et al.,* 2011), which is a desirable feature for a potential new gene and enhancer trap tool.

The low frequency of *Muta1* germline remobilization (0.167%) observed in these experiments is at odds with the high germline remobilization frequencies reported for the

P element system (Bellen *et al.,* 1989). However, it is known that different transposons have different patterns of mobilization, some elements proliferate during meiosis and other elements appear to be more active in somatic cells (Kazazian 2011). Another limiting factor that is not explored in this thesis is the potential for overproduction inhibition. Overproduction inhibition occurs when the production of a transposase reaches some concentration threshold that effectually decreases the overall transposition frequency. This has been observed for the resurrected *Mariner* family element, *Sleeping Beauty* (*SB*), and has been described as a major limitation of the *SB* system for use in gene therapy (Lohe and Hartl, 1996, Wilson *et al.,* 2007). There are is currently no literature reporting overproduction inhibition in *MULE* family elements.

One possibility for increasing the germline mobility of *Muta1* in *D. melanogaster* would be the use of a germline specific promoter for driving transposase expression. The transgenic fly lines showed increased *Muta1* expression after heat-shock treatment and leaky expression without heat-shock treatment (Table 3.3). However, the once-daily heat-shock protocol may not have driven *Muta1* expression in the germline tissues at a consistent level or at the right developmental stage to promote a high frequency of germline transposition events. There are a multitude of germline specific promoters available for use in *D. melanogaster,* including the *vasa* and *nanos* regulatory sequences (Sano *et al.,* 2002, Van Doren *et al.,* 1998). As there is no need for new enhancer trap systems in *D. melanogaster,* I did not remake a helper line with a germline-specific promoter. However, this observation helped inform the enhancer trap experiments designed for *Aedes* and *Anopheles*, which are discussed in Chapters 4 and 5.

3.5 References

Arensburger, Peter, Robert H. Hice, Liqin Zhou, Ryan C. Smith, Ariane C. Tom, Jennifer A. Wright, Joshua Knapp, David A. O'Brochta, Nancy L. Craig, and Peter W. Atkinson. 2011. "Phylogenetic and Functional Characterization of the hAT Transposon Superfamily." *Genetics* 188 (1): 45 LP – 57.

Bellen, H. J., C. J. O'Kane, C. Wilson, U. Grossniklaus, R. K. Pearson, and W. J. Gehring. 1989. "P-Element-Mediated Enhancer Detection: A Versatile Method to Study Development in Drosophila." *Genes & Development* 3 (9): 1288–1300.

Bhatt, Samir, Peter W. Gething, Oliver J. Brady, Jane P. Messina, Andrew W. Farlow, Catherine L. Moyes, John M. Drake, et al. 2013. "The Global Distribution and Burden of Dengue." *Nature*. https://doi.org/10.1038/nature12060.

Chalvet, Fabienne, Christine Grimaldi, Fiona Kaper, Thierry Langin, and Marie-Josée Daboussi. 2003. "Hop, an Active Mutator-like Element in the Genome of the Fungus Fusarium Oxysporum." *Molecular Biology and Evolution* 20 (8): 1362–75.

Cresse, A. D., S. H. Hulbert, W. E. Brown, J. R. Lucas, and J. L. Bennetzen. 1995. "Mu1-Related Transposable Elements of Maize Preferentially Insert into Low Copy Number DNA." *Genetics* 140 (1): 315.

Crooks, G. E. 2004. "WebLogo: A Sequence Logo Generator." *Genome Research*. https://doi.org/10.1101/gr.849004.

D'Souza, J., P. Y. Cheah, P. Gros, W. Chia, and V. Rodrigues. 1999. "Functional Complementation of the Malvolio Mutation in the Taste Pathway of Drosophila Melanogaster by the Human Natural Resistance-Associated Macrophage Protein 1 (Nramp-1)." *The Journal of Experimental Biology* 202 (14): 1909.

Dupeyron, Mathilde, Kumar S. Singh, Chris Bass, and Alexander Hayward. 2019. "Evolution of Mutator Transposable Elements across Eukaryotic Diversity." *Mobile DNA*. https://doi.org/10.1186/s13100-019-0153-8.

Ferguson, Ann A., and Ning Jiang. 2012. "Mutator-like Elements with Multiple Long Terminal Inverted Repeats in Plants." *Comparative and Functional Genomics* 2012 (March): 695827.

Feschotte, Cédric, and Ellen J. Pritham. 2007. "DNA Transposons and the Evolution of Eukaryotic Genomes." *Annual Review of Genetics* 41 (1): 331–68.

Fraser, M. J., T. Clszczon, T. Elick, and C. Bauser. 1996. “Precise Excision of TTAA-Specific Lepidopteran Transposons piggyBac (IFP2) and Tagalong (TFP3) from the Baculovirus Genome in Cell Lines from Two Species of Lepidoptera.” *Insect Molecular Biology* 5 (2): 141–51.

Fu, Xue-Qian, F. U. Xue-qian, Jing Feng, Y. U. Bin, G. A. O. You-jun, Yong-Lian Zheng, and Y. U. E. Bing. 2013. “Morphological, Biochemical and Genetic Analysis of a Brittle Stalk Mutant of Maize Inserted by Mutator.” *Journal of Integrative Agriculture*. https://doi.org/10.1016/s2095-3119(13)60200-2.

Fu, Yu, Akira Kawabe, Mathilde Etcheverry, Tasuku Ito, Atsushi Toyoda, Asao Fujiyama, Vincent Colot, Yoshiaki Tarutani, and Tetsuji Kakutani. 2013. “Mobilization of a Plant Transposon by Expression of the Transposon-Encoded Anti-Silencing Factor.” *The EMBO Journal* 32 (17): 2407–17.

Garrels, Wiebke, Lajos Mátés, Stephanie Holler, Anna Dalda, Ulrike Taylor, Björn Petersen, Heiner Niemann, Zsuzsanna Izsvák, Zoltán Ivics, and Wilfried A. Kues. 2011. “Germline Transgenic Pigs by Sleeping Beauty Transposition in Porcine Zygotes and Targeted Integration in the Pig Genome.” *PloS One* 6 (8): e23573.

Guimond, N., D. K. Bideshi, A. C. Pinkerton, P. W. Atkinson, and D. A. O’Brochta. 2003. “Patterns of Hermes Transposition in Drosophila Melanogaster.” *Molecular Genetics and Genomics: MGG* 268 (6): 779–90.

Hickman, Alison B., Andrea Regier Voth, Hosam Ewis, Xianghong Li, Nancy L. Craig, and Fred Dyda. 2018. “Structural Insights into the Mechanism of Double Strand Break Formation by Hermes, a hAT Family Eukaryotic DNA Transposase.” *Nucleic Acids Research* 46 (19): 10286–301.

Jiang, Ning, Ann A. Ferguson, R. Keith Slotkin, and Damon Lisch. 2011. “Pack-Mutator-like Transposable Elements (Pack-MULEs) Induce Directional Modification of Genes through Biased Insertion and DNA Acquisition.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (4): 1537–42.

Kazazian, Haig H. 2011. “Mobile DNA Transposition in Somatic Cells.” *BMC Biology* 9 (1): 62.

Kokoza, V., A. Ahmed, E. A. Wimmer, and A. S. Raikhel. 2001. “Efficient Transformation of the Yellow Fever Mosquito Aedes Aegypti Using the piggyBac Transposable Element Vector pBac[3xP3-EGFP Afm].” *Insect Biochemistry and Molecular Biology*. https://doi.org/10.1016/s0965-1748(01)00120-5.

Li, Xianghong, Hosam Ewis, Robert H. Hice, Nirav Malani, Nicole Parker, Liqin Zhou, Cédric Feschotte, Frederic D. Bushman, Peter W. Atkinson, and Nancy L. Craig. 2013. "A Resurrected Mammalian hAT Transposable Element and a Closely Related Insect Element Are Highly Active in Human Cell Culture." *Proceedings of the National Academy of Sciences* 110 (6): E478 LP – E487.

Li, Yubin, Linda Harris, and Hugo K. Dooner. 2013. "TED, an Autonomous and Rare Maize Transposon of the Mutator Superfamily with a High Gametophytic Excision Frequency." *The Plant Cell* 25 (9): 3251–65.
Lisch, Damon. 2002. "Mutator Transposons." *Trends in Plant Science* 7 (11): 498–504.

Lisch, Damon. 2015. "Mutator and MULE Transposons." *Microbiology Spectrum* 3 (2): MDNA3–0032 – 2014.

Liu, Geyi, Elena L. Aronovich, Zongbin Cui, Chester B. Whitley, and Perry B. Hackett. 2004. "Excision of Sleeping Beauty Transposons: Parameters and Applications to Gene Therapy." *The Journal of Gene Medicine* 6 (5): 574–83.

Liu, Kun, and Susan R. Wessler. 2017a. "Functional Characterization of the Active Mutator-like Transposable Element, Muta1 from the Mosquito Aedes Aegypti." *Mobile DNA* 8 (January): 1.

Liu, Kun, and Susan R. Wessler. 2017b. "Transposition of Mutator–like Transposable Elements (MULEs) Resembles hAT and Transib Elements and V(D)J Recombination." *Nucleic Acids Research* 45 (11): 6644–55.

Liu, Sanzhen, Cheng-Ting Yeh, Tieming Ji, Kai Ying, Haiyan Wu, Ho Man Tang, Yan Fu, Dan Nettleton, and Patrick S. Schnable. 2009. "Mu Transposon Insertion Sites and Meiotic Recombination Events Co-Localize with Epigenetic Marks for Open Chromatin across the Maize Genome." *PLoS Genetics* 5 (11): e1000733.

Lohe, A. R., and D. L. Hartl. 1996. "Autoregulation of Mariner Transposase Activity by Overproduction and Dominant-Negative Complementation." *Molecular Biology and Evolution* 13 (4): 549–55.

Marquez, Claudia P., and Ellen J. Pritham. 2010. "Phantom, a New Subclass ofMutatorDNA Transposons Found in Insect Viruses and Widely Distributed in Animals." *Genetics*. https://doi.org/10.1534/genetics.110.116673.

Mátés, Lajos, Marinee K. L. Chuah, Eyayu Belay, Boris Jerchow, Namitha Manoj, Abel Acosta-Sanchez, Dawid P. Grzela, et al. 2009. "Molecular Evolution of a Novel Hyperactive Sleeping Beauty Transposase Enables Robust Stable Gene Transfer in Vertebrates." *Nature Genetics* 41 (May): 753.

McCarty, Donald R., and Robert B. Meeley. n.d. “Transposon Resources for Forward and Reverse Genetics in Maize.” *Handbook of Maize*. https://doi.org/10.1007/978-0-387-77863-1_28.

Nene, Vishvanath, Jennifer R. Wortman, Daniel Lawson, Brian Haas, Chinnappa Kodira, Zhijian Jake Tu, Brendan Loftus, et al. 2007. “Genome Sequence of Aedes Aegypti, a Major Arbovirus Vector.” *Science* 316 (5832): 1718–23.

Raizada, Manish N., Maria-Ines Benito, and Virginia Walbot. 2001. “The MuDR Transposon Terminal Inverted Repeat Contains a Complex Plant Promoter Directing Distinct Somatic and Germinal Programs.” *The Plant Journal*. https://doi.org/10.1046/j.1365-313x.2001.00939.x.

Robertson, Donald S. 1978. “Characterization of a Mutator System in Maize.” *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. https://doi.org/10.1016/0027-5107(78)90004-0.

Sano, Hiroko, Akira Nakamura, and Satoru Kobayashi. 2002. “Identification of a Transcriptional Regulatory Region for Germline-Specific Expression of Vasa Gene in Drosophila Melanogaster.” *Mechanisms of Development* 112 (1): 129–39.

Sarkar, Abhimanyu, Kurt Yardley, Peter W. Atkinson, Anthony A. James, and David A. O’brochta. 1997. “Transposition of the Hermes Element in Embryos of the Vector Mosquito, Aedes Aegypti.” *Insect Biochemistry and Molecular Biology* 27 (5): 359–63.

Singer, Tatjana, Cristina Yordan, and Robert A. Martienssen. 2001. “Robertson’s Mutator Transposons in A. Thaliana Are Regulated by the Chromatin-Remodeling Gene Decrease in DNA Methylation (DDM1).” *Genes & Development* 15 (5): 591–602.

Smith, Ryan C., and Peter W. Atkinson. 2011. “Mobility Properties of the Hermes Transposable Element in Transgenic Lines of Aedes Aegypti.” *Genetica* 139 (1): 7–22.

Van Doren, Mark, Anne L. Williamson, and Ruth Lehmann. 1998. “Regulation of Zygotic Gene Expression in Drosophila Primordial Germ Cells.” *Current Biology: CB* 8 (4): 243–46.

Wilson, Matthew H., Craig J. Coates, and Alfred L. George. 2007. “PiggyBac Transposon-Mediated Gene Transfer in Human Cells.” *Molecular Therapy: The Journal of the American Society of Gene Therapy* 15 (1): 139–45.

Woodard, Lauren E., Xianghong Li, Nirav Malani, Aparna Kaja, Robert H. Hice, Peter W. Atkinson, Frederic D. Bushman, Nancy L. Craig, and Matthew H. Wilson. 2012. “Comparative Analysis of the Recently Discovered hAT Transposon TcBuster in Human Cells.” *PloS One* 7 (11): e42666.

Xu, Zhennan, Xianghe Yan, Steve Maurais, Huihua Fu, David G. O'Brien, John Mottinger, and Hugo K. Dooner. 2004. "Jittery, a Mutator Distant Relative with a Paradoxical Mobile Behavior: Excision without Reinsertion." *The Plant Cell* 16 (5): 1105–14.

Yuan, Yao-Wu, and Susan R. Wessler. 2011. "The Catalytic Domain of All Eukaryotic Cut-and-Paste Transposase Superfamilies." *Proceedings of the National Academy of Sciences of the United States of America* 108 (19): 7884–89.

Zhao, Dongyan, Ann Ferguson, and Ning Jiang. 2015. "Transposition of a Rice Mutator-Like Element in the Yeast Saccharomyces Cerevisiae." *The Plant Cell* 27 (1): 132.

3.6 Figures and Tables

Experiment	Donors screened	True events	Frequency per donor
Muta1 enhancer trap construct	150,506	24	1.59E-04
piggyBac positive control donor	81,447	5	6.14E-05

Table 3.1 – The result of interplasmid transposition assays performed in HeLa culture. The *Muta1* element shows a higher transposition frequency than the *piggyBac* control element.

Fertile G0 crosses	*piggyBac* lines	*piggyBac* transformation rate	*Muta1* lines	*Muta1* transformation rate
61	9	14.8%	3	4.9%

Table 3.2 – The germline transformation frequencies of the *piggyBac* element (pBac3ChspMuta1) and the *Muta1* element (pMutaENT3) in *Drosophila melanogaster* embryos. The frequency was calculated by the number of fertile G_0 crosses giving rise to transgenic G_1 progeny divided by the total number of fertile crosses.

	Parental line after heat shock compared to no heat shock	F_1 progeny after heat shock compared to no heat shock	F_1 progeny compared to parent, no heat shock	F_1 progeny compared to parent after heat shock
Relative fold change in *Muta1* transposase expression	**90.76 fold increase**	**28.01 fold increase**	**4.0 fold increase**	**2.41 fold increase**

Table 3.3 – Quantitative real-time PCR data for *Muta1* transposase expression in *D. melanogaster* parental line and F_1 enhancer trap cross, before and after heat shock protocol.

Fly line	Transformation	Plasmid	TSD	Integration location	Features
1126D	*piggyBac*	pBac3ChspMuta1	TTAA	Chr X @base 14265760-	Inserted into intron of CG13404, 154 amino acid unknown function
1109J	*piggyBac*	pBac3ChspMuta2	TTAA	Chr 2R @base 10806761-	Inserted into intron of windei, a chromatin associated protein
1109A	*piggyBac*	pBac3ChspMuta3	TTAA	Chr 2 @base 7661823+	Inserted into intron of Slowpoke binding protein
1126N	*Muta1*	pMutaENT3	ATACGTTTT	Chr 3R @base 17714022+	Inserted into intron of osa, a chromatin associated protein

Table 3.4 – Construct integration locations of selected enhancer trap and helper lines of *D. melanogaster*. Integration locations were determined by inverse PCR or by plasmid rescue and by subsequent sequencing. All lines exhibited a clean integration of the element flanked by identical target site duplications at both ends.

Fly line	Generation sampled	Novel adult phenotype	Putative TSD	Integration location	Remaining end	Features
Parental line 1126N	na	na	AAAACGTAT	Chr 3R @base 17714022+	na	Integrated into intron of osa
1A	25	na	GCGGCTGGC	Chr 3L @base 16117962	Right	Integrated into exon of taf4
1A	25	na	GTTTCGTTT	Chr 2R @base 17846679	Right	Integrated into intron of olf186-F
1B	15	Loss of EGFP eye	CCCTAATC	Chr 3R @base 17713750	Left, possibly absent	Integrated into intron of osa
2	21	na	CCACGAATC	Chr 2L @base 265614	Left, possibly absent	Inserted into intron of CG3645, exon of CG17075
2	21	na	GATTCCTGA	Chr 2L @base 19118240	Right	Inserted into intron of dopa cecarboxylase
6B	15	na	ATATATATA	Chr 3R @base 25058215	Right	Inserted into intron of staccato

Table 3.5 – Table of remobilization events in *D. melanogaster* enhancer trap crosses for which new integration locations have been fully or partially characterized. The parental integration location of the enhancer trap construct is included at the top of the table.

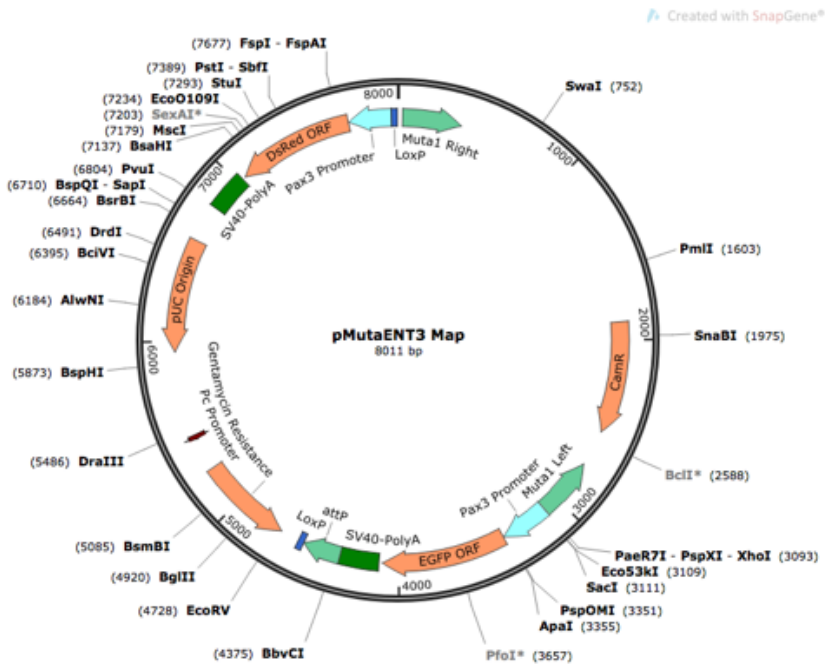

Figure 3.1 – The *Muta1* enhancer trap plasmid containing two bi-directional fluorescent eye markers, EGFP and DsRed, between the *Muta1* right and left ends.

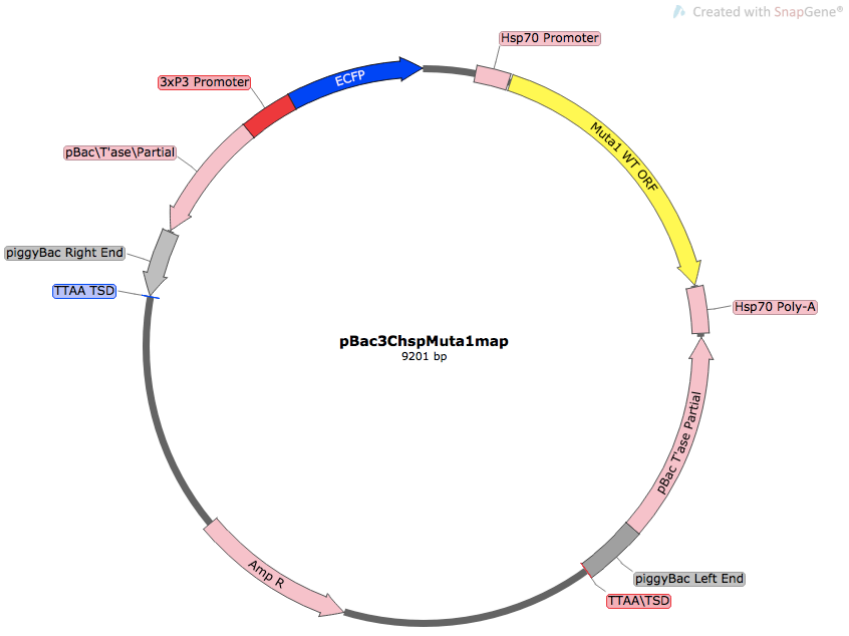

Figure 3.2 – The *Muta1* helper plasmid containing the *Muta1* ORF under control of the *hsp-70* promoter, containing an ECFP eye marker for screening.

Figure 3.3 – A shorter version of the *Muta1* enhancer trap plasmid containing two bi-directional fluorescent eye markers, EGFP and DsRed, between the *Muta1* right and left ends.

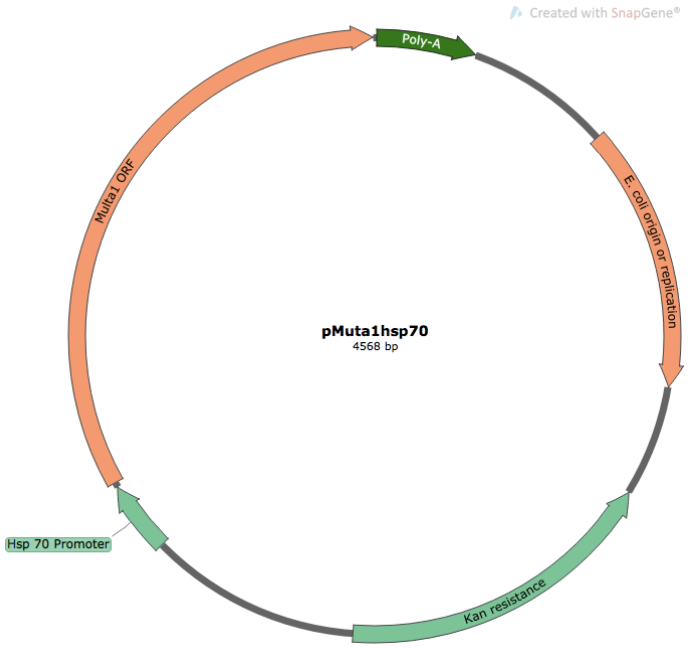

Figure 3.4 – A *Muta1* helper plasmid containing the *Muta1* coding sequence under control of the *hsp-70* promoter.

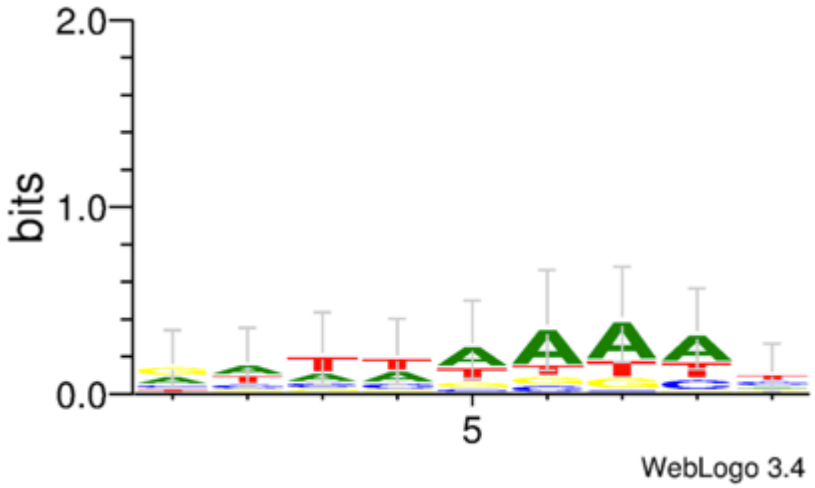

Figure 3.5 – A WebLogo of *Muta1* 9-base-pair target site duplications recovered from interplasmid transposition assays in HeLa cell culture.

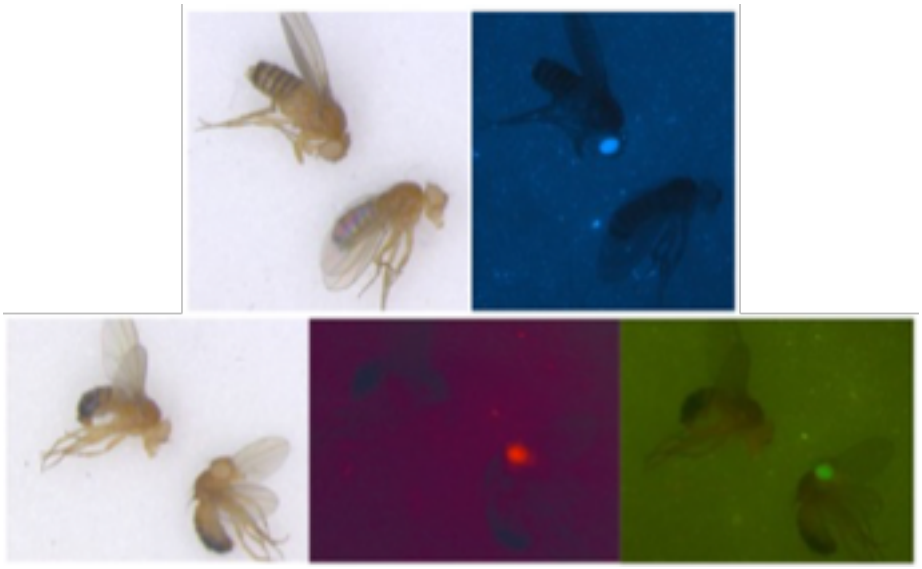

Figure 3.6 –*Muta1* and *piggyBac* transgenic lines of *D. melanogaster*.
Top row: One wild-type CSW female (bottom right in each shot) alongside one female (top left) from a line transformed with *piggyBac* from the helper construct, pBac3ChspMuta1. The transgenic female shows 3PX3 driven expression of ECFP.
Bottom row: One wild-type CSW male (top left in each shot) alongside one male (bottom right) from a line transformed with *Muta1* from the enhancer trap construct, pMutaENT3. The transgenic male shows 3PX3 driven expression of DsRed and EGFP.

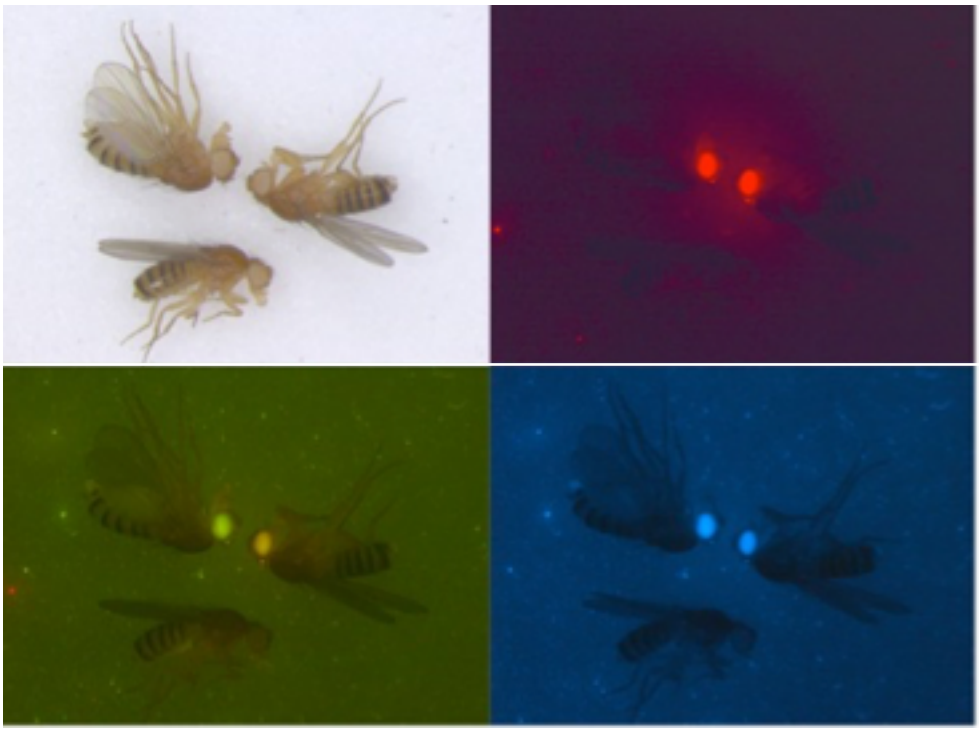

Figure 3.7 –Enhancer trap crossed line compared to enhancer trap crossed line "1B" exhibiting loss of 3PX3 EGFP expression. In all photographs the female in the top left corner is from a crossed enhancer trap line showing no changes in eye marker expression from both constructs (pMutaENT3 and pBac3ChspMuta1). The female fly on the right side of every panel is from crossed enhancer trap line "1B" showing retention of DsRed and ECFP eye expression but a loss of EGFP eye expression. The bottom fly in every panel is a female wild type CSW fly.

Chapter 4 – A *Muta1* transposon-based enhancer trap system in *Aedes aegypti*

4.1 Introduction

4.1.1 *Aedes aegypti* is a major global arboviral vector

The mosquito species *Aedes aegypti* is found in tropical, subtropical, and various temperate regions throughout the world, including the United States and California. *Ae. aegypti* is the major vector of significant arboviruses including those responsible for dengue, yellow, chikungunya, and Zika fevers. The According to the World Health Organization, about half of the world population is estimated to be at risk for contracting dengue. This is due to many contributing factors, including globalization and the spread of *Aedes*, poverty and lack of resources, climate change, the increase of insecticide resistance, *Ae. aegypti* feeding habits, and the adaptably of *Aedes* to thrive in urban and semi-urban environments alongside humans (Jansen and Beebe, 2010).

The burden of mosquito vector-borne diseases and the emergence of new diseases will continue to be a problem with increasing globalization and weather-related phenomena arising from global climate change. For example, the Centers for Disease Control report that the incidence of dengue fever infection did not occur in the Americas until 1981 and it is now endemic in these countries. *Aedes aegypti* and *Aedes albopictus*, which are the two major vectors of the dengue fever virus, have recently been reported as having achieved persistent populations in California, including in Los Angeles, Orange, Kern, and Alameda counties, but prior to 2011 these species had never been reported to have persistent populations anywhere in California (Porse *et al.,* 2015). The *Ae. aegypti*

mosquito is likely to have originated in Africa from a tree-hole dwelling species called *Ae. aegypti formosus* and was first spread to the Americas during the Atlantic slave trade genocide that began in the 16th century, and likely moved into Asia at a later time in the 19th century (Tabachnick 1991, Brown *et al.,* 2013).

Traditional vector control methods for controlling mosquito populations and decreasing disease transmission are numerous, from environmental control to treated bed nets. Insecticides are used to kill the adult or larval life stages of the mosquito, but the application of persistent insecticides comes with environmental and human health concerns (Weaver *et al.,* 2016). Additionally, increasing insecticide resistance is a constant challenge as this trait is positively selected for after exposure. The primary neurotoxic insecticides used to control *Aedes* populations, pyrethroids, carbamates, organochlorines, and organophosphates, have all seen a buildup of resistance in all locations where they have been deployed across the globe (Moyes *et al.,* 2017).

In additional to traditional vector control methods, genetic strategies also provide methods for population control. Sterile insect technique (SIT), which generally involves the large-scale release of sterile males to compete for mates in a wild-type population, has been used successfully in a variety of species, including *Ae. aegypti* (Harris *et al.,* 2012, Carvalho *et al.,* 2015). SIT is costly and complicated, however, due to the necessity of mass rearing and release over long-term periods. The concept of gene drive, the ability of a gene to increase its copy within a population at a pace faster than expected by Mendelian inheritance, has been discussed for decades as a promising tool for insect population modification (Curtis, 1968). The development of targeted CRISPR/Cas9 gene

editing and the experimental success of the mutagenic chain reaction method (CRISPR/Cas9-based gene drive system) has made the possibility of gene drive in *Aedes* a realistic ambition (Gantz and Bier, 2015, Adelman and Tu, 2016). Successful and heritable CRISPR/Cas9 genome editing in *Aedes* has already been demonstrated in numerous reports (Kistler *et al.,* 2015, Li *et al.,* 2017, Buchman *et al.,* 2019). The ability to use genetic methods for population control or the modulation of *Aedes* vectoral capacity, however, requires an in-depth knowledge of genetic development, including tissue, sex and developmental specific genes and promoters, as well as the physiological relationship between *Ae. aegypti* and arboviruses.

One robust tool for studying the developmental and functional genomics of arthropods is transposable element-based enhancer and gene trapping systems, which have been used in the model organism, *Drosophila melanogaster*, for thirty years now (Bellen *et al.,* 1989). These forward-based genetics techniques do not currently exist in *Ae. aegypti*. Germline transformation and transgene delivery in *Ae. aegypti* has been accomplished using several exogenous transposable elements, however, none of these elements have been shown to remobilize in the genome (Sarkar *et al.,*1997, Jasinskiene *et al*., 1998, Kokoza *et al.,* 2001). The piRNA pathway, which is described in greater detail in Section 4.1.3, has been suggested as the cause of post-integrational silencing in *Aedes* (Palavesam *et al.,* 2013). One of the aims of this thesis is to employ a *Muta1*-based enhancer trap system in *Ae. aegypti* as a new tool to allow the identification of promoters and genes that will be of interest for downstream genetic modification and gene drive research. The *Muta1* element, described in Section 3.1.3, is endogenous to *Ae. aegypti*

and shows evidence of evolutionarily recent proliferation and activity. It may be possible that the *Muta1* element is capable of evading the same host-silencing mechanisms that prohibit exogenous transposons, like *piggyBac* or *Hermes,* from remobilizing within the genome.

4.1.2 *Aedes aegypti* genomic features

The *Ae. aegypti* genome is 1.38 Gb in size and it is comprised of 47% transposable element sequence (Nene *et al.,* 2007). For perspective, the genome of *Anopheles gambiae*, which is the major malaria parasite vector in Africa, has a genome size of ~278 Mb, and is comprised of 20% transposable element sequence (Neafsey *et al.,* 2015). The sheer size of the *Ae. aegypti* genome and propensity of transposons and other simple repeats have posed a challenge in the complete assembly and annotation of the genome (Nene *at al.,* 2007, Timoshevskiy *et al.,* 2014). These genome characteristics also make it a difficult task to recover and study genomic mutations.

Another barrier to expanding forward based genetics tools in *Ae. aegypti* is the unusual inability of exogenous transposons to remobilize after integration. The *Ae. aegypti* genome has been successfully transformed using three transposable elements: *Hermes* from the housefly *Musca domestica, piggyBac* from the cabbage looper moth *Trichoplusia ni,* and *Mos1* from *Drosophila mauritiana* (Jasinskiene *et al.,*1998, Kokoza *et al.,* 2001, Coates *et al.,* 1998, Sethuraman *et al.,* 2007). None of these transposons have shown the ability to remobilize in *Aedes*, effectively making transposon-based trapping experiments a current impossibility for this species. Several publications have

shown evidence of piRNA targeting of exogenous transposons, post integration (Arensburger *et al.,* 2011, Palavesam *et al.,* 2013). A study using *Ae. aegypti* cell culture showed that a *piggyBac* element can excise from a donor plasmid in interplasmid transposition assays, but chromosomally integrated *piggyBac* is immobile even in the presence of functional transposase (Palavesam *et al.,* 2013). In another study, piRNA libraries generated for transformed *Ae. aegypti* transgenic lines revealed piRNAs mapping to exogenous transposons, *Hermes*, *Mos1*, and *piggyBac*, as well as other foreign sequences from the introduced transposon construct (Arensburger *et al.,* 2011).

The *Ae. aegypti* genome contains 14 different MULE family elements including the active *Muta1* element (Liu and Wessler, 2017). The *Muta1* element has eight full-length copies, all containing 8 base pair or 9 base pair target site duplications, and hundreds of copies of non-autonomous MITEs within the *Ae. aegypti* genome (Liu and Wessler, 2017). Due to the observed remobilization activity of *Muta1* in yeast, *D. melanogaster* and *Ae. aegypti* embryos, and human cell culture (see Chapter 3 results), coupled with the evolutionarily recent proliferation of *Muta1* in the *Ae. aegypti* genome, it is reasonable to speculate that the endogenous *Muta1* element may able, to some extent, to evade host silencing mechanisms (Liu and Wessler, 2017, Shah, 2015).

4.1.3 *Aedes aegypti* has an expanded piRNA pathway

The piRNA pathway, which maintains germline stability by silencing transposable element activity, has been well characterized and studied in the model organism, *D. melanogaster* (Aravin *et al.,* 2001, Brennecke *et al.,* 2007, Malone *et al.,*

2009). In *D. melanogaster,* the expression of the three canonical PIWI proteins, Piwi, Aubergine (Aub), and Ago3, is confined to the germline and somatic ovarian support cells (nurse cells). There are two piRNA silencing pathways, including the primary piRNA pathway and the ping-pong amplification pathway (Malone *et al.,* 2009). During piRNA biogenesis, piRNA precursors transcribed from distinct genomic loci called piRNA clusters, which exist as dual-strand (bidirectional) or uni-strand (unidirectional), depending on the direction of transcription (Brennecke *et al.,* 2007). The uni-strand piRNA clusters, which are the main piRNA clusters transcribed in ovarian somatic support cells, are transcribed in a single direction from a distinct promoter by RNA pol II and follow canonical mRNA processing in the nucleus (Goriaux *et al.,* 2014). The uni-strand piRNA precursor is subsequently processed into mature anti-sense primary piRNAs within cytoplasmic granules referred to as nuage (Lim and Kai, 2007). In the primary silencing pathway, the mature anti-sense primary piRNAs, which are generally 26 – 30 nucleotides in length, are loaded into the protein PIWI to form a piRISC (piRNA induced silencing complex) and transported back into the nucleus where the complementary transposable elements are silenced through repressive histone marks (Sienski *et al.,* 2012). Bidirectional piRNA clusters, which are only transcribed in germ cells, can be transcribed from both strands by RNA pol II. Unlike uni-strand piRNA clusters, bidirectional clusters do not have promoters and, instead, require the presence of H3K9me3 repressive histone marks. The deposition of H3K9me3 repressive marks, which are both inherited and deposited *de novo,* is facilitated by nuclear PIWI-piRISC (Sato and Saomi, 2018). The piRNA precursors from bidirectional clusters are also

processed in cytoplasmic nuage, where the sense piRNAs can bind Piwi or they can bind Aub. The piRNAs that bind Aub feed into the secondary piRNA pathway, called the ping-pong amplification loop. Aubergine and Ago3 generate secondary piRNAs within the cytoplasm by cleaving sense and antisense transposable element transcripts, creating additional piRNAs, which continue targeting active transposon transcripts through Aub and Ago3 binding (Brennecke *et al.,* 2007).

Ae. aegypti has a notable expansion of the PIWI proteins compared to *Drosophila* (Campbell *et al.,* 2008). There is a one-to-one, or near one-to-one ratio of PIWI orthologs between *Drosophila* and *Anophelines*, including the malaria vectors, *An. gambiae* and *An. stephensi* (Campbell *et al.,* 2008, Macias *et al.,* 2014). However, there are eight PIWI genes in *Ae. aegypti*, including AeAgo3 and AePiwi1 through AePiwi7 (Campbell *et al.,* 2008). While the expression of PIWI proteins is observed in the *Drosophila* germline and ovarian somatic cells (Brennecke *et al.,* 2007, Sato and Siomi, 2015), the collection of *Ae. aegypti* PIWIs shows a varied expression profile with some genes expressed in the germline and during embryogenesis and others with greater expression in the soma (Akbari *et al.,* 2013, Schnettler *et al.,* 2013, Han, 2017).

In addition to these differences, *Ae. aegypti* piRNAs target more than transposons. Approximately 16% of the *D. melanogaster* genome is comprised of transposable element sequences (Arensburger *et al.,* 2011), and 89% of sequenced piRNAs map to known piRNA clusters and transposon sequences in the genome (Song *et al.,* 2014). However, while *Ae. aegypti* piRNA clusters were found to map to more than half of the

genome, only 19% of sequenced piRNAs mapped to transposons, with the majority mapping to genes, and non-coding sequences (Arensburger *et al.,* 2011).

The evolutionarily expanded piRNA pathway constituents in *Ae. aegypti* have been implicated in antiviral immunity (Morazzani *et al.,* 2012, Miesen *et al.,* 2015, Miesen *et al.,* 2016). *Ae. aegypti* has a functioning antiviral siRNA pathway, that is conserved among insects, and utilizes small interfering RNAs that form RNA induced silencing complexes with the AGO clade of Argonaute proteins to target exogenous viral genes (reviewed by Bronkhorst and van Rij, 2014).

4.1.4 Chapter aims

The goal of the research described in this chapter is to generate stable transgenic enhancer trap lines in *Ae. aegypti* using the endogenous *Muta1* transposable element. There is no exogenous transposon that has been shown to remobilize in *Ae. aegypti*, therefore no enhancer or promoter trap systems have been developed in *Ae. aegypti.* The evidence presented in Chapter 3 led me to believe that *Muta1* may have a greater potential to evade *Aedes* silencing mechanisms than the exogenous elements that have been previously used in transformation experiments. Based on the results of the *D. melanogaster* enhancer trap crosses described in Chapter 3, which exhibited a low germline remobilization frequency, I reasoned that using a germline specific promoter, instead of a heat-shock inducible promoter, may increase the likelihood of germline remobilization events. I made a new helper construct using the *Ae. aegypti Exuperantia* regulatory sequence to drive *Muta1* transposase expression. To generate transgenic *Aedes*

lines, embryo microinjections were performed in Dr. Atkinson's laboratory and at the Insect Transformation Facility at the University of Maryland Institute for Bioscience and Biotechnology Research. I reared injected eggs and established adult crosses to screen for transgenic progeny, backcrossed transgenic lines, and used a combination of inverse PCR and Splinkerette PCR to characterize transgene integrations. I used qRT-PCR to verify the level of *Muta1* transposase expression in transgenic helper lines

4.2 Materials and methods

4.2.1 Plasmid construction

Construction of plasmids pMutaENT3 and pBac3ChspMuta1 is outlined in Chapter 3.

Construction of helper pBac3CExuMuta1 (Figure 4.1): The plasmid pAAEL010097-RS-UTR-Opie2-dsRed was a kind gift from Dr. Omar Akbari (Akbari *et al.,* 2014). This plasmid contains the 5' and 3' regulatory sequences for the *Ae. aegypti Exuperantia* homolog, AAEL010097, and has expression in *Ae. aegypti* pre-vitellogenic ovaries, in post-blood fed oocytes, and male testis (Akbari *et al.,* 2014). Plasmid pBac3ChspMuta1 was linearized with ThermoFisher FastDigest enzyme *Avr*II and purified with Qiagen QIAquick PCR purification kit according to the manufacturer's protocol. The following primers were used to amplify 1,564 bp of the *Muta1* CDS and add homology arms. The homology sequence to *exu* is in bold.

Muta Exu Hom F:

5' – **TCGAGGCTTAATTAAGTTT**CTGCAGACTAGTATGGACTC – 3'

Muta Exu Hom R:

5’ – **GAATTCTGCCGCGTT**CTCGAGTTATTTTGATTTTGATCC – 3’

PCR was performed using New England Biolabs Q5 High Fidelity Polymerase and the PCR fragment was purified with QIAquick PCR Purification Kit using manufacturer protocol.

The plasmid pAAEL010097-RS-UTR-Opie2-dsRed was digested with ThermoFisher FastDigest enzyme *Mss*I to obtain a linear plasmid and purified with Qiagen QIAquick PCR purification kit according to manufacturer protocol. The first assembly reaction using the ThermoFisher GeneArt Seamless Assembly Kit joins the pAAEL010097-RS-UTR-Opie2-dsRed plasmid, linearized with blunt cutter MssI, to the *Muta1* CDS amplified from pBac3ChspMuta1. The plasmid obtained was named pVector-AAEL010097-RS-Muta1-UTR-Opie2-dsRed.

The following primers were used to amplify a 4,742 bp fragment from the assembled plasmid contain the *Exu* regulatory sequences flanking the *Muta1* CDS.

Exu Muta F:

5’ – GGTTAATTCGAGCTCGCCCGGGGTCC – 3’

Exu Muta R:

5’ – CCACCGAGTATGGGCGCGCCAAC – 3’

A 5,526 bp fragment was amplified from pBac3ChspMuta1 using New England Biolabs Q5 High Fidelity Polymerase and the primers below. The homology sequence to *exu* is in bold.

PBSmall Exu Hom F:

5’ - **GCCCATACTCGGTGG**CATCGTCTAAAGAACTACCC – 3 ‘

PBSmall Exu Hom R:

5’ – **GGGCGAGCTCGAATT**CCGCGTATCGATAAGCTTTAAG – 3’

The fragment was purified with QIAquick PCR Purification Kit using the manufacturer protocol.

The second assembly reaction using the ThermoFisher GeneArt Seamless Assembly Kit joins the pVector-AAEL010097-RS-Muta1-UTR-Opie2-dsRed Exu/*Muta1*fragment with the pBac3ChspMuta1 vector fragment containing the Pax3-ECFP marker, left and right *piggyBac* transposon ends, and an ORI-AmpR vector sequence. The plasmid, pBac3CExuMuta1, was maxi-prepped using Zymo Zyppy Maxi-Prep Kit according to manufacturer protocol.

4.2.2 *piggyBac* and *Muta1* mRNA production

piggyBac and *Muta1* mRNA was synthesized for in-house *Ae. aegypti* embryo injections. First, DNA template was made using New England Biolabs Q5 High Fidelity Polymerase Kit with GC-enhancer manufacturer protocol, using the following primers, with the T7 initiation sequence in bold.

pBac mRNA F:

5’ -

GAAACTAATACGACTCACTATAGGGAGAGCCGCCACATGGGTAGTTCTTT

AGACGATG – 3’

pBac mRNA R:

5’ – CTTATTAGTCAGTCAGAAACAAC – 3’

Muta mRNA F:

GAAACTAATACGACTCACTATAGGGAGAGCCGCCACATGGACTCGGACAG

CGATAG – 3’

Muta mRNA R:

5’ – TTTTGATTTTGATCCTAAGTGAGCTGCA – 3’

The template PCR products were purified using an Agencourt RNAClean XP Kit (catalog# A63987). A 1.5X volume of XP Bead solution was added to the PCR products, vortexed, and incubated at room temperature for 5 minutes. The tubes were placed in a magnetic stand for 5 minutes until the XP Bead solution cleared and the supernatant was removed without disturbing the magnetic beads. A wash step was preformed twice using 500ul of a freshly prepared 80% RNase-free ethanol solution. The templated DNA was eluted from the magnetic XP Beads using 20ul RNase-free water and the template concentration was checked by nanodrop reading.

The mRNA was synthesized using the New England Biolabs HiScribe T7 ARCA mRNA Kit (with tailing, catalog# E2060S) following manufacturer protocol. The quality and size of the transcribed mRNA was verified using the Agilent 2100 Bioanalyzer and the Bioanalyzer RNA 6000 nano Kit.

4.2.3 *Ae. aegypti* transformation

The *Ae. aegypti* Higgs white-eye strain was used as the background strain for injections. (Shah, 2015, Handler *et al.,* 1998). A three-plasmid injection mix was

prepared as described (Sarkar *et al.,* 1997) using 500ng/ul of pMutaENT3 (Figure 3.1), 250ng/ul pBac3ChspMuta1 (Figure 3.2), and phsp70pBac (Shah, 2015, Handler *et al.,* 1998), and a series of embryo microinjections were performed at the Insect Transformation Facility at the University of Maryland Institute for Bioscience and Biotechnology Research (Table 4.1). Embryo microinjections were also performed in Dr. Peter Atkinson's lab at the University of Riverside, California. For creating transgenic helper lines, embryo injections used a mix of 300ng/ul of the helper plasmid, pBac3CExuMuta1 (Figure 4.1) and 300ng/ul of synthesized *piggyBac* mRNA. For creating transgenic enhancer trap lines, embryo injections used a mix of 300ng/ul pMutaENT3 (Figure 3.1) and 300ng/ul synthesized *Muta1* mRNA. For injections performed at both facilities, the injected G_0 were reared to adulthood and crossed *en masse*. The G_1 progeny were screened for fluorescent eye marker expression to identify germline-transformation events.

4.2.4 Molecular verification of transgenic *Ae. aegypti* lines

Splinkerette PCR (Potter and Luo, 2010) was used to characterize *piggyBac* integrations in *Ae. aegypti*. The Qiagen DNeasy Blood and Tissue kit was used for genomic DNA extracts. Qiagen provides a modified protocol for genomic DNA extraction from insects and this protocol was followed as written.

Genomic DNA digests used BstYI (ThermoFisher Fast Digest enzymes). For PCR reactions, New England Biolabs Q5 High-Fidelity DNA Polymerase was used

following manufacturer protocol for 50ul reactions. The following primers were used for mapping *piggyBac* insertions by Splinkerette PCR.

SPLNK-GATC-Top:

5' – GATCCCACTAGTGTCGACACCAGTCTCTAATTTTTTTTTTCAAAAAAA – 3'

SPLNK-Bot:

5' –

CGAAGAGTAACCGTTGCTAGGAGAGACCGTGGCTGAATGAGACTGGTGTCGA

CACTAGTGG – 3'

SPLNK R1:

5' – CGAAGAGTAACCGTTGCTAGGAGAGACC – 3'

SPLNK R2:

5' – GTGGCTGAATGAGACTGGTGTCGAC – 3'

3' SPLNK R1:

5' – CACTCAGACTCAATACGACAC – 3'

3' SPLNK R2:

5' – GGATGTCTCTTGCCGAC – 3'

For verification of *Muta1* transformed lines, several PCR reactions were performed to recover pMutaENT3 fragments from genomic DNA. The Qiagen DNeasy Blood and Tissue kit was used for genomic DNA extracts. Qiagen provides a modified protocol for genomic DNA extraction from insects and this protocol was followed as written.

Standard PCR was performed using New England Biolabs Q5 High-Fidelity DNA Polymerase was used following manufacturer protocol for 50ul reactions. The following primers were used for obtaining pMutaENT3 fragments (Figures 4.5 through 4.7).

EGFP 115 For:

5’ – TCAAGATCCGCCACAACATC – 3’

EGFP 115 Rev:

5’ – GTGCTCAGGTAGTGGTTGTC – 3’

RPS7 new For:

5’ – CAGACCACCATTGAACACAA – 3’

RPS7 new Rev:

5’ – ATGCACACCCTAGTTCCGTA – 3’

IM LE FP1:

5’ – CTTTGTCACGATCCATTAGTCACTG – 3’

DsRed 2-1 F:

5’ – CACGTACACCTTGGAGCCGTAC – 3’

DsRed 2-1 R:

5’ – TGCTCCACGATGGTGTAGTCCT – 3’

IM RE RP1:

5’ – GAGAGAGGTTTTTCATTACATTTCTATGTATAC – 3’

pUC to RE:

5’ – GGTATCTTTATAGTCCTGTCGGG – 3’

Gent to LE:

5’ – CGAGATCATAGATATAGATCTCACTACGC – 3’

In preparation for the inverse PCR reactions for *Muta1* integrations, genomic DNA was digested with TaqI and Bsu15I (ThermoFisher Fast Digest enzymes). Digested genomic DNA was column purified with the Qiaquick PCR Purification Kit, following kit protocol. Purified products were self-ligated with T4 DNA Ligase (Thermo Fisher Scientific) and chloroform extracted following the ThermoFisher kit protocol. Inverse PCR reactions used Bioneer AccuPower PCR PreMix tubes and the following primers were used for *Muta1* integrations:

IM LE FP1:

5’ – CTTTGTCACGATCCATTAGTCACTG – 3’

IM LE RP1:

5’ – CATAACAGTGTGAGAAGCGTACG – 3’

IM LE FP2:

5’ – GCCTTATACTAGTTTATTGTATATTTGTACTACG – 3’

IM LE RP2:

5’ – GTACGACTAGATAAAGATGTTCATCATG – 3’

IM RE FP1:

5’ – TGACTCATGTGAACAACGGTAAC – 3’

IM RE RP1:

5’ – GAGAGAGGTTTTTCATTACATTTCTATGTATAC – 3’

IM RE FP2:

5’ – GTTTTAAAAATACGATTTCTGGTTATGGTTATGC – 3’

IM RE RP2:

5’ – TCCTAAGTGAGCTGCAATTGC – 3’

4.2.5 Molecular verification of transgene expression in *Ae. aegypti* lines

Quantitative RT-PCR was used to assess the levels of *Muta1* transposase expression in the transgenic lines. Three biological replicates were used for each sample, male adult or 72-hour post-blood fed female adult, and each biological replicate comprised three adults. This was to prevent variation between samples due to simple stochastic differences between individual adults. All adults were between 10 and 14 days old. RNA from each biological replicate was extracted using ThermoFisher TRIzol reagent following the manufacturer’s protocol for RNA isolation from 50-100mg tissue samples. The RNA samples were treated with the ThermoFisher TURBO DNA-free Kit according to manufacturer protocol, with a longer incubation period of 1 hour. All RNA samples were diluted to 50ng/ul stocks before cDNA synthesis using the New England Biolabs ProtoscriptII First Strand cDNA Synthesis Kit, using half-reactions.
All cDNA for each biological replicate was diluted to 100ng/ul stocks for subsequent qRT-PCR reactions.

Quantitative RT-PCR was performed using Bio-Rad iQ SYBR Green Supermix following the protocol for half reactions. Thermocycling reactions were performed on a Bio-Rad MyiQ Dectection system.

For the RPS7 housekeeping gene, a standard curve was generated using 450nM primer concentrations and an annealing temperature of 58.3ºC. The standard curve for this primer set had an E value of 99.7 and an R^2 value of 0.992.

RPS7 58 For:

5' – CATCCTGGAGGATCTGGTCTTCC – 3'

RPS7 58 Rev:

5' – GCTTCTTGTACACTGACGTGAAGG – 3'

For the *Muta1* transgene, a standard curve was generated using 400nM primer concentrations and an annealing temperature of 57ºC. The standard curve for this primer set had an E value of 95.3 and an R^2 value of 0.996.

Muta1 2015 For:

5' – GCGTATGGTAACGTTCAAGGC – 3'

Muta1 2015 Rev:

5' – GTACTATTTTCGCTGGCGTTG – 3'

All quantitative RT-PCR data was analyzed using the Pfaffl equation (Pfaffl, 2001).

4.3 Results

4.3.1 Transgenic helper lines were established in *Ae. aegypti*

Three rounds of *Ae. aegypti* embryo microinjections were performed at the Insect Transformation Facility at the University of Maryland Institute for Bioscience and Biotechnology Research and the injected eggs were sent to me for rearing. These injection sets used the *Muta1* enhancer trap plasmid, pMutaENT3, the *Muta1* helper plasmid, pBac3ChspMuta1, and the *piggyBac* helper plasmid phsp70pBac. Table 4.1 shows the adult emergence rate from the injected eggs and the transformation frequency of the plasmids. A total of 878 surviving G_0 adults were recovered from 6,134 injected embryos, giving an adult emergence rate of 15.3%. The pBac3ChspMuta1 transformation rate, seven transformed lines out of 878 surviving G_0 adults, giving a 0.8% *piggyBac* transformation frequency.

Ae. aegypti embryo microinjections were also performed in our laboratory. My injection protocol used synthesized *piggyBac* transposase mRNA in the injection mix instead of a three-plasmid injection mix, as was used at the University of Maryland. I also used the helper plasmid with the *Exuperantia* promoter, pBac3CExuMuta1. For establishing the *Muta1* helper line, 2,632 embryos were injected, and 526 surviving G_0 adults were recovered, providing an adult emergence rate of 20%. A total of fifteen *piggyBac* transgenic lines were recovered from 526 G_0 adults, giving a *piggyBac* transformation rate of 2.8% (Table 4.2).

The *Ae. aegypti* lines transformed with pBac3CExuMuta1 containing the *Muta1* coding sequence under germline promoter control were preferable over the helper lines transformed with pBac3ChspMuta1 from the University of Maryland, as the *exuperantia* promoter sequence was selected to drive *Muta1* expression specifically in germline tissues. Out of the fifteen pBac3CExuMuta1 transgenic lines, nine lines survived after subsequent backcrossing. The integration locations were recovered from five of the nine helper lines (Table 4.4). Out of the 5 characterized lines, two of the lines, J and Q, contained two separate chromosomal integrations. All of the characterized lines had canonical *piggyBac* 5'-TTAA-3' target site duplications and perfect integrations.

There were no observed fitness effects, such as decreased fecundity or larval survival, in any of the helper lines.

4.3.2 Transgenic *Ae. aegypti* lines showed an increase in *Muta1* transposase expression

Quantitative RT-PCR was used to quantify levels of *Muta1* transposase in adult males and blood-fed females for the transgenic lines and for the wild-type lab strains, as a control. Adult *Ae. aegypti* males showed a different expression profile compared to females. Males from the lines G, R, and N showed a high standard deviation in the expression data (Figure 4.2). Males from lines A, Q, C, J, and T, all showed better consistency between data points and displayed about a 5 to 6.5-fold increase in *Muta1* transposase levels compared to the Higgs background strain.

Interestingly, the wild type lab strains Orlando and Liverpool also had about a 5.5 to 6-fold increase in *Muta1* transposase levels compared to the Higgs background strain. The expression results for 72-hour post-blood fed females were quite different (Figure 4.3). The lines A and R showed a high standard deviation in the expression data. Despite the ECFP marker expression in the transgenic lines, none of the lines saw a marked increase in *Muta1* transposase expression level compared to the background Higgs strain. The R line and J line females displayed a slightly reduced *Muta1* expression profile compared to Higgs.

4.3.3 The establishment of transgenic enhancer trap lines in *Ae. aegypti* is ongoing

Three rounds of *Ae. aegypti* embryo microinjections were performed at the Insect Transformation Facility at the University of Maryland Institute for Bioscience and Biotechnology Research. Of the 6,134 embryos injected, no *Muta1* enhancer trap lines were recovered.

Ae. aegypti embryo microinjections were also performed in our lab. My injection protocol used synthesized *Muta1* transposase mRNA in the injection mix. For establishing the *Muta1* enhancer trap line, 5,830 embryos were injected, and 847 surviving G_0 adults were recovered, providing an adult emergence rate of 20.8% (Table 4.3). The 847 recovered G_0 adults were back-crossed *en masse* and G_1 progeny were screened as larvae for fluorescence. Six lines were recovered displaying non-canonical phenotypes, with faint ubiquitous EGFP or ECFP expression. Genomic DNA was isolated from these larvae and PCR analysis showed the presence of the pMutaENT3

construct within Line I (Figure 4.5 through Figure 4.7). Line I showed faint ubiquitous ECFP expression and PCR analysis indicated the presence of the pMutaENT3 transposon, including the *Muta1* ends. Figure 4.4 displays the primers used to verify the presence of pMutaENT3 in the I Line genomic DNA. Inverse PCR was used, as described in Section 4.2.3, to attempt to recover the integrated *Muta1* ends and the genomic junction sequences, however, only PCR artifacts were recovered.

4.4 Discussion

Aedes aegypti is the major vector of serious human and zoonotic viruses. Insights into *Ae. aegypti* genetics and host-pathogen interactions are of critical medical importance. One robust tool for studying the developmental and functional genomics of arthropods is transposable element-based enhancer and promoter trapping systems (Bellen, 1999). However, the ability of *Aedes* to target and silence integrated exogenous transposons (Palavesam *et al.,* 2013, Arensburger *et al.,* 2011) has resulted in the lack of a robust enhancer trapping system that could help identify tissue and sex specific promoters and enhancers in *Aedes*. The goal of the research described in this chapter was to generate stable transgenic enhancer trap lines in *Ae. aegypti* using the endogenous *Muta1* transposable element, as the evidence presented in this chapter and in Chapter 3 indicates that this endogenous transposon may be able to evade host silencing mechanisms.

The application of a *Muta1*-based enhancer trap system will require the generation of two separate lines that can be maintained as homozygous lines and crossed

for experimentation. One line, designated as the helper line, contains a helper construct driving *Muta1* transposase under the control of a germline promoter, in this case I used the *Ae. aegypti Exuperantia* homolog (Akbari *et al.,* 2014). This helper line provides the *Muta1* transposase to remobilize the construct contained on the enhancer trap line. I was able to initially generate 15 *piggyBac* transformed helper lines at a 2.8% transformation rate (Table 4.2). A recent meta-analysis explains that the mean *piggyBac* transformation efficiency reported in *Ae. aegypti* embryo microinjections is approximately 7%, with a wide range of variation reported, from 0% germline transformation up to about 17% (Gregory *et al.,* 2016). The germline transformation efficiency of 2.8% is a reasonable figure compared to other publications and requires a feasible number of embryo microinjections to assure the creation of at least one transformed line. Publications reporting *piggyBac* transformation in *Ae. aegypti* typically require embryo injections of several thousand or more (Kokoza *et al.,* 2001, Gregory *et al.,* 2016).

The *piggyBac* transformation frequency of the embryos injected at the University of Maryland Insect Transformation Facility was quite low at 0.8% (Table 4.1). However, the G_0 injected eggs must were shipped from Maryland to California for rearing. One can reasonably postulate that the stress of motion, temperature and pressure changes can have a negative impact on the embryos. Indeed, the adult emergence rate of the shipped G_0 embryos was 15.3% compared to 20% for embryos injected in our laboratory. Another factor that may contribute to the higher transformation rate I achieved was the use of synthesized *piggyBac* mRNA in the injection mix, instead of using a plasmid injection

mix, as is standard protocol for the University of Maryland Insect Transformation Facility.

No *Muta1* transformed lines were produced from the injections performed at the University of Maryland Insect Transformation Facility during the course of this project, out of 6,134 injected embryos. This is not a curious outcome, as the transformation efficiency for *piggyBac* was quite low in the same experiments (Table 4.1). However, we have been able to obtain *Muta1* transformed lines from the Insect Transformation Facility previously. The University of Maryland Insect Transformation Facility performed a set of *Ae. aegypti* Orlando strain embryo microinjection experiments for our lab in 2014 and obtained a *Muta1* transformation rate of 4% and a *piggyBac* transformation rate of 17.3% (Shah, 2015).

The germline transformation frequencies in *D. melanogaster* reported in Chapter 3 are 14.8% in *piggyBac* and 4.9% for *Muta1*, which means *Muta1* has a 3-fold lower rate of transformation compared to *piggyBac*, in this case (Table 3.2). The previous embryo microinjection experiment performed at the University of Maryland Insect Transformation Facility showed that *Muta1* had a 4.3-fold lower transformation frequency compared to *piggyBac* (Shah, 2015). I achieved a 2.8% *piggyBac* transformation frequency in *Ae. aegypti* embryo microinjections (Table 4.2) in our laboratory. If the expected transformation frequency of *Muta1* is about 3-fold to 4.3-fold lower than the *piggyBac* element, I could reasonably assume that would translate into a 0.6-0.9% *Muta1* transformation frequency for embryo injections performed under our laboratory conditions. Based on the number of surviving G_0 adults, 847 (Table 4.3), from

the *Muta1* embryo microinjections, I should have recovered about five transgenic lines, assuming a conservative transformation frequency of 0.6%. Out of the screened G_1 form the *Muta1* injection experiments, six lines with non-canonical fluorescent phenotypes were recover, and a single line, Line I, appeared to contain the pMutaENT3 construct (Table 4.3). The techniques typically used to characterize transposon integrations, including inverse PCR, Splinkerette PCR, and genome walking, require at least one known sequence for which primers can be designed, and these primers are typically made as close to the transposon ends as possible. There are over 300 non-autonomous *Muta1* MITEs within the *Ae. aegypti* genome (Liu and Wessler, 2017). This makes it challenging to recover junction fragments as the presence of so many *Muta1* transposon end sequences typically produce many products that are not specific to the pMutaENT3 construct (Liu and Wessler, 2017). Using inverse PCR, I was not able to recover the integrated plasmid, rather I recovered endogenous *Muta1* from *Ae. aegypti.* I have tried to design primers further out of the *Muta1* end sequences of the pMutaENT3 plasmid, even using vector sequence. However, PCR methods for recovering junction fragments, including inverse PCR, Splinkerette PCR, and genome walking, are typically more robust when using primers as close to the junction sequence (transposons ends) as possible (Ochman, *et al.,* 1988, Potter and Luo, 2010, Shyamala and Ames, 1989).

A possible but unexplored explanation could be the there is a difference in host silencing activity between the lab strains Orlando, used for previous experiments (Shah, 2015), and Higgs, the strain used in the experiments described in this chapter. To explore this hypothesis, control transformation experiments can be performed using the *piggyBac*

element, which is amenable to transformation in *Aedes* (Gregory *et al.,* 2016), and directly comparing transformation frequencies between the Orlando and Higgs strains.

The quantitative RT-PCR results for transformed males showed that there is a 5 to 6-fold lower level of *Muta1* transposase expression in the Higgs strain compared to the Orlando and Liverpool lab strains. In addition, the *Muta1* expression levels in the Orlando and Liverpool lab strains were comparable to the *Muta1* expression levels in the transformed Higgs lines (Figure 4.5). These data suggest that the Higgs strain may have more robust transposon silencing mechanisms. If this is indeed the case, the strains Liverpool or Orlando may provide preferable background strains for these enhancer trap experiments.

The *Muta1* expression data for 72-hour post-blood fed females is different than the data for males. In these data, all of the lab and transgenic strains have a comparable *Muta1* expression profile compared to the Higgs strain (Figure 4.6). It is feasible that females are sometimes capable of silencing transgenes. For example, in a study that generated piRNA libraries from transgenic *Ae. aegypti* lines, piRNAs were found to uniquely map to transgene sequences including exogenous transposase sequences, exogenous transposon ends, and plasmid vector sequences (Arensburger *et al.,* 2011). We have observed differential transgene expression in males and females previously (Han, 2017). In this case, a transgenic line with an EGFP fluorescent eye marker only had visible expression in transgenic males. For females, PCR analysis was used to confirm the presence of the transgenes in the absence of the fluorescent marker

expression, although the mechanism of the observed female silencing is still unknown (Han, 2017).

In the transgenic *Ae. aegypti* helper lines described in this chapter, ECFP eye marker expression was not distinctively different between transgenic lines, even though the *Muta1* transposase expression was measurably different. However, it is difficult to make any assumptions about transgene silencing based on eye marker expression alone. There are stochastic differences between mosquito larvae that could make the eye marker expression appear to be slightly different between transgenic lines. Additionally, the observed brightness of ECFP is only about 40% as bright as EGFP - it is a convenient screening marker, but not observably bright enough to detect expression variability by eye (Day and Davidson, 2009). Differences in fluorescent marker expression that could be attributed to some form of biological silencing would also need to be quantified molecularly by qRT-PCR.

There were no observed fitness effects, such as decreased fecundity or larval survival, in any of the helper lines. One of the *Ae. aegypti* helper lines, Line Q, displayed the largest increase in *Muta1* transposase in adult males as compared to the background strain (Figure 4.2). This line has been maintained with fluorescent marker expression for approximately 12 generations but has no apparent fitness problems. There are eight full-length endogenous copies of *Muta1* and hundreds of *Muta1* MITEs in the *Ae. aegypti* genome (Liu and Wessler, 2017). One can reasonably assume that germline expression of the *Muta1* transposase transgene would be equally capable of remobilizing endogenous *Muta1* elements. In fact, non-autonomous *Muta1* MITEs, which have

sometimes very little intervening sequence, may be more amenable to remobilization compared to the pMutaENT3 element, which carries 4946bp of cargo sequence (Geurts *et al.,* 2003). The appearance of reduced fecundity or sterile adults in helper Line Q could indicate a high level of germline transposition activity (Wright *et al.,* 2013), but this has not been observed. The Q helper line will be used for a targeted sequencing experiment to identify potentially remobilized endogenous *Muta1* elements (Chapter 6).

4.5 References

Adelman, Zach N., and Zhijian Tu. 2016. "Control of Mosquito-Borne Infectious Diseases: Sex and Gene Drive." *Trends in Parasitology* 32 (3): 219–29.

Aravin, A. A., N. M. Naumova, A. V. Tulin, V. V. Vagin, Y. M. Rozovsky, and V. A. Gvozdev. 2001. "Double-Stranded RNA-Mediated Silencing of Genomic Tandem Repeats and Transposable Elements in the D. Melanogaster Germline." *Current Biology: CB* 11 (13): 1017–27.

Arensburger, Peter, Robert H. Hice, Jennifer A. Wright, Nancy L. Craig, and Peter W. Atkinson. 2011a. "The Mosquito Aedes Aegypti Has a Large Genome Size and High Transposable Element Load but Contains a Low Proportion of Transposon-Specific piRNAs." *BMC Genomics* 12 (1): 606.

Arensburger, Peter, Robert H. Hice, Liqin Zhou, Ryan C. Smith, Ariane C. Tom, Jennifer A. Wright, Joshua Knapp, David A. O'Brochta, Nancy L. Craig, and Peter W. Atkinson. 2011. "Phylogenetic and Functional Characterization of the hAT Transposon Superfamily." *Genetics* 188 (1): 45 LP – 57.

Bellen, H. J. 1999. "Ten Years of Enhancer Detection: Lessons from the Fly." *The Plant Cell* 11 (12): 2271–81.

Bellen, H. J., C. J. O'Kane, C. Wilson, U. Grossniklaus, R. K. Pearson, and W. J. Gehring. 1989. "P-Element-Mediated Enhancer Detection: A Versatile Method to Study Development in Drosophila." *Genes & Development* 3 (9): 1288–1300.

Brennecke, Julius, Alexei A. Aravin, Alexander Stark, Monica Dus, Manolis Kellis, Ravi Sachidanandam, and Gregory J. Hannon. 2007. "Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila." *Cell* 128 (6): 1089–1103.

Bronkhorst, Alfred W., and Ronald P. van Rij. 2014. "The Long and Short of Antiviral Defense: Small RNA-Based Immunity in Insects." *Current Opinion in Virology* 7 (August): 19–28.

Brown, Julia E., Benjamin R. Evans, Wei Zheng, Vanessa Obas, Laura Barrera-Martinez, Andrea Egizi, Hongyu Zhao, Adalgisa Caccone, and Jeffrey R. Powell. 2014. "Human Impacts have Shaped Historical and Recent Evolution in Aedes Aegypti, the Dengue and Yellow Fever Mosquito." *Evolution; International Journal of Organic Evolution* 68 (2): 514–25.

Buchman, Anna, Stephanie Gamez, Ming Li, Igor Antoshechkin, Hsing-Han Li, Hsin-Wei Wang, Chun-Hong Chen, et al. 2019. “Engineered Resistance to Zika Virus in Transgenic Aedes Aegypti Expressing a Polycistronic Cluster of Synthetic Small RNAs.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (9): 3656–61.

Campbell, Corey L., William C. Black, Ann M. Hess, and Brian D. Foy. 2008. “Comparative Genomics of Small RNA Regulatory Pathway Components in Vector Mosquitoes.” *BMC Genomics* 9 (1): 425.

Carvalho, Danilo O., Andrew R. McKemey, Luiza Garziera, Renaud Lacroix, Christl A. Donnelly, Luke Alphey, Aldo Malavasi, and Margareth L. Capurro. 2015. “Suppression of a Field Population of Aedes Aegypti in Brazil by Sustained Release of Transgenic Male Mosquitoes.” *PLoS Neglected Tropical Diseases* 9 (7): e0003864.

Cary, Lynne Csiszar, Michael Goebel, Bartholomew G. Corsaro, Hwei-Gene Wang, Elliot Rosen, and M. J. Fraser. 1989. “Transposon Mutagenesis of Baculoviruses: Analysis of Trichoplusia Ni Transposon IFP2 Insertions within the FP-Locus of Nuclear Polyhedrosis Viruses.” *Virology* 172 (1): 156–69.

Coates, Craig J., Nijole Jasinskiene, Linda Miyashiro, and Anthony A. James. 1998. “Mariner Transposition and Transformation of the Yellow Fever Mosquito, Aedes Aegypti.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (7): 3748.

Curtis, C. F. 1968. “Possible Use of Translocations to Fix Desirable Genes in Insect Pest Populations.” *Nature* 218 (5139): 368–69.

Lohe, A. R., and D. L. Hartl. 1996. “Autoregulation of Mariner Transposase Activity by Overproduction and Dominant-Negative Complementation.” *Molecular Biology and Evolution* 13 (4): 549–55.

Day, Richard N., and Michael W. Davidson. 2009. “The Fluorescent Protein Palette: Tools for Cellular Imaging.” *Chemical Society Reviews* 38 (10): 2887–2921.

Gantz, Valentino M., and Ethan Bier. 2015. “Genome Editing. The Mutagenic Chain Reaction: A Method for Converting Heterozygous to Homozygous Mutations.” *Science* 348 (6233): 442–44.

Geurts, Aron M., Ying Yang, Karl J. Clark, Geyi Liu, Zongbin Cui, Adam J. Dupuy, Jason B. Bell, David A. Largaespada, and Perry B. Hackett. 2003. “Gene Transfer into Genomes of Human Cells by the Sleeping Beauty Transposon System.” *Molecular Therapy: The Journal of the American Society of Gene Therapy* 8 (1): 108–17.

Goriaux, Coline, Sophie Desset, Yoan Renaud, Chantal Vaury, and Emilie Brasset. 2014. "Transcriptional Properties and Splicing of the Flamenco piRNA Cluster." *EMBO Reports* 15 (4): 411–18.

Gregory, M., L. Alphey, N. I. Morrison, and S. M. Shimeld. 2016. "Insect Transformation with piggyBac: Getting the Number of Injections Just Right." *Insect Molecular Biology* 25 (3): 259–71.

Han, Michael. 2017. "The piRNA System in Aedes Aegypti." UC Riverside.

Handler, A. M., S. D. McCombs, M. J. Fraser, and S. H. Saul. 1998. "The Lepidopteran Transposon Vector, piggyBac, Mediates Germ-Line Transformation in the Mediterranean Fruit Fly." *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.95.13.7520.

Harris, Angela F., Andrew R. McKemey, Derric Nimmo, Zoe Curtis, Isaac Black, Siân A. Morgan, Marco Neira Oviedo, et al. 2012. "Successful Suppression of a Field Mosquito Population by Sustained Release of Engineered Male Mosquitoes." *Nature Biotechnology* 30 (September): 828.

Jansen, Cassie C., and Nigel W. Beebe. 2010. "The Dengue Vector Aedes Aegypti: What Comes next." *Microbes and Infection / Institut Pasteur* 12 (4): 272–79.

Jasinskiene, Nijole, Craig J. Coates, Mark Q. Benedict, Anthony J. Cornel, Cristina Salazar Rafferty, Anthony A. James, and Frank H. Collins. 1998. "Stable Transformation of the Yellow Fever Mosquito, Aedes Aegypti, with the Hermes Element from the Housefly." *Proceedings of the National Academy of Sciences of the United States of America* 95 (7): 3743.

Kistler, Kathryn E., Leslie B. Vosshall, and Benjamin J. Matthews. 2015. "Genome Engineering with CRISPR-Cas9 in the Mosquito Aedes Aegypti." *Cell Reports* 11 (1): 51–60.

Kokoza, V., A. Ahmed, E. A. Wimmer, and A. S. Raikhel. 2001. "Efficient Transformation of the Yellow Fever Mosquito Aedes Aegypti Using the piggyBac Transposable Element Vector pBac[3xP3-EGFP Afm]." *Insect Biochemistry and Molecular Biology*. https://doi.org/10.1016/s0965-1748(01)00120-5.

Li, Ming, Michelle Bui, Ting Yang, Christian S. Bowman, Bradley J. White, and Omar S. Akbari. 2017. "Germline Cas9 Expression Yields Highly Efficient Genome Engineering in a Major Worldwide Disease Vector, Aedes Aegypti." *Proceedings of the National Academy of Sciences of the United States of America* 114 (49): E10540–49.

Macias, V., J. Coleman, M. Bonizzoni, and A. A. James. 2014. “piRNA Pathway Gene Expression in the Malaria Vector Mosquito Anopheles Stephensi.” *Insect Molecular Biology* 23 (5): 579–86.

Malone, Colin D., Julius Brennecke, Monica Dus, Alexander Stark, W. Richard McCombie, Ravi Sachidanandam, and Gregory J. Hannon. 2009. “Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary.” *Cell* 137 (3): 522–35.

Miesen, Pascal, Erika Girardi, and Ronald P. van Rij. 2015. “Distinct Sets of PIWI Proteins Produce Arbovirus and Transposon-Derived piRNAs in Aedes Aegypti Mosquito Cells.” *Nucleic Acids Research* 43 (13): 6545–56.

Morazzani, Elaine M., Michael R. Wiley, Marta G. Murreddu, Zach N. Adelman, and Kevin M. Myles. 2012. “Production of Virus-Derived Ping-Pong-Dependent piRNA-like Small RNAs in the Mosquito Soma.” *PLoS Pathogens* 8 (1): e1002470.

Moyes, Catherine L., John Vontas, Ademir J. Martins, Lee Ching Ng, Sin Ying Koou, Isabelle Dusfour, Kamaraju Raghavendra, et al. 2017. “Contemporary Status of Insecticide Resistance in the Major Aedes Vectors of Arboviruses Infecting Humans.” *PLoS Neglected Tropical Diseases* 11 (7): e0005625.

Neafsey, Daniel E., Robert M. Waterhouse, Mohammad R. Abai, Sergey S. Aganezov, Max A. Alekseyev, James E. Allen, James Amon, et al. 2015. “Highly Evolvable Malaria Vectors: The Genomes of 16 Anopheles Mosquitoes.” *Science* 347 (6217): 1258522.

Nene, Vishvanath, Jennifer R. Wortman, Daniel Lawson, Brian Haas, Chinnappa Kodira, Zhijian Jake Tu, Brendan Loftus, et al. 2007. “Genome Sequence of Aedes Aegypti, a Major Arbovirus Vector.” *Science* 316 (5832): 1718–23.

Ochman, H., A. S. Gerber, and D. L. Hartl. 1988. “Genetic Applications of an Inverse Polymerase Chain Reaction.” *Genetics* 120 (3): 621.

Palavesam, Azhahianambi, Caroline Esnault, and David A. O’Brochta. 2013. “Post-Integration Silencing of piggyBac Transposable Elements in Aedes Aegypti.” *PloS One* 8 (7): e68454.

Pfaffl, Michael W. 2001. “A New Mathematical Model for Relative Quantification in Real-Time RT–PCR.” *Nucleic Acids Research* 29 (9): e45–e45.

Porse, Charsey Cole, Vicki Kramer, Melissa Hardstone Yoshimizu, Marco Metzger, Renjie Hu, Kerry Padgett, and Duc J. Vugia. 2015. “Public Health Response to Aedes Aegypti and Ae. Albopictus Mosquitoes Invading California, USA.” *Emerging Infectious Diseases* 21 (10): 1827–29.

Potter, Christopher J., and Liqun Luo. 2010. “Splinkerette PCR for Mapping Transposable Elements in Drosophila.” *PloS One* 5 (4): e10168.

Sarkar, Abhimanyu, Kurt Yardley, Peter W. Atkinson, Anthony A. James, and David A. O’brochta. 1997. “Transposition of the Hermes Element in Embryos of the Vector Mosquito, Aedes Aegypti.” *Insect Biochemistry and Molecular Biology* 27 (5): 359–63.

Sato, Kaoru, and Mikiko C. Siomi. 2018. “Two Distinct Transcriptional Controls Triggered by Nuclear Piwi-piRISCs in the Drosophila piRNA Pathway.” *Current Opinion in Structural Biology* 53 (December): 69–76.
Sethuraman, Nagaraja, Malcolm J. Fraser Jr, Paul Eggleston, and David A. O’Brochta. 2007. “Post-Integration Stability of piggyBac in Aedes Aegypti.” *Insect Biochemistry and Molecular Biology* 37 (9): 941–51.

Shah, Presha Vijaykumar. 2015. *Mobilization of Newly Identified Transposon Muta1 in Aedes Aegypti and Drosophila Melanogaster*. University of California, Riverside.

Shyamala, Venkatakrishna, and Giovanna Ferro-Luzzi Ames. 1989. “Genome Walking by Single-Specific-Primer Polymerase Chain Reaction: SSP-PCR.” *Gene* 84 (1): 1–8.
Sienski, Grzegorz, Derya Dönertas, and Julius Brennecke. 2012. “Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression.” *Cell* 151 (5): 964–80.

Tabachnick, Walter J. 1991. “Evolutionary Genetics and Arthropod-Borne Disease: The Yellow Fever Mosquito.” *American Entomologist* 37 (1): 14–26.

Timoshevskiy, Vladimir A., Nicholas A. Kinney, Becky S. deBruyn, Chunhong Mao, Zhijian Tu, David W. Severson, Igor V. Sharakhov, and Maria V. Sharakhova. 2014. “Genomic Composition and Evolution of Aedes Aegyptichromosomes Revealed by the Analysis of Physically Mapped Supercontigs.” *BMC Biology* 12 (1): 27.

Weaver, Scott C., Federico Costa, Mariano A. Garcia-Blanco, Albert I. Ko, Guilherme S. Ribeiro, George Saade, Pei-Yong Shi, and Nikos Vasilakis. 2016. “Zika Virus: History, Emergence, Biology, and Prospects for Control.” *Antiviral Research* 130 (June): 69–80.

Wright, Jennifer A., Ryan C. Smith, Xianghong Li, Nancy L. Craig, and Peter W. Atkinson. 2013. “IPB7 Transposase Behavior in Drosophila Melanogaster and Aedes Aegypti.” *Insect Biochemistry and Molecular Biology* 43 (10): 899–906.

4.6 Figures and Tables

Injection date	Embryos injected	G_0 adult emergence	Emergence percent	*piggyBac* lines	*piggyBac* transformation rate	*Muta1* lines
11/2016	n=1,602	n=216	13.4%	n=2	0.9%	0
05/2017	n=2,299	n=339	14.7%	n=3	0.9%	0
10/2017	n=2,233	n=323	14.4%	n=2	0.6%	0

Table 4.1 – Data from the *Ae. aegypti* embryo microinjections performed at the Insect Transformation Facility at the University of Maryland Institute for Bioscience and Biotechnology Research

Injection date	Embryos injected	G_0 adult emergence	Emergence percent	G_0 adults with progeny screened	*piggyBac* lines	*piggyBac* transformation rate
1/2018 to 3/2018	n=2,632	n=629	23.9%	n=526	n=15	2.8%

Table 4.2 – Data from the *Ae. aegypti* embryo microinjections for *piggyBac* transformation performed in Dr. Peter Atkinson's laboratory.

Injection date	Embryos injected	G_0 adult emergence	Emergence percent	G_0 adults with progeny screened	*Muta1* lines	*Muta1* transformation rate
3/2018 to 5/2018	n=5,830	n=1,211	20.8%	N=847	n=1*	0.1%

Table 4.3 – Data from the *Ae. aegypti* embryo microinjections for *Muta1* transformation performed in Dr. Peter Atkinson's laboratory.

Helper Line	Transformation	Plasmid	Phenotype	# of Integrations	TSD	Integration Location	Features
A	*piggyBac*	pBac3CexuMuta1	Canonical 3PX3-ECFP expression plus anal papillae	1	TTAA	Chr1 @base 187244235+	Intergenic region - insertInserted into last exon, E5, of AAEL027751 (from 187243856 - 187246068), putative tubulin polyglutamylase
C	*piggyBac*	pBac3CexuMuta1	Canonical 3PX3-ECFP expression	1	TTAA	Chr1 @base 40501467-	Upstream ~6kb is AAEL014196
G	*piggyBac*	pBac3CexuMuta1	Ubiquitous ECFP expression				
J	*piggyBac*	pBac3CexuMuta1	Canonical 3PX3-ECFP expression plus anal papillae	2	TTAA	Chr3 @ base 25910148-	Inserted 39 bases into 5'UTR E1 of AAEL019562 (5'UTR 25,910,109 to 25,911,232)
					TTAA	Chr3 @ base 365751941-	Inserted into a 5853bp Bel element within the intron 2 3 of AAEL004348 (1160aa protein unk function)
M	*piggyBac*	pBac3CexuMuta1	Canonical 3PX3-ECFP expression				
N	*piggyBac*	pBac3CexuMuta1	Canonical 3PX3-ECFP expression plus anal papillae	1	TTAA	Chr2 @ base 278006719-	Upstream ~30kb is AAEL011071, possible snrnp
Q	*piggyBac*	pBac3CexuMuta1	Canonical 3PX3-ECFP expression	2	TTAA	Chr3 @ base 235166441-	Inserted into a 265bp Sola element within the first intron of AAEL017395 (309aa protein unk function, putative mitochondrial proten)
					TTAA	Chr3 @ base 238105159-	Intergenic region - 309kb downstream of AAEL013518 (341aa protein unk function)
R	*piggyBac*	pBac3CexuMuta1	Canonical 3PX3-ECFP expression				
T	*piggyBac*	pBac3CexuMuta1	Canonical 3PX3-ECFP expression				

Table 4.4 – Integration data from the *piggyBac* transformed *Ae. aegypti* lines produced in Dr. Peter Atkinsons lab.

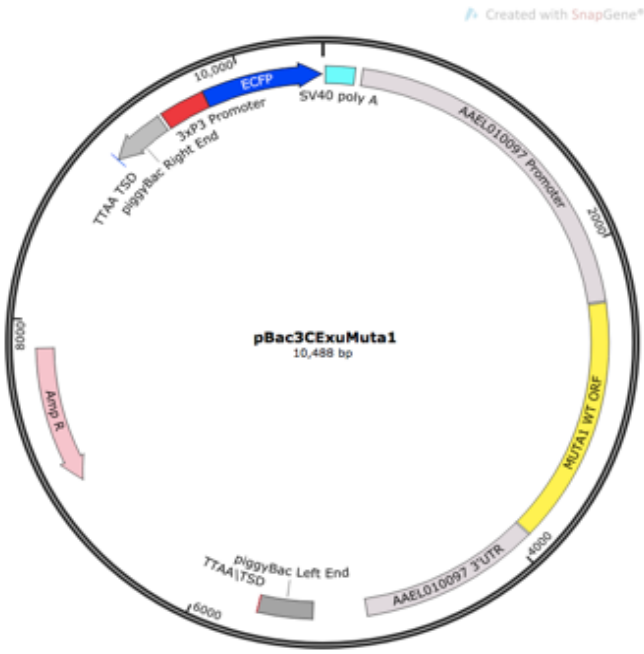

Figure 4.1 – The *Muta1* helper plasmid containing the *Muta1* coding sequence, flanked by *Aedes Exuperantia* homolog 5' and 3' regulatory sequence and selectable ECFP eye marker. Minimal *piggyBac* left and right ends flank the transgenes.

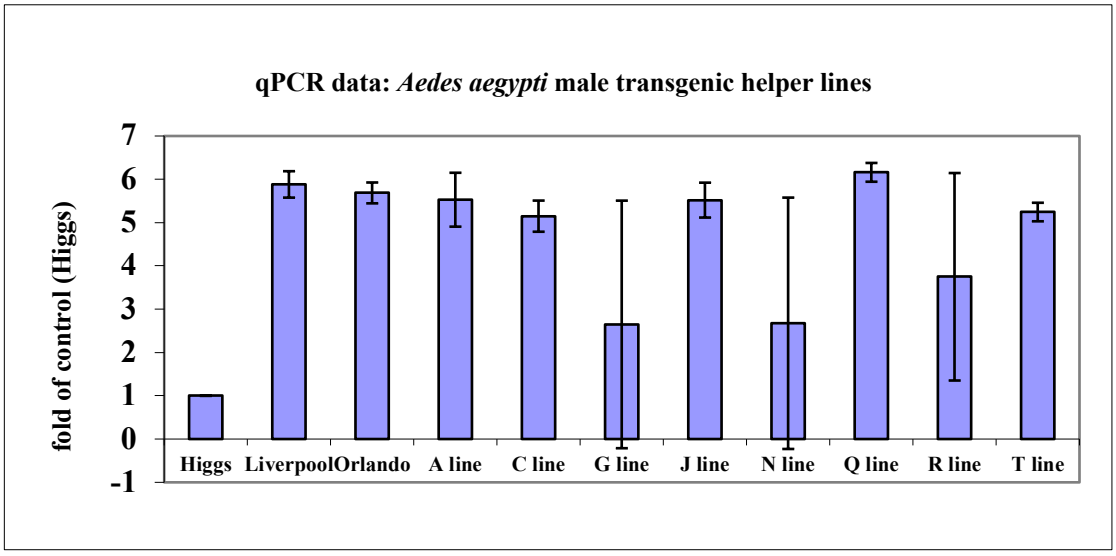

Figure 4.2 – Quantitative RT-PCR expression data in adult males. The *Muta1* expression data is compared to the Higgs background strain.

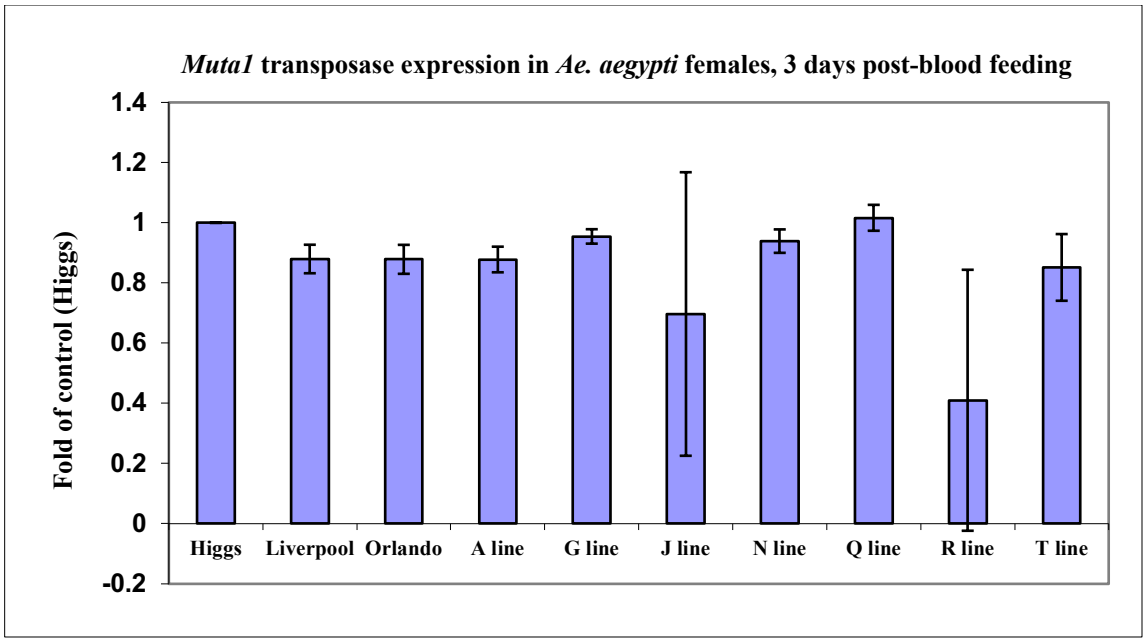

Figure 4.3 – Quantitative RT-PCR expression data in adult females. The *Muta1* expression data is compared to the Higgs background strain.

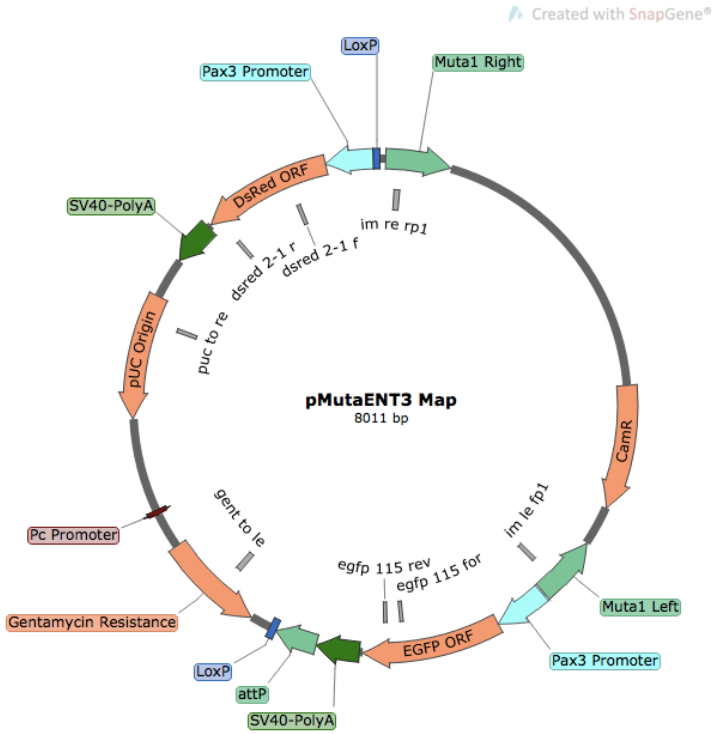

Figure 4.4 – The *Muta1* enhancer trap plasmid, pMutaENT3, with primers labeled for the molecular verification of the putative *Muta1* transformed mosquitoes from I Line ENT3.

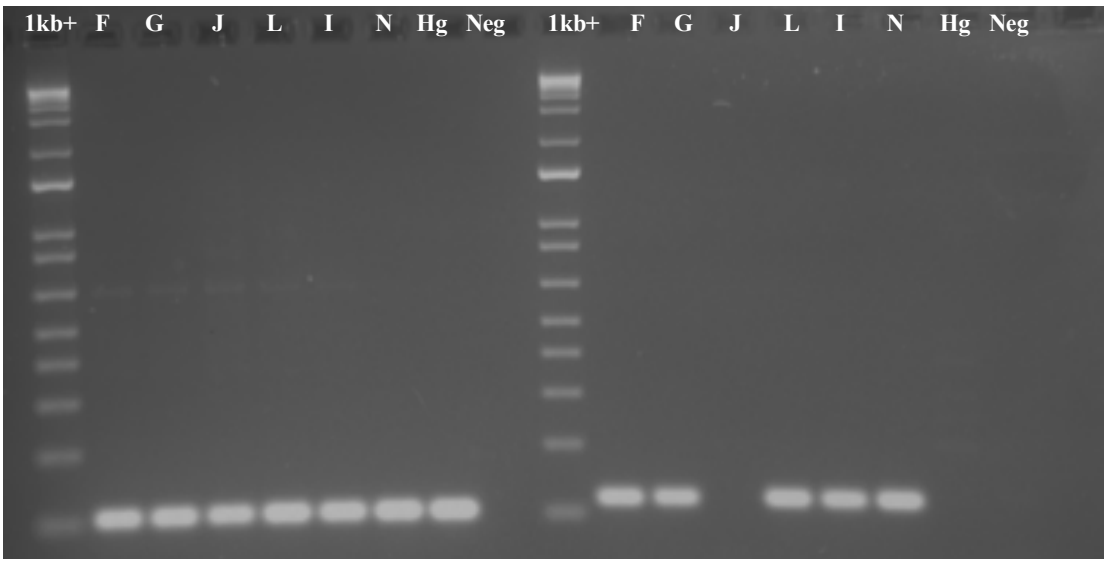

Figure 4.5 – The agarose gel on the left shows the PCR products using primers RPS7 New For and RPS7 New Rev (107bp) and genomic DNA from 6 putative transgenic lines and the Higgs background strain.
The agarose gel on the right shows the PCR products using primers EGFP 115 F and EGFP 115 R (115bp) and genomic DNA from 6 putative transgenic lines and the Higgs background strain.

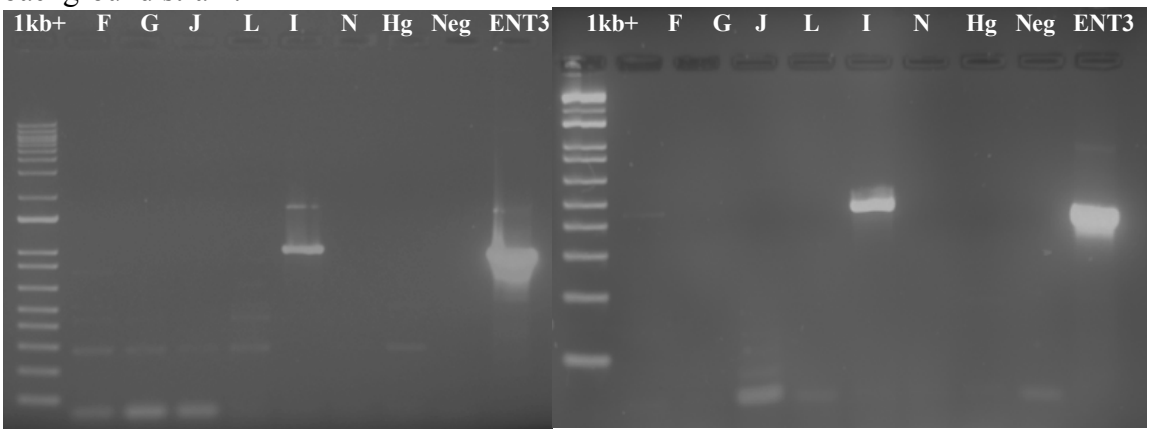

Figure 4.6 – The agarose gel on the left shows the PCR products using primers IM LE FP1 and EGFP 115 R (933bp) and genomic DNA from 6 putative transgenic lines and the Higgs background strain. The positive control template used a dilution of the pMutaENT3 plasmid.
The agarose gel on the right shows the PCR products using primers DsRed2-1 For and DsRed2-1 Rev (441bp) and genomic DNA from 6 putative transgenic lines and the Higgs background strain. The positive control template used a dilution of the pMutaENT3 plasmid.
In both gels, only the positive control pMutaENT3 plasmid and genomic DNA from Line I produce the correct size PCR fragments.

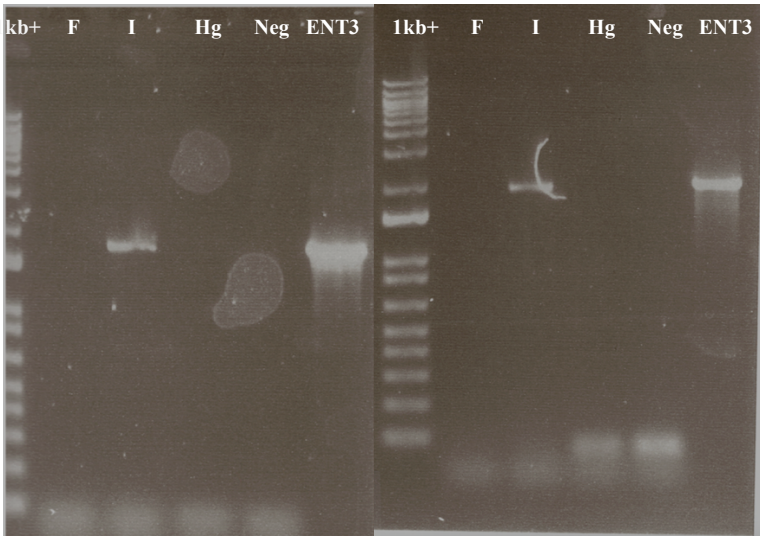

Figure 4.7 – The agarose gel on the left shows the PCR products using primers IM RE RP1 and pUC to RE (1572bp) and genomic DNA from 2 putative transgenic lines and the Higgs background strain. The positive control template used a dilution of the pMutaENT3 plasmid.
The agarose gel on the right shows the PCR products using primers IM LE FP1 and LE to Gent (1939bp) and genomic DNA from 2 putative transgenic lines and the Higgs background strain. The positive control template used a dilution of the pMutaENT3 plasmid.
In both gels, only the positive control pMutaENT3 plasmid and genomic DNA from Line I produce the correct size PCR fragments.

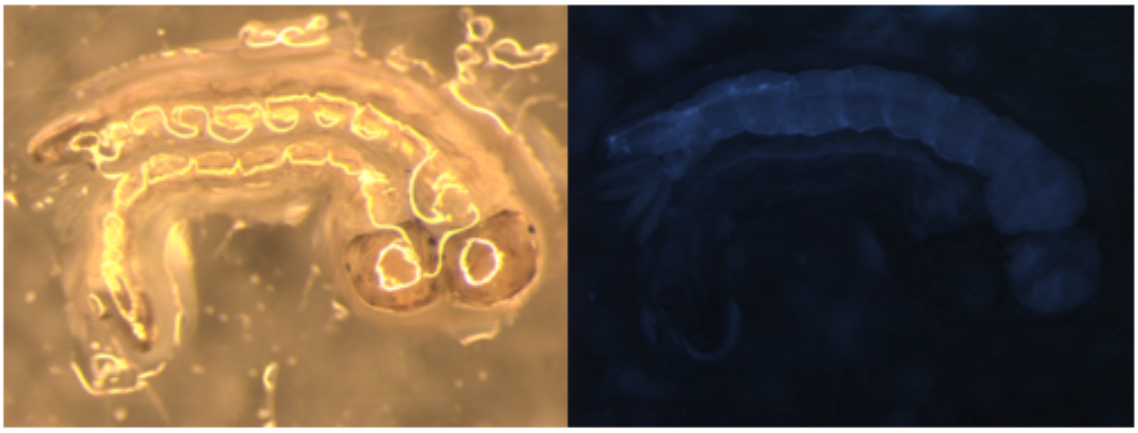

Figure 4.8 – *Aedes aegypti* larvae: The top larva is from the pMutaENT3 injected I Line that displays faint ubiquitous ECFP expression. The bottom larva is from the wild type background strain and shows no fluorescent phenotype.

Chapter 5 – A *Muta1* transposon-based enhancer trap system in *Anopheles coluzzii* and *Anopheles stephensi*

5.1 Introduction

5.1.1 *Anopheles stephensi and Anopheles coluzzii* are major global vectors of malaria

Anopheles mosquito species are widespread across the globe. The genus includes almost 500 described species, and close to 100 of these species can vector malaria parasites that cause human disease (Sinka *et al.,* 2012). Malaria is a life-threatening disease caused by five parasites of the genus *Plasmodium*, which infect humans, as well as a few other vertebrate species (Molina-Cruz, *et al.,* 2016). Despite coordinated global efforts, the World Health Organization reported over 219 million cases of malaria in 2017, and an estimated 435,000 deaths, mostly in children under five years old. The major burden of malaria rests on India and 15 different countries in sub-Saharan Africa, where *Plasmodium falciparum* is the predominant parasite (WHO). Members of the *An. gambiae* complex, including *An. gambiae, An. coluzzii,* and *An. arabiensis,* plus *An. funestus*, are the main malaria vectors in Africa, while *An. stephensi* is the main malaria vector in India and Western Asia (Sinka *et al.,* 2012). *Anopheles* species also transmit the roundworms that cause lymphatic filariasis, which is a neglected tropical disease that is predicted to effect 120 million people a year in 73 different countries (Rebollo and Bockarie, 2014).

The *Anopheles gambiae* complex refers to several highly genetically introgressed and morphologically indistinguishable *Anopheles* strains that were once thought to be a

single species (Davidson, 1962, Fontaine *et al.,* 2015). There are difficulties in differentiating between recent speciation events and genetic hybridization (Wolf and Ellegren, 2017), but recent genetic analysis has shown that the so called "S" and "M" forms of the *Anopheles gambiae* mosquito are two separate species, now known as *An. gambiae* and *An. coluzzii*, respectively, and they have likely separated into different clades only ~500,000 years ago (Coetzee *et al.,* 2013, Fontaine *et al.,* 2015).

Just like the *Ae. aegypti* mosquito, traditional vector control methods for controlling *Anopheles* mosquito populations and decreasing disease transmission are varied and complex. Environmental control, outdoor use of pesticides, indoor residual insecticide spraying, and the use of insecticide treated bed nets are all employed in concert in areas at risk of malaria transmission (WHO). Increasing insecticide resistance in mosquito populations is a constant challenge as this trait is positively selected for after exposure. A buildup of resistance to the primary insecticide used in treated bed nets, pyrethroids, has recently been observed in all *Anopheles* species surveyed across Africa (Ranson and Lissenden, 2016). Additionally, *Plasmodium falciparum*, which is the most prominent and deadly of the five malaria-causing *Plasmodium* species, is also becoming increasingly resistant to the artemisinin class of drugs that are used as a first-line treatment to malaria parasite infection (Blasco *et al.,* 2017). *Plasmodium falciparum* resistance to artemisinin combination therapies (ACTs) was first reported in 2008 and has since become a new complication in the fight against malaria (Ouji *et al.,* 2018).

Just as with *Aedes* mosquito species, the increasing prevalence of pyrethroid and insecticide resistance in *Anopheline* malaria vectors has driven the search for conceivable

genetic methods of mosquito population control. The development of targeted of method has made the possibility of gene drive within a species a realistic ambition (Gantz and Bier, 2015). In recent research, laboratory strains of *An. stephensi* have been modified to express an anti-malaria antibody cassette using CRISPR/Cas9 gene drive technology (Gantz *et al.,* 2015), and CRISPR/Cas9 gene drive has been used for targeting reproductive fitness and sex determination and *An. gambiae* laboratory strains (Hammond *et al.,* 2016, Kyrou *et al.,* 2018). However, it has been noted that, particularly within the highly genetically introgressed *Gambiae* complex of *Anopheles*, the occurrence of genetic variation and selective evolutionary pressures add an extra layer of complexity when trying to translate gene-drive research from the laboratory to field applications (Fouet *et al.,* 2018).

5.1.2 Genomic content and piRNA pathway in *An. coluzzii* and *An. stephensi*

The genomes of sequenced *Anopheles* species have a mean size of 275 Mb, not much larger than the 144 Mb genome of *D. melanogaster* (Smith *et al,* 2007, Neafsey *et al.,* 2013). The *An. stephensi* genome is about 221 Mb and transposable elements make up 7.1% of the annotated genome. Most annotated transposon sequences reside in heterochromatic regions (Jiang *et al.,* 2014), which is also the case for *An. gambiae* (Holt *et al.,* 2002). Phylogenetic analysis has shown that both *An. coluzzii* and *An. stephensi* have a one-to-one ratio of PIWI orthologs compared to *D. melanogaster*, and do not display the evolutionary expansion of PIWIs observed in *Ae. aegypti* (Macias *et al.,* 2014). However, *An. coluzzii* and *An. stephensi* are not major arboviral vectors. The

relationship between an arbovirus and its vector host is complex – an arthropod must be able to control arboviral infection at a minimal fitness cost while the arbovirus must be able to replicate and maintain a stable infection long enough to be passed along to another host. Similar to *Aedes*, the West Nile Fever virus vector, *Culex pipiens,* also has an expansion of PIWI proteins, containing seven PIWI orthologs (Olsen and Blair, 2015).

The culicine mosquitoes, including *Culex* and *Aedes*, are critical arbovirus vectors compared to anopheline mosquitoes, which are only known to transmit the virus responsible for o'nyong-nyong fever (Keene, *et al.,* 2004, Waldock *et al.,* 2012). Recent research demonstrated that the exogenous siRNA antiviral pathway only conferred limited protective effects in o'nyong-nyong infected *An. gambiae* females (Carissimo *et al.,* 2015). The evolutionary difference in the piRNA pathway makeup between culicines and anophelines may be due to the difference in arbovectoral capacity between these two mosquito genera. Indeed, the mosquito innate immune response, particularly in *Anopheles gambiae* females, has been well studied in response to *Plasmodium* infection, and the primary immune response is controlled by the Toll and Imd cell signaling pathways that are able to recognize "pathogen associated molecular patterns" (PAMPs) and subsequently initiate the release of anti-pathogen peptides (Clayton *et al.,* 2014). There does not appear to be a link between the piRNA pathway and parasitic infection in *Anopheles* mosquitoes.

5.1.3 Germline transformation and forward genetic tools in *An. coluzzii* and *An. stephensi*

Transposable elements have been used for transformation and transgene delivery in *An. coluzzii* and *An. stephensi* for almost twenty years (Catteruccia *et al.,* 2000). *An. stephensi* has been transformed by the *Minos Mariner* family element from *Drosophila hydei, piggyBac* from *Trichoplusia ni* (Catteruccia *et al.,* 2000, Nolan *et al.,* 2002). The *An. gambiae* mosquito has only been transformed by the *piggyBac* element (Grossman *et al.,* 2001). *An. gambiae* is also known to have an active *hAT* element, *Herves*, that can transform *D. melanogaster*, however, the possibility of germline remobilization of *Herves* within its host is unknown (Arensburger *et al.,* 2005).

Although it has been almost twenty years since *An. stephensi* was first transformed by a transposon, only the *piggyBac* element has been successfully deployed for enhancer trapping experiments in *An. stephensi* (O'Brochta *et al.,* 2011, O'Brochta *et al.,* 2012, Reid *et al.,* 2018). While the *Minos* element was one of the first to be used to transform *An.* stephensi, it does not appear to be capable of germline remobilization in the mosquito (Scali *et al.,* 2007). The first experiments using *piggyBac* mediated enhancer detection recovered germline remobilization frequencies of up to 6% in *An. stephensi*, with a greater frequency of remobilization originating from transgenic females versus males (O'Brochta *et al.,* 2011, O'Brochta *et al.,* 2012). A synthetic autonomous *piggyBac* element that utilizes the germline promoter *nanos* has been employed for enhancer detection in *An. stephensi*, however, this experimental system showed low germline mobility, even with the use of the *nanos* regulatory sequence (Macias *et al.,*

2018). While the *piggyBac* element has provided a foundation for the creation of forward genetic screens in *An. stephensi*, no other element has been able to be deployed for trapping experiments. Additionally, there is no literature showing successful enhancer trapping lines in *An. coluzzii* with any available transposon. Similar to *Ae. aegypti,* there is a need for additional forwards genetics and functional genomics tools in *Anopheles*, particularly with respect to increasing the repertoire of knowledge on promoters and gene regulation (O'brochta *et al.,* 2012).

5.1.4 Chapter aims

The goal of the research described in this chapter was to generate stable transgenic enhancer trap lines in the major malaria vectors *An. coluzzii* and *An. stephensi* using the *Muta1* transposable element. To date, only the *piggyBac* transposon, which sees the widest use in insects, has been used to generate enhancer trap lines in *An.* stephensi. No element has been successfully used for trapping experiments in *An. gambiae* or any other *Anopheles* species. Based on the results of the *D. melanogaster* enhancer trap crosses described in Chapter 3, which exhibited a low germline remobilization frequency, I reasoned that using a germline specific promoter, instead of a heat-shock inducible promoter, may increase the likelihood of germline remobilization events. I therefore constructed an *Anopheles* helper plasmid using the *An. gambiae vasa* regulatory sequence (Papathanos *et al.,* 2009), which is active in female ovaries and male testis, to drive expression of *Muta1* transposase in germline tissues. To generate transgenic *Anopheles* lines, embryo microinjections were performed at the Insect

Transformation Facility at the University of Maryland Institute for Bioscience and Biotechnology Research. I reared G_1 eggs from enhancer trap crosses and screened transgenic progeny as larvae, pupae and adults to identify novel fluorescent expression patterns as evidence of *Muta1* remobilization. I used a combination of inverse PCR and Splinkerette PCR to characterize transgene integrations. I used qRT-PCR to verify the level of *Muta1* transposase expression in transgenic helper lines. *Muta1* transgenic lines were successfully created for both *Anopheles* species.

5.2 Materials and methods

5.2.1 Plasmid construction

Construction of pBac3ChspMuta1SMALL (Figure 5.3): The plasmid pBac3ChspMuta1SMALL (Figure 5.2) was used as template for a vector sequence that contained ampicillin resistance and an *E. coli* origin of replication and minimal length *piggyBac* ends. To remove extra sequence and create minimal *piggyBac* arms, two PCR fragments were amplified from pBac3ChspMuta (Figure 3.2). Before PCR, pBac3ChspMuta was digested with restriction enzymes to avoid carryover of the plasmid into subsequent reactions. For the first fragment, pBac3ChspMuta1 was digested with PvuI (ThermoFisher Fast Digest) and an 8305bp fragment was gel purified with Qiagen QIAquick Gel extraction kit following the manufacturer's protocol.

The gel purified fragment was used a template to create a 3286bp PCR fragment using the following primers.

Helper NheI For:

5’ – TACGCTAGCTCTAGAGATCGCACGGTTAATTCGAG – 3’

Helper SnaBI Rev:

5’ – TCATACGTACTCGAGGTATAAGTTCGAGATCGGCC – 3’

New England Biolabs Q5 High-Fidelity DNA Polymerase was used following manufacturer protocol for 50ul reactions.

For the second fragment, pBac3ChspMuta1 was digested with AvrII and NotI (ThermoFisher Fast Digest) and an 6962bp fragment was gel purified with Qiagen QIAquick Gel extraction kit following the manufacturer’s protocol.

The gel purified fragment was used a template to create a 4243bp PCR fragment using the following primers.

PBac Ends SnaBI For:

5’ – TCATACGTAACTAGTCTCATCGTCTAAAGAACTACC – 3’

PBac Ends NheI Rev:

5’ – TACGCTAGCACTAGTCCAAAGTTGTTTCTGACTGA – 3’

New England Biolabs Q5 High-Fidelity DNA Polymerase was used following manufacturer protocol for 50ul reactions.

Both of the PCR products were purified using the Qiagen QIAquick PCR purification kit according to the manufacturer’s protocol and quantified. The purified PCR fragments were digested with restriction enzymes NheI and SnaBI (Thermo Fisher Fast Digest) and purified again using the Qiagen QIAquick PCR purification kit according to the manufacturer’s protocol. The two fragments were ligated using Thermo Scientific T4 DNA Ligase following the manufacturer protocol.

Construction of pBac3CAnVasMuta1 (Figure 5.3): The plasmid Vas2inpCR4BluntTopo-Maxi1 (Figure 5.1) containing the *An. gambiae vasa* promoter sequence (Papathanos *et al.,* 2009) was a kind gift from Robert Harrell of University of Maryland Institute for Bioscience and Biotechnology Research. This plasmid was used as template to amplify the *An. gambiae vasa* regulatory sequence using the primers below.

Vasa M13 For:

5' – CAGTCACGACGTTGTAAAACG – 3'

Vasa M13 Rev:

5' – CAGGAAACAGCTATGACCATG – 3'

The plasmid pBac3ChspMuta1SMALL was used as template to amplify a large vector fragment excluding the HSP70 promoter sequence using the primers below.

Helper Assembly For:

5' – GGTCATAGCTGTTTCCTGACAATCTGCAGACTAGTATGG – 3'

Helper Assembly Rev:

5' – GCTGTACAAGTAAAGCGGC CGTTTTACAACGTCGTG – 3'

For all PCR reactions, New England Biolabs Q5 High-Fidelity DNA Polymerase was used following manufacturer protocol for 50ul reactions.

The two PCR fragments were assembled using the GeneArt Seamless Cloning and Assembly Kit (Life Technologies), transferred to TOP10 cells, and plated onto LB agar containing ampicillin.

5.2.2 *An. coluzzii* and *An. stephensi* rearing and crosses

The *An. coluzzii* and *An. stephensi* colonies were all reared under standard conditions following the same protocol used by the Insect Transformation Facility at the University of Maryland (O'brochta *et al.,* 2011), with a modified membrane blood-feeding protocol. For remobilization crosses, 20 females from the enhancer trap line were crossed to 20 males of the helper line. The reciprocal crosses were performed as well. The heterozygous progeny was outcrossed *en masse* to the wildtype background strains, SDA-500 for *An. stephensi,* and NGS for *An. coluzzii* (Table 5.3). The resulting progeny were counted and screened in the larval, pupal, and adult stages for novel fluorescent phenotypes.

5.2.3 Molecular verification of transgenic lines

The Qiagen DNeasy Blood and Tissue kit was used for genomic DNA extracts. Qiagen provides a modified protocol for genomic DNA extraction from insects and this protocol was followed as written.

In preparation for the inverse PCR reactions for *piggyBac* integrations, genomic DNA was digested with BamHI and BstYI (ThermoFisher Fast Digest enzymes). Digested genomic DNA was column purified with the Qiaquick PCR Purification Kit, following kit protocol. Purified products were self-ligated with T4 DNA Ligase (Thermo Fisher Scientific) and chloroform extracted following the ThermoFisher kit protocol. Inverse PCR reactions used Bioneer AccuPower PCR PreMix tubes and the following primers were used for *piggyBac* integrations:

PB RE RP1:

5’-CAACATGACTGTTTTTAAAGTACAAA-3’

PB RE FP1:

5’-GTCAGAAACAACTTTGGCACATATC-3’

PB RE RP2:

5’-CCTCGATATACAGACCGATAAAAC-3’

PB RE FP2:

5’-TGCATTTGCCTTTCGCCTTAT-3’

Splinkerette PCR (Potter and Luo, 2010) was used to characterize *Muta1* integrations in *Anopheles* lines and crosses. Genomic DNA digests used BstYI (ThermoFisher Fast Digest enzymes). For PCR reactions, New England Biolabs Q5 High-Fidelity DNA Polymerase was used following manufacturer protocol for 50ul reactions. The following primers were used for mapping *Muta1* insertions.

Muta1 splnk RE 1:

5’ – GGCAATTGCAGCTCACTTAGGATCAAAATC – 3’

Muta1 splnk RE 2:

5’ – CTTTTCATTGACTCATGTGAACAACGGTAACA – 3’

Muta1 splnk LE 1:

5’ – CCTCGAGCATTTCCAGCTTCGTAGTACAAATATC – 3’

Muta1 splnk LE 2:

5’ – CAGTGACTAATGGATCGTGACAAAGTGACCC – 3’

Quantitative RT-PCR was used to assess the levels of *Muta1* transposase expression in the transgenic lines. Three biological replicates were used for each sample, male adults or 48-hour post-blood fed female adults, and each biological replicate comprised three adults (note that data for males and females of line F3B represent two biological replicates). This was to prevent variation between samples due to simple stochastic differences between individual adults. All adults were between four and eight days old. RNA from each biological replicate was extracted using ThermoFisher TRIzol reagent following manufacturer protocol for RNA isolation from 50-100mg tissue samples. The RNA samples were treated with the ThermoFisher TURBO DNA-free Kit according to manufacturer protocol, with a modified longer incubation period of 1 hour. All RNA samples were diluted to 50ng/ul stocks before cDNA synthesis using the New England Biolabs ProtoscriptII First Strand cDNA Synthesis Kit, using half-reactions. All cDNA for each biological replicate was diluted to 100ng/ul stocks for subsequent qRT-PCR reactions.

Quantitative RT-PCR was performed using Bio-Rad iQ SYBR Green SuperMix following protocol for half-volume reactions. Thermocycling reactions were performed on a Bio-Rad MyiQ Detection system with the annealing temperatures for each primer set listed below.

For the *An. stephensi RPS7* housekeeping gene, a standard curve was generated using 400nM primer concentrations and an annealing temperature of 60ºC. The standard curve for this primer set had an E value of 98.6 and an R^2 value of 0.991.

Ste 58-RPS7 For:

5’ – GAGGTTGTCGGCAAGCGTATC – 3’

Ste 58-RPS7 Rev:

5’ – CGATGGTGGTCTGCTGGTTC – 3’

For the *Muta1* transgene in *An. stephensi* and *An. coluzzii*, a standard curve was generated using 400nM primer concentrations and an annealing temperature of 57ºC. The standard curve for this primer set had an E value of 95.3 and an R^2 value of 0.996.

Muta1 2015 For:

5’ – GCGTATGGTAACGTTCAAGGC – 3’

Muta1 2015 Rev:

5’ – GTACTATTTTCGCTGGCGTTG – 3’

All quantitative RT-PCR data was analyzed using the Pfaffl equation (Pfaffl, 2001).

5.3 Results

5.3.1 Transgenic helper lines were established in *An. coluzzii* and *An. stephensi*

Embryo microinjections in *An. stephensi* and *An. coluzzii* were performed at the Insect Transformation Facility at the University of Maryland Institute for Bioscience and Biotechnology Research. These injection sets used the *Muta1* enhancer trap plasmid, pMutaETS (Figure 3.3), the *Anopheles Muta1* helper plasmid, pBac3CAnVasMuta1 (Figure 5.2), and the *piggyBac* helper plasmid phsp70pBac. The surviving G_0 adults were crossed en masse and G_1 larvae were initially screened for fluorescence at the insect transformation facility. The identified transgenic lines were sent to our laboratory as embryos and were subsequently homozygosed by selecting for fluorescent larvae every

generation. The lines were homozygosed, which took approximately five to six generations, before characterizing *piggyBac* and *Muta1* construct integrations. One *piggyBac* transformed helper line was recovered in *An. coluzzii,* called M2B (Table 5.1), and one *piggyBac* transformed helper line was recovered in *An. stephensi,* called F3B (Table 5.2). In the line F3B, a single perfect insertion of the pBac3CAnVasMuta1 *piggyBac* construct was recovered and showed an integration location immediately upstream (>200bp) from the 5' UTR of ASTI03495, a homolog of *D. melanogaster deltex,* which is a regulator of notch signaling. In the line M2B, a single insertion was recovered in an intergenic region within a truncated (522bp) *Mariner*-like transposon. Both of lines, M2B and F3B, contained only a single insertion of the helper construct.

5.3.2 Transgenic *An. coluzzii* and *An. stephensi* lines showed in increase in *Muta1* transposase expression

Quantitative RT-PCR was used to quantify levels of *Muta1* transposase in adult males and 48-hour post blood-fed females for the transgenic lines and for the background strains, as a point of reference. In the *An. coluzzii* helper line, M2B, males showed up to a 7-fold increase in *Muta1* expression (Figure 5.6) compared to the NGS background strain, and two-day post-blood fed females showed almost a 7-fold increase in *Muta1* transposase expression (Figure 5.5). In the *An. stephensi* helper line, F3B, males showed about a 4.5-fold increase in *Muta1* expression (Figure 5.8), compared to the NGS background strain, and two-day post-blood fed females showed almost a 5-fold increase

in *Muta1* transposase expression (Figure 5.9). Both of lines, M2B and F3B, contained only a single insertion of the helper construct.

5.3.3 Transgenic *Muta1* enhancer trap lines were established in *An. coluzzii* and *An. stephensi*

Three *Muta1* transformed enhancer trap lines were recovered including two lines in *An. stephensi,* called MIRG and F3RG (Table 5.2), and one line in *An. coluzzii,* called ENT3 (Table 5.1). The *An. coluzzii* line ENT3 has a single insertion of the enhancer trap construct within an intergenic region, downstream (<200bp) of ACOM041868. The *An. stephensi* line F3RG has a single insertion of the enhancer trap construct within an intergenic region upstream (>1000bp) upstream of ASTEI10162, which is an uncharacterized ABC-like transporter, and (>800bp) downstream of an uncharacterized protein ASTEI10161, which contains two ubiquitin-associated domains. The *An. stephensi* line M1RG contains two integrations of the enhancer trap construct. One insertion is immediately downstream (7bp) of ASTEI06582, which is a Piezo-type mechanosensitive ion-channel component protein, and the other insertion is within a 381-base-pair *Gypsy*-like element.

5.3.4. *Muta1* remobilizes in *An. coluzzii*

One *Muta1* remobilization event was recovered after screening experiments in *An. coluzzii* (Table 5.7). Progeny with the novel fluorescent phenotype showed brighter EGFP expression near the larval eyes and two bright areas of both DsRed and EGFP

expression near the larval anal papillae (Figure 5.4). This phenotype was carried into the pupal stage with an increased EGFP signal from the pupal head (Figure 5.4). The new insertion location was within a different contig compared to the parental line, ENT3. The fluorescent progeny that contained the novel insertion also retained the same parental integration at the same location. The new enhancer trap insertion was in an intergenic region, upstream (<300bp) of an uncharacterized protein, ACOM027426, and downstream (<600bp) of protein ACOM027425, which is a homolog of *D. melanogaster jing*. Due to the direction of transcription, factors influencing expression of ACOM027425 could also affect transcription of the enhancer trap fluorescent markers. The *D. melanogaster* gene, *jing,* encodes a C2H2-zinc finger containing transcription factor protein with transcript expression during embryonic development and within adult mushroom bodies, which are structures in arthropod brains (Strausfeld *et al.,* 1995). The uncharacterized protein, ACOM027426, which contains a G-patch domain, is transcribed on the reverse strand, and could also impact transcription of the enhancer trap fluorescent markers. A protein-protein BLAST search of ACOM027426 has no close hits in *D. melanogaster*, but has several related hits in other mosquito species, including an *Ae. aegypti* gene that is implicated as an NF-kappa-β-repressing factor, AAEL011738. The transcript is expressed throughout development with a strong peak of expression in the 1st larval stage (Akbari *et al.,* 2013).

5.4 Discussion

This chapter presents the first record of germline transformation of *An. stephensi* and *An. coluzzii* with a *MULE* family element, *Muta1* from the genome of *Ae. aegypti*. This is also the first instance of transposon transformation in *An. gambiae* with a transposon other than the *piggyBac* element (Grossman *et al.,* 2001). The *piggyBac* element is well established as an effective transgene delivery tool in various arthropod species (Gregory *et al.,* 2016), and the data presented in this chapter indicate that *Muta1* can serve as an alternative to *piggyBac*, particularly in *Anopheles.* Like the *piggyBac* element, the *Muta1* element can excise precisely, without leaving a target site duplication footprint behind in the host genome (Wilson, *et al.,* 2007, Liu and Wessler, 2017).

In the *An. stephensi* line F3B, a single pBac3CAnVasMuta1 insertion was recovered and showed an integration location immediately upstream (>200bp) from the 5' UTR of ASTI03495, a homolog of *D. melanogaster deltex,* which is a regulator of notch signaling. In the line M2B, a single insertion was recovered in an intergenic region within a truncated (522bp) *Mariner*-like transposon. Previous research has shown that the *piggyBac* element shows a non-random insertion site preference near transcriptional start sites and within long terminal repeats (Wilson *et al.,* 2007, Li *et al.,* 2013). The F3B line insertion 5' to *deltex* agrees with reported patterns of *piggyBac* integrations. I have not found literature showing that *piggyBac* preferentially integrates into DNA transposon derived sequences, as is the case for the M2B insertion.

Three *Muta1* transformed enhancer trap lines were recovered including two lines in *An. stephensi,* called MIRG and F3RG (Table 5.1), and one line in *An. coluzzii,* called

ENT3 (Table 5.2). One *Muta1* remobilization event was recovered after screening experiments in *An. coluzzii* (Table 5.3). This is the first report of a transposon-based enhancer trap system being successfully employed in *An.* coluzzii. The new enhancer trap insertion was in an intergenic region, upstream (<300bp) of an uncharacterized protein, ACOM027426, and downstream (<600bp) of protein ACOM027425, which is a homolog of *D. melanogaster jing.* Due to the direction of transcription, either gene could influence expression of the DsRed and EGFP markers within the enhancer trap construct and cause the observed novel phenotype.

The remobilization frequency of the *Muta1* element in *An. coluzzii* can be calculated different ways (Macias *et al.,* 2018, O'Brochta *et al.,* 2012). The frequency of enhancer detection can be calculated as one enhancer phenotype per 3,202 progeny screened, or a 0.03% frequency of *Muta1* enhancer detection in *An. coluzzii* (O'Brochta *et al.,* 2012). The remobilization frequency can also be calculated by the number of novel phenotypes per heterozygous adults, in this case one novel phenotype out of [(45*3 females) + (5*3 males)], or a 0.7% remobilization frequency (Macias *et al.,* 2018). For the *An. coluzzii* crosses, a total of 150 heterozygous adults were outcrossed. The crosses for *An. coluzzii* were performed at the Insect Transformation Facility at the University of Maryland and the eggs were sent to our laboratory for me to screen, resulting in 3,202 progeny being screened for novel fluorescent patterns (Table 5.3). However, for the *An. stephensi* crosses, I performed these myself in our laboratory, and only a total of 80 heterozygous *An. stephensi* parents have been crossed and progeny screened (1,213 progeny). Enhancer trapping with the *piggyBac* element has been successful in *An.*

stephensi and it has been transformed by two more elements than *An. coluzzii*. It is therefore reasonable to assume that *Muta1* remobilization events can also be recovered in *An. stephensi*, as long as enough adults are crossed. If the *An. stephensi* lines have a comparable remobilization frequency to what was observed in the *An. coluzzii* experiments, 0.03%-0.7% remobilization frequency depending on method of calculation, it is reasonable to assume that enhancer phenotypes will also be detected in *An. stephensi*. The remobilization crosses will be repeated with *An. stephensi* in order to double the number of heterozygous adults screened, from 80 to 160 adults.

The projected success of CRISPR-Cas9 mediated genome editing and gene-drive technologies for mosquitoes is contingent upon a robust knowledge of genetic regulation of development and response to pathogens. The *Muta1*-based enhancer trap system, which has been employed in both *An. stephensi* and *An. coluzzii*, can be used to help identify tissue, developmental and sex-specific promoter and enhancer sequences with these species. Another benefit of having *An. coluzzii* and *An, stephensi* lines transformed by the *Muta1* enhancer construct is that these lines contain attP docking sites for the addition of transgenes through recombination-mediated cassette exchange (Bateman *et al.,* 2006). Recombination-mediated cassette exchange can be done by injecting directly into embryos. There have been no observable fitness effects in the *Muta1* transformed lines, ENT3 or M1RG. Additionally, *An. coluzzii* and *An. stephensi* do not contain any *Muta1*-like elements in their genome, so there can be no cross-remobilization of the *Muta1* construct by any potentially active endogenous elements, providing a stable docking site for transgenes.

5.5 References

Akbari, Omar S., Igor Antoshechkin, Henry Amrhein, Brian Williams, Race Diloreto, Jeremy Sandler, and Bruce A. Hay. 2013. "The Developmental Transcriptome of the Mosquito Aedes Aegypti, an Invasive Species and Major Arbovirus Vector." *G3: Genes|Genomes|Genetics* 3 (9): 1493.

Blasco, Benjamin, Didier Leroy, and David A. Fidock. 2017. "Antimalarial Drug Resistance: Linking Plasmodium Falciparum Parasite Biology to the Clinic." *Nature Medicine* 23 (August): 917.

Carissimo, Guillaume, Emilie Pondeville, Melanie McFarlane, Isabelle Dietrich, Christian Mitri, Emmanuel Bischoff, Christophe Antoniewski, et al. 2015. "Antiviral Immunity of Anopheles Gambiae Is Highly Compartmentalized, with Distinct Roles for RNA Interference and Gut Microbiota." *Proceedings of the National Academy of Sciences of the United States of America* 112 (2): E176.

Catteruccia, Flaminia, Tony Nolan, Thanasis G. Loukeris, Claudia Blass, Charalambos Savakis, Fotis C. Kafatos, and Andrea Crisanti. 2000. "Stable Germline Transformation of the Malaria Mosquito Anopheles Stephensi." *Nature* 405 (6789): 959–62.

Clayton, April M., Yuemei Dong, and George Dimopoulos. 2014. "The Anopheles Innate Immune System in the Defense against Malaria Infection." *Journal of Innate Immunity* 6 (2): 169–81.

Coetzee, Maureen, Richard H. Hunt, Richard Wilkerson, Alessandra Della Torre, Mamadou B. Coulibaly, and Nora J. Besansky. 2013. "Anopheles Coluzzii and Anopheles Amharicus, New Members of the Anopheles Gambiae Complex." *Zootaxa* 3619: 246–74.

Davidson, G. 1962. "Anopheles Gambiae Complex." *Nature* 196 (4857): 907–907.

Fontaine, Michael C., James B. Pease, Aaron Steele, Robert M. Waterhouse, Daniel E. Neafsey, Igor V. Sharakhov, Xiaofang Jiang, et al. 2015. "Extensive Introgression in a Malaria Vector Species Complex Revealed by Phylogenomics." *Science* 347 (6217): 1258524.

Fouet, Caroline, Peter Atkinson, and Colince Kamdem. 2018. "Human Interventions: Driving Forces of Mosquito Evolution." *Trends in Parasitology* 34 (2): 127–39.

Gantz, Valentino M., and Ethan Bier. 2015. "Genome Editing. The Mutagenic Chain Reaction: A Method for Converting Heterozygous to Homozygous Mutations." *Science* 348 (6233): 442–44.

Gantz, Valentino M., Nijole Jasinskiene, Olga Tatarenkova, Aniko Fazekas, Vanessa M. Macias, Ethan Bier, and Anthony A. James. 2017. “Highly Efficient Cas9-Mediated Gene Drive for Population Modification of the Malaria Vector Mosquito Anopheles Stephensi.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (49): E6736.

Gregory, M., L. Alphey, N. I. Morrison, and S. M. Shimeld. 2016. “Insect Transformation with piggyBac: Getting the Number of Injections Just Right.” *Insect Molecular Biology* 25 (3): 259–71.

Grossman, G. L., C. S. Rafferty, J. R. Clayton, T. K. Stevens, O. Mukabayire, and M. Q. Benedict. 2001. “Germline Transformation of the Malaria Vector, Anopheles Gambiae, with the piggyBac Transposable Element.” *Insect Molecular Biology* 10 (6): 597–604.

Hammond, Andrew, Roberto Galizi, Kyros Kyrou, Alekos Simoni, Carla Siniscalchi, Dimitris Katsanos, Matthew Gribble, et al. 2015. “A CRISPR-Cas9 Gene Drive System Targeting Female Reproduction in the Malaria Mosquito Vector Anopheles Gambiae.” *Nature Biotechnology* 34 (December): 78.

Holt, Robert A., G. Mani Subramanian, Aaron Halpern, Granger G. Sutton, Rosane Charlab, Deborah R. Nusskern, Patrick Wincker, et al. 2002. “The Genome Sequence of the Malaria Mosquito Anopheles Gambiae.” *Science* 298 (5591): 129.

Jiang, Xiaofang, Ashley Peery, A. Brantley Hall, Atashi Sharma, Xiao-Guang Chen, Robert M. Waterhouse, Aleksey Komissarov, et al. 2014. “Genome Analysis of a Major Urban Malaria Vector Mosquito, Anopheles Stephensi.” *Genome Biology* 15 (9): 459.

Keene, Kimberly M., Brian D. Foy, Irma Sanchez-Vargas, Barry J. Beaty, Carol D. Blair, and Ken E. Olson. 2004. “RNA Interference Acts as a Natural Antiviral Response to O’nyong-Nyong Virus (Alphavirus; Togaviridae) Infection of Anopheles Gambiae.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (49): 17240–45.

Kyrou, Kyros, Andrew M. Hammond, Roberto Galizi, Nace Kranjc, Austin Burt, Andrea K. Beaghton, Tony Nolan, and Andrea Crisanti. 2018. “A CRISPR–Cas9 Gene Drive Targeting Doublesex Causes Complete Population Suppression in Caged Anopheles Gambiae Mosquitoes.” *Nature Biotechnology* 36 (September): 1062.

Li, Meng Amy, Stephen J. Pettitt, Sabine Eckert, Zemin Ning, Stephen Rice, Juan Cadiñanos, Kosuke Yusa, Nathalie Conte, and Allan Bradley. 2013. “The piggyBac Transposon Displays Local and Distant Reintegration Preferences and Can Cause Mutations at Noncanonical Integration Sites.” *Molecular and Cellular Biology* 33 (7): 1317.

Liu, Kun, and Susan R. Wessler. 2017. “Transposition of Mutator–like Transposable Elements (MULEs) Resembles hAT and Transib Elements and V(D)J Recombination.” *Nucleic Acids Research* 45 (11): 6644–55.

Macias, V., J. Coleman, M. Bonizzoni, and A. A. James. 2014. “piRNA Pathway Gene Expression in the Malaria Vector Mosquito Anopheles Stephensi.” *Insect Molecular Biology* 23 (5): 579–86.

Macias, Vanessa M., Alyssa J. Jimenez, Bianca Burini-Kojin, David Pledger, Nijole Jasinskiene, Celine Hien Phong, Karen Chu, et al. 2017. “Nanos-Driven Expression of piggyBac Transposase Induces Mobilization of a Synthetic Autonomous Transposon in the Malaria Vector Mosquito, Anopheles Stephensi.” *Insect Biochemistry and Molecular Biology* 87 (August): 81–89.

Neafsey, Daniel E., George K. Christophides, Frank H. Collins, Scott J. Emrich, Michael C. Fontaine, William Gelbart, Matthew W. Hahn, et al. 2013. “The Evolution of the Anopheles 16 Genomes Project.” *G3: Genes|Genomes|Genetics* 3 (7): 1191.

Nolan, Tony, Tom M. Bower, Anthony E. Brown, Andrea Crisanti, and Flaminia Catteruccia. 2002. “piggyBac-Mediated Germline Transformation of the Malaria Mosquito Anopheles Stephensi Using the Red Fluorescent Protein dsRED as a Selectable Marker.” *The Journal of Biological Chemistry* 277 (11): 8759–62.

O’Brochta, David A., Robert T. Alford, Kristina L. Pilitt, Channa U. Aluvihare, and Robert A. Harrell. 2011. “piggyBac Transposon Remobilization and Enhancer Detection in Anopheles Mosquitoes.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (39): 16339.

O’Brochta, David A., Kristina L. Pilitt, Robert A. Harrell, Channa Aluvihare, and Robert T. Alford. 2012. “Gal4-Based Enhancer-Trapping in the Malaria Mosquito Anopheles Stephensi.” *G3: Genes|Genomes|Genetics* 2 (11): 1305.

Olson, Ken E., and Carol D. Blair. 2015. “Arbovirus-Mosquito Interactions: RNAi Pathway.” *Current Opinion in Virology* 15 (December): 119–26.

Ouji, Manel, Jean-Michel Augereau, Lucie Paloque, and Françoise Benoit-Vical. 2018. “Plasmodium Falciparum Resistance to Artemisinin-Based Combination Therapies: A Sword of Damocles in the Path toward Malaria Elimination.” *Parasite* 25: 24–24.

Papathanos, Philippos A., Nikolai Windbichler, Miriam Menichelli, Austin Burt, and Andrea Crisanti. 2009. “The Vasa Regulatory Region Mediates Germline Expression and Maternal Transmission of Proteins in the Malaria Mosquito Anopheles Gambiae: A Versatile Tool for Genetic Control Strategies.” *BMC Molecular Biology* 10 (1): 65.

Pfaffl, Michael W. 2001. “A New Mathematical Model for Relative Quantification in Real-Time RT–PCR.” *Nucleic Acids Research* 29 (9): e45–e45.

Potter, Christopher J., and Liqun Luo. 2010. “Splinkerette PCR for Mapping Transposable Elements in Drosophila.” *PloS One* 5 (4): e10168.

Ranson, Hilary, and Natalie Lissenden. 2016. “Insecticide Resistance in African Anopheles Mosquitoes: A Worsening Situation That Needs Urgent Action to Maintain Malaria Control.” *Trends in Parasitology* 32 (3): 187–96.

Rebollo, Maria P., and Moses J. Bockarie. 2013. “Toward the Elimination of Lymphatic Filariasis by 2020: Treatment Update and Impact Assessment for the Endgame.” *Expert Review of Anti-Infective Therapy* 11 (7): 723–31.

Scali, Christina, Tony Nolan, Igor Sharakhov, Maria Sharakhova, Andrea Crisanti, and Flaminia Catteruccia. 2007. “Post-Integration Behavior of a Minos Transposon in the Malaria Mosquito Anopheles Stephensi.” *Molecular Genetics and Genomics: MGG* 278 (5): 575–84.

Sinka, Marianne E., Michael J. Bangs, Sylvie Manguin, Yasmin Rubio-Palis, Theeraphap Chareonviriyaphap, Maureen Coetzee, Charles M. Mbogo, et al. 2012. “A Global Map of Dominant Malaria Vectors.” *Parasites & Vectors* 5 (April): 69–69.

Smith, Christopher D., Shengqiang Shu, Christopher J. Mungall, and Gary H. Karpen. 2007. “The Release 5.1 Annotation of Drosophila Melanogaster Heterochromatin.” *Science* 316 (5831): 1586.

Strausfeld, N. J., E. K. Buschbeck, and R. S. Gomez. 1995. “The Arthropod Mushroom Body: Its Functional Roles, Evolutionary Enigmas and Mistaken Identities.” In *The Nervous Systems of Invertebrates: An Evolutionary and Comparative Approach: With a Coda Written by T.H. Bullock*, edited by O. Breidbach and W. Kutsch, 349–81. Basel: Birkhäuser Basel.

Waldock, Joanna, Kenneth E. Olson, and George K. Christophides. 2012. “Anopheles Gambiae Antiviral Immune Response to Systemic O’nyong-Nyong Infection.” *PLoS Neglected Tropical Diseases* 6 (3): e1565.

Wilson, Matthew H., Craig J. Coates, and Alfred L. George. 2007. “PiggyBac Transposon-Mediated Gene Transfer in Human Cells.” *Molecular Therapy: The Journal of the American Society of Gene Therapy* 15 (1): 139–45.

Wolf, Jochen B. W., and Hans Ellegren. 2016. “Making Sense of Genomic Islands of Differentiation in Light of Speciation.” *Nature Reviews. Genetics* 18 (November): 87.

5.6 Figures and tables

Parental line	Transformation	Plasmid	Phenotype	Integrations	TSD	Integration location	Features
M2B	*piggyBac*	pBac3CAnVasMuta1	Canonical 3PX3-ECFP expression	1	TTAA	Contig ABKP02000993.1 scf 1925491347 @base 323,579	Intergenic region - inserted into Class II element, TcMar-Tc1
ENT3	*Muta1*	pMutaETS	Canoncal 3PX3-EGFP and 3PX3-DsRed exression	1	GCGCGCCAA	Contig ABKP02026141.1 scf 1925491365 @base 1,075,421+	Intergenic region, downstream of uncharcterized protein ACOM041868
Single generation *Muta1* remobilization events							
			Larval: In addition to cannonical expression, brighter EGFP/dsRed near annal papillae	1	TTTGCTTAC	Contig ABKP02022086.1 scf 1925491374 @base 878,949+	Intergenic region, downstream of uncharcterized protein ACOM027425 (*jing* homolog), upstream of uncharcaterized protein ACOM027426 (reverse strand)

Table 5.1 – Integration data for the *An. coluzzii* helper and enhancer trap lines and the integration location of a novel *Muta1* remobilization event.

Parental Line	Transformation	Plasmid	Phenotype	# of integrations	TSD	Integration Location	Features
F3B	*piggyBac*	pBac3CAnVasMuta1	Canonical 3PX3-ECFP expression plus expression in anal papillae	1	TTAA	scaffold_00115 @ base 353,961	Less than 200bp upstream of ASTI03495, *deltex* homolog
F3RG	*Muta1*	pMutaENT3	Canoncal 3PX3-EGFP and 3PX3-DsRed exression	1	CTTTAAAA	scaffold_00115 @ base 203,789	Intergenic region, downstream of uncharacterized protein ASTEI10161, upstream of uncharacterized protein ASTEI10162
M1RG	*Muta1*	pMutaENT3	Canoncal 3PX3-EGFP and 3PX3-DsRed exression	2	CGACAGGG	scaffold_00042 @ base 203,343	Intergenic region, immediately 3' of protein ASTEI06582 (ion channel component)
					GTTTAGAG	scaffold_00003 @ base 544,990	Intergenic region - inserted into a Class I Gypsy element

Table 5.2 – Integration data for the *An. stephensi* helper and enhancer trap lines.

An. coluzzii outcross	# outcrosses	Progeny screened	Total screened	Remobilization events
♀ (45) pMuta ETSa (dsRed/EGFP): pBac3CAnVasMuta1 (ECFP) :: ♂ (05) wt	3	2228	3202	1
♂ (05) pMuta ETSa (dsRed/EGFP): pBac3CAnVasMuta1 (ECFP) :: ♀ (45) wt	3	974		

Table 5.3 –*An. coluzzii* enhancer trap cross data. One remobilization event was recovered out of a total 3202 progeny from 6 backcrossed matings.

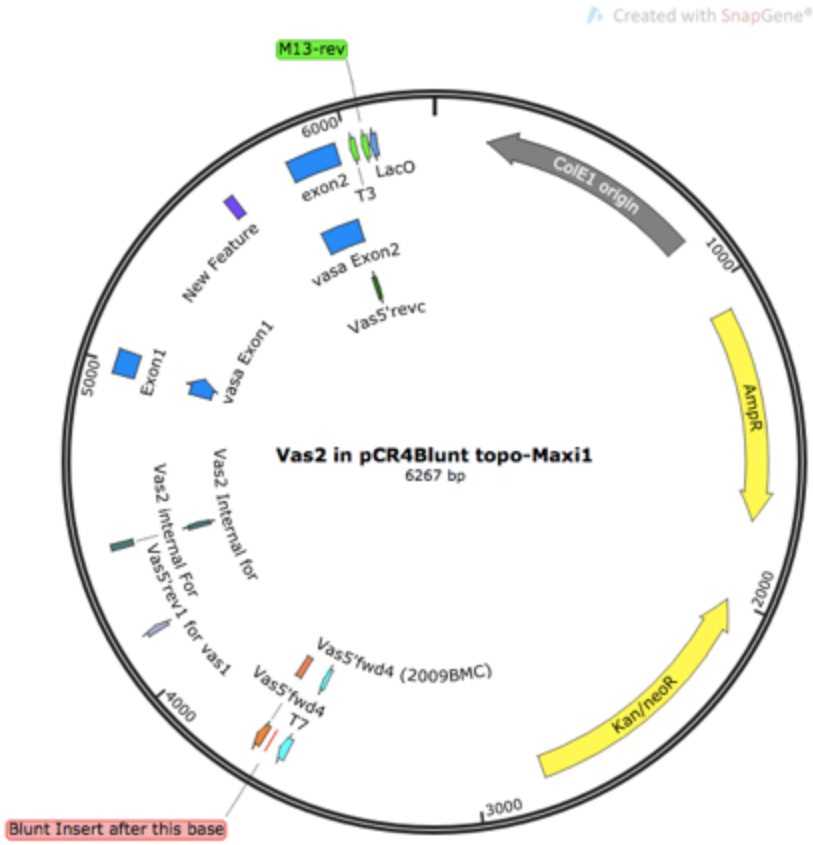

Figure 5.1 – The plasmid Vas2inpCr4BluntTopoMaxi1, a gift from Robert Harrell of University of Maryland Institute for Bioscience and Biotechnology Research containing the *An. gambiae vasa* regulatory sequence.

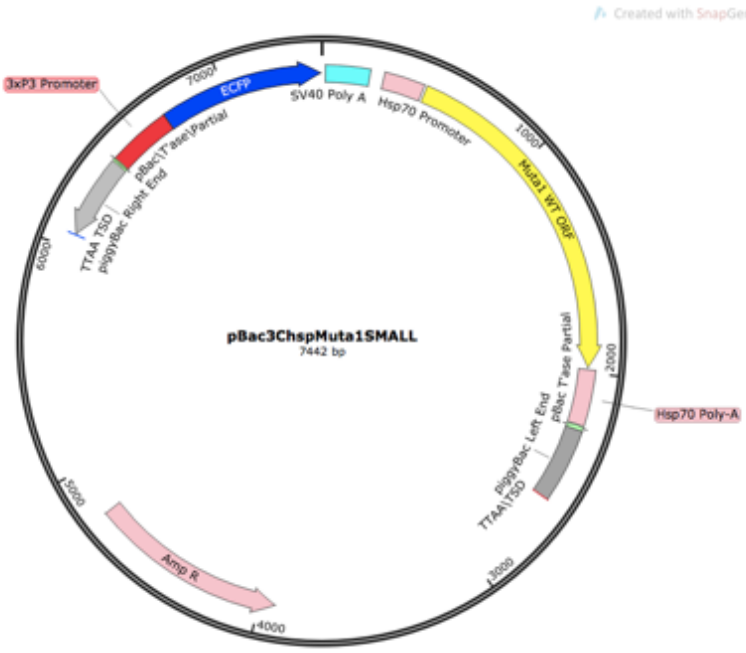

Figure 5.2 – The plasmid pBac3ChspMuta1SMALL, a version of pBac3ChspMuta1 (described in Chapter 3) that contains minimal length *piggyBac* transposon ends

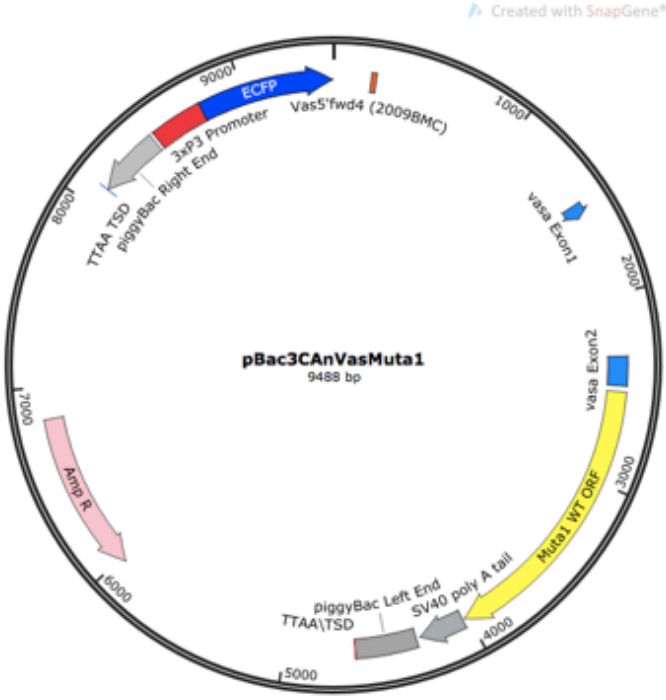

Figure 5.3 – The *Muta1* helper plasmid containing the *Muta1* coding sequence, proceeded by *An. gambiae vasa* 5' sequence and selectable ECFP eye marker. Minimal *piggyBac* left and right ends flank the transgenes.

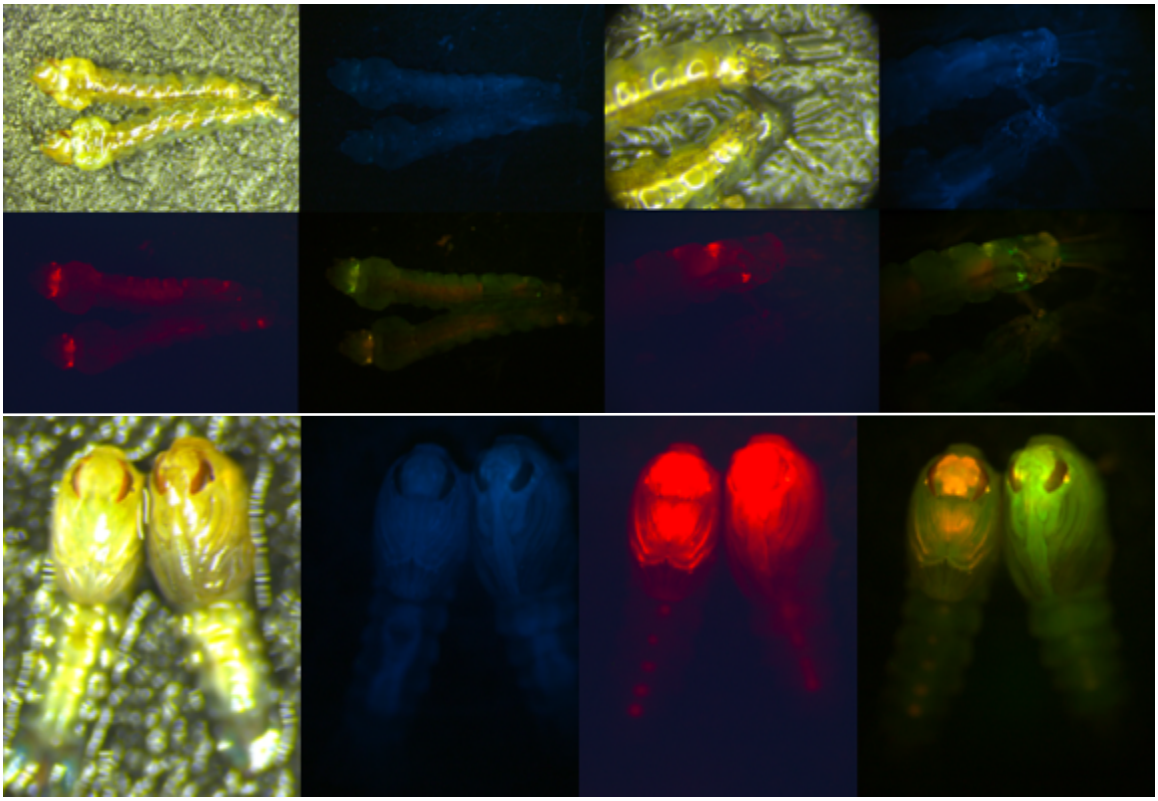

Figure 5.4 – Enhancer trap remobilization patterns in *An. coluzzii.*
Top panel: larvae from remobilization cross. The top larva shows the remobilized phenotype with brighter EGFP visible in the eyes and two bright areas of DsRed and EGFP near the anal papillae. The bottom larva shows the parental, non-remobilized phenotype, in which 3PX3-EGFP is not as bright and there are no bright areas of fluorescence near the anal papillae.
Lower panel: Pupae from remobilization cross: The right pupa shows the remobilized phenotype with brighter EGFP visible in the eyes - the two bright areas of DsRed and EGFP near the anal papillae are not visible in the pupae. The left pupa shows the parental, non-remobilized phenotype, in which 3PX3-EGFP is not as bright.

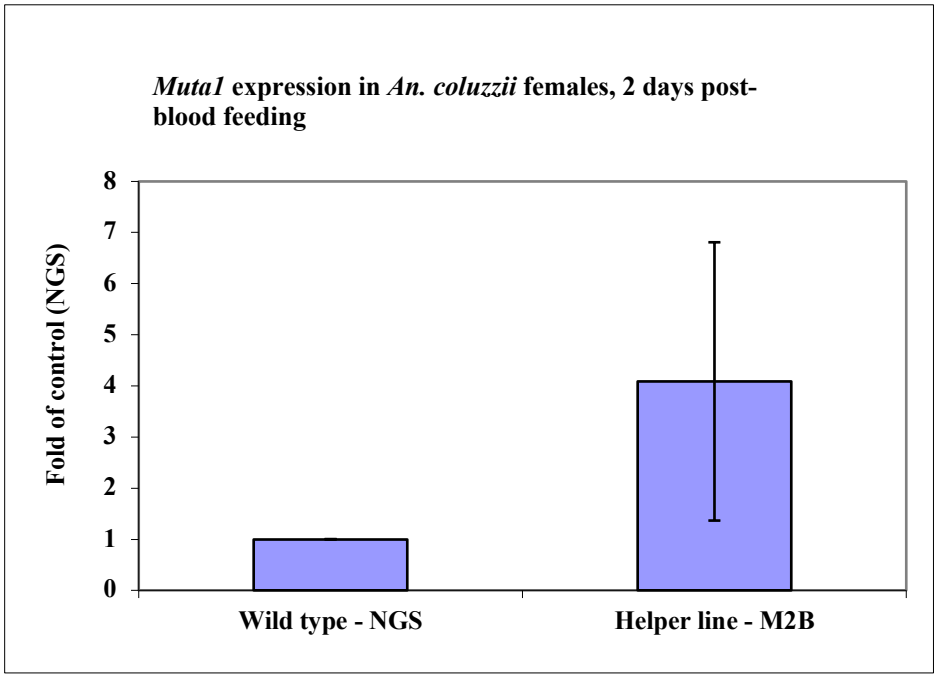

Figure 5.5 – Quantitative RT-PCR expression data in adult 48-hour post blood-fed M2B line females. The *Muta1* expression data is compared to the *An. coluzzii* background strain, NGS.

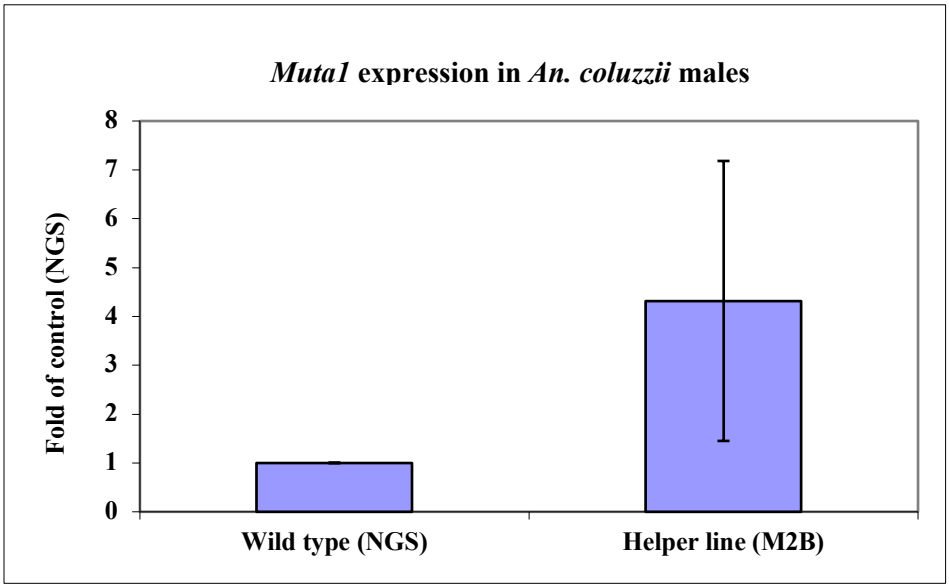

Figure 5.6 – Quantitative RT-PCR expression data in adult M2B line males. The *Muta1* expression data is compared to the *An. coluzzii* background strain, NGS.

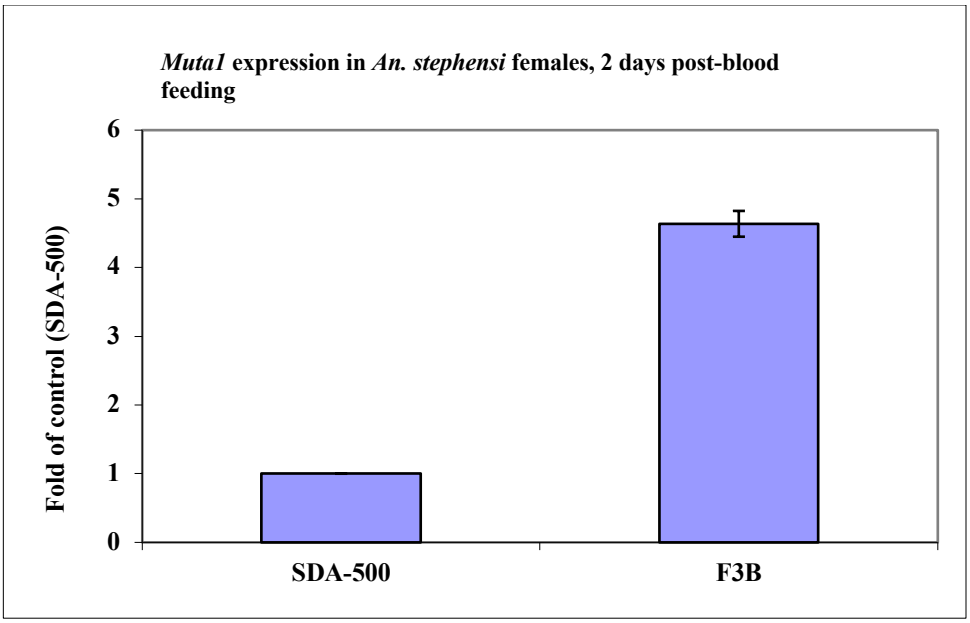

Figure 5.7 – Quantitative RT-PCR expression data in adult 48-hour post blood-fed F3B line females. The *Muta1* expression data is compared to the *An. stephensi* background strain, SDA-500.

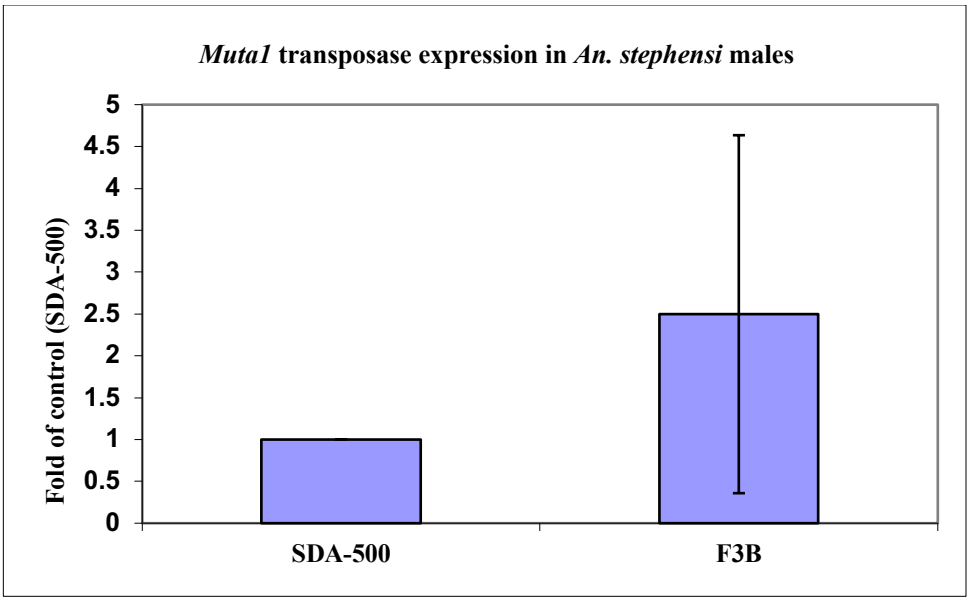

Figure 5.8 – Quantitative RT-PCR expression data in adult F3B line males. The *Muta1* expression data is compared to the *An. stephensi* background strain, SDA-500.

Chapter 6 – Future directions with *Muta1*

In this thesis I have described the successful development of the *MULE*-like *Muta1* transposable element into a gene vector in mosquitoes. *Muta1* is only the second transposable element to genetically transform *An. coluzzii* and the third to do so in *An. stephensi*. I have showed that *Muta1* can integrate into its host, *Ae. aegypti*. Notably, I have demonstrated the first example of enhancer trapping in *An. coluzzii.* As such, *Muta1* is new important platform for the study of mosquito functional genomics. Below I detail further experiments with *Muta1* that should be conducted in each of these species in order to further refine the performance parameters of *Muta1* in them.

6.1 *Aedes aegypti*

The adult female *Muta1* transgene expression data results were unanticipated and led to the question of possible silencing in females (Figure 4.3). Our laboratory has observed transgene silencing in females previously, although the mechanism remains unknown (Han, 2017). Of particular interest is the observation that adult males of the Orlando and Liverpool lab strains show a higher level of endogenous *Muta1* expression compared to Higgs (Figure 4.2). This suggests that there may a greater level of silencing in the Higgs strain, and provides also provides a possible explanation for the poor transformation rates achieved at the Insect Transformation Facility for the Higgs injections (Table 4.1) versus the high Orlando transformation rates obtained previously (Shah, 2015). A relatively straightforward approach to test this hypothesis would be to

execute a fixed amount of embryo injections for each lab strain, Higgs, Orlando, and Liverpool, using *piggyBac*, which is amenable to transformation in *Aedes* (Gregory *et al.,* 2016). Based on the 23.9% adult survival and 2.8% transformation frequency achieved in Higgs (Table 4.2), I could perform a total of 500 embryo injections per strain and expect to obtain at least three transgenic lines per strain, unless there are strain differences that negatively impact transformation. There currently is no research that evaluates the transformation frequency in different lab strains of *Ae. aegypti*. If there is a lab strain that is more amenable to genetic transformation, this would help inform any future transformation experiments in *Aedes*.

The putative enhancer trap transformed line, Line I, does not show canonical expression of the eye markers (Figure 4.4 to 4.8). As the junction fragment has been difficult to obtain, it cannot be confirmed where the pMutaENT3 construct has integrated and whether or not there are positional effects causing the observed phenotype. Depending on the outcome of the *piggyBac* transformation experiment for the three lab strains, the *Muta1* injections with pMutaENT3 will be repeated using the most amenable strain.

There were no observed fitness effects, such as decreased fecundity or larval survival, in any of the helper lines expressing *Muta1* transposase. A stable helper line that can be easily maintained and homozygosed is desirable for enhancer trap experiments. However, the *Ae. aegypti* enhancer trap system described in this thesis employs an endogenous active transposon, *Muta1*. There are eight full-length copies of *Muta1* and over 300 non-autonomous *Muta1* MITEs within the *Ae. aegypti* genome (Liu

and Wessler). In the helper lines expressing *Muta1* under germline promoter control (Table 4.2), there is the potential for these endogenous elements to be remobilized. If these endogenous elements are remobilized at a high enough frequency, I would expect to see an observable fitness cost, like sterility and (Wright *et al.,* 2013). To assess endogenous element remobilization, I will generate PCR-based sequencing libraries using Oxford Nanopore sequencing technology. I will use transposon display, a technique in which genomic DNA is digested and ligated to adapter sequences, then PCR is performed using transposon-specific primers (Remekar *et al.,* 2018). I will use degenerate primers to the *Muta1* TIRs, to allow for some mismatch, and the transposon display library generated will be sequenced using Oxford Nanopore long-read technology. Transposon display libraries can be created for the Higgs background strain and for the helper Q Line to assess whether there has been remobilization of endogenous elements (Miller *et al.,* 2018).

6.2 *Anopheles stephensi* and *Anopheles coluzzii*

The *Muta1*-based enhancer trap system, which has been employed in both *An. stephensi* and *An. coluzzii*, can be used to help identify tissue, developmental and sex-specific promoter and enhancer sequences with these species. In *An. coluzzii,* one *Muta1* remobilization event was recovered after screening experiments from 150 back-crossed adults (Table 5.3), giving a remobilization frequency of 0.7%. This is the first report of a transposon-based enhancer trap system being successfully employed in *An.* coluzzii. The single remobilization event originated from a female germline. Previous enhancer trap

experiments in *An. stephensi* have shown a higher remobilization frequency in the female germline compared to the male germline (O'Brochta *et al.,* 2011, O'Brochta *et al.,* 2012). Now that it is established that *Muta1* can transform and remobilize in *An. coluzzii*, further modifications can be made to the *Muta1* enhancer trap system. *An. coluzzii* has only been transformed by *piggyBac* and *Muta1*, and only *Muta1* remobilizes (Grossman *et al.,* 2001). It is feasible to use both transposons to create a *Muta1* and GAL4-based bipartite enhancer detection system, similar to the GAL4-based enhancer trap that has been used in *An. stephensi* with the *piggyBac* and *Minos* transposable elements (O'Brochta *et al.,* 2012).

I have not yet identified a remobilization event in the *An. stephensi* enhancer trap crosses. However, the total number of heterozygous adults backcrossed and the total number of resulting progeny screened for *An. stephensi* is fewer than what I was able to generate for *An. coluzzii*. *An. coluzzii* has a 0.7% remobilization frequency, depending on the method of calculation (Macias *et al.,* 2018). If *Muta1* has a similar remobilization frequency in *An. stephensi* as for *An. coluzzii*, it is reasonable to assume that enhancer phenotypes will also be detected in *An. stephensi*. The remobilization crosses will be repeated with *An. stephensi* in order to double the number of heterozygous adults screened, from 80 to 160 adults.

6.2 References

Gregory, M., L. Alphey, N. I. Morrison, and S. M. Shimeld. 2016. "Insect Transformation with piggyBac: Getting the Number of Injections Just Right." *Insect Molecular Biology* 25 (3): 259–71.

Grossman, G. L., C. S. Rafferty, J. R. Clayton, T. K. Stevens, O. Mukabayire, and M. Q. Benedict. 2001. "Germline Transformation of the Malaria Vector, Anopheles Gambiae, with the piggyBac Transposable Element." *Insect Molecular Biology* 10 (6): 597–604.

Han, Michael. 2017. "The piRNA System in Aedes Aegypti." UC Riverside.

Liu, Kun, and Susan R. Wessler. 2017. "Functional Characterization of the Active Mutator-like Transposable Element, Muta1 from the Mosquito Aedes Aegypti." *Mobile DNA* 8 (January): 1.

Macias, Vanessa M., Alyssa J. Jimenez, Bianca Burini-Kojin, David Pledger, Nijole Jasinskiene, Celine Hien Phong, Karen Chu, et al. 2017. "Nanos-Driven Expression of piggyBac Transposase Induces Mobilization of a Synthetic Autonomous Transposon in the Malaria Vector Mosquito, Anopheles Stephensi." *Insect Biochemistry and Molecular Biology* 87 (August): 81–89.

Miller, Danny E., Cynthia Staber, Julia Zeitlinger, and R. Scott Hawley. 2018. "Highly Contiguous Genome Assemblies of 15 Drosophila Species Generated Using Nanopore Sequencing." *G3: Genes|Genomes|Genetics* 8 (10): 3131.

O'Brochta, David A., Robert T. Alford, Kristina L. Pilitt, Channa U. Aluvihare, and Robert A. Harrell. 2011. "piggyBac Transposon Remobilization and Enhancer Detection in Anopheles Mosquitoes." *Proceedings of the National Academy of Sciences of the United States of America* 108 (39): 16339.

O'Brochta, David A., Kristina L. Pilitt, Robert A. Harrell, Channa Aluvihare, and Robert T. Alford. 2012. "Gal4-Based Enhancer-Trapping in the Malaria Mosquito Anopheles Stephensi." *G3: Genes|Genomes|Genetics* 2 (11): 1305.

Ramekar, Rahul Vasudeo, Kyong-Cheul Park, Kyu Jin Sa, and Ju Kyong Lee. 2018. "Mutator-Based Transposon Display: A Genetic Tool for Evolutionary and Crop-Improvement Studies in Maize." *Molecular Biotechnology* 60 (11): 799–809.

Shah, Presha Vijaykumar. 2015. *Mobilization of Newly Identified Transposon Muta1 in Aedes Aegypti and Drosophila Melanogaster*. University of California, Riverside.

Wright, Jennifer A., Ryan C. Smith, Xianghong Li, Nancy L. Craig, and Peter W. Atkinson. 2013. “IPB7 Transposase Behavior in Drosophila Melanogaster and Aedes Aegypti.” *Insect Biochemistry and Molecular Biology* 43 (10): 899–906.