ABSTRACT

Title of Document: THE BEHAVIOR AND EVOLUTION OF CLASS II TRANSPOSABLE ELEMENTS IN THE MALARIAL MOSQUITO, ANOPHELES GAMBIAE

Ramanand Arun Subramanian, Doctor of Philosophy, 2008

Directed By: Dr. David A. O'Brochta, Professor, University of Maryland Biotechnology Institute

Transposable elements are DNA sequences with a unique ability to change their genomic location. Transposable elements are fascinating because of their ability to move, and their ubiquitous presence and contribution to the evolution of all prokaryotic and eukaryotic genomes. Their mobility properties have made them extremely useful as molecular tools in the laboratory. Transposable elements have also been proposed to be useful as genetic drive agents to introduce phenotype-altering genes in natural populations of mosquitoes, to control vector-borne diseases such as malaria.

Presented in this thesis are studies on the behavior and evolution of two endogenous Class II transposable elements, *Herves* and *Topi* in natural populations of *Anopheles gambiae*, a species seriously being considered for population modification using genetic manipulation. In Chapters 2 and 4, results from the analysis of copy number, activity, and nucleotide sequence as well as structural diversity of *Herves* and

Topi elements, respectively in 5-6 *An.gambiae* populations in Africa are described. In Chapter 3, studies to identify and assess the activity of the natural variants of *Herves* transposase in *An.gambiae* are described.

The results from these studies show that both *Herves* and *Topi* elements have long histories in *An.gambiae* with *Topi* present in *An.gambiae* earlier than *Herves*. *Herves,* but not *Topi*, is still active in natural populations of *An.gambiae* with more than one active form of *Herves* transposase responsible for its activity. Both the elements, despite their long history in *An.gambiae*, have a very high percentage of individuals with complete forms of the element. This observation is an unusual feature of these elements, which would not be predicted for elements with such a long history. The presence of complete and active forms of *Herves* and *Topi*, elements with long histories in *An.gambiae,* argues against the possibility of rapid accumulation of deleted forms of transposable elements as a general feature of their evolution.

Better understanding of the behavior and evolution of Class II transposable elements in *An.gambiae* shows that Class II transposable elements still hold promise as a genetic drive agent for this particular species.

THE BEHAVIOR AND EVOLUTION OF CLASS II TRANSPOSABLE ELEMENTS IN THE MALARIAL MOSQUITO, ANOPHELES GAMBIAE

By

Ramanand Arun Subramanian

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Professor David A. O'Brochta, Chair
Professor Eric Baehrecke
Professor Leslie Pick
Professor Gerald S.Wilkinson
Professor Louisa P. Wu

Preface

Declaration of author's intent to use own previously published text.

The main text, tables, figures and figure legends in their entirety for:

Chapter 2: Transposable element dynamics of the *hAT* element *Herves* in the human malaria vector, *Anopheles gambiae s.s*.

were used, and only modified to meet the formatting requirements of this dissertation. Full citation is given below:

Ramanand A. Subramanian, Peter Arensburger, Peter W. Atkinson, David A. O'Brochta. Transposable element dynamics of the *hAT* element *Herves* in the human malaria vector *Anopheles gambiae s.s*. *Genetics*. 2007 Aug; 176(4):2477-87.

Dedication

I dedicate this thesis to my parents

Acknowledgements

I would like to express my sincere gratitude to Dr. David A. O'Brochta, my research advisor, for his guidance, support and constant encouragement, which has made this possible. When an advisor like Dave becomes your role model, everything else falls in place. I have learned and immensely benefited from his scientific intellect, curiosity, discipline, patience and perseverance.

My sincere thanks to my advisory committee members: Dr. Eric Baehrecke, Dr. Leslie Pick, Dr. Gerald S. Wilkinson and Dr. Louisa P. Wu for their guidance, valuable ideas and, their support and belief in my abilities.

I would like to thank Dr. Harold E. Smith for his time and effort to teach me molecular techniques in the laboratory.

Thanks to all the former and present members of Dave's lab: Channa, Biyi, Ed, Nicci, Dan, Tram, Nambi, Kristina, Robert, Barbara, Rob and Valerie, for their help and friendship. I specially thank Valerie and Kristina, for reading and suggesting changes in my manuscript.

Thanks to Dr. Nancy L. Craig, for encouraging as well as permitting me to work in her lab. I would like to specially thank Xianghong, for spending time teaching me biochemical techniques during my few weeks in Nancy's lab. Thanks to

Dr. Zvi Kelman and Dr. Phil Bryan for helpful discussions and guidance with my biochemical experiments. I thank Dr. Lori Urban for allowing me to use equipment in her lab for my experiments. Thanks to Dr. Donald Nuss and people in his lab, especially Qihong, Fu and Xuemin for their help with some of my experiments.

I take this opportunity to say how immensely grateful I am to my Mom, Shantha and Dad, Subramanian, for always being there for me. I owe everything I am today to them. I thank my sister, Sangeetha, for her encouragement through out my life and also my brother-in-law for his support in the past year.

Thanks to all my wonderful friends, Girish, Srikanth, Sundaram, Bhanumathy, Rama, Uma, Alpana, Prathima,Wasia, Abhi, Bhuvana and Sudeshna for all the good times we have had and keeping my spirits up. Thanks to Carla and Aswani, for being great house-mates and friends during my graduate school.

I would like to thank God, for all the opportunities and experiences in my life.

Table of Contents

List of Tables

List of Figures

Chapter 1: Transposable elements and their application as genetic drive agents to control vector-borne diseases

Transposable elements and genome evolution

Transposable elements are DNA sequences that have the ability to change their genomic location. Transposable elements are ubiquitous in both prokaryotic and eukaryotic genomes and can contribute substantially to genome content. Transposable elements, for instance, comprise 15% of the *Drosophila* genome, 45 % of the human genome, and 50% of the *Aedes aegypti* genome (KIDWELL 2002; NENE *et al.* 2007). They can be broadly classified into two kinds based on the presence of a DNA/RNA intermediate during transposition. Transposable elements that have an RNA intermediate and require a reverse transcription step are called Class I or retrotransposons. Class II or DNA transposable elements have a DNA intermediate and move through a cut-and-paste mechanism, where a transposase catalyzes the excision of the transposon from one site, and insertion into another site in the genome. Transposable elements containing sequences that encode proteins necessary for their transposition (transposase in the case of DNA transposons) are called autonomous elements. Non-autonomous elements cannot catalyze their own transposition but are capable of transposing using proteins from another source in the genome.

Transposable elements are potent mutagenic agents because of their ability to excise from one location and insert into other parts of the genome. Mutations caused by transposable element insertions are often deleterious but can also serve as a source of genetic variation contributing significantly to the evolution of genomes

(BROOKFIELD 2005; KIDWELL and LISCH 2001). Transposable element insertions can have a variety of consequences, such as altering the levels and patterns of gene expression, causing chromosome breakage, illegitimate recombination and genome rearrangement (KIDWELL and LISCH 2001). Transposable element insertions are not always deleterious, for example, strong selection for an *S*-element insertion in a heat-shock protein gene in *D.melanogaster* (MASIDE et al. 2002) and a *Doc* element insertion upstream from the transcription start of a cytochrome P450 gene, *Cyp6g1* (SCHLENKE and BEGUN 2004) in *D.simulans* has led to their spread and fixation in the natural populations. Presence of the *Doc* insertion is correlated with increased *Cyp6g1* transcript which is associated with insecticide resistance in *D.melanogaster* (SCHLENKE and BEGUN 2004). There are also examples of transposable elements being co-opted to perform host functions. For example, RAG1 and RAG2 are genes that are involved in V(D)J recombination in vertebrates and have evolved from ancient transposable elements (KAPITONOV and JURKA 2005). Two retrotransposons, HeT-A and TART function as telomeres in *D.melanogaster* (LEVIS *et al.* 1993).

Another fascinating feature of transposable elements is their ability to cross species boundaries and enter new genomes by horizontal transfer. There have been a number of reports of horizontal transfer events involving transposable elements in both prokaryotes and eukaryotes (reviewed in (SILVA *et al.* 2004)). A recent study in *D.melanogaster* suggested that a large proportion of the transposable elements had a relatively recent origin as a result of horizontal transfer (SANCHEZ-GRACIA *et al.* 2005). *mariner*-like elements (MLEs) are a family of transposable elements that have a wide host range; they have been found in plants, insect genomes, other invertebrates

and vertebrates including humans. Many examples of horizontal transfer involving members of this family of elements have been reported (HARTL *et al.* 1997a). The widespread presence of transposable elements, their mutagenic properties, deleterious and beneficial effects of their insertions as well as their ability to invade new genomes are some of the features of transposable elements for continued interest in studying them. Transposable elements are also studied because they have proven to be very useful tools in the lab with a wide range of applications.

Application of transposable elements

The mobility properties of transposable elements that have made them fascinating and significant components of genomes have also led to their widespread use as laboratory tools. Class II transposable elements have been used for insertional mutagenesis and are now essential tools for performing functional genomics studies in a wide range of species. For example, *Tn7* transposon in yeast (KUMAR *et al.* 2004a) *P*-elements in *Drosophila* (COOLEY et al. 1988; ZHANG and SPRADLING 1994), *Mos-1* in *C.elegans* (BESSEREAU *et al.* 2001)**,** *Minos* element in *Ciona intestinalis* (SASAKURA *et al.* 2003), *piggyBac* element in malaria parasite, *Plasmodium* (BALU *et al.* 2005), red flour beetle, *Tribolium castaneum* (LORENZEN et al. 2007) and also vertebrates (MISKEY *et al.* 2005). One of the advantages of using transposons for mutagenesis is that the genes mutated by transposon insertions are molecularly tagged and can be easily identified and isolated. Another advantage is that vectors can be engineered such that a reporter molecule is expressed in a context-dependent manner. Such vectors have been used to identify enhancer regions of genes, 5' promoter regions and 3' regions of genes. For example, *P*-elements in *Drosophila*

melanogaster (DUFFY 2002) and *Tol2* transposon in Zebra fish (PARINOV *et al.* 2004) have been used to identify enhancer regions of genes.

Transposable elements have been very useful as gene vectors for germ-line transgenesis in invertebrates, vertebrates and plants. Insect biologists now have at least six transposable element- based gene vectors from which to choose when considering the creation of a transgenic insect (*P*, *hobo*, *Tn5, mariner, Minos, piggyBac*, and *Hermes*) (ATKINSON *et al.* 2001). For instance, the *Hermes* element has been useful in creating transgenic *D.melanogaster* (O'BROCHTA and ATKINSON 1996), *Aedes aegypti* (JASINSKIENE *et al.* 1998), *Ceratitis capitata* (MICHEL *et al.* 2001), *Stomoxys calcitrans* (O'BROCHTA *et al.* 2000), *Tribolium castaneum* (BERGHAMMER *et al.* 1999) and butterfly, *Bicyclus anynana* (MARCUS *et al.* 2004). Other commercially useful insects such as silk worm, *Bombyx mori,* have been transformed using *piggyBac* (TAMURA *et al.* 2000), *P*-elements (KIM *et al.* 2003) and *Minos* (UCHINO et al. 2007) transposable elements. Besides insects, transposable elements have been useful to transform plants (BAKER *et al.* 1986; VANSLUYS *et al.* 1987) and vertebrates (LARGAESPADA 2003). *Sleeping Beauty* and *Tol2* transposable elements from fish, *piggyBac* element from moth and *Frog Prince* element from frog have all been used to transform human and mouse cell lines (DING *et al.* 2005; IVICS *et al.* 1997; KAWAKAMI and NODA 2004; MISKEY *et al.* 2003). *Sleeping Beauty* has also been used to achieve stable chromosomal integrations and long term transgene expression in mice (HORIE *et al.* 2001; YANT *et al.* 2000).

The successful use of transposable elements to transform vertebrate cells has led to research towards developing transposable element mediated gene-therapy in

humans (IZSVAK and IVICS 2004). There are various transposable elements such as *L1* elements, *Tol2, Tc1, Tc3, Himar1, Mos1, Minos* and *Sleeping Beauty* that have been found to be active in human and mouse cell lines (LARGAESPADA 2003). *Sleeping Beauty* is a synthetic *Tc1/mariner* element derived from defective elements in Salmonid fish genome. *Sleeping Beauty* is active in both mouse and human cell lines and has been useful for germ line as well as somatic cell transgenesis in mice. The idea of using transposable elements as gene delivery vectors for therapeutic purposes has been tested using the *Sleeping Beauty* transposon in mice. Five percent of hepatocytes expressed the *lacZ* gene when a plasmid containing the *lacZ* gene within the transposon was administered into living mice (IZSVAK and IVICS 2004). In another experiment, a *Sleeping Beauty* vector containing a human factor IX expression cassette when administered to hemophilic mice resulted in partial correction of a bleeding disorder. In Fumaryl Acetoacetate Hydrolase (FAH) deficient mice a *Sleeping Beauty*- FAH expressing construct has been able to correct a lethal recessive hereditary disease in 62 % of the experimental animals (IZSVAK and IVICS 2004). Even though work needs to be done to increase the transposition efficiency of transposable elements and also to control the target specificity of their insertion, they hold promise as a non-viral gene transfer tool for gene therapy in humans (FESCHOTTE 2006; IZSVAK and IVICS 2004).

While transposable elements have been used largely as tools to modify and study individual organisms they are now being considered as tools to manipulate natural populations. Transposable elements not only enter new genomes by horizontal transfer but also become ubiquitous in the natural populations of the newly invaded

species by transposition and vertical transmission alone. One of the most convincing examples is the horizontal transfer of *P*-elements from *Drosophila willistoni* to *Drosophila melanogaster* and the spread of *P*-elements throughout world populations of *D.melanogaster* in a few decades (ANXOLABEHERE *et al.* 1988). This ability of transposable elements to increase in frequency is the basis of their proposed use as a genetic drive agent to spread transmission-blocking genes in natural populations of mosquitoes to control vector-borne diseases such as malaria. Vector-borne diseases, as the name suggests, are the ones in which the disease-causing pathogen needs an organism (vector) to be transmitted from infected to uninfected individuals. The proposed strategy is to spread gene(s) that render vectors incapable of transmission through natural populations of the vector using transposable elements in a fairly short time as a way to control vector-borne diseases.

Burden of vector- borne diseases especially malaria

Vector-borne diseases such as malaria, trypanosomiasis, encephalitis, leishmaniasis, filariasis, onchocerciasis and dengue collectively account for more than 1.5 million deaths per annum around the world (HILL *et al.* 2005) (Figure 1-1)**.** Malaria, the most important vector-borne disease is estimated to cause around one million deaths per year. Malaria is the third highest pathogen-specific cause of death in the world after HIV/AIDS and tuberculosis. However, the morbidity caused by a disease determines its true impact and can be assessed using DALY (Disability-Adjusted Life Years), where one DALY is defined as one lost year of healthy life, and is a measurement of the difference between the current health of a population and an ideal situation where everyone in a population lives into old age in full health.

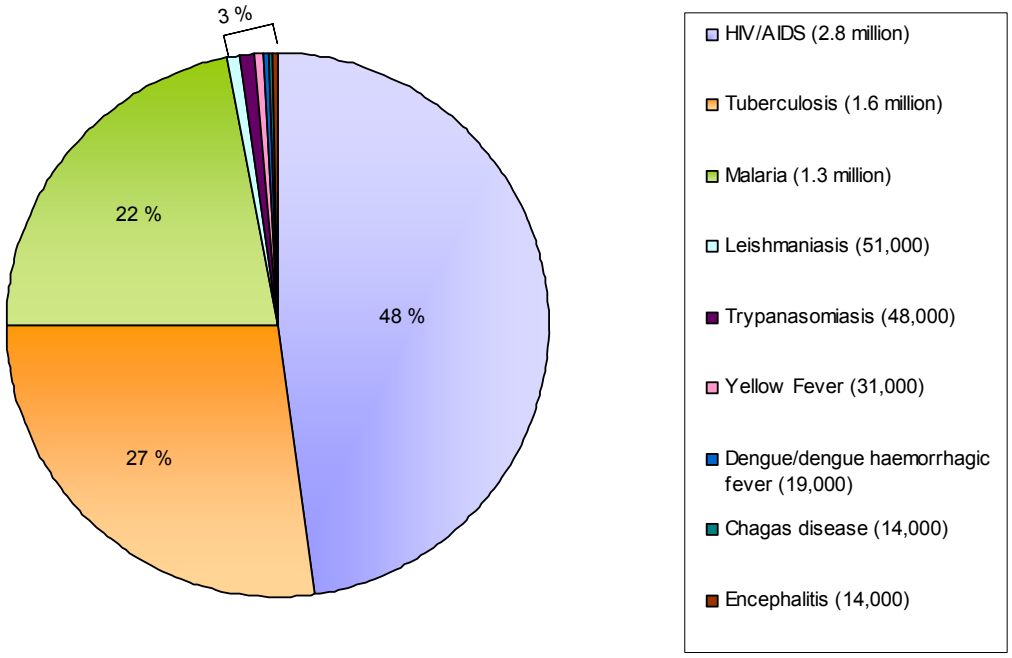


FIGURE 1- 1: Global estimates of human mortality caused by vector-borne diseases compared to HIV/AIDS and Tuberculosis.
The total number of human deaths due to various vector-borne diseases in comparison to two non-vector borne diseases-HIV/AIDS and tuberculosis is shown. **[data and idea from Hill *et al.*]**

Malaria causes severe fever, anemia, fatigue and other serious complications affecting the normal health and life of the people with the disease. Thus, when judged in terms of loss of normal health, the disease burden due to malaria which infects around 500 million people a year is much more than tuberculosis (HILL *et al.* 2005).

Malaria

Malaria is caused by the protozoan parasite *Plasmodium*, and is transmitted from infected individuals by the female *Anopheles* mosquito. There are approximately 430 species of *Anopheles* of which only 30-40 transmit malaria. *An.gambiae* is the most potent vector in sub-Saharan Africa where malaria is most prevalent. There are more than 100 species of *Plasmodium* that can infect animals of which only four; *P.vivax, P. ovale, P.malariae* and *P.falciparum* infect humans. *P.falciparum* causes, cerebral malaria, the most severe and fatal malaria. Some of the symptoms of cerebral malaria are abnormal behavior, seizures, coma, severe anemia and cardiovascular shock and it can lead to death if not treated within 24 -72 hours. *P.falciparum* is common in Africa where it is responsible for 90% of the deaths caused by malaria. *P.vivax* is found mostly in Asia, Latin America and some parts of Africa and rarely causes death, but it can be incapacitating and contribute to the disease burden of malaria. The other two species of *Plasmodium* are less frequently encountered.

The etiology of malaria has been understood for over a century and yet it still remains one of the most deadly diseases in the world. Mosquitoes are absolutely required for the development and transmission of the parasite; without the mosquitoes there would be no transmission and no disease. Thus, the battle against malaria has

been most successful when vector-control was implemented and successful. Preventing human contact with vectors has been effectively done by two important vector-control strategies, pesticide-treated bed nets and insecticides. Mosquito control programs using insecticides were successful for a short time during 1950's and 1960's when they were well implemented, after which there was a re-emergence of malaria by the 1970's (GUBLER 1998). Multiple factors-such as insecticide resistant mosquitoes, drug resistant parasites, unavailability of vaccines and also socio-economic factors in the endemic regions have been responsible for the re-emergence of the disease. India and Sri Lanka are examples of two countries that are seeing a resurgence of malaria due to the discontinuation of vector-control programs, complacency and reduction of financial and political support for control/elimination programs started in the 1950's (GUBLER 1998). The magnitude of complexity involved in the control/elimination of this disease makes it unlikely that there will be one solution to this problem. The loss of effectiveness of promising tools (for example, insecticides) requires the development of new approaches and complementary strategies to be taken to control this disease.

Genetically modified mosquitoes and population modification: A new approach

Mosquitoes are obligate hosts for the development and transmission of the malarial parasite, *Plasmodium.* Thus, eliminating this host (mosquito) is a highly effective way of eradicating malaria. The use of insecticides, such as DDT, has been very successful in reducing malaria transmission in the past by eliminating mosquitoes. However, sporadic use of insecticides has resulted in insecticide resistant mosquitoes that have contributed to the re-emergence of the disease. Even though

there are confounding problems of insecticide resistance and environmental concerns, DDT still remains an efficient strategy to provide protection and safety to an enormous number of people at very low costs (TREN and BATE 2001). Although killing of mosquitoes using insecticides is effective, it may not be necessary in order to control the disease. The actual problem is not the mosquito but its ability to transmit the parasite. So, eliminating this ability of mosquitoes without actually killing them is an ecologically better solution. Craig (1963) suggested using genetic technology to create refractory mosquitoes that are unable to transmit the parasite and then modify wild populations of mosquitoes such that all of them acquire this property (WHITTEN 1985). Genetic manipulation of insects was made possible by the genetic transformation of *Drosophila melanogaster* in 1982 (RUBIN and SPRADLING 1982). However, two major hurdles have to be accomplished before we can use this new approach to disrupt malaria transmission in nature, 1) create a refractory strain of mosquitoes and 2) develop a method to modify existing populations of mosquitoes with the desired properties.

Generation of refractory mosquitoes

The nature of the life cycle of the parasite in the mosquito presents us with multiple opportunities to interfere with its development and transmission. The life cycle of *Plasmodium* in mosquitoes starts when the mosquito ingests an infected blood meal. *Plasmodium* gametocytes present in the infected blood mature to form male and female gametes which fuse and become diploid zygotes. The zygotes quickly develop into motile ookinetes that penetrate the mid-gut epithelium and differentiate into oocysts on the basal surface of the gut epithelium. In two weeks, the

oocyst ruptures releasing thousands of haploid sporozoites into the mosquito haemocoel. The sporozoites invade and emerge in the ducts of the salivary gland tissues. Upon feeding on a vertebrate host the mosquito injects saliva along with parasites thereby infecting another host (Figure 1-2). Of all the gametocytes ingested by the mosquito with the blood meal, only 10% develop into ookinetes and of these only 20% mature into oocysts (BLANDIN *et al.* 2004). Each oocyst produces thousands of sporozoites, some of which (10%) invade the salivary gland (SINDEN 2002) and are transmitted to a new host. Because the parasite population is reduced right after it enters the mosquito and remains low until the oocyst stage, efforts to modify mosquitoes to impair their transmission abilities have focused on pre-sporozoite stages of *Plasmodium* development (RIEHLE *et al.* 2003).

Multiple strategies are being considered to interfere with the parasite development in the mosquito (ITO *et al.* 2002; KIM *et al.* 2004; MOREIRA *et al.* 2000). These strategies involve either the expression of novel effector molecules or altering the expression of endogenous effector molecules that result in inhibition of the parasite development. Some effector molecules are toxic to the parasite while others block the activity of parasite-expressed proteins that are important for the parasite invasion of different tissues in the mosquito. There are other effector molecules that interfere with parasite and mosquito receptor interactions. Also, altering the expression of certain innate immune effector molecules has resulted in the inhibition of parasite development. Transgenic *Anopheles* mosquitoes with reduced vector competence have been generated with at least three effector molecules. Cecropin A is an innate immune effector that is synthesized in response to *Plasmodium* infection in

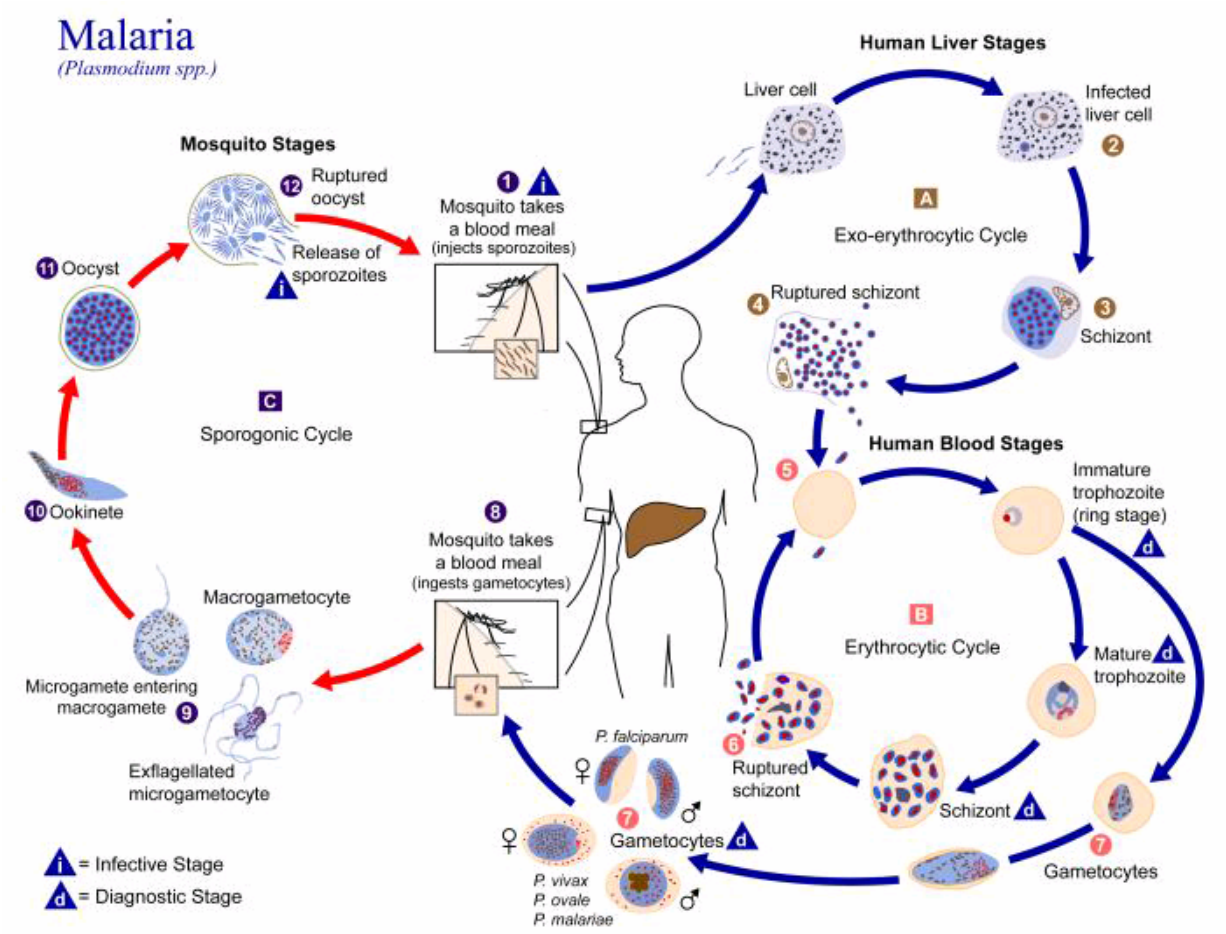


FIGURE 1- 2: Life cycle of *Plasmodium* parasite in mosquito and human hosts.
From CDC public domain: Content provider -Alexander J. da Silva, PhD and Melanie Moser

mosquitoes (DIMOPOULOS *et al.* 2001; WATERHOUSE *et al.* 2007). The parasite is able to escape the effect of Cecropin A and other innate immune effectors by its ability to invade tissues where these molecules are not synthesized. Altering the expression of one such immuno-peptide, Cecropin A (*cecA*) such that it is expressed 24 h after a blood meal in the posterior mid-gut using an *Aedes aegypti carboxy peptidase* promoter resulted in ~61% reduction in the oocyst number of *P.berghei* in transgenic *An.gambiae* (KIM *et al.* 2004). *Bee venom phospholipase A2 (PLA2)* is another effector; when this gene is expressed using a gut specific and blood meal-inducible *An.gambiae* promoter in transgenic *An.stephensi* has an 87% reduction of *P.berghei* oocyst number (MOREIRA *et al.* 2002). SM1 (Salivary gland and mid-gut binding protein) is a synthetic peptide that when also expressed using the same promoter in transgenic *An.stephensi* resulted in 81.6% reduction in *P.berghei* oocyst number (ITO *et al.* 2002).

In the three cases described above, two different strategies were used to disrupt the development of the *Plasmodium* parasite. *PLA2* and SM1, for instance, interfere with the interaction of the parasite and the mid-gut cell surface while the expression of Cecropin A, an immune response effector, was altered to inhibit the development of the parasite. Other effector strategies that affect parasite gene expression or act as anti-parasite toxins are also being tested for their anti-*Plasmodium* capacity (NIRMALA and JAMES 2003) and generation of a refractory strain without the ability to transmit seems achievable.

Population modification

Identification and use of effector genes to generate transgenic insects with

reduced vector competence is encouraging. However, successful use of genetically modified mosquitoes to control vector-borne diseases depends upon introducing reduced vector competence into wild populations of mosquitoes, which can be achieved in two ways – population replacement and population modification. Population replacement would involve an inundated release of the refractory mosquitoes following a significant reduction in the natural population of mosquitoes (using insecticides). This approach requires the production of a large number of refractory mosquitoes and it may not be possible to produce sufficient number of mosquitoes to achieve population replacement for a country or a continent. In contrast, population modification requires only the production of a manageable number of refractory mosquitoes as it relies on a (genetic drive) mechanism to rapidly increase the frequency of the refractory transgene in natural populations of the mosquitoes.

A few mechanisms - such as meiotic drive (segregation distorter) (WOOD *et al.* 1978), use of homing endonuclease genes (BURT 2003), bacterial symbionts like *Wolbachia* (BEARD *et al.* 1998), and linking transgenes to autonomous Class II transposable elements (KIDWELL and RIBEIRO 1992)- have been suggested to rapidly increase the frequency of refractory genes in natural populations of mosquitoes. Unfortunately, the critical step of linking the transgenes to the drive mechanisms has not been demonstrated except for transposable elements. Transposable elements are used as a gene vector to transform mosquitoes but their ability to drive refractory genes to fixation in populations is yet to be demonstrated. There have been natural cases of expansion in the frequency of transposable elements, such as the rapid

increase in frequency of *P*-elements and their spread across the world populations of *Drosophila melanogaster* in a few decades (ANXOLABEHERE *et al.* 1988). However, there never has been a deliberate attempt to achieve this.

Models that simulate the spread of transposable elements in populations predict that transposable elements can be used to spread refractory genes to fixation and achieve the required impact on the disease under certain conditions. Ribeiro and Kidwell (1994) developed a simple population model to describe the expected change in frequency of transposons with a specific intergeneration transmission rate, *i*, after their introduction into a population (RIBEIRO and KIDWELL 1994). *i* is a measure of the infectivity, or the ability of a transposable element to jump to another chromosome, and can have a value between 0 and 1. When $i=1$, the frequency of transposon-bearing gametes derived from a cross of an individual carrying a transposable element (T) and a wild-type individual (W) increases from 0.5 to 1. The fitness of transposon bearing individuals can be lower than the wild-type individuals because of the deleterious effects caused by the transposon jumping. Ribeiro and Kidwell (1994) found that the element spread rapidly and became fixed if the transmission rate (*i)* was greater than 45% of the fitness cost to individuals bearing the elements. In other words, if the infertility caused by the transposition is less than 45%, the element spreads to fixation. They also found that a release ratio of $\leq 1\%$ of a large population was sufficient for spread under these conditions.

Kiszewski and Spielman (1998) used a spatially explicit model to reexamine the expected dynamics of transposon spread. Their model had about 300 villages with each containing about 100 mosquito breeding sites and they assumed a transmission

rate of i=1 (100% transmission) in all their simulations. They found that a transposon needs to have less than 30% fitness cost in order to spread and become fixed in each of the mosquito subpopulations. They also found that environmental factors played a role in the spread of transposons; when they assumed a short dry season with a continuous level of breeding, the transposon cost on fitness can not be more than 20% to achieve a spread. Large releases did not promote fixation, especially when breeding seasons were long. When the transposon bearing individuals were randomly released throughout the modeled regions, they found that fixation was achieved more rapidly than if they were released in an aggregated fashion. A more recent model that combined both population genetic and epidemiological ideas concluded that the efficacy of the genetic drive can be fairly low (~ 40%) for refractoriness to reach fixation; however, the efficacy of refractoriness needs to be 100% for this strategy to eradicate malaria (BOETE and KOELLA 2003). Nevertheless, these models do suggest that if certain requirements are met then transposable elements can be used to spread refractory genes to fixation to disrupt malaria transmission.

Transposable elements - a promising genetic drive system

The ability of transposable elements to move and increase in copy number makes them good candidates for a genetic drive system (CURTIS 2003; KIDWELL and RIBEIRO 1992; RIBEIRO and KIDWELL 1994). Class II transposable elements move by a cut-and-paste mechanism without any RNA intermediate; increase in copy number in this case is brought out by the DNA repair mechanisms of the cell which uses the homologous chromosome as a template (Figure 1-3). Because transposable elements move and increase in copy number, they are inherited in frequencies greater than the

expected Mendelian ratios. For instance, the cross between an individual heterozygous for a gene and an individual without the gene results in 50% of offspring that have the gene (Mendelian ratio). However, if the gene is a transposable element and it jumps to the homologous chromosome then all the offspring resulting from the cross have the element. So, the frequency of inheritance is greater than the Mendelian ratio (50% in this case) and the transposable elements are therefore said to have a transmission advantage (Figure 1-4a). Class II transposable elements based on their transposition rate, pattern of jumping and the timing of jump can have a high transmission advantage. For instance, if the transposition event is pre-meiotic as opposed to post-meiotic, the transmission advantage would be much higher (Figure 1-4b). Transposable elements will thus increase in frequency in a population as long as their transmission advantage is greater than the fitness cost due to their random insertion into genes.

A notable example of such an increase in frequency in a population is the rapid increase of the frequency and spread of *P*-elements in the world populations of *D.melanogaster (*ANXOLABEHERE *et al.* 1988). Studies indicate that *P*-elements have been introduced into *D.melanogaster* from *D.willistoni* by horizontal transfer and they have spread by transposition and vertical transmission alone to become ubiquitous in natural populations within a few decades (ANXOLABEHERE et al. 1988). Another example is the spread of the *hobo* element in *D.melanogaster*; *hobo* elements were probably introduced into the *D.melanogaster* genome in the 1950s (PASCUAL and PERIQUET 1991; PERIQUET *et al.* 1989a; PERIQUET *et al.* 1989b) and

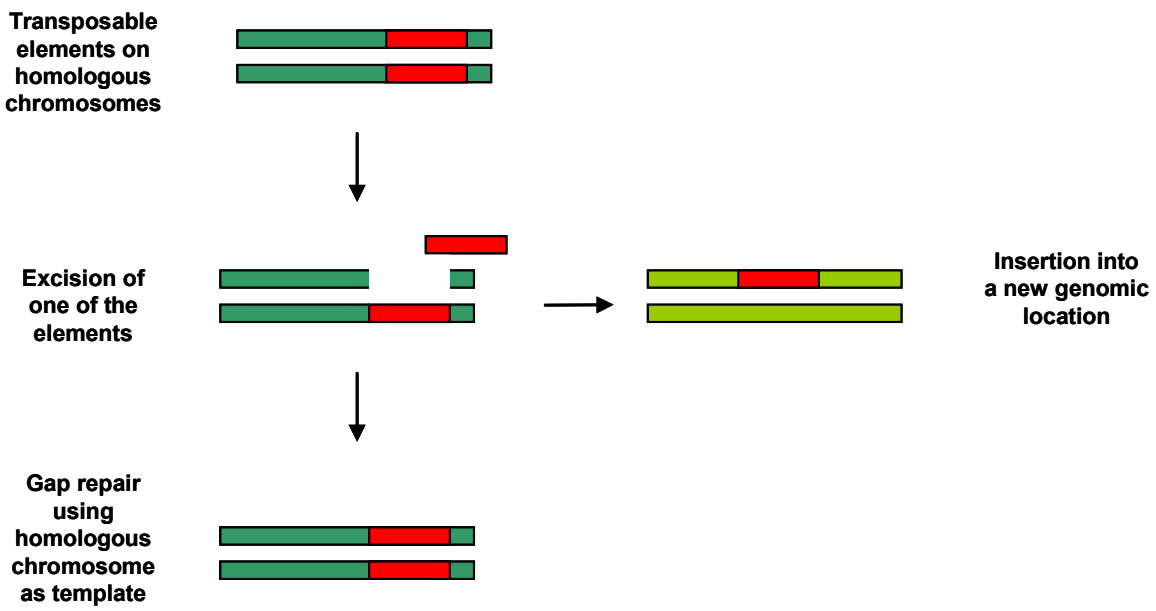


FIGURE 1- 3: "Replication" of a DNA transposon.
Excision of a DNA transposable element, results in a chromosomal break, which is repaired by the DNA repair mechanism of the cell that results in an increase in copy number.

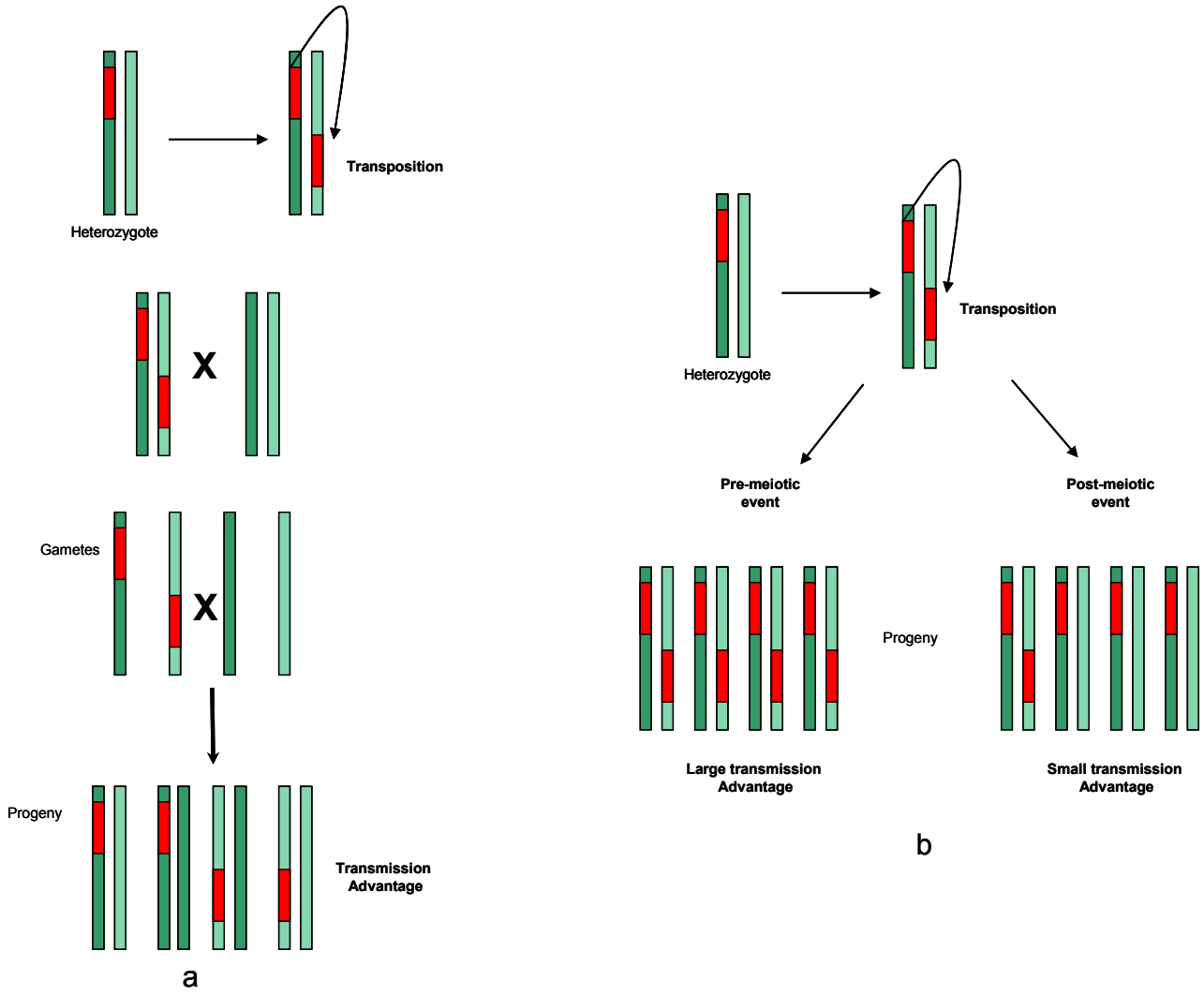


FIGURE 1- 4: Transmission advantage of transposable elements.

a. Transposition to the homologous chromosome results in the transposable element being inherited by all the offspring as opposed to only 50% if there was no transposition
b. Depending on the timing of the transposition event, transposable elements can have a bigger transmission advantage

have spread through the world's populations within the last 50 years. Because of the ability of transposable elements to spread through populations, any gene - such as PLA2 or *cecA* that targets the parasite - can be "driven" to high frequencies by linking it to an appropriate transposable element.

Evolution of transposable elements

According to Hartl *et al* (1997), transposable element evolution in an organism has three phases. In the first phase, right after an element is introduced into a genome, the element increases in copy number as a result of high rates of replicative transposition (Invasive phase). As a result of a number of forces, such as natural selection and evolution of repression systems, the activity of the transposable element is regulated and the copy number tends to reach an equilibrium (Equilibrium phase). During this phase, the rate of loss of elements due to excision is equal to the increase in number of elements due to replicative transposition. This phase is followed by inactivation of functional elements (autonomous elements) due to deletions and mutations leading to the gradual loss of elements (that are now fixed) and eventual extinction due to drift (Stochastic loss phase) (Figure 1-5).

Even though this model of transposable element evolution seems to apply to all transposable elements studied so far, the specifics such as the length of each phase can vary depending on the element and species under question. For instance, even though *P*-elements have been in *Drosophila melanogaster* for less than a century, most of the elements are internally deleted (ENGELS *et al.* 1990). The accumulation of internally deleted defective forms of the element may not be rapid as seen by the widespread occurrence of intact forms of *Hermes* elements in *Musca domestica*

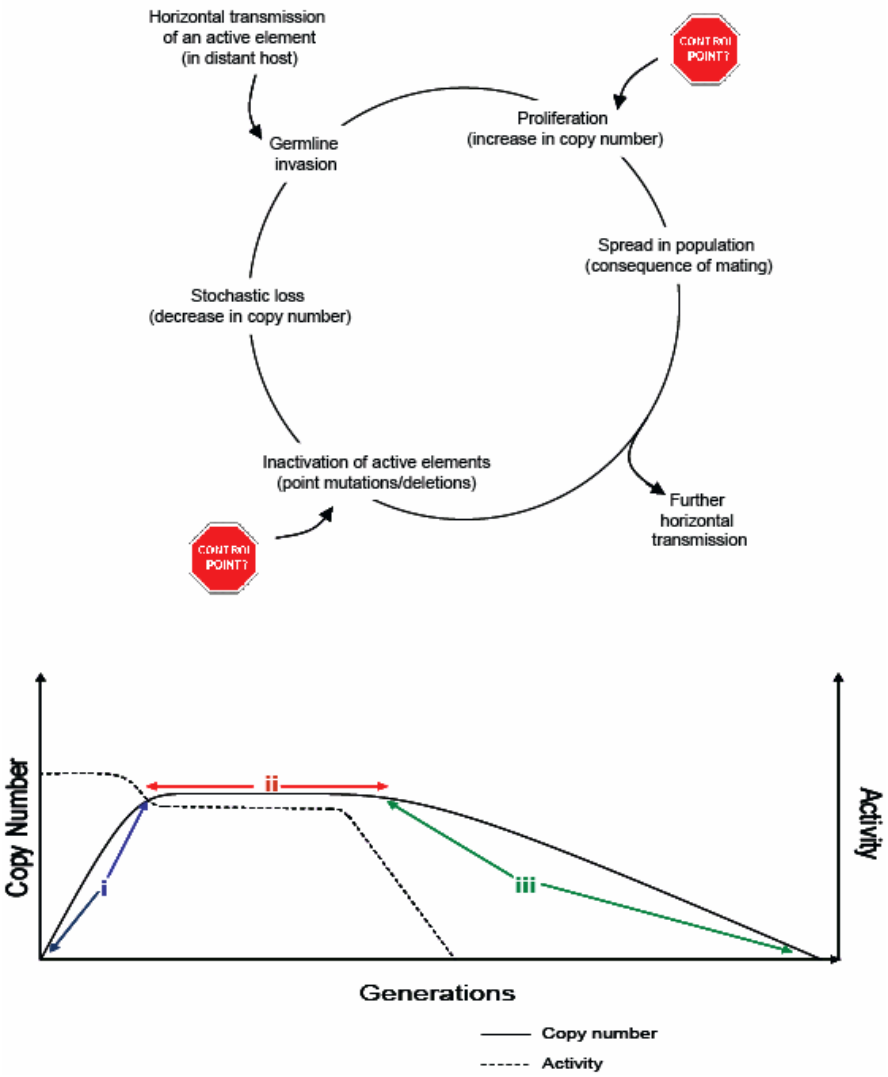


FIGURE 1- 5: A model of the life cycle of transposable elements
Horizontal transfer of an active transposable element into an organism results in increase in copy number initially (i. Invasive Phase) due to high activity but after some time there is a decrease in activity due to repressive forces resulting in equilibrium (ii. Equilibrium Phase) when the increase in copy number due to transposition is equal to loss of elements by excision, this is followed by loss of functional elements from the population which leads to less or no activity leading to eventual extinction of the element from the population (iii. Stochastic loss Phase) (idea from (HARTL *et al.* 1997b)

(L A. Cathcart, E S. Krafsur, P W. Atkinson, D A. O'Brochta and R A. Subramanian, unpublished) or *hobo* elements in *Drosophila melanogaster* (GALINDO *et al.* 1995; YAMASHITA *et al.* 1999). Some host genomes may be more accessible to transposable element invasions than others and also the host regulatory mechanism may vary depending on the transposable element in question. *P*-elements seem to have evolved a self-regulatory mechanism by deleted forms of the elements called *KP*-elements. Given the possibility that Class II transposable elements may serve as genetic drive agents it is important to understand their evolution in the target species, *An.gambiae*.

Transposable elements in *An.gambiae*

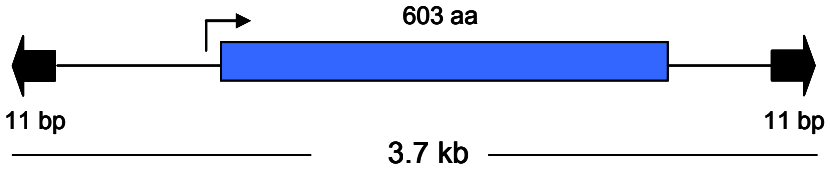
A large portion of the *Anopheles gambiae* genome is composed of transposable elements. Transposable elements form 16% of the euchromatin and 60% of the heterochromatic regions of the genome (HOLT *et al.* 2002). At least 50 different families of transposable elements have been identified in the *An.gambiae* genome and represent all major families of transposable elements (ARENSBURGER *et al.* 2005; BESANSKY *et al.* 1996; BIEDLER and TU 2003; BIESSMANN *et al.* 1999; DE CARVALHO *et al.* 2004; GROSSMAN *et al.* 1999; QUESNEVILLE *et al.* 2003; TU and COATES 2004; TU 2001). But, there have been no studies at a population level to understand the evolution and behavior of these elements in the mosquito. Studies of transposable elements at a population level are critical for our understanding of the consequences of using transposable elements as genetic drive agents in this species. These studies will also be helpful in understanding the requirements that have to be met for the successful spread of refractory genes using transposable elements in the natural populations of this mosquito species.

Questions addressed in the thesis

The ability of transposable elements to rapidly increase in frequency and spread in natural populations (ANXOLABEHERE et al. 1988; KIKUNO *et al.* 2006; PERIQUET *et al.* 1989b) make them good candidates for a genetic drive system to spread refractory genes in mosquito populations to control vector-borne diseases. However, the limited number of studies in *Drosophila* and none in the target vector, *Anopheles gambiae* at the population level does raise some concern. Before attempting an intentional release of genetically manipulated mosquitoes with a genetic drive system into a natural population, the consequences of such an approach needs to be fully explored. Even though transposable elements have shown the ability to spread in natural populations, the requirements and circumstances in which a successful spread can occur needs to be understood. I have attempted to understand this by studying the contemporary activity of endogenous elements in the species, *Anopheles gambiae*. I studied the dynamics of two Class II transposable elements *Herves* and *Topi* in the natural populations of *Anopheles gambiae* to gain a better understanding of the evolution of DNA transposable elements in this medically important insect.

Herves was discovered as a result of an effort to identify active *hAT* family of transposable elements (ARENSBURGER *et al.* 2005) (which includes ***h****obo* from *D.melanogaster*, ***A****c* from maize and ***T****am3* from *Antirrhinum majus*). *Herves* has a typical structure of a Class II transposable element, i.e. it is 3.7 kb long with 11bp inverted terminal repeats flanking an open reading frame coding for a 603 amino acid transposase protein (Figure 1-6). It was active in transposition assays in both

a. *Herves* transposable element



b. *Topi* transposable element

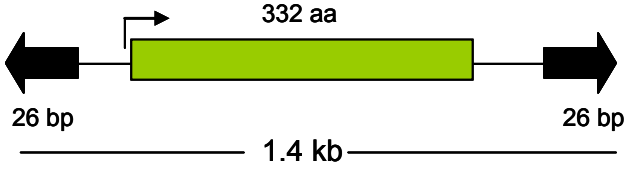


FIGURE 1- 6: Class II transposable elements in *An.gambiae*

a. *Herves* transposable element: Has a size of ~3.7 kb, 11 bp inverted terminal repeats and a 603 amino acid transposase

b. *Topi* transposable element: Has a size of ~1.4 kb, 26 bp inverted terminal repeats and a 332 amino acid transposase

Drosophila S2 cells and embryos, as well as *Aedes aegypti* embryos (ARENSBURGER *et al.* 2005); P Arensburger and P W Atkinson, unpublished results). *An.gambiae* is a species complex with six morphologically indistinguishable species: *An.gambiae s.s, An.arabiensis, An.merus, An.melas, An. bwambae* and *An.quadrianulatus. Herves* was detected in all the members of the *An.gambiae* species complex except *An.bwambae* for which data is not available. *Topi* belongs to a *Tc1/mariner* superfamily of transposable elements (that includes *Tc1* from *C.elegans* and *mariner* from *Drosophila mauritiana*. It has 26 bp inverted terminal repeats and a coding region encoding a 332 amino acid full-length transposase enzyme (Figure 1-6). It was found to be in 17-31 sites in the genome (GROSSMAN *et al.* 1999).

In Chapters 2 and 4, I have tried to understand the dynamics of *Herves* and *Topi* transposable elements in *An.gambiae* by addressing the following questions:

- Is the element active in the natural population?
- Is the element currently invading the natural population in Africa?
- How long has the element been in the species?
- What is the frequency of intact forms of the element?
- Is the evolution of the *Topi* element similar to *Herves*?

I took a population genetics approach to address these questions. *An.gambiae s.s* samples from 6 different locations in Africa (mostly in East Africa and one in West Africa) were used for the analysis. Site-occupancy frequency distribution was used to determine the distribution, copy number and activity parameters. PCR of the

internal region of the element was used to assess the structure of the element. Nucleotide sequence data from both coding and non-coding regions of the element were used to analyze the patterns of geographic distribution, diversity, and residence time and also the selection pressure in the transposase coding regions.

The basis of the study of the *Topi* transposable element described in Chapter 4 was mainly to obtain some comparative data and to assess if the findings of Chapter 2 were general features of all Class II transposable elements in *An.gambiae*.

Some of the important features that I observed in Chapter 2 with *Herves* were evidence of recent activity, high frequency of complete forms of the element, higher level of conservation of coding region of transposase and also evidence of purifying selection in this region. This led me to the questions addressed in Chapter 3

- Is there a source of functional *Herves* transposase in natural populations of *An.gambiae*?
- Has only one form of active transposase been selected for in the natural populations?
- Are there any shared forms of *Herves* transposase between different members of the species complex?

I took a biochemical approach to address these questions. Different forms of *Herves* transposase were identified in three members of *An.gambiae* species complex. The variant *Herves* transposases were expressed and purified from *E.coli* and their activity assessed by an *in vitro* assay.

Chapter 2: Transposable element dynamics of the *hAT* element *Herves* in the human malaria vector, *Anopheles gambiae s.s.*

Ramanand A. Subramanian,* Peter Arensburger, † Peter W. Atkinson,† David A. O'Brochta*[1]

**Center for Biosystems Research; University of Maryland Biotechnology Institute; 9600 Gudelsky Drive, Rockville, Maryland, 20850, †Department of Entomology, University of California, Riverside, California, 92521-0314*

Genetics 2007 Aug; 176: 2477-87

ABSTRACT

Transposable elements are being considered as genetic drive agents for introducing phenotype-altering genes into populations of vectors of human disease. The dynamics of endogenous elements will assist in predicting the behavior of introduced elements. Transposable element display was used to estimate the site occupancy frequency distribution of *Herves* in six populations of *Anopheles gambiae s.s.* The site occupancy distribution data suggest that the element has been recently active within the sampled populations. All 218 individuals sampled contained at least one copy of *Herves* with a mean of 3.6 elements per diploid genome. No significant differences in copy number were observed among populations. Nucleotide polymorphism within the element was high ($\pi = 0.0079$ in non-coding sequences and 0.0046 in coding sequences) relative to that observed in some of the more well-studied elements in *D. melanogaster*. In total, 33 distinct forms of *Herves* were found based on the sequence of the first 528 bp of the transposase open reading frame. Only 2 forms were found in all six study-populations. Although *Herves* elements in *An. gambiae* are quite diverse, 85% of the individuals examined had evidence of

complete forms of the element. Evidence was found for the lateral transfer of *Herves* from an unknown source into the *An. gambiae* lineage prior to the diversification of the *An. gambiae* species complex. The characteristics of *Herves* in *An. gambiae* are somewhat unlike those of *P* elements in *D. melanogaster*.

INTRODUCTION

hAT elements comprise a large and prevalent group of Class II transposable elements found in a wide range of plants and animals (KEMPKEN and WINDHOFER 2001; KUNZE and WEIL 2002; RAY *et al.* 2007). *hAT* elements are not only of interest for their role in genome evolution but also as tools for genetically modifying organisms, with the elements *Hermes* and *hobo* being two examples of *hAT* element-derived insect gene vectors (BLACKMAN *et al.* 1989; O'BROCHTA *et al.* 1996). Transposable elements from other families such as *piggyBac*, *Mos I* and *Minos* have also been developed into effective insect gene vectors that are now employed in a variety of applications (ATKINSON *et al.* 2001b). Using these relatively new gene-integration tools, a novel form of biological control is being considered to stem the transmission of certain arboviruses (e.g. Dengue) and parasites (e.g. *Plasmodium*) by mosquitoes and other arthropod vectors (ADELMAN *et al.* 2002; ALPHEY *et al.* 2002; BEARD *et al.* 2002). This strategy involves the introduction of transgenic insects into natural populations of a target species with the intent of replacing the native population with genetically modified con-specifics (ANONYMOUS 1991; CRAIG 1963; JAMES 1992; MILLER 1992). Introduced transgenic mosquitoes will contain transgenes conferring incompatibility (refractoriness) or resistance to the target pathogen or parasite. An increase in the frequency of the transgene within natural

populations of the vector will, under certain conditions, lead to a reduction or elimination of vector-borne disease transmission (BOETE and KOELLA 2002).

Designing gene vectors and effector transgenes for refractoriness such that they will increase in natural populations and eventually reach fixation is a considerable challenge and transposable elements may provide a means by which this can be accomplished (BRAIG and YAN 2001). The replicative nature of transposable element movement (even by elements that move by a cut-and-past fashion i.e. Class II elements) results in elements acquiring a transmission advantage, resulting in their gradual increase in frequency in populations (KISZEWSKI and SPIELMAN 1998; RIBEIRO and KIDWELL 1994). The magnitude of that transmission advantage is determined by the rate of transposition, the degree to which transposition is conservative or replicative, the spatial patterns of element transposition within a genome, the biology of the transposable element and its interactions with the host insect, and the size, structure and characteristics of the target population (RASGON and GOULD 2005).

While intra-species spreading of transposable elements through transposition has been observed in nature following recent horizontal transfer events involving transposable elements (e.g. *P* and *hobo* elements), population modification has never been attempted by the deliberate and intentional release of an active autonomous transposable element into natural populations of insects (ROBERTSON 2002). Predicting the outcome of such an intentional release of transgenic insects containing active autonomous transposable element gene vectors is an enormous challenge but one that must be successfully met if population replacement biological control using

transposable elements is to be successful (ALPHEY *et al.* 2002). Data that might inform those predictions include an understanding of the dynamics of endogenous Class II transposable elements within the host insect. Endogenous elements are likely to reveal temporal and spatial patterns of spread as well as how population structure has influenced those patterns. Currently our understanding of the population dynamics of Class II transposable elements in insects is based almost entirely on studies of *P* and *hobo* elements in *D. melanogaster* and closely related species (ANXOLABEHERE *et al.* 1990; ANXOLABEHERE *et al.* 1988b; BUCHETON *et al.* 1992; SILVA and KIDWELL 2004; SIMMONS 1992). These studies have documented the ability of these elements to spread rapidly through populations and for the elements to become structurally modified over time, most often by internal deletion. The propensity of these elements to accumulate internal deletions rapidly has raised a serious concern about using transposable elements as transgene spreading agents, namely, the frequent loss of transgenes. Maintaining tight linkage between the anti-parasite effector gene and the associated gene drive system has been repeatedly stated as an essential characteristic of this biological control strategy (CURTIS 2003; JAMES 2005). To what extent these characteristics of *P*, *hobo* and *mariner* elements are general characteristics of Class II elements remains to be fully explored. Because a proposed target species for this novel population replacement-based biological control strategy is the human malaria vector, *Anopheles gambiae,* the study of Class II transposable element dynamics in this species is particularly relevant.

Recently, a functional *hAT* element, *Herves*, was discovered in *An. gambiae*, providing an opportunity to examine the dynamics of an active Class II transposable

element in this insect (ARENSBURGER *et al.* 2005). *Herves* is notably different at the sequence level from the well-studied *hobo* element from *D. melanogaster* and *Hermes* from *Musca domestica*, sharing only about 20% amino acid identity with these elements (ARENSBURGER *et al.* 2005). A *Herves* element isolated from the RSP strain of *An. gambiae* that was established as a laboratory colony in the early 1990s (VULULE *et al.* 1994) was shown to be transpositionally active in laboratory-based mobility assays in *D. melanogaster* (ARENSBURGER *et al.* 2005) and *Aedes aegypti* (P. Arensburger and P. Atkinson, unpublished). A recent study of the element's abundance and site-occupancy frequency in natural populations of *An. gambiae s.s.*, *An. merus*, and *An. arabiensis* in Mozambique revealed that it was present in all three species at approximately 5 copies per diploid genome and site-occupancy frequency distributions suggested that *Herves* had been recently active in the three species examined (O'BROCHTA *et al.* 2006). In the population of *An. gambiae* examined in Mozambique, 95% of the individuals tested contained intact (non-deleted) forms of the element, which is quite unlike *P* elements in *D. melanogaster* in which most elements are internally deleted derivatives of the canonical element (O'HARE *et al.* 1992). Here *Herves* has been investigated in six populations of *An. gambiae* using a variety of methods to see if the characteristics of the element observed in Mozambique were general features of the element and how it compares to other well-studied Class II elements.

MATERIAL AND METHODS

Collection Site: *Anopheles gambiae s.s.* from six populations were used in this study with sample sizes ranging from 15-94 individuals (Figure 2-1). Samples

from Asembo Bay (hereafter referred to as Asembo), Kisian and Malindi have been described (LEHMANN *et al.* 2003b). Asembo and Kisian are located in western Kenya and were sampled in 1994 and 1996 respectively (LEHMANN *et al.* 2003b). Malindi, located in eastern Kenya, was sampled in 1996 (LEHMANN *et al.* 2003b). The northeastern region of Tanzania was sampled in 2004 in the region in and around the village of Zenet (MEERAUS *et al.* 2005). Samples from southern Mozambique (Furvela) were collected in 2003 as described (O'BROCHTA *et al.* 2006). Samples from north-central Nigeria (Bakin Kogi) were collected in 1999 (LEHMANN *et al.* 2003b).

DNA Isolation: Genomic DNA was isolated from individual mosquitoes as described (O'BROCHTA *et al.* 2006) and resuspended in 100 µl of distilled water and stored at -80ºC.

Species Identification: Species identification was performed using the method of Scott et al. (1993) as described (O'BROCHTA *et al.* 2006) using $1/100^{th}$ of the total genomic DNA from a single mosquito in a volume of 1µl (SCOTT *et al.* 1993). This method permits the identification of species-specific polymorphisms in the intergenic spacer region of ribosomal RNA genes using PCR. Only *An. gambiae s.s.* samples yielding unambiguous species identification results were used in subsequent analyses.

Transposable element display: Transposable element display is a PCR-based DNA fingerprinting method derived from the Amplified Fragment Length Polymorphism (AFLP) method (VOS *et al.* 1995). It was performed here as described previously with only minor modifications (O'BROCHTA *et al.* 2006). Transposable

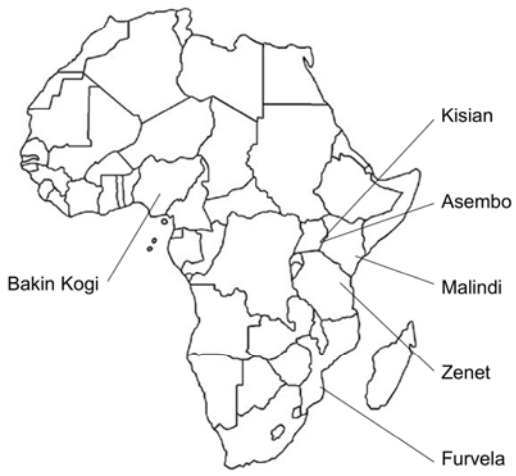


FIGURE 2- 1: Political map of Africa showing locations of sample populations.

element display was performed in triplicate using 2-5 μl (approximately 200ng) of genomic DNA for each replicate. Genomic DNA was digested for 4 hours in a volume of 40 μl at 37°C with 4 units of the restriction endonuclease *Mse*I using conditions recommended by the manufacturer (New England Biolabs). Sixty picomoles of adapters were ligated to the *Mse*I digestion products by adding 10 μl of 1X restriction enzyme buffer containing 5 mM ATP, 50 mM DTT (dithiothreitol), 10 μg BSA (bovine serum albumin), 4 units of *Mse*I, 1 Weiss unit of T4 DNA ligase and incubated at 37°C overnight. The adapters were prepared by mixing equimolar amounts of oligonucleotides HhaIa (5' GAT GAG TCC TGA GTA CG 3') and MseIb2 (5' TAC GTA CTC AGG ACT CAT CAA G 3'), heating them to 100°C for 10 minutes and then allowing the mixture to very slowly cool to room temperature. The design of the adapters and the digestion/ligation reaction conditions result in the efficient creation of only monomeric *Mse*I-cut genomic DNA fragments with terminal adapters.

Five microliters of the restriction/ligation reaction were used as the template in a polymerase chain reaction ("preselective PCR") performed in a 50 μl reaction volume containing 1X PCR Buffer II (Applied Biosystems), 0.2 mM dNTPs (an equimolar mixture of dATP, dTTP, dCTP, dGTP), 2.5 mM $MgCl_2$, 1 unit AmpliTaq® DNA polymerase (Applied Biosystems), and 24 pmoles of primer HhaIa and primer HervTEDAL1a (5' ATT TCG ACG GGT TCC TAC C 3'). HervTEDAL1a is a *Herves*-specific primer that anneals to sequences approximately 150 bp from the 5' end of the element. The DNA polymerase was added as a complex with TaqStart™ Antibody (ClonTech) as described by the manufacturer for

the purpose of "hot-starting" the reaction. The reaction conditions were 95°C/3 mins followed by 25 cycles of 95°C/15 sec, 63°C/30 sec, 72°C/1.0 min and a final cycle of 72°C/5 min. A second PCR was performed ("selective PCR") using 5 μl of the preselective PCR products as template in a 20 μl reaction containing 1X PCR Buffer II, 0.2 mM dNTPs, 2.5 mM $MgCl_2$, 1 unit AmpliTaq® DNA polymerase (bound to TaqStart™ Antibody as above), 9 pmoles each of primers HhaIa and Cy5™-labeled HervTEDAL2 (5' GTT GAT TAG ATG AAC GTA GG 3'). The Cy5™-labeled primers were purified by HPLC prior to their use. HervTEDAL2 anneals to sequences approximately 80 bp from the 5'end of the element. Following a denaturation step at 95ºC for 3 minutes "touchdown" PCR conditions were created in which during the first 5 cycles the annealing temperature was decreased 1°C after each cycle with the first of these cycles being 95°C/15 sec, 64°C/30 sec, 72°C/1.0 min. Following these 5 cycles an additional 25 cycles were performed at 95°C/15 sec, 60°C /30 sec, 72°C /1.0 min with a final cycle of 72°C/5 min.

To visualize products of transposable element display 5 μl of selective PCR products were mixed with 5μl of loading buffer (95% deionized formamide, 10mM EDTA) heated to 95°C for 3 minutes, cooled quickly on ice and 6 μl were loaded on a 6% polyacrylamide gel (19:1 acrylamide : bisacrylamide) containing 6.7 M urea in 1X TBE buffer (90 mM Tris-borate, 2 mM EDTA). ALFExpress™Sizer™50-500 (Amersham/Pharmacia) was used as a size standard. Electrophoresis was performed at 70 watts (constant) for 2.5 hours at which time the gel was transferred to 3MM filter paper and dried. The dried gel was scanned on a STORM 860 phosphoimager (Molecular Dynamics). The products obtained from the three independent replicate

reactions of the same sample were run on the same gel to assist with determining the presence of bands. Based on the combined results of three transposable element display experiments a band was called as present or absent if it was unambiguously present in at least 2 of the 3 replicates. Determining the presence of bands in this way resulted in a single scoring matrix that was then used in subsequent analyses.

Site-occupancy frequency distributions were estimated using transposable element display data. Using the frequency distributions and assuming the model of Charlesworth and Charlesworth (1983) the model parameter β, that measured, in part, the forces removing insertions from natural populations, was estimated. The model parameter β is equal to the product of four times the effective population size and the rate of element loss. Estimation of β and the copy number of *Herves* per diploid genome were performed as described by Wright *et al.* (2001) who considered the dominant nature of transposable element display signals and the application of the parameter estimation methods of Charlesworth and Charlesworth (1983) to diploid organisms. Note that although each sample was analyzed three times for transposable element display these replicates were used to produce a single scoring matrix. The advantage of this procedure is that it increased the accuracy of determining the presence of bands and minimized errors that tend to result in overestimations of β.

Transposase Open Reading Frame Detection: To assess *Herves* open reading frames for the presence of deletions and insertions, PCR primers were designed that were complementary to sequences flanking the transposase ORF: 1372f (5' CCA CAA ATT GAT CTA CGC TCC 3') and 3469r (5' GAT GCA TCT ATT

ATG ATT AAG GC 3'). One fiftieth of the genomic DNA from one mosquito (2 μl) was used as template in a 50μl reaction containing 1X ThermalAce™ (Invitrogen), 0.2 mM dNTPs (an equimolar mixture of dATP, dTTP, dCTP, dGTP), 2.5 mM $MgCl_2$, 2 units ThermalAce™ DNA polymerase (Invitrogen), and 24 pmoles of primer1372f and 3469r. Amplification reactions were performed under the following conditions: 95°C/3 min followed by 30 cycles of 95°C/30 sec, 48°C/30 sec, 72°C/3.0 min and a final cycle of 72°C/10 min. Reaction products were fractionated on a 1% agarose gel. PCR products of the samples that failed to produce a detectable product following one round of PCR were used as templates (5μl) in a second PCR under the same conditions described above but with primers 1407f (5' GAT CAA AGG TAA CAT TAG TCT TG 3') and 3294r (5' CCA TGT TAC AAA TTT TGC AAC G 3') and rechecked on a 1% agarose gel. Open reading frames free of deletions and insertions yielded PCR products 2100 bp after the first PCR and 1900 bp after the second PCR. We estimate that elements with deletions as small as 100 bp would be detectable using this strategy.

Sampling and PCR for population analysis: Transposable element display permitted occupied sites to be identified and these data were used in determining the composition of the subset of individual mosquito genomic DNAs that would be used in the analysis of sequence diversity of 1474 bp of the non-coding region and the first 528 bp of the transposase open reading frame. This selected subset of individual mosquito genomic DNAs was such that *Herves* elements at most occupied sites, as determined by transposable element display, were included in the PCR template pool. So, a total of 49 individuals containing elements at the 130 different sites identified

by transposable element display were included in the PCR template pool to give us an opportunity to amplify *Herves* elements inserted at different genomic sites within the populations. Using this subset of genomic DNAs a portion of the left end of the element was amplified using a nested PCR strategy. Five microliters of genomic DNA from each of the 49 individuals were used as template in separate 20µl reactions containing 5X Phusion HF Buffer (NEB), 0.2µM dNTPs (an equimolar mixture of dATP, dTTP, dCTP, dGTP), 0.4 units Finnzymes Phusion™ DNA polymerase (New England Biolabs; error rate = 4.4×10^{-7}), and 24 pmoles of primer 24F (5' TAG AGT TGT GCC TCA AGA ACC AGA 3') and primer 2035R (5' TGG TTC AGG TTT GTC CAT CC 3'). Amplification reactions were preformed under the following conditions: 98°C/1 min followed by 25 cycles of 98°C/10 sec, 65°C/30 sec, 72°C/1 min 30 sec and a final cycle of 72°C/10 min. Reaction products were fractionated in a 1% agarose gel. PCR products from samples that failed to produce detectable products on an agarose gel following one round of PCR were used as templates (5µl) in a second PCR under the same conditions described above using primers 24F (5' TAG AGT TGT GCC TCA AGA ACC AGA 3') and 2002r (5'GCT ATA GCT TTG GCG GTC G 3') and rechecked on a 1% agarose gel. The 2kb amplification product was eluted from the gel, precipitated, resuspended in 20 µl dH_2O and cloned into the pCR®-Blunt II TOPO vector (Invitrogen). Up to five clones per individual were sequenced and these sequences were used in subsequent analyses. From samples "Zenet", "Asembo", "Bakin-Kogi", "Kisian", "Furvela" and "Malindi" a total of 57 (GenBank accessions EF588609-EF588665), 51 (EF588428-EF588478), 40 (EF588479-EF588518), 29 (EF588552-EF588580), 33 (EF588519-EF588551) and

28 (EF588581-EF588608) sequences, respectively, were obtained. Note, the methods used to obtain the sequences for this analysis did not permit these elements to be assigned to specific sites identified in the site-occupancy (transposable element display) analysis.

Sequence Analysis: Sequences were aligned using AlignX, a ClustalW-base alignment program in VectorNTI Advance 10.0.1 (Invitrogen). Nucleotide diversity was estimated from average pair-wise number of differences between elements, π (NEI and LI 1979) and from the number of polymorphic sites, θ (WATTERSON 1975). π and θ were estimated using DnaSP 3 (ROZAS and ROZAS 1995; ROZAS *et al.* 2003). Estimates of the observed silent site diversity in the first 528 bp of the 5' end of the transposase coding region was computed using the Kumar method (NEI and KUMAR 2000) as implemented in MEGA 3.1 (KUMAR *et al.* 2004b). Expected values of silent site diversity were calculated following Sanchez-Gracia et al. (2005) and were the product of the haploid copy number and the average synonymous diversity (0.0209) from a sample of 35 nuclear genes (MORLAIS *et al.* 2004). Tajima's D was calculated using DnaSP 3. Further analysis was performed on the first 528 bp of the 5' end of the transposase open reading frame. Unique variants of elements were identified (referred to as forms), their frequencies determined and the relationship of the forms determined using TCS1.21 (CLEMENT *et al.* 2000). Alignment gaps were treated as missing data in this analysis. Estimates of the number of synonymous substitutions per synonymous site (dS) and of non-synonymous substitutions per non-synonymous site (dN) and their ratio, ω= dN/dS, were obtained using maximum likelihood (ML) methods employed by CODEML in PAML 3.13 (YANG 1997) using the alignment of

the 33 different forms for the analysis (Supplemental Figure2-1). PAML permits an assessment of the observed substitution data after assuming different codon substitution models that differ in the way selection pressure is distributed within the gene. Here we have examined our data in light of three simple models: a single ratio model (M0) that assumes one ω for all sites, a neutral model (M1) that assumes that there are two classes of sites within the gene; those that are conserved (p_0) with $\omega_0=0$ and those that are neutral ($p_1=1- p_0$) with $\omega_1=1$, and finally, a discrete model (M3) that assumes three classes of sites each having a unique value of ω that is estimated from the data (YANG 1997). A likelihood ratio test (LRT), which is twice the log-likelihood difference between two models being compared, was used to determine which model best reflected the observed data. The LRT statistic has a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters between the two models (YANG *et al.* 2000).

RESULTS

Site Occupancy: Transposable element display has been a useful tool for assessing the number and position of transposable elements within the genome of individual organisms (BIEDLER *et al.* 2003; GUIMOND *et al.* 2003; WRIGHT *et al.* 2001). As performed here, templates longer than 1 kb are likely to be poorly represented because the length of the extension reactions during PCR was only one minute. Because the *An. gambiae* genome is composed of 64.8% adenines and thymines and we produced PCR templates by digesting the genomic DNA with *Mse*I (TTAA) we expected only 0.004% of the resulting fragments to be 1 kb or more in length. (We estimated this by determining what percent of the fragments greater than

80 bp were over 1kb in length. Eighty base pairs is the invariable amount of *Herves* DNA contained in each PCR product. We assumed fragment sizes following *Mse*I digestion would have an exponential distribution with $\lambda = 0.324^4$. Therefore, 0.415 of all fragments were calculated to be greater than 80 bp and 0.0017 of all fragments were greater than1 kb.) Consequently, few elements will be undetected because they are on excessively long templates. Restriction site polymorphism can result in increased estimates of site occupancy diversity since an element at one site would be displayed as two bands of different lengths resulting in those bands being scored as two elements occupying two sites. While restriction site polymorphism will have this effect on the analysis, the frequency of such polymorphisms is expected to be very low based on the known level of nucleotide polymorphism in *An. gambiae* s.s. (MORLAIS *et al.* 2004) and our failure to detect the same element in two different positions in transposable element displays following band isolation, reamplification and DNA sequencing (GUIMOND *et al.* 2003) and (R. A. Subramanian and D. O'Brochta, unpublished). Confounding effects of restriction site polymorphism will be small and are not a significant source of variation in transposable element display.

In this study all individuals in this study that were analyzed by transposable element display (n = 218) contained at least one *Herves* element (Table 2-1). Element copy numbers within the six populations analyzed ranged from 2.9-4.4 elements per diploid genome as calculated using the method of Wright et al (2001). No individuals were found in any population that contained more than 7.0 elements. In all populations there was an abundance of occupied sites that were observed in only small numbers of individuals (Figure2-2). In Zenet, Malindi and Furvela elements

TABLE 2- 1: Site occupancy of *Herves* elements

Location	N[a]	κ[b]	dcn[c]	β[d]
Asembo	24	25	3.5	9.5
Kisian	15	14	2.9	2.9
Malindi	25	17	3.4	11.0
Zenet	73	31	3.8	2.1
Furvela[e]	49	23	4.4	1.9
Bakin-Kogi	32	20	3.3	2.3

[a] Individuals analyzed by transposable element display

[b] Number of unique chromosomal sites containing *Herves*

[c] Diploid copy number of *Herves* (WRIGHT *et al.* 2001).

[d] $4N_e(\nu+s)$ from Charlesworth and Charlesworth (1983) where N_e is the effective population size, ν is the excision rate and s is the strength of selection against element insertions.

[e] Data from O'Brochta *et al.* (2006)

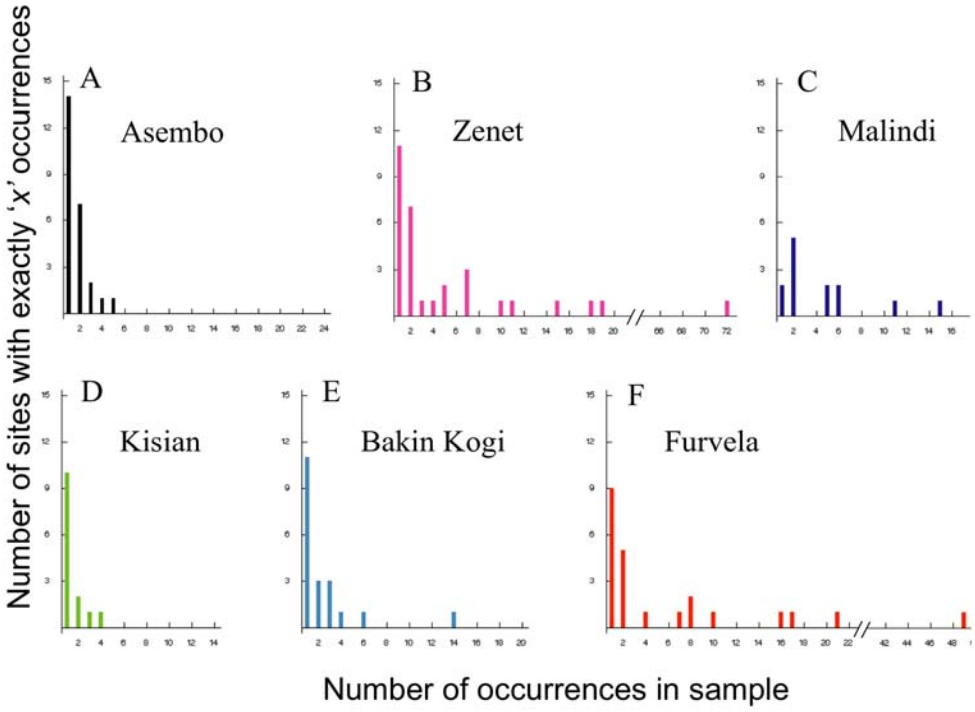


FIGURE 2- 2: Site occupancy frequency distribution.
A-F. The number of sites that were found in a sample exactly "*x*" times is plotted on the *x*-axis and the site occupancy is plotted on the *y*-axis.

with high site occupancy frequencies were observed although none of these elements were shared among these populations (Figure 2-2).

Charlesworth and Charlesworth (1983) and Langley *et al*. (1983) provided theoretical frameworks for understanding site occupancy frequency distributions, which could also be used to estimate element mobility rates under certain conditions. Both models can be expressed using a single parameter (β), assume that the elements are at equilibrium and that there are an infinite number of insertion sites within the genome. According to the models (CHARLESWORTH and CHARLESWORTH 1983; LANGLEY *et al.* 1983) parameter values greater than one indicate the existence of forces other than drift (mobility and/or selection) that have played a major role in shaping the observed distribution. In this study estimates of β were, in all cases, greater than one suggesting that element mobility and/or selection played a significant role in shaping the observed distribution (Table 2-1).

Nucleotide Polymorphism: Approximately 2 kb of sequence beginning at the left (5') inverted terminal repeat and through the first 528 bp of the transposase open reading frame was amplified, cloned and sequenced (Figure 2-3). A total of 238 sequences containing the first 528 bp of the transposase open reading frame were analyzed from six different locations. The average nucleotide polymorphism in the 1474 bp of non-coding sequence ($\pi = 0.0079$) was significantly different from the polymorphism observed in the coding region ($\pi = 0.0046$; $P < 0.001$) (Table 2-2). Within the non-coding region the observed polymorphisms were non-uniformly distributed in a 666 bp region beginning at nucleotide 568 having a highly reduced level of polymorphism (Figure 2-3). This region corresponds to a large stretch of

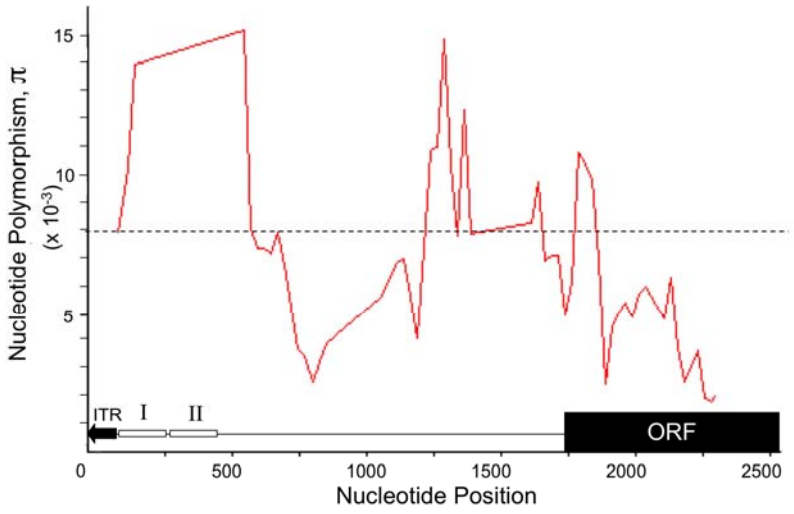


FIGURE 2- 3: Nucleotide polymorphism in *Herves*.
The results of a sliding window analysis (100 bp window in 25 bp steps) showing the levels of nucleotide polymorphism, π, as a function of position within the element. The horizontal dotted line represents the average nucleotide polymorphism reported for 35 *An. gambiae* nuclear genes (MORLAIS *et al.* 2004). ITR, inverted terminal repeat; I, II, subterminal direct repeats; ORF, transposase open reading frame.

TABLE 2- 2 : Nucleotide sequence polymorphism in *Herves*

		Non-coding			Coding		
Location	**Seqs**[a]	**Poly**[b]	π[c]	θ[d]	**Poly**[b]	π[c]	θ[d]
Asembo	51	44	0.0056 (0.0037)	0.0076 (0.0023)	15 (3+12)	0.0034 (0.0042)	0.0063 (0.0023)
Kisian	29	60	0.0086 (0.0009)	0.0128 (0.0043)	7 (1+6)	0.0024 (0.0004)	0.0034 (0.0016)
Malindi	28	44	0.0076 (0.0006)	0.0084 (0.0029)	7 (2+5)	0.0033 (0.0005)	0.0034 (0.0016)
Zenet	57	109	0.0084 (0.0008)	0.0177 (0.0050)	21 (7+14)	0.0057 (0.0009)	0.0104 (0.0035)
Furvela	33	35	0.0091 (0.0004)	0.0079 (0.0027)	8 (5+3)	0.0056 (0.0032)	0.0037 (0.0017)
Bakin-Kogi	40	53	0.0086 (0.0006)	0.0095 (0.0030)	6 (1+5)	0.0015 (0.0003)	0.0028 (0.0014)
Combined	238	124	0.0079 (0.0003)	0.0216 (0.0049)	35 (14+21)	0.0046 (0.0004)	0.0134 (0.0035)

[a] Number of sequences analyzed

[b] Number of polymorphic positions; Numbers in parenthesis = synonymous + non-synonymous sites

[c] Pairwise nucleotide diversity (NEI and LI 1979); standard deviation in parenthesis

[d] Nucleotide diversity based on segregating sites (WATTERSON 1975); standard deviation in parenthesis;

DNA with unknown function 5' of the transposase-coding region and just 3' of a pair of 100 bp sub-terminal tandem repeats (ARENSBURGER *et al.* 2005).

Levels of silent site diversity in *Herves* elements were compared to the average silent site diversity for single-copy host genes (Table 2-3) as part of an effort to look for evidence of lateral introduction of *Herves* into the *An. gambiae* lineage (BROOKFIELD 1986; SANCHEZ-GRACIA *et al.* 2005b). The observed levels of silent diversity among *Herves* elements ranged from 3 to 125-fold less than the silent site diversity seen on average in 35 nuclear genes (MORLAIS *et al.* 2004). In addition, Tajima's D statistic was calculated and found to be insignificant for each location although when calculated based on the pooled data it was significant (1.91; $P<0.05$; Table 2-3) indicating an excess of low frequency variants (TAJIMA 1989).

Structural Integrity: Class II transposable elements can be autonomous or non-autonomous. Autonomous elements code for functional transposase and can undergo transposition. Non-autonomous elements cannot code for functional transposase usually as a result of deletions that remove some or all of the coding region. *P* elements in *Drosophila,* for example, often exist in forms that contain large deletions of internal sequences leaving only terminal and sub-terminal sequences resulting in non-autonomous elements (ENGELS 1989). The complete *Herves* open reading frame is approximately 1.8 kb in length and the structural integrity of *Herves* elements was assessed by amplifying this region using primers flanking it. *Herves* elements without any deletions resulted in PCR products of 2 kb in length and elements with deletions 100 bp or more produced distinct products less than 2 kb. Of the 218 individuals tested from six locations 85% showed evidence of the presence of

TABLE 2- 3: Genetic diversity of *Herves* elements from different locations

Locations	Haploid copy number	π_s [a]			Tajima's D
		Observed	Expected[b]	Observed/Expected	
Asembo	1.8	0.002	0.038	0.053	-1.40 [d]
Kisian	1.55	0.001	0.032	0.031	-0.86 [d]
Malindi	1.7	0.002	0.036	0.056	-1.32 [d]
Zenet	1.9	0.005	0.040	0.126	-1.53 [d]
Furvela	2.15	0.015	0.045	0.334	1.51 [d]
Bakin-Kogi	1.7	0.0003	0.036	0.008	-1.36 [d]
All	1.8[c]	0.006	0.038	0.158	-1.91*

[a] π_s represents the average pairwise nucleotide diversity at synonymous sites.

[b] see Material and Methods.

[c] Average haploid copy number from all locations

[d] $P > 0.05$

* $P < 0.05$

TABLE 2- 4: Frequency of *Herves* Open Reading Frames

Location	*N*[a]	Complete ORF[b]
Asembo	24	1.00
Kisian	15	0.90
Malindi	25	0.88
Zenet	73	0.84
Furvela	49	0.95
Bakin-Kogi	32	0.44

[a] Number of mosquitoes analyzed

[b] Frequency of mosquitoes with evidence of an intact *Herves* ORF (2.1 kb PCR product).

complete open reading frames (Table 2-4). Individuals with complete elements were least abundant in Nigeria (Bakin Kogi) where only 44% showed evidence of complete open reading frames ($N = 32$). In western Kenya intact forms of the element were found in 100% of the individuals from Asembo ($N = 24$) and 90% of the individuals from Kisian ($N = 15$). In eastern coastal Kenya (Malindi, $N = 25$) and northeastern coastal Tanzania (Zenet, $N = 73$) approximately 85% of the individuals tested contained intact forms of the element. In southern Mozambique (Furvela, N=49) 95% of the individuals sampled contained intact elements.

Genealogical Relationships: A genealogical analysis of the *Herves* elements, based on the first 528 bp of coding sequence, was performed and resulted in the identification of 33 forms among the 238 sequences that were analyzed (Table 2-5, Figure 2-4). Form-diversity (the equivalent of haplotype diversity and measured using the same algorithm) varied among locations and ranged from a low of 0.565 in Bakin Kogi to a high of 0.903 in Zenet (Table 2-5). Of the 33 forms, only 2 (Form 1 and Form 2) were found at all six sampling locations (Figure 2-4 and 2-5) and these comprised 51% (n = 238) of the elements analyzed. Twenty-four forms were found at only single locations (Figure 2-5, Table 2-6). Form 2 was the most abundant form in Bakin Kogi, Asembo, Malindi and Kisian (Figure 2-1). In northeastern Tanzania (Zenet) where form-diversity was highest the most abundant form was Form 5, a form that is closely related to Form 2 (Figure 2-4). In southern Mozambique (Furvela) however, a unique form (Form 30) was most abundant and accounted for 21% of the 57 sequences analyzed from this location. Form 30 was highly diverged from the abundant Form 2 and consequently was one of the most unusual elements

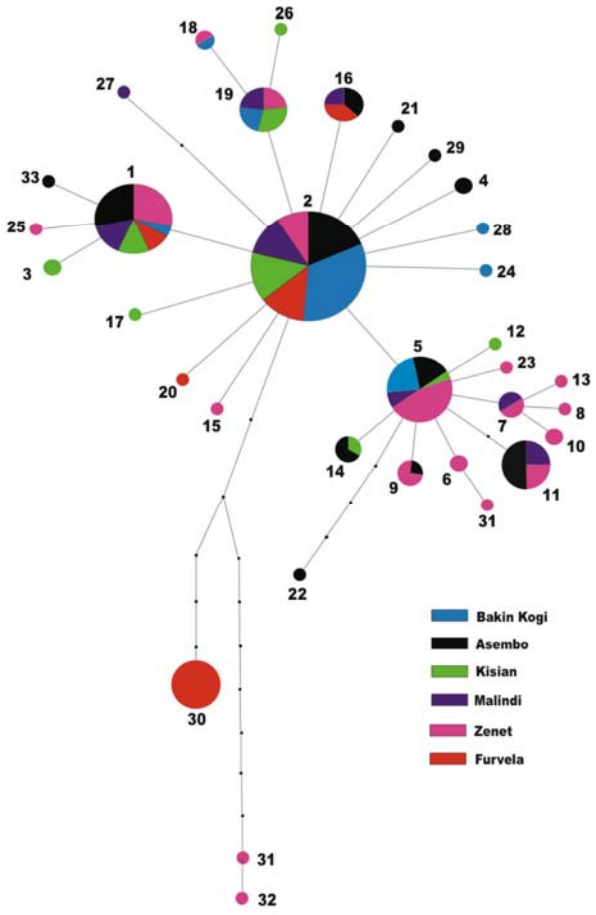


FIGURE 2- 4: Network of genealogical relationships of forms of *Herves* ORFs based on statistical parsimony.

The abundance and relationship of individual forms are shown. Each node represents a single mutational step. The area of the circles is proportional to the form frequency class. Shading refers to the region in which forms were found. In cases where forms are shared among regions, shading is proportional to the frequency of the form in each region. Small black dots represent missing forms. (TEMPLETON *et al.* 1992)

TABLE 2- 5: *Herves* ORF Form diversity

Location	Seqs[a]	Forms	Form diversity [b]
Asembo	51	12	0.857 (0.028)
Kisian	29	9	0.820 (0.055)
Malindi	28	8	0.841 (0.044)
Zenet	57	17	0.903 (0.022)
Furvela	33	5	0.706 (0.049)
Bakin-Kogi	40	7	0.565 (0.088)
Combined	238	33	0.833 (0.018)

a Sequences analyzed

b Standard deviation in parenthesis

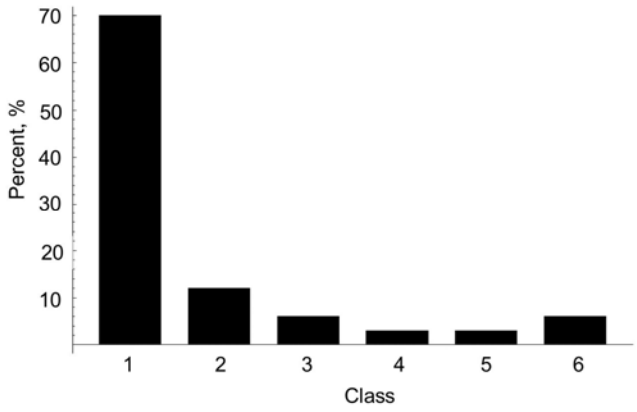


FIGURE 2- 5: Frequency of classes of *Herves* forms.
Herves forms were classified based on the number of locations at which they were found (1-6). The percentage of forms in each class is plotted on the *y*-axis.

TABLE 2- 6: Shared Forms between locations[a]

	Asembo	Kisian	Malindi	Zenet	Furvela	Bakin-Kogi
Asembo	5[b]					
Kisian	4	4[b]				
Malindi	5	4	1[b]			
Zenet	5	4	6	10[b]		
Furvela	3	2	3	2	2[b]	
Bakin-Kogi	3	4	4	5	2	2[b]

[a] Number of Forms shared between locations

[b] Number of Forms found at only this location

encountered in this analysis; only Form 31 and Form 32 from Zenet were more divergent (Figure 2-4). Zenet was unusual among the locations analyzed because it had the greatest number of forms (17), 10 of which were unique to this location. Not only were there a large number of element forms at this location but also the diversity of elements was very high. On average each location had 9.67 forms (± 4.27) and shared 3.6 forms (± 1.4) with other locations (Table 2-6).

Natural Selection: We tested for evidence of selective constraints within the transposase open reading frame by estimating ω (the ratio d_N/d_S) using maximum likelihood. The ω ratios ranged from 0.41-0.71 under all models (M0, M1 and M3; see Material and Methods) revealing evidence of purifying selection (YANG 1997). The neutral model (M1) was rejected when compared to the discrete model (M3) that allows for 3 classes of sites with different values of ω. The LRT statistic, 2Δl (2Δl = 2(-1037.77 - (-1028.00)), for this comparison was 19.54, which was greater than the critical value of $\chi^2_{[0.001,2]} = 13.816$.

DISCUSSION

Understanding the dynamics of active transposable elements in *An. gambiae* will inform predictions concerning the outcomes of biological control efforts by population replacement using transposable elements as gene drive agents. While there have been studies that have looked at the evolutionary history of Class II transposable elements in insects, few studies involving insects other than *Drosophila* have attempted to examine the dynamics of Class II transposable elements at the population level, making the current study of *Herves* in *An. gambiae* somewhat unique.

Here we examined the dynamics of *Herves* by measuring the site-occupancy frequency, nucleotide-sequence diversity and by performing a genealogical analysis of the element. The rare occurrence of locally fixed, *Herves*-occupied sites and the widespread abundance of sites that are occupied in only a few individuals are consistent with there being recent activity of *Herves* within *An. gambiae*. The site-occupancy levels observed in this study (β_{Herves} = 1.9-11.0) were similar or somewhat lower than those reported for putatively active transposable elements in *D. melanogaster*: $\beta_{P\ element}$ = 16.6 (AJIOKA and EANES 1989), $\beta_{P\ element}$ = 5.85 (BIEMONT *et al.* 1994), β_{copia} = 9.79 (BIEMONT *et al.* 1994), β_{copia} = 16.9 (LEIGH-BROWN and MOSS 1987), β_{copia} = 48.3 (KAPLAN and BROOKFIELD 1983).

An. gambiae is distributed almost continuously throughout its range in Africa and demes are likely to be large and diffuse (LEHMANN *et al.* 1998). Little population differentiation between populations separated by up to 50 km has been reported (LEHMANN *et al.* 1997) and this has also been found over distances of 6000 km (LEHMANN *et al.* 1996). Lehmann et al. (1998) suggest that Wright's isolation by distance model may best describe the relationships among populations (WRIGHT 1951). Population admixture might be contributing to the pattern of site-occupancy observed in this study. However, consistent with the idea that *Herves* is currently capable of transposing in natural populations of *An. gambiae* is the finding that *Herves* elements isolated from *An. gambiae* collected from the field within the last 20 years are active when introduced into other insects in the laboratory (ARENSBURGER *et al.* 2005b).

A number of pieces of data indicate that *Herves* entered the *An. gambiae* lineage via a horizontal gene transfer. A comparison of the silent site diversity among *Herves* elements and 35 nuclear genes (MORLAIS *et al.* 2004) revealed less diversity within *Herves* transposable elements than expected assuming similar mutation rates apply to Class II transposable elements and nuclear genes (SANCHEZ-GRACIA *et al.* 2005b). Others have used intra- and inter-specific diversity comparisons to infer the introduction of transposable elements into host genomes (SANCHEZ-GRACIA *et al.* 2005b; SILVA and KIDWELL 2000) and the diversity data for *Herves* is qualitatively similar to those data. Second, when elements are horizontally transferred to a new host species there is a period of time when natural selection will favor active autonomous elements and this will leave a distinct molecular signature within the elements in the form of a skewed ratio of synonymous and non-synonymous substitution rates (ROBERTSON and LAMPE 1995). In this study a comparison of the synonymous and non-synonymous substitution rates within the *Herves* transposase-coding region detected evidence of purifying selection and is consistent with the hypothesis that *Herves* was laterally introduced into this lineage from an unknown source.

Although *Herves* displays evidence of being horizontally introduced into the *An. gambiae* lineage, the timing of this event remains uncertain. The intensity of the molecular signals indicating horizontal transfer suggests that this event was not in the very recent past. Sanchez-Gracia et al (2005) recently examined 14 transposable elements in *D. melanogaster* and, based on silent site diversity, concluded that 13 were products of horizontal transfer that probably occurred approximately 5-12

million years ago. Sanchez-Gracia et al. (2005) observed levels of silent diversity within the transposable elements studied approximately 100-fold less than that observed in 21 nuclear genes while in this study silent site diversity was only 6-fold less than expected when the data were pooled, and ranged from 3-fold to 125-fold less than expected depending on the location from which the samples were collected. These data appear consistent with an historical lateral transfer event, although not one that has occurred recently.

The form diversity observed in this study is also consistent with *Herves* having an extended residence time within the *An. gambiae* lineage. Interestingly however, while the number of forms of *Herves* as determined by the sequence of the 5' end of the transposase gene totaled 33, the frequency of individuals with at least one copy of an element that had either no internal deletions or deletions less than 100 bp (the limits of the detection method) was over 90%. Internally deleted elements can arise quickly following the introduction of a transposable element as has been displayed by the well-studied *P* element in Drosophila species (O'HARE *et al.* 1992). This is distinctly not the case for *Herves* and may be due to a number of factors. First, if deleted elements are preferentially removed from the genome then one would see a relative abundance of intact forms as observed here. Currently there are no data for the differential removal of smaller, internally deleted forms of an element and indeed, smaller non-autonomous elements can have an activity advantage in the presence of functional transposase (LAMPE *et al.* 1998; SPRADLING 1986). An alternative possibility is that *Herves* elements may have reduced opportunities to form internally deleted elements. Internal deletions of Class II transposable elements arise

in some cases during the double-stranded DNA gap repair process following element excision. For example, following *P* element excision in *D. melanogaster* the resulting double-stranded gap is filled during a homology-dependent recombination process in which homologous or ectopic copies of a *P* element are copied into the gap (ENGELS *et al.* 1990). Premature resolution of these recombination products before this templated gap repair process is complete results in the creation of incomplete elements. The extent to which post-excision repair involves homology-dependent recombination or non-homologous end joining will determine, to some extent, how often internally deleted elements are created within a genome (RIO 2002). A preference for *Herves* excision products to be repaired using non-homologous end joining mechanisms could explain two aspects of *Herves* observed in *An. gambiae* – the relative abundance of intact elements and their low copy number.

hAT element excision results in double-stranded breaks in the chromosome in which the ends of each chromosome are sealed by hairpin structures (ZHOU *et al.* 2004). These hairpin structures are resolved by a nicking event followed by end-joining. The hairpin structures that arise on the empty donor site following *hAT* element excision are not seen following *P* element excision. We speculate that this predisposes *Herves* post-excision repair to occur via non-homologous end-joining and thereby reduces the frequency with which internally deleted elements are created.

Herves is present at low copy numbers within *An. gambiae* and the data suggest that copy-number equilibrium has not been reached (Tajima's D statistic for pooled data = -1.91). The low copy number of *Herves*, while not unique among Class II transposable elements, tends to be somewhat unexpected if the element was

introduced into this lineage in the distant past. Class II transposable elements tend to increase in copy number when they are active within a genome. This increase in copy number occurs despite the conservative cut-and-paste nature of Class II element movement because the double-stranded breaks that arise following element excision can be repaired using homology dependent repair processes that result in a copy of the element being inserted into the gap (RIO 2002). Alternatively, an increase in copy number can occur as a result of Class II transposable elements moving from replicated regions of the genome to unreplicated regions of the genome during S-phase (WILSON *et al.* 2003). Although the mechanisms of copy number increase may vary, it seems well established that element copy-number is expected to increase during periods of element activity. The low number of *Herves* elements in all individuals sampled therefore seems at odds with the diversity data that points to an extended residence time in the *An. gambiae* lineage. The tendency of different Class II transposable elements to increase in copy number has never been systematically compared although it is reasonable to think that some elements might be more "replicative" than others. *hAT* elements, and *Herves* in particular, may have a relatively low replication potential because of the presence of hairpin-containing intermediates following excision.

The structure of the population of *An. gambiae* in Africa has been studied and it has been proposed that there are two main divisions of the gene pool – a northwestern division including Senegal, Ghana, Nigeria, Cameroon, Gabon, Democratic Republic of Congo and western Kenya, and a southeastern division including Kenya, Tanzania, Malawi and Zambia (LEHMANN *et al.* 2003b). It has been

proposed that there has been a recent bottleneck in the southeast division resulting in reduced genetic diversity followed by colonization from the northwest division. (LEHMANN *et al.* 2003b). The data presented here shows little evidence of geographical variation and is inconsistent with the above model. Samples from Mozambique showed the highest levels of silent site diversity and no reduction in the diversity of forms as might be expected following a bottleneck. In fact, samples from Nigeria not only showed the least silent site diversity but also had the least amount of form diversity. Further sampling of *Herves* from populations in western Africa is needed to confirm the modest trends revealed in this study.

ACKNOWLEDGEMENTS

The generosity of Tovi Lehmann, Derek Charlwood, Christopher Curtis and Wilhelmine Meeraus for providing samples is gratefully acknowledged. Matthew Hare, Sky Lesnick, Floyd Reed and Subhamoy Pal provided us with valuable advice and useful discussions. The National Institutes of Health, R01GM48102, supported this work.

Chapter 3: Biochemical analysis of natural variants of *Herves* transposase in *An.gambiae*

ABSTRACT

Class II transposable elements have been proposed for use as genetic drive agents to introduce malaria transmission-blocking genes into natural populations of *An.gambiae*. We have studied earlier, *Herves* transposable element in *An.gambiae* as part of our efforts to understand the evolution and behavior of Class II transposable elements in this species. We found that *Herves* was present in all six analyzed locations in Africa, at a low copy number that ranged from 2.9-4.4 per diploid genome. Insertion-site frequency distribution data of *Herves* elements indicated that the elements have been recently active. We found a high frequency (>85%) of individuals with complete forms of the element in most of the locations. Nucleotide sequence diversity analysis showed that the transposase coding region was more conserved than the non-coding region. Also, there was evidence of purifying selection in the *Herves* transposase coding region. All these observations led to the hypothesis that functional sources of *Herves* transposase should be present in natural populations of *An.gambiae*. We tested this hypothesis by sampling *Herves* transposase coding regions in three closely related members of the *An.gambiae* species complex, *An.gambiae s.s, An.arabiensis* and *An.merus.* We found 13 forms that were capable of encoding a full-length *Herves* transposase protein from a total of 67 sequences analyzed. We expressed and purified 9 out of the 13 variant forms of *Herves* transposase in *E.coli.* We found that 7 of the 9 variant *Herves* transposase proteins were active in an *in vitro* strand-transfer reaction. Of the 7 active forms, 4 were

isolated from a sample of 9 individual *An.gambiae s.s* mosquitoes, indicating that 45% of the individuals have a source of functional *Herves* transposase. Despite the availability of transposase, the copy number and the apparent transposition activity of *Herves* are low; suggesting that *Herves* elements in *An.gambiae* might be under the control of a host - regulatory mechanism.

INTRODUCTION

The mobility properties of transposable elements have made them very useful tools with a wide range of applications in the laboratory. Besides their use in the lab, Class II transposable elements have been proposed for use as a genetic drive mechanism to spread refractory genes in natural populations of mosquitoes to control vector-borne diseases such as malaria. Even though the spread of *P*-elements in *Drosophila melanogaster* shows that transposable elements are capable of rapidly increasing in frequency in natural populations (ANXOLABEHERE *et al*. 1988) there never has been a deliberate attempt to achieve this. The outcome of such an attempt to spread refractory genes using Class II transposable elements in mosquitoes is not clear. This is in part due to our limited understanding of the behavior of Class II transposable elements in the target species for such a control, *An.gambiae*.

We have attempted to better understand the behavior of Class II transposable elements in *An.gambiae* by studying the *Herves* transposable element in natural populations of this species in Africa. *Herves* belongs to the *hAT* family of transposable elements that includes *hobo* from *D.melanogaster, Ac* from maize, *Tam3* from *Antirrhinum majus* and *Hermes* from *Musca domestica* (ARENSBURGER *et al.* 2005). We have studied the dynamics of the *Herves* transposable element in 6

different locations of Africa by determining their presence/absence, insertion site-frequency distribution, frequency of complete open reading frames as well as the nucleotide and form ("haplotype") diversity of the *Herves* elements. We observed that *Herves* was present in all of the mosquitoes analyzed with a low average copy number of 3.6 per diploid genome. We observed that there was a high frequency of complete open reading frames (>85%) of *Herves* transposase in most of our locations. Sequence diversity in the transposase coding region ($\pi = 0.0046$) was lower than in the non-coding region ($\pi = 0.0079$) and we detected evidence for purifying selection in the transposase coding region. The insertion site frequency distribution showed an abundance of sites that were rare implying that the elements have been recently active. These findings together with the previously described transpositionally active *Herves* element isolated from the RSP strain of *An. gambiae* that was established as a laboratory colony in the early 1990s (ARENSBURGER *et al.* 2005) led to the hypothesis that functional sources of *Herves* transposase should be present in natural populations of *An.gambiae*.

In this study we tested this hypothesis by sampling transposase coding regions from three members of *An.gambiae* species complex, *An.gambiae s.s, An.arabiensis* and *An.merus.* We identified 13 *Herves* transposase forms that were intact without any pre-mature stop codons in all the three species and expressed and purified 9 out of the 13 proteins in *E.coli.* We tested these variant *Herves* transposase proteins using an *in vitro* strand-transfer assay. Strand-transfer is a step in the transposition reaction, where the transposase catalyzes the joining of the 3'-OH ends of the excised transposable elements to the target DNA. We supplied pre-cleaved *Herves* L-ends

(that have their 3'-OH ends already exposed) together with a target plasmid DNA to the variant *Herves* transposase proteins and tested if they were capable of performing the strand-transfer reaction. This study, besides investigating the presence/absence of a functional transposase in the natural population of *An.gambiae* will also contribute to the structure-function studies of the transposase proteins.

The mechanism of transposition for various bacterial DNA transposons, such as *Tn5*, *Tn7, Tn10,* have been studied both *in vitro* and *in vivo* (CRAIG 1997; HANIFORD 2006; KLECKNER *et al.* 1996; PETERS and CRAIG 2001; REZNIKOFF 2003). The mechanisms of transposition of eukaryotic transposable elements, such as *P*-elements, *hobo*, *mariner* and *Minos*, have been extensively studied in *Drosophila*. Other eukaryotic transposable elements, such as *Mos1* from *Drosophila mauritiana, Hermes* from *Musca domestica*, *Tc1* and *Tc3* from *C.elegans* have also been studied (AUGE-GOUILLOU *et al.* 2005; AUGE-GOUILLOU *et al.* 2001; MICHEL and ATKINSON 2003; MICHEL *et al.* 2002; MICHEL *et al.* 2003; VANLUENEN *et al.* 1994). Additional insights into the mechanism of transposition and the activity of transposases has been gained from the crystal structures of *Mos1*, *Hermes,Tc3, IS200, Tn5* and TnSA (catalytic component of *Tn7* system) proteins (DAVIES *et al.* 1999; HICKMAN *et al.* 2005; LEE *et al.* 2006; RICHARDSON *et al.* 2006; VANPOUDEROYEN *et al.* 1997).

The results obtained from this study would be helpful to identify functional forms of *Herves* transposase as well as to assess their frequency in the natural populations of *An.gambiae*. The sequences of these forms could be compared to the known and predicted characteristics of the *hAT* transposase proteins contributing to our knowledge of the structure and function of this family of transposases.

MATERIALS AND METHODS

Samples: Nine individuals from *An.gambiae* s.s, 4 from *An.arabiensis* and 5 individuals from *An.merus* were randomly selected. Of the 9 individuals from *An.gambiae s.s*, 3 were from Furvela, Mozambique, 4 were from Kisumu, Kenya, and 2 from Malindi, Kenya. All the *An.arabiensis* and *An.merus* were from Furvela, Mozambique. All of these samples have previously been used and described in our earlier studies (O'BROCHTA *et al.* 2006; SUBRAMANIAN *et al.* 2007).

DNA Isolation: Genomic DNA was isolated from individual mosquitoes as described (O'BROCHTA *et al.* 2006) and resuspended in 100 µl of distilled water and stored at -80ºC.

Species Identification: Species identification was performed using the method of Scott et al. (1993) as described (O'BROCHTA *et al.* 2006) using $1/100^{th}$ of the total genomic DNA from a single mosquito (SCOTT *et al.* 1993). This method permits the identification of species-specific polymorphisms in the intergenic spacer region of ribosomal RNA genes using PCR.

Screen for variant *Herves* transposase forms: To screen for variant *Herves* transposase open reading frames, the region containing the transposase was amplified using PCR primers that were complementary to sequences flanking the transposase ORF: 1372f (5'-CCA CAA ATT GAT CTA CGC TCC-3') and 3469r (5'-GAT GCA TCT ATT ATG ATT AAG GC-3'). One fiftieth of the genomic DNA from one mosquito (2 µl) was used as template in a 50µl reaction containing 1X ThermalAce™ (Invitrogen), 0.2 mM dNTPs (an equimolar mixture of dATP, dTTP, dCTP, dGTP), 2.5 mM $MgCl_2$, 2 units ThermalAce™ DNA polymerase (Invitrogen), and 24 pmoles

of primer1372f and 3469r. Amplification reactions were performed under the following conditions: 95°C/3 min followed by 30 cycles of 95°C/30 sec, 48°C/30 sec, 72°C/3.0 min and a final cycle of 72°C/10 min. Reaction products were fractionated on a 1% agarose gel. The ~2100 bp amplification product was eluted from the gel, precipitated, resuspended in 20 µl dH_2O and cloned into the pCR®-Blunt II TOPO vector (Invitrogen). 1-5 clones were sequenced depending on the cloning efficiency. The *Herves* transposase open reading frame sequences were then translated using the "Translator" tool available on www.fr33.net to identify sequences that did not have any pre-mature stop codons and were capable of encoding full-length proteins.

***Herves* transposase expression and purification:** The variant *Herves* transposase forms that were capable of producing a full-length *Herves* transposase were then PCR amplified from the respective pCR®-Blunt II TOPO plasmids and cloned between NcoI and Hind III sites of pBAD/Myc-HisA (Invitrogen) to generate a Herves-Myc-His fusion construct. Note that only 9 of the 13 forms were cloned, the other four forms were not cloned because of cloning difficulties. Also, the *Herves* transposase form (595) which had previously tested positive for transposition activity in *Drosophila* (ARENSBURGER *et al.* 2005) was cloned and used as a positive control in the subsequent analysis. Each pBAD/*Herves*-Myc-HisA plasmid was transformed into *Escherichia coli* strain Top10 (Invitrogen), grown overnight in LB medium containing 100 mg/ml of ampicillin in a shaker at 37 °C. The overnight culture (1:100) was used to inoculate 1L of fresh LB containing ampicillin and cells were grown to an absorbance (A_{260}) of 0.6 at 37°C. The culture was then induced with 0.1% L-arabinose and grown for an additional 16 h at 16 °C. The induced cells were

then washed and centrifuged at 4 ^{o}C with Binding buffer (5mM Imidazole, 500 mM NaCl, 20mM Tric-Cl (pH 7.8), 10 % Glycerol). The pelleted cells were then resuspended in 20 ml of Binding buffer and lysed by sonication. After centrifugation of the sonicated cells, the supernatant was loaded onto a pre-equilibrated Ni^{2+} Sepharose column (Amersham). The column was washed with 4 column volumes of Binding buffer, followed by 6 column volumes of Wash buffer (60mM Imidazole, 500mM NaCl, 20mM Tric-Cl (pH 7.8), 10% Glycerol). The *Herves*-Myc-His fusion protein was eluted using 2 column volumes of Elution buffer (200mM Imidazole, 500mM NaCl, 20mM Tric-Cl (pH 7.8), 10% Glycerol) and dialyzed in three steps against dialysis buffer containing 20mM Tric-Cl (pH 7.8) and 10% Glycerol. The first dialysis step was 1 h long with the dialysis buffer alone; the second step was with fresh dialysis buffer containing 2 mM DTT for another 1 h; followed by a third overnight dialysis in fresh dialysis buffer containing 2 mM DTT and 0.5 mM PMSF. The protein was then stored at -20 °C.

Strand-transfer Assay: The assay was performed as described in Zhou *et al* (ZHOU *et al.* 2004) and adapted for *Herves*. Pre-cleaved *Herves* L-ends were made by annealing oligonucleotides *Herves*LT (5'-TAG AGT TGT GCC TCA AGA ACC AGA ACT GTA CG -3') and *Herves*LB (5'- GTA CAG TTC TGG TTC TTG AGG CAC AAC TCT A -3') radiolabeled at the 5' end with γ- P^{32}-dATP and was used as a substrate for the strand-transfer reaction with 300 ng of pUC19 target DNA. The reaction was carried out in a 10 µl volume containing 25mM MOPS (pH = 7.6), 100mM NaCl, 10mM $MgCl_2$, 5% glycerol, 10 mM DTT, 1mg/ml BSA and 200 ng of *Herves* transposase protein. Reactions were performed at 30 °C for 2 h. The reactions

were stopped by addition of SDS and EDTA to a final concentration of 1 % SDS and 20 mM EDTA and incubating the mixture at 65 °C for 30 minutes. Half of the mixture was loaded onto 1 % TBE agarose gel run at 80 volts for 1h and then dried onto a DE81 filter paper and exposed to a phosphor screen for 45 minutes and scanned on a STORM 860 phosphoimager (Molecular Dynamics). The results were verified by repeating the procedure. The *Herves* transposase form, 595, that is active in *D.melanogaster* and *Aedes aegypti* embryos was used as a positive control in all reactions. A no-protein control was also included and contained distilled water instead of the *Herves* transposase protein.

RESULTS

Intact *Herves* transposase in *An.gambiae s.l:* To identify functional forms in the natural populations of *An.gambiae* in Africa, the *Herves* transposase open reading frame region was amplified, cloned and sequenced from three closely related members of the *An.gambiae* species complex, *An.gambiae s.s, An.arabiensis* and *An.merus.* A total of 67 sequences were obtained, 30 from *An.gambiae s.s,* 17 from *An.arabiensis* and 20 from *An.merus.* Of the 67 sequences, 58 were complete (~1.8 kb) without any deletions and 9 sequences (eight from *An.arabiensis* and one from *An.merus*) had deletions. Out of a total of 58 complete sequences, 5, 2 and 6 sequences from *An.gambiae s.s, An.arabiensis* and *An.merus* respectively did not have any pre-mature stop codons, and were, therefore, presumably capable of producing a full-length *Herves* transposase protein (Table 3-1).

Analysis of *Herves* transposase sequences: The nucleotide sequence diversity of the *Herves* transposase sequences was highest in *An.arabiensis* (π =

0.0092) and lowest in *An.merus* ($\pi = 0.0053$). The *Herves* sequences from *An.gambiae s.s* had a $\pi = 0.0073$. We found 55 different forms among 58 complete sequences from the three members of the *An.gambiae* species complex. There were 3 forms from *An.gambiae s.s* that were recovered twice; however, in each instance the two identical forms were recovered from the same individual making it possible that they were the PCR amplification products of the same *Herves* element. All the other forms were different from each other. A greater number of non-synonymous changes compared to synonymous changes were observed in all three species (Table 3-2). A total of 129 mutations in the transposase coding region in *An.gambiae s.s* (31 synonymous and 98 non-synonymous), 71 mutations in *An.arabiensis* (17 synonymous and 54 non-synonymous) and 78 mutations in *An.merus* (24 synonymous and 54 non-synonymous) were observed (Table 3-2).

The alignment of the 13 "intact" forms of the *Herves* transposase that did not have any pre-mature stop codons with the sequence of a known functional *Herves* transposase revealed some patterns. There were at least six mutations (Thr to Ser, Ile to Val, Ileu to Val, Val to Ala, Ileu to Thr, Tyr to Phe) that were shared between all the forms obtained from *An.merus* (Figure 3-1). There were ten mutations in region B and four mutations in region D that correspond to the catalytic and α-helical domain of *Hermes* transposase respectively (Figure 3-1). Five of the mutations in region A are in a region of *Hermes* transposase that has been shown to be important for the binding of the transposase to the ends of transposon. A tryptophan to cysteine mutation in region C was also seen; the tryptophan residue has been shown to be important for DNA hairpin formation in *Hermes* and *Tn5* transposition reactions

TABLE 3- 1: Summary of the samples used for the analysis

Species	Number of Individuals[a]	Number of Sequences			
		Total[b]	Deleted[c]	Complete[d]	Intact[e]
An.gambiae s.s	9	30	0	30	5
An.arabiensis	4	17	8	9	2
An.merus	5	20	1	19	6
Total	18	67	9	58	13

[a] number of mosquitoes used to amplify the *Herves* open reading frame region

[b] total number of sequences obtained

[c] number of sequences that had deletions in the open reading frame and were smaller than ~1.8kb

[d] number of sequences that were complete with a length of ~1.8 kb

[e] number of complete sequences that had no pre-mature stop codons and could encode a full-length *Herves* transposase protein

TABLE 3- 2: Diversity of *Herves* transposase region in *An.gambiae*

Species	Number of Sequences	Sequence diversity		Form diversity	
		Poly[a]	π[b]	No. of Forms	Form diversity[c]
An.gambiae s.s	30	129 (31+98)	0.0073 (0.0007)	27	0.99 (0.011)
An.arabiensis	9	71 (17+54)	0.0092 (0.0007)	9	1.0 (0.052)
An.merus	19	78 (24+54)	0.0053 (0.0005)	19	1.0 (0.017)

[a] Number of polymorphic positions; Numbers in parenthesis = synonymous + non-synonymous sites

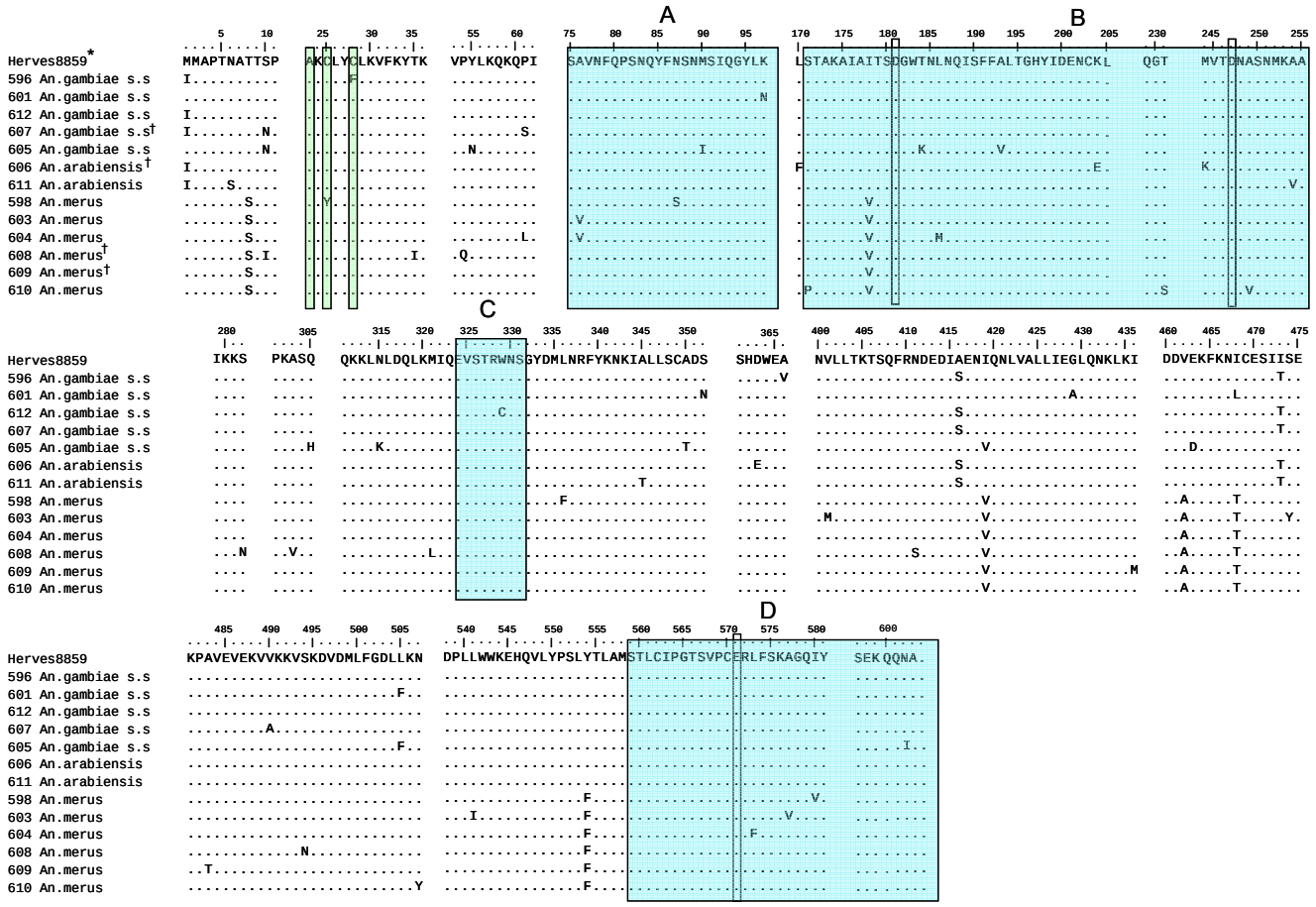
[b] Pairwise nucleotide diversity (NEI and LI 1979); standard deviation in parenthesis

[c] Standard deviation in parenthesis

(ASON and REZNIKOFF 2002; HICKMAN *et al.* 2005). There were two mutations, cysteine to lysine and cysteine to phenyl alanine; involving conserved residues that form a BED-finger domain thought to be important for DNA binding of the transposase proteins (ARAVIND 2000). There were a number of other mutations in regions not known to play a role in catalysis and DNA binding based on our understanding from other *hAT* transposases (Figure 3-1).

***Herves* transposase and Strand-transfer activity:** Only 9 out of the 13 variant *Herves* transposase forms, 4 from *An.gambiae s.s* and *An.merus* each and one from *An.arabiensis* that were capable of producing full-length transposase were used for the biochemical studies. The other four were not used because they proved difficult to clone. The 9 variant *Herves* transposases were expressed in *E.coli* and a ~67 kDa transposase protein was purified in each case (Figure 3-2).

The excision of the transposon from the donor site is followed by a transposase mediated joining of the 3'-OH ends of the transposon to the target DNA. The activity of the variant *Herves* transposases was determined by examining their ability to join pre-cleaved *Herves* left ends including the inverted terminal repeat to a target plasmid *in vitro*. Depending on if one or two *Herves*-L ends are joined to the target plasmid DNA, they can be seen as a Single End Joining (SEJ) or a Double End Joining (DEJ) product. We tested the activity of the 9 variant *Herves* transposase proteins in this assay and 7 forms were able to transfer and join the *Herves*-L ends to the target plasmid DNA (Figure 3-3). The other two forms (598 and 610) did not show any strand transfer products.



* Amino acid sequence of the *Herves* transposase that has shown to be active in *Drosophila melanogaster* and *Aedes aegypti*

† *Herves* transposase variants that have not been tested in this analysis

FIGURE 3- 1: Alignment of amino acid sequence of the 13 variant *Herves* transposases from *An.gambiae*

The alignment of amino acid sequences of the 13 different *Herves* transposase isolated from 3 members of the *An.gambiae* species complex, *An.gambiae s.s, An.arabiensis* and *An.merus* with the sequence of the active *Herves* transposase is shown. The mutations in different proteins are shown. The conserved residues are shown as dots (.) and a break in the alignment where there was conservation among all variant proteins is shown as empty spaces. The conserved DDE triad that forms the active site of the *hAT* family of transposases is shown using boxes. Blue shaded region A, corresponds to the N-terminal domain, regions B, C and D correspond to the regions in the catalytic and the α-helical domain of the *Hermes* protein that have been shown to be critical for its function. The green shaded regions show the conserved residues of the BED-domain predicted to be important for DNA binding.

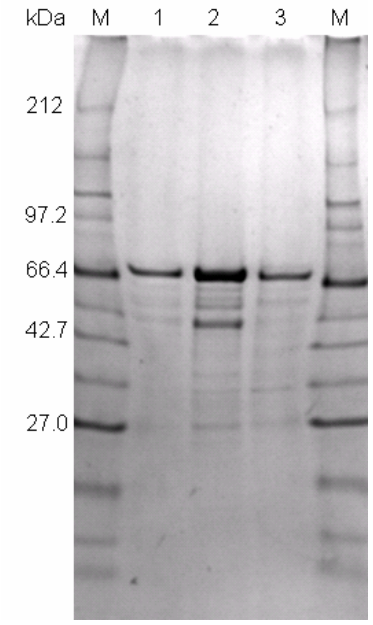


FIGURE 3- 2: Purified *Herves* transposase protein.
Three variant *Herves* transposase proteins of ~67 kDa after purification on a 4-14% PAGE gel. Molecular weight markers (M) are shown on the left side in kilodaltons.

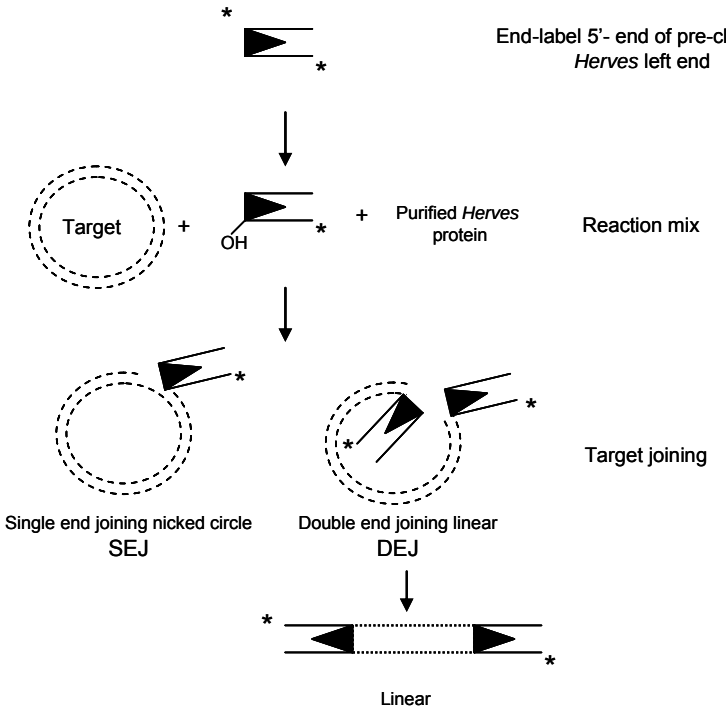
DISCUSSION

Class II transposable elements have proven to be useful in a wide range of applications in the laboratory. Apart from their use as tools for molecular studies in the laboratory, they are also being considered for use as genetic drive agents to spread genes through mosquito populations that would disrupt vector-borne disease transmission (KIDWELL and RIBEIRO 1992). We have used *Herves* to understand the behavior of Class II transposable elements in natural populations of *An.gambiae*, a species being seriously considered for control by such a genetic modification strategy (SUBRAMANIAN *et al.* 2007). Based on our previous studies that examined insertion-site frequency distribution, frequency of complete *Herves* transposase open reading frames, nucleotide sequence diversity and selection pressures on the transposase coding region, we predicted the presence of functional sources of *Herves* transposase in natural populations of *An.gambiae*.

In this study, we tested the above hypothesis by sampling *Herves* transposase coding regions from three closely related members of *An.gambiae* species complex, *An.gambiae s.s, An.arabiensis* and *An.merus.* We sequenced a total of 58 complete open reading frames that encode for *Herves* transposase, of which 13 were found to be "intact" with no pre-mature stop codons. As predicted, all nine forms that were expressed in *E.coli* produced a full length protein of ~67kDa. When the activity of the nine variant proteins was tested using an *in vitro* strand-transfer assay, seven of them showed activity.

The transposition of the *hAT* family of transposases is initiated by a transposase mediated nick, one nucleotide into the donor strand flanking the 5'-end of

a



b

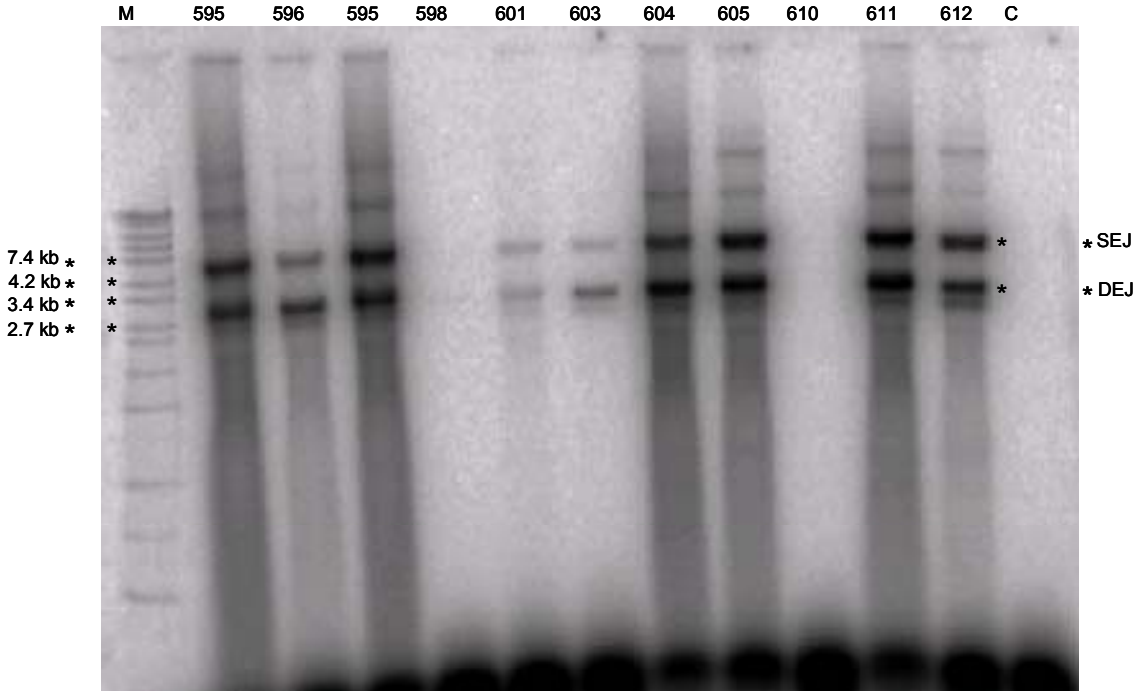


FIGURE 3- 3: Strand-transfer reaction using variant *Herves* transposase proteins.

a. Schematic of the strand transfer reaction
b. Results of the strand transfer reaction using nine variant *Herves* transposase proteins. 595 is the *Herves* transposase form that is active in *Drosophila* used as a positive control in the assay. The molecular weight markers are shown on the left side in kilo basepairs. Single End-Joining (SEJ) and Double End-Joining (DEJ) are indicated on the right side.

the transposon, leaving a nucleotide from the donor strand attached to the 5' end of the transposon (Figure 3-4). This generates a 5'- phosphate at the end of the transposon and a 3'-OH at the end of the flanking donor DNA. The 3'-OH end of the flanking DNA acts as a nucleophile and attacks the other strand (non-transferred strand) at the junction of the transposon and the flanking donor DNA. This results in formation of a hairpin structure on the donor DNA and the release of the transposon with a single unpaired nucleotide from the donor site attached at the 5'-end of the transposon. After excision from the donor site, the 3'-OH ends of the transposon attack the phosphodiester backbone of a target DNA molecule in a staggered transesterification reaction called strand transfer. This creates two complementary 8-bp single stranded regions in the target DNA flanking the transposon insertion. The 8-bp gaps are filled in by DNA repair mechanisms to create a characteristic 8-bp target site duplication observed for the *hAT* transposable elements (Figure 3-4) (CRAIG *et al.* 2002).

The crystal structure of the *Hermes* transposase reveals that there are three domains: an N-terminal domain (residues 79-150) and a catalytic domain that is divided by an α-helical domain (265-552) which is inserted into the catalytic domain (HICKMAN *et al.* 2005). The catalytic domain brings three essential amino acids (Aspartate, Aspartate and Glutamate) Asp181 (D), Asp247 (D) and Glu571 (E) in close proximity, so that they can coordinate Mg^{2+} ions that are essential for the catalysis. These three residues form the characteristic DDE motif that has been observed to be conserved in all transposases of the *hAT* family (RUBIN *et al.* 2001). It was shown that when these residues were mutated, the *Hermes* transposase, even

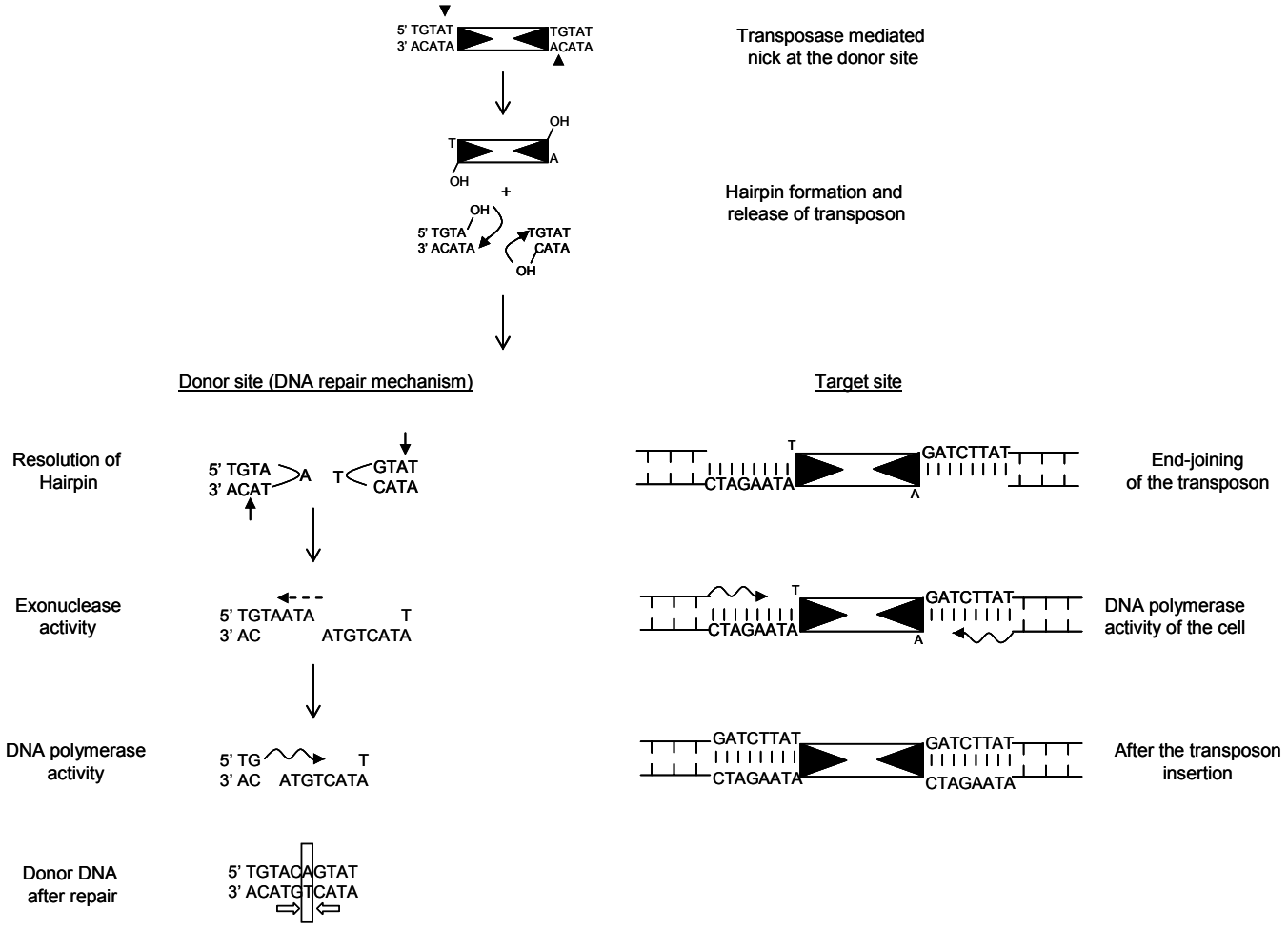


FIGURE 3- 4: Mechanism of transposition of *hAT* elements.
A transposase mediated nick at the donor site results in a 3'-OH in the flanking donor DNA that attacks the other DNA strand at the junction of the transposon and the flanking DNA. This results in hairpin formation at the donor site and release of the transposon. The donor site is repaired by the DNA repair mechanism of the cell generating palindromes that are footprints of excision events. The transposase mediates the end-joining of the transposon at the 3'-OH ends. The gaps are filled in by the DNA polymerase resulting in 8-bp target site duplications at the insertion sites.

though capable of binding DNA, had greatly reduced activity in all DNA cleavage and joining steps. This indicates that these three acidic residues are critical for the catalysis of the transposition reaction (ZHOU *et al.* 2004). The N-terminal domain of *Hermes* was shown to be involved in specific DNA binding to the transposon ends. A truncated version of the *Hermes* protein that did not contain these residues failed to bind to *Hermes* ends while the untruncated version bound specifically to a 30-nucleotide fragment of the *Hermes* L-end and not to non-specific DNA (HICKMAN *et al.* 2005). The inserted α-helical domain projects a tryptophan residue, Trp319, into the active site of the enzyme and has been shown to be required for DNA cleavage and hairpin formation through a base-flipping mechanism. Base-flipping is a mechanism where a single nucleotide base is rotated through 180° into an extra-helical location so that the enzyme can get access to a base that is usually buried in the double helix. This mechanism has been described for a number of enzymes such as DNA methylases, glucosyltransferases, glycosylases as well as for transposases (DAVIES *et al.* 1999; ROBERTS and CHENG 1998). The crystal structure of the *Tn5* transposase post-cleavage intermediate revealed that the thymidine 2 from the non-transferred strand is flipped out and stacked against the indole ring of the tryptophan (W298) and this interaction was shown to be critical for hairpin formation (ASON and REZNIKOFF 2002; DAVIES *et al.* 1999). One important difference to note in the mechanisms of *Tn5* and *Hermes* (*hAT)* transposition is that the hairpin formation occurs in the transposon in the case of *Tn5*, while it happens in the flanking donor DNA in the case of *Hermes* transposition. This difference is due to the first cleavage step, which generates a 5'-phosphate at the end of the *Hermes* transposon while it

generates a 3'-OH at the end of *Tn5* transposon. A tryptophan to alanine mutant (W319A) of the *Hermes* transposase was defective in DNA cleavage and hairpin formation but showed activity in strand-transfer reactions when provided with pre-cleaved ends (HICKMAN *et al.* 2005). This observation together with the understanding from the *Tn5* transposase mechanism, strongly suggests that Trp319 is involved in DNA cleavage and hairpin formation also in *Hermes* transposition. This tryptophan residue is another conserved feature of all other *hAT* transposases which reiterates its importance in the transposition reaction (RUBIN *et al.* 2001).

Based on our knowledge of the crystal structure of the *Hermes* transposase and the characteristic features of other *hAT* transposase proteins, we identified the regions in *Herves* transposase that may be important for its function. Region A corresponds to the N-terminal domain, Regions B, C and D correspond to the regions in the catalytic and the α-helical domain of the *Hermes* protein that have been shown to be critical for its function (HICKMAN *et al.* 2005; ZHOU *et al.* 2004). Regions B and D contain the conserved DDE amino acids critical for the catalysis of transposition. Region C contains the tryptophan that is important for cleavage and DNA hairpin formation. Sequences from bp position 1-75, 5' to the region A, even though was not important for binding of *Hermes* transposase to the *Hermes* L-ends, has been proposed to contain a BED-finger domain predicted to be involved in DNA binding (ARAVIND 2000). It has also been shown to contain the nuclear localization signal for *Hermes* transposase (MICHEL and ATKINSON 2003).

In this study, we tested the activity of variant *Herves* transposase proteins using strand-transfer assay. The variant *Herves* transposase proteins were supplied

with pre-cleaved *Herves* L-ends (with their 3'-OH ends already exposed) and tested for their ability to join the 3'-OH ends of the *Herves* L-ends to a target plasmid DNA. We observed that 7 out of the 9 variants that we tested were functional and showed the ability to strand-transfer. The two proteins 598 and 610 have some unique mutations that are not present in the other variants that may be responsible for their inactivity. *Herves* transposase variant 610 has three mutations, serine to proline (S171P), threonine to serine (T231S) and alanine to valine (A249V) in Region B that corresponds to the region that forms the catalytic domain in *Hermes* transposase. These mutations are likely to be responsible for the inactivity of the protein in the assay. Similarly, an Asparigine to Serine mutation (N87S) in region A in *Herves* variant 598 might affect the DNA binding activity of the protein, which is critical for element transposition. The activity of these proteins was tested simultaneously using identical conditions and the contents of the reactions were distributed from a common master mix. In addition, the protein concentrations of the variant *Herves* transposases used in the assay were also the same. Even though the experiment as performed here is not quantitative, the experimental set up enables us to make some inferences about the relative activity of these proteins. The strand-transfer results (Figure 3-3) indicate that variants 601 and 603 may have a lower activity compared to the other variant *Herves* transposase proteins. This observation was consistent between experiments. A lysine to asparagine (K97N) substitution in region A for 601 and an alanine to valine (A577V) substitution in region D for 603 together with other unique mutations not within the described regions may be contributing to their lower activity.

There are a number of other mutations that are observed in the four regions that do not seem to affect the ability of the proteins to end-join the *Herves* L-ends to the target plasmid DNA. Strand-transfer activity does not necessarily mean that the protein is capable of catalyzing a complete transposition reaction. For instance, from the structure and function of *Hermes* transposase we can predict that protein 612 should be defective for DNA cleavage and hairpin formation steps due to the tryptophan to cysteine substitution at position 329. The tryptophan to alanine (W319A) mutant of *Hermes* transposase was, as described earlier defective in DNA cleavage and hairpin formation but was able to produce single-and double-end-joining products (SEJ & DEJ). Two other *Herves* variants, 596 and 598, have a cysteine to lysine (C25Y) and cysteine to phenyl alanine (C28F), respectively, in two conserved residues in the BED-domain which is predicted to be critical for DNA binding (ARAVIND 2000). Even though the cysteine substitution at position 25 did not seem to affect the strand transfer for 596, it did affect the strand transfer for 598.

Additional experiments testing the ability of the *Herves* variants to perform the full transposition reaction are necessary to confirm that these proteins can catalyze the complete transposition reaction. The predictions for the possible inactivity of these variant *Herves* proteins can be tested by changing the mutated residues to the corresponding residues seen in active forms by mutagenesis methods and testing for activity. From this analysis and also from comparing the sequences with the known structure-function features of *Hermes* we can predict that at least six of the seven proteins that showed activity in the strand-transfer are capable of catalyzing the complete transposition of *Herves* elements.

Based on the results of this study the frequency of individuals with *Herves* transposase coding regions capable of making fully functional transposase is high. Of the 7 variant *Herves* transposase proteins that were functional, 4 were from *An.gambiae s.s* and were isolated from 4 individual mosquitoes. Only 9 individuals of *An.gambiae s.s* were used in this study, indicating that approximately 45% of the individuals have source of functional *Herves* transposase. Despite the availability of *Herves* transposase, the copy number and the apparent transposition activity of *Herves* were low. This suggests the presence of host repression systems that regulate the activity of these elements. Our failure to detect RNA transcripts of *Herves* transposase in mosquitoes from natural, as well as lab, populations of *An.gambiae* using RT-PCR supports this hypothesis (O. A. Akala and R. A. Subramanian, unpublished).

In summary, we found 13 variant *Herves* transposase proteins that are capable of producing full-length protein. We expressed and purified nine out of the 13 variant proteins and tested them using an *in vitro* assay. Seven out of the nine proteins showed ability to end-join the *Herves* L-ends to a target plasmid DNA. Even though these results need to be corroborated with further experimental evidence, based on their activity in strand-transfer assay, we can conclude that there is a source of functional *Herves* transposase in natural populations of *An.gambiae*. However, a host repression system seems to regulate the activity of these transposase proteins resulting in the low observed activity of the elements.

Chapter 4: The population genetics of *Topi,* a *Tc1/mariner* family of transposable element in the malarial mosquito, *An.gambiae s.s.*

ABSTRACT

Class II transposable elements have been successfully used as gene vectors to transform a number of insect species. Besides their use as gene vectors in insects they are also being considered as genetic drive agents to spread refractory genes into natural populations of mosquitoes to control vector-borne diseases such as malaria. We have studied *Herves*, an active endogenous element in *An.gambiae* earlier, to understand the evolution and behavior of Class II transposable elements in this species. Here, we study *Topi*, a *Tc1/mariner* element to determine if the natural history of *Herves* is shared by other Class II transposable elements in *An.gambiae*. We examined the dynamics of *Topi* elements in five populations in Africa by measuring site-occupancy frequency and nucleotide sequence diversity, as well as by analyzing the structure of the elements in these locations. We found no evidence of recent activity based on the site-occupancy distribution data. All 74 individuals sampled from five different locations had *Topi* elements with a high copy number that ranged from 10 - 34 per diploid genome. Nucleotide sequence diversity in the coding region of *Topi* elements was higher (π= 0.051) than *Herves* indicating that *Topi* was present in the *An.gambiae* genome longer than *Herves*. Further evidence for this was observed from the analysis of the silent-site diversity of these elements. Silent-site diversity of *Topi* elements were only 3 to 5-fold lower than expected. Despite their long history in *An.gambiae*, all samples analyzed had a complete form of the element ~ 1kb in size as well as a deleted form of ~ 600bp. We found 14 forms, of *Topi*

transposase in the sampled 58 sequences (which were capable of encoding a full-length transcript). Lack of evidence for recent activity based on insertion-site frequency distribution data suggests that either these forms are not functional or that they are under host regulation. The evolution of the *Topi* transposable element seems similar to the *Herves* transposable element in *An.gambiae*.

INTRODUCTION

Class II transposable elements have been used successfully as gene vectors in a number of insect species (ATKINSON *et al.* 2001). A collection of Class II transposable elements that includes *P*-elements, *hobo*, *Tn5, mariner, Minos, piggyBac*, and *Hermes* have been used to transform insects such as *D.melanogaster* (O'BROCHTA and ATKINSON 1996), *Stomoxys calcitrans* (O'BROCHTA *et al.* 2000), *Tribolium castaneum* (BERGHAMMER *et al.* 1999), *Ceratitis capitata* (MICHEL *et al.* 2001) and butterfly, *Bicyclus anynana* (MARCUS *et al.* 2004). Medically important insects such as *Aedes aegypti* (JASINSKIENE *et al.* 1998), *Anopheles gambiae* (GROSSMAN *et al.* 2001; KIM *et al.* 2004), *Anopheles stephensi* (CATTERUCCIA *et al.* 2000) and also commercially useful insects such as silk worm, *Bombyx mori,* have been transformed using Class II transposable elements (TAMURA *et al.* 2000). Class II transposable elements have also been used successfully as gene vectors for stable chromosomal integration of transgenes that, when expressed appropriately, impair the development of malaria parasites, *Plasmodium,* in *Anopheles* mosquitoes (ITO *et al.* 2002; KIM *et al.* 2004; MOREIRA *et al.* 2002). Genetically modified mosquitoes and population modification using a genetic drive agent to spread the refractory genes are being considered to control vector-borne diseases such as malaria. Transposable

elements with their ability to move and also rapidly increase in copy number have been proposed for use as genetic drive agents to rapidly increase the frequency of refractory genes in mosquito populations (KIDWELL and RIBEIRO 1992). The most extensively documented example of such a rapid increase in frequency of transposable elements is the spread of *P*-elements in *D.melanogaster*. *P*-elements after their introduction into *D.melanogaster* from a closely related species, *D.willistoni*, rapidly increased in frequency and became distributed throughout world populations of *D.melanogaster* within a few decades (ANXOLABEHERE *et al*. 1988).

The potential use of Class II transposable elements as genetic drive agents to spread refractory genes through mosquito populations to control vector-borne diseases such as malaria has led to studies designed to understand the behavior of these elements in the target species for such a control, *An.gambiae*. We have studied *Herves*, a Class II transposable element that belongs to the *hAT* family of transposable elements in natural populations of *An.gambiae* from six different locations in Africa. We used insertion-site frequency distribution data to assess the copy number and activity of the element in natural populations of this species. We looked at the sequence diversity by analyzing both the coding and non-coding regions of the element. In addition, we assessed the structural diversity of these elements by analyzing the frequency of complete open reading frames in these populations.

We found that *Herves* was present in all of the populations analyzed but at a low copy number; the average element copy numbers in the six populations analyzed ranged from 2.9 - 4.4. Even though the copy number was low, there was evidence for recent activity in all of the analyzed populations (ARENSBURGER *et al.* 2005). The

element was found in all the members of the *An.gambiae* species complex indicating that this element was probably present prior to the evolution of the species complex. We cannot, however, rule out the possibility of horizontal transfer among these species as some introgression has been observed between at least two members of the species complex, *An.arabiensis* and *An.gambiae s.s* (BESANSKY *et al.* 1997). The hypothesis of a long residence time in the species was supported by the high sequence diversity and form ("haplotype") diversity in these populations (SUBRAMANIAN *et al.* 2007). Even though the element was present in the species for an extended amount of time we observed several characteristics that would not be predicted for an element with a long species history. We found a high frequency of complete open reading frames (>85 %) in most of the populations of *An.gambiae*. In addition, we found a higher conservation of the coding than the non-coding regions of the *Herves* transposase as well as evidence for purifying selection in the coding region. These results indicate that *Herves* is likely to still be active in natural populations of *An.gambiae*.

As part of an effort to determine if the natural history of *Herves* is shared with other Class II transposable elements, in this study we investigated *Topi*, a Class II transposable element that belongs to *Tc1/mariner* family of transposable elements in five locations in Africa. We tried to understand the evolution of *Topi* by analyzing the same features that we had previously studied in the *Herves* element, giving us an opportunity to compare and contrast the behavior and evolution of these two elements in *An.gambiae*.

Even though studies on two elements may not reflect the evolution of all the Class II transposable elements in *An.gambiae*, results from them would contribute to the development of a model to predict the outcomes of a Class II transposable element invasion in this species. One of the concerns for using Class II transposable elements as genetic drive agents is loss of the refractory transgenes before their fixation in natural populations due to accumulation of deletions in the transposable elements carrying them. This concern is largely due to the observation that *P*-elements in *Drosophila melanogaster* rapidly evolved forms of the element that contain internal deletions (O'HARE *et al.* 1992). If the features observed in *Herves* such as maintenance of structural integrity (few deleted forms) and activity for an extended period of time, are general features of Class II transposable element evolution in *An.gambiae*, then these elements may be well - suited to spread refractory genes in this species to control malaria.

MATERIAL AND METHODS

Sample: *Anopheles gambiae s.s.* from five populations were used in this study with a sample size of 16 individuals each from Kisumu, Malindi and Zenet,15 from Furvela and 10 individuals from Bakin Kogi populations (Table 4 -1). Samples from Malindi, Bakin Kogi, Zenet and Furvela have been previously described (SUBRAMANIAN *et al.* 2007). Malindi is located in eastern Kenya and was sampled in 1996 (LEHMANN *et al.* 2003). Bakin Kogi is in north-central Nigeria and samples were collected in 1999 (LEHMANN *et al.* 2003). Zenet is a village in northeastern region of Tanzania and was sampled in and around the village in 2004 (MEERAUS *et al.* 2005). Samples from southern Mozambique (Furvela) were collected in 2003 and

were earlier described (O'BROCHTA *et al.* 2006). Samples from Kisumu were collected in 2005 from two villages Iguhu and Kombewa in Western Kenya.

DNA Isolation and Whole genome amplification: Genomic DNA was isolated from individual mosquitoes as described (O'BROCHTA *et al.* 2006) and resuspended in 100 μl of distilled water and stored at -80ºC. One hundredth of the genomic DNA from one mosquito (1 μl) was used in the whole genome amplification using GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Piscataway, NJ) following the manufacturer's recommendations. Amplified genomic DNA was resuspended in 20 μl of distilled water and stored at -80ºC.

***Topi* transposable element display:** The procedure used for transposable element display has previously been described (GUIMOND *et al* 2003, O'BROCHTA *et al* 2006) and was modified for the analysis of *Topi* transposable element as described below. Transposable element display was performed in triplicate using one eighth (2.5 μl) of the DNA obtained after the whole genome amplification of $1/100^{th}$ of the genomic DNA obtained from one mosquito (see below) for each replicate. Genomic DNA was digested for 4 hours in a volume of 20 μl at 37^{o}C with 2 units of the restriction endonuclease *Dpn*II using conditions recommended by the manufacturer (New England Biolabs). *Dpn*II digestion products were ligated to 30 picomoles of adapters by adding 5 μl of 1X restriction enzyme buffer containing 5 mM ATP, 50 mM DTT (dithiothreitol), 10 μg BSA (bovine serum albumin), 4 units of *Dpn*II, 1 Weiss unit of T4 DNA ligase and incubated at 37^{o}C overnight. To prepare the adapters, equimolar amounts of oligonucleotides MspIa (5' GAC GAT GAG TCC TGA G 3') and DpnIIb (5' GAT CCT CAG GAC TCA TC 3') were heated to 100^{o}C

for 10 minutes and then allowed to very slowly cool to room temperature. The conditions used for the digestion/ligation reactions and also the design of the adapters allow the creation of only monomeric *Dpn*II-cut genomic DNA fragments with terminal adapters.

The next step was a polymerase chain reaction ("preselective PCR") with five microliters of the restriction/ligation reaction as the template in a 25 μl reaction volume containing 1X PCR Buffer II (Applied Biosystems), 0.2 mM dNTPs (an equimolar mixture of dATP, dTTP, dCTP, dGTP), 2.5 mM $MgCl_2$, 1 unit AmpliTaq® DNA polymerase (Applied Biosystems), and 24 pmoles of primer MspIa and primer TETopiR1 (5' GTT AGA ATG TGT TTT CG C 3'). The DNA polymerase was added as a complex with TaqStart™ Antibody (ClonTech) as described by the manufacturer for the purpose of "hot-starting" the reaction. The reaction conditions were 95°C/3 mins followed by 25 cycles of 95°C/15 sec, 54°C/30 sec, 72°C/1.0 min and a final cycle of 72°C/5 min. A second PCR was performed ("selective PCR") using 5 μl of the 20 times diluted preselective PCR products as a template in a 20 μl reaction containing 1X PCR Buffer II, 0.2 mM dNTPs, 2.5 mM $MgCl_2$, 1 unit AmpliTaq® DNA polymerase (bound to TaqStart™ Antibody as above), 9 pmoles each of primers MspIa and Cy5™-labeled TETopiR2 (5' TAA ACA GTC CTT TTC AGG 3'). The Cy5™-labeled primers were purified by HPLC prior to their use. Following an initial denaturation step at 95ºC for 3 minutes, "touchdown" PCR conditions were used in which during the first 5 cycles the annealing temperature was decreased 1°C after each cycle with the first of these cycles being 95°C/15 sec, 59°C/30 sec, 72°C/1.0 min. Following these 5 cycles an

additional 25 cycles were performed at 95°C/15 sec, 54°C /30 sec, 72°C /1.0 min with a final cycle of 72°C/5 min. TETopiR1 and TETopiR2 are *Topi* element specific primers that anneal to sequences approximately 150 bp and 90 bp from the 3'end of the element.

Five micro liters of selective PCR products were mixed with 5µl of loading buffer (95% deionized formamide, 10mM EDTA) and the mixture heated to 95°C for 3 minutes, cooled quickly on ice and 6 µl loaded on a 6% polyacrylamide gel (19:1 acrylamide : bisacrylamide) containing 6.7 M urea in 1X TBE buffer (90 mM Tris-borate, 2 mM EDTA). ALFExpress™Sizer™50-500 (Amersham/Pharmacia) was used as a size standard. Electrophoresis was performed for 2.5 hours at a constant voltage of 70 watts. The gel was then transferred to 3MM filter paper and dried. The dried gel was scanned on a STORM 860 phosphoimager (Molecular Dynamics) to visualize the products of the transposable element display. The selective PCR products from the three independent replicates of a sample were run on the same gel to assist unambiguous calling of bands. A band was called as present or absent if it was present in at least 2 of the three replicates. From the three replicates, a single scoring matrix was obtained that was used in subsequent analyses. The advantage of this procedure is that it increased the accuracy of determining the presence of bands and minimized errors in subsequent analyses.

Transposable element display data was used to estimate the site-occupancy frequency distribution of *Topi* and by assuming the models of Charlesworth and Charlesworth (1983) these data were used to estimate the parameter β. The model parameter β measures the forces removing the elements from natural populations

(drift, excision and selection). Because the model used in this analysis assumes that the copy number is in equilibrium, it also reflects the forces that tend to add elements to the population (replicative transposition). Estimation of β and the copy number of *Herves* per diploid genome were performed as described by Wright *et al.* (2001) who considered the dominant nature of transposable element display signals and the application of the parameter estimation methods of Charlesworth and Charlesworth (1983) to diploid organisms. A one way- ANOVA and Tukey's HSD test was used to compare the average diploid copy number among locations for statistical differences between different locations.

***Topi* transposase detection and sequencing:** To analyze the structure and sequence of *Topi* elements, *Topi* transposase open reading frame was amplified using a Topi277F (5'-ATG GGT CGC GGA AAG CAC TG-3') primer that annealed to the 5' end of the open reading frame and a Topi1302R primer (5'- GCG GTG TTC CAC TGA GCG-3') that annealed to the DNA flanking the open reading frame. One fiftieth of the genomic DNA from one mosquito (2 μl) was used as the template in a 50μl reaction containing 1X ThermalAce™ (Invitrogen), 0.2 mM dNTPs (an equimolar mixture of dATP, dTTP, dCTP, dGTP), 2.5 mM $MgCl_2$, 2 units ThermalAce™ DNA polymerase (Invitrogen), and 24 pmoles of primer Topi277F and Topi1302R. The following conditions were used for the amplification reactions: 95°C/3 min followed by 30 cycles of 95°C/30 sec, 55°C/30 sec, 72°C/1min 30 secs and a final cycle of 72°C/10 min. Reaction products were fractionated in a 1% agarose gel. The 1kb amplification products from all samples and the approximately 600 bp products from 8 samples were eluted from the gel, precipitated, resuspended

in 20 µl dH_2O and cloned into the pCR®-Blunt II TOPO vector (Invitrogen). One clone per individual was sequenced and these sequences were used in subsequent analyses. From samples "Kisumu" (12), "Malindi" (8), "Zenet" (10), "Furvela" (11) and "Bakin Kogi" (8) a total of 49 sequences were obtained.

Sequence Analysis: Sequence alignments were done using AlignX, a ClustalW-base alignment program in VectorNTI Advance 10.0.1 (Invitrogen). Nucleotide diversity was estimated from average pair-wise number of differences between elements, π (NEI AND LI 1979) and from the number of polymorphic sites, θ (WATERSON 1975) π and θ were estimated using DnaSP 3 (ROZAS AND ROZAS 1995). The silent-site diversity estimates were calculated using the Kumar method (NEI AND KUMAR 2000) implemented in MEGA 3.1 (KUMAR *et al*. 2004b). Expected values of silent-site diversity were calculated described in Sanchez-Gracia *et al* (2005) and were the product of the haploid copy number and the average synonymous diversity (0.0209) from a sample of 35 nuclear genes (MORLAIS *et al* 2004). The average nucleotide-sequence diversity, π , and, the average expected and observed silent-site diversity estimates were compared among locations using a one way-ANOVA. Post-hoc comparisons were made using Tukey's HSD test, $p < 0.05$ denoted significant difference. An alignment of 14 sequences that did not have any pre-mature stop codons were used for estimating the number of synonymous substitutions per synonymous site (dS) and of non-synonymous substitutions per non-synonymous site (dN) and their ratio, ω= dN/dS using maximum likelihood (ML) methods employed by CODEML in PAML 3.13 (YANG 1997). PAML can be used to examine the data using various codon substitution models that make different assumptions about the

way selection pressure is distributed within the gene. We examined the data using three simple models: a one-ratio model (M0) that assumes one ω for all sites, a neutral model (M1) that assumes that there are two classes of sites within the gene; those that are conserved (p_0) with $\omega_0=0$ and those that are neutral ($p_1=1- p_0$) with $\omega_1=1$, and finally, a discrete model (M3) that assumes three classes of sites each having a unique value of ω that is estimated from the data (YANG 1997). In each case, a likelihood ratio was calculated which was used to compare and determine which model best reflected the observed data using a likelihood ratio test (LRT). The LRT statistic is twice the log-likelihood difference between two models being compared and has a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters between the two models (YANG *et al* 2000).

RESULTS

Methods validation: Transposable element display is a DNA finger-printing method used to assess the copy number and position of transposable elements in the genome (BIEDLER *et al.* 2003b; GUIMOND *et al.* 2003; SUBRAMANIAN *et al.* 2007). We adapted the technique for the analysis of *Topi* transposable elements and estimated the copy number and site-occupancy distributions in five different populations of *An.gambiae s.s* in Africa. Because of the limited amount of genomic DNA available for analysis, a whole genome amplification method was employed to produce adequate amounts of DNA. Whole genome amplification is a method of uniformly producing microgram quantities of genomic DNA from small quantities of genomic DNA. Although shown by others to faithfully reproduce genomic DNA (GORROCHOTEGUI-ESCALANTE and BLACK 2003), we confirmed the findings by

comparing the results of transposable element display obtained using whole genome amplified DNA with those that were obtained using original, non-amplified genomic DNA. An analysis of 11 samples verified that the amplified genomic DNA reproduced the patterns of *Topi* insertion and copy number obtained from the original genomic DNA sample.

Transposable element (TE) display, as performed in this study does not result in the efficient amplification of fragments longer than 1kb because the extension time in the PCR reactions was only 1 minute. Because the *An.gambiae* genome is composed of 64.8% adenines and thymines and PCR templates for TE display were produced by digesting genomic DNA with *DpnII* (GATC), we expected only 7% of the resulting fragments to be 1 kb or more in length. We estimated this by calculating the percentage of fragments greater than 90 bp that were longer than 1 kb. Ninety base pairs is the invariable amount of *Topi* DNA contained in each PCR product. We assumed fragment sizes following *DpnII* digestion followed an exponential distribution ($\lambda e^{-\lambda x}$) with λ = (0.176) (0.324) (0.324) (0.176). Therefore, 0.746 of all fragments were greater than 90 bp and 0.052 of all fragments were greater than 1 kb. Thus, 7 % (0.052/ 0.746 * 100) of all fragments were greater than 1 kb. The specificity of the *Topi* TE display was confirmed by eluting and sequencing 10 randomly selected bands from the gel. All the sequenced bands contained *Topi* elements as expected (Data not shown).

Copy number / Site Occupancy: In this study all individuals analyzed (n=74) had at least 2 copies of the *Topi* element and one sample from Malindi had 37 copies of the element. Mean element copy numbers in the five populations analyzed

TABLE 4- 1: Site occupancy

Location	N^{*}	κ^{Δ}	dcn†	$\beta^{\dagger\dagger}$
Kisumu (k)	16	72	31.3^{zfb}	0.6
Malindi (m)	16	78	33.8^{zfb}	0.7
Zenet (z)	16	56	18.2^{kmf}	0.5
Furvela (f)	15	59	10.22^{kmzb}	0.8
Bakin Kogi (b)	10	63	18.4^{kmf}	1.5

* Individuals analyzed by transposable element display

$^{\Delta}$ Number of unique chromosomal sites containing *Topi*

† Diploid copy number of *Topi* (WRIGHT *et al.* 2001)

†† $4N_e(\nu+s)$ from Charlesworth and Charlesworth (1983)

$^{k\,m\,z\,f\,b}$ the copy number was significantly different from the indicated location at a significance level of 0.05

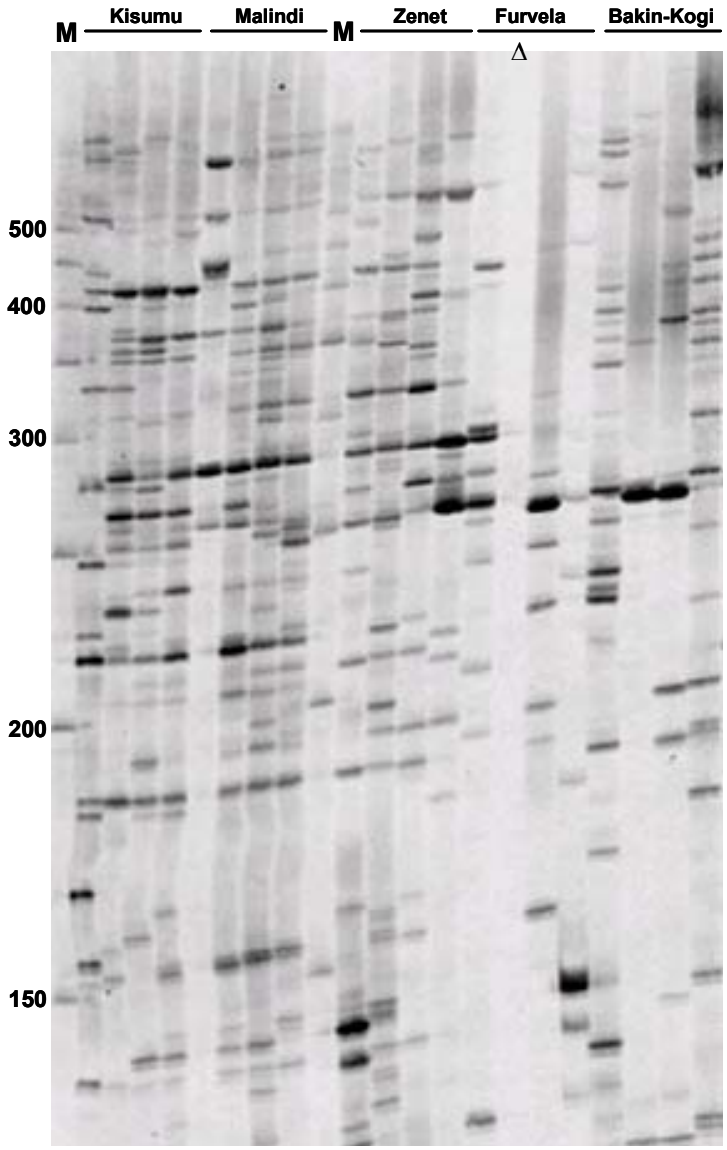


FIGURE 4 - 1: Transposable element display of the right end of *Topi* elements. A sample of transposable element display results obtained from five different locations is represented. Molecular weight markers (M) in base pairs are shown on the left side.

[Δ] Empty lane

ranged from 10.2 – 33.8 per diploid genome. There was a statistically significant difference in copy numbers between all the locations ($p < 0.05$, Tukey's HSD test) except between Kisumu and Malindi, and Zenet and Bakin-Kogi (Table 4-1) (Figure 4-1). The copy number in Furvela was significantly lower than all the other locations analyzed. There were 19 and 17 elements with high site-occupancy frequencies that were present in more than 10 individuals in Malindi and Kisumu respectively. Furvela had the least number of high frequency occupied sites with only one that was present in 9 of 15 individuals.

We used the model of Charlesworth and Charlesworth (1983) to analyze the observed site-occupancy distributions of the *Topi* element in *An.gambiae*. The model assumes the elements are at equilibrium and that there are infinite insertion-sites within the genome. The model parameter β reflects the effects of forces other than drift that might be playing a role in shaping the observed distribution. According to the models, β values greater than one indicate that the forces of mobility and/or selection are responsible for the observed frequency distribution. We observed that all the locations except Bakin Kogi ($\beta=1.5$) showed β values less than one indicating that there has been little recent activity of *Topi* in *An.gambiae s.s* (Table 4-1).

Structure of *Topi* elements: Autonomous Class II transposable elements code for functional transposase and can undergo transposition. Non-autonomous elements are usually deleted forms of the element which depend on transposase expressed from other elements in the genome. Class II elements like *P-elements* in *Drosophila* often exist in forms that have large internal deletions (ENGELS 1989),

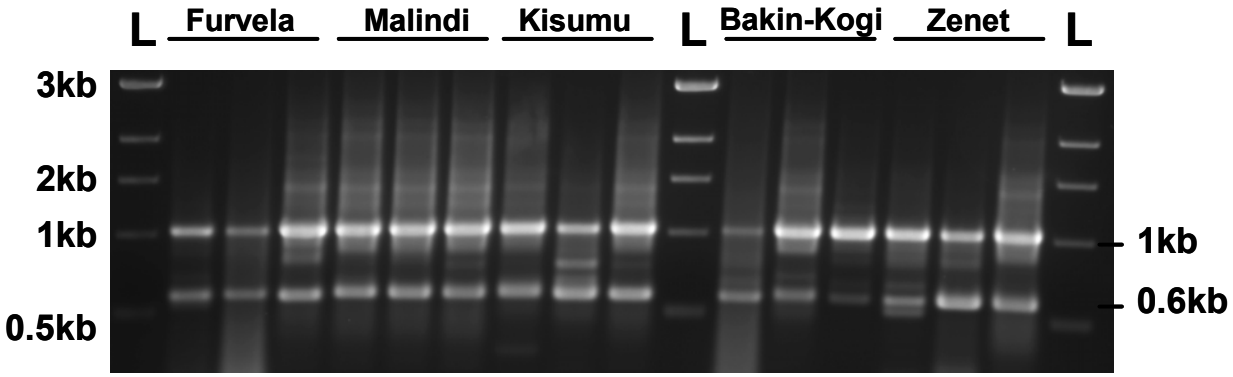


FIGURE 4 - 2: Structure of *Topi* elements.
PCR products of a sample of individuals from five different locations used to analyze the structure of *Topi* elements are shown. Molecular weight markers (M) are shown on the left side in kilobase pairs. The ~1kb complete *Topi* transposase open reading frame is indicated on the right side. Approximately 0.6 kb deleted form observed in all individuals is also indicated on the right side

however, *hAT* elements such as *Herves* in *An.gambiae* (SUBRAMANIAN *et al.* 2007) and *Hermes* in *Musca domestica* (L. A. Cathcart, E. S. Krafsur, P. W. Atkinson, D. A. O'Brochta and R. A. Subramanian, unpublished) are rarely found with deletions. We analyzed the structure of *Topi* elements by amplifying the internal ~ 1kb *Topi* transposase coding region using PCR. We observed that all individuals analyzed (n= 74) had at least one copy of the 1 kb complete open reading frame and a ~600 bp deleted form (Figure 4-2). There were other less prevalent deleted elements of other sizes present in some of the individuals analyzed (Figure 4-2).

Nucleotide diversity of *Topi* elements: The 1kb complete *Topi* transposase coding region amplified was cloned and sequenced from 49 individuals to analyze the sequence diversity of the *Topi* elements in five different populations. Only one sequence per individual was obtained to give us the opportunity of sampling as many different elements as possible. All of the 49 sequences sampled were different from each other. The nucleotide sequence polymorphism ranged from $\pi = 0.029$ to $\pi = 0.062$ with the average being $\pi = 0.051$ (Table 4-2). The π values were only significantly different between Malindi and Furvela, and Zenet and Furvela ($p < 0.05$, Tukey's HSD test). Eight deleted forms of *Topi* were recovered and analyzed. Two sequences each were of Form A (828 bp), Form B (785 bp) and Form E (572 bp); one each of Form C (758 bp) and Form D (732 bp) (Figure 4-3). Deleted forms had ~ 200 bp to ~400 bp deletions in different regions of the *Topi* open reading frame (Figure 4-3). Form E had ~600 bp deletion when compared to the "canonical *Topi* ORF", however both the sequences of Form E had an extra 175 bp which was not similar to the canonical element (Figure 4-3).

TABLE 4- 2: Nucleotide sequence polymorphism in *Topi* open reading frame

Location	Seqs*	Poly†	π Δ	θ ¶
Kisumu (k)	12	168	0.0452 (0.009)	0.056 (0.022)
Malindi (m)	8	149	0.062 (0.0077)[f]	0.060 (0.026)
Zenet (z)	10	145	0.0521 (0.0099)[f]	0.053 (0.022)
Furvela (f)	11	130	0.029 (0.0106)[mz]	0.049 (0.019)
Bakin-Kogi (b)	8	122	0.047 (0.013)	0.067 (0.029)
Combined	49	227	0.051 (0.0032)	0.086 (0.024)

* Number of sequences analyzed

† Number of polymorphic positions

Δ Pairwise nucleotide diversity (NEI and LI 1979); standard deviation in parenthesis

¶ Nucleotide diversity based on segregating sites (WATTERSON 1975); standard deviation in parenthesis

f m z π was significantly different from the indicated location at a significance level of 0.05

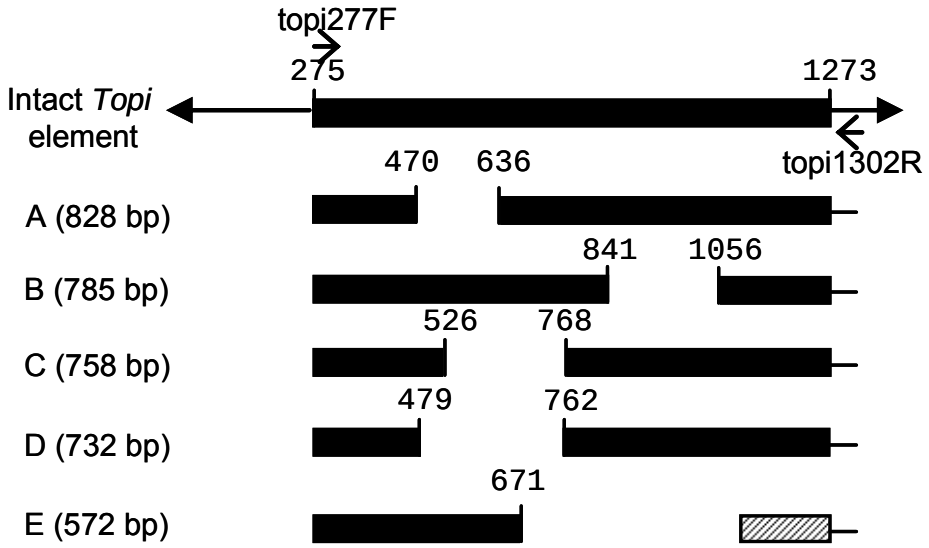


FIGURE 4 - 3: Structure of deleted forms of *Topi* elements.
The position of the deletion corresponding to the full length *Topi* element is shown for each form. The position of deletion and the additional 175 bp of sequence that is not similar to the full length element is also shown for Form E. The position of the primers, topi277F and topi1302R, that were used to amplify these forms are also shown.

Comparing the levels of silent-site diversity of transposable elements with that of single-copy host genes can be useful when looking for evidence of a lateral transfer event sometime in the history of the element and to understand when such an event might have occurred (SANCHEZ-GRACIA *et al.* 2005). Here, we compared the silent-site diversity (π_s) of *Topi* elements with the average silent diversity of 35 nuclear genes in *An.gambiae* (MORLAIS *et al.* 2004). The observed silent-site diversity of *Topi* elements was significantly lower than the expected average silent-site diversity seen in 35 nuclear genes (MORLAIS *et al.* 2004). The observed π_s was 3 to 5-fold lower than the expected π_s in all locations analyzed (Table 4-3). Comparisons between populations showed that the observed π_s in Furvela was significantly lower than the π_s observed in Kisumu, Malindi and Zenet ($p < 0.05$, Tukey's HSD test); and while the expected π_s in Bakin-Kogi was significantly lower than that in Malindi ($p < 0.05$, Tukey's HSD test). The expected silent-site diversity in Furvela was significantly lower than π_s at all other locations and the expected π_s in Bakin-Kogi was lower than all the other locations except Zenet ($p < 0.05$, Tukey's HSD test). The expected silent-site diversity in Kisumu and Malindi, and Zenet and Bakin-Kogi were not significantly different from each other.

Natural Selection: We tested for evidence of selective constraints within the *Topi* transposase coding region of the 14 sequences that had no pre-mature stop codons by estimating $\omega = d_N/d_S$ using ML (YANG 1997; YANG *et al.* 2000). The ω values ranged from 0.45 to 0.51 under all models (M0, M1 and M3) showing evidence of purifying selection. Even though the discrete model (M3) fit the data better than the neutral model (M1), the LRT statistic, $2\Delta l$ ($2\Delta l = 2(-2371.47 - (-$

2366.7)), for this comparison was 9.54, which was less than the critical value of $\chi^2_{[0.001, 2]} = 13.816$.

DISCUSSION

Class II transposable elements are being considered for use as genetic drive agents to spread transmission-blocking genes through mosquito populations to control vector-borne diseases such as malaria. Understanding the behavior of transposable elements in *An.gambiae,* a target species for such a control strategy will be helpful in predicting the outcomes of such an approach. We recently described the dynamics of an active Class II transposable element, *Herves* in *An.gambiae* populations in Africa (SUBRAMANIAN *et al.* 2007). We found that *Herves* was able to maintain its structural integrity for a longer time than what has been observed with other elements, such as the *P*-element in *Drosophila* (O'HARE *et al.* 1992). We found higher conservation of the *Herves* transposase coding region and also evidence for purifying selection in this region (SUBRAMANIAN *et al.* 2007). Here, we studied the dynamics of *Topi*, a *Tc1/mariner* family transposable element in 5 different populations of *An.gambiae s.s* in Africa to understand the evolution and behavior of the transposon in *An.gambiae s.s*. We have used the results from this study together with our earlier study of the *Herves* element to gain a better understanding of the general features of the evolution of Class II transposable elements in *An.gambiae*.

We examined the dynamics of the *Topi* element by measuring the site-occupancy frequency, nucleotide sequence diversity and also by analyzing the structure of the element. The element copy number was higher (10.2- 33.8) and the site-occupancy levels lower ($\beta = 0.5$- 1.5) than those reported for the *Herves* element. Assuming that

TABLE 4- 3: Silent-site diversity of *Topi* elements from different locations

Locations	Haploid copy number	π_s †		
		Observed	**Expected** Δ	**Observed/Expected***
Kisumu (k)	15.65	0.063 (0.041)^{f}	0.33 (0.064)^{zfb}	0.193
Malindi (m)	16.9	0.081 (0.044)^{fb}	0.353 (0.075)^{zfb}	0.227
Zenet (z)	9.1	0.063 (0.042)^{f}	0.190 (0.082)^{f}	0.332
Furvela (f)	5.11	0.022 (0.023)^{kmz}	0.107 (0.066)^{kmzb}	0.220
Bakin Kogi (b)	9.2	0.044 (0.043)^{m}	0.192 (0.087)^{kmf}	0.229
All	11.13 #	0.051	0.238	0.219

† π_s represents the average pairwise nucleotide diversity at synonymous sites

Δ see Materials and Methods

Average haploid copy number from all locations

k m z f b π_s was significantly different from the indicated location at a significance level of $p < 0.05$

* the observed π_s was significantly lower than the expected π_s at all locations at a significance level of $p < 0.05$

the elements are at copy number equilibrium, β values greater than 1 have been interpreted as evidence for recent activity (CHARLESWORTH and CHARLESWORTH 1983; CHARLESWORTH and LANGLEY 1989; LANGLEY *et al.* 1983). β values for *Topi* were less than 1 in most locations except in Bakin-Kogi which was only slightly higher (β =1.5) indicating no recent activity of the element in this species. This is consistent with a previous report on *Topi* in which, the authors found no variation in insertion sites based on *in situ* hybridization on 4 individual mosquito samples from the *An.gambiae* PEST strain (GROSSMAN *et al.* 1999). They found *Topi* in all members of the *An.gambiae* species complex suggesting that this element was present even before the diversification of this species (GROSSMAN *et al.* 1999). However, the possibility of transfer of the element between species can not be ruled out because there was introgression reported between *An.gambiae s.s* and *An.arabiensis* (BESANSKY *et al.* 1997). Consistent with a hypothesis of an extended residence time is the higher levels of nucleotide sequence polymorphism (average $\pi = 0.051$) compared to the *Herves* elements (average $\pi = 0.0046$). Assuming similar mutation rates for transposable elements and the rest of the genome, an element is expected to accumulate more mutations the longer it is in the species and hence nucleotide sequence diversity can be used to understand the age of an element in a particular species. The *Topi* elements that we analyzed were highly polymorphic with an average $\pi = 0.051$, which is ten times higher than observed for *Herves* ($\pi = 0.0046$).

Assuming that transposable elements have similar mutation rates as nuclear genes, the silent-site diversity can also be used to assess if there was a lateral transfer event and also when it may have occurred. In other words, it would be helpful to

predict the age of an element in the species. The longer an element has been in the species, the silent-site diversity would be closer to a nuclear gene in the species. The observed silent-site diversity among *Topi* elements revealed less diversity than expected indicating that *Topi* may have entered the *An.gambiae* via horizontal gene transfer. However, the observed diversity was only 3 to 5- fold lower than expected as compared to *Herves* where there was a higher fold difference (3 to 125-fold) indicating again that *Topi* entered the *An.gambiae* genome earlier than *Herves*. Even though the element has been in the species longer than *Herves* it seems to retain at least one copy of the complete element which if active will provide a transposase source for the other non-autonomous *Topi* elements in the genome. The presence of a ~ 600 bp deleted form that encodes for a truncated protein of 179 amino acids (in every individual) might suggest some regulatory potential for this form of *Topi* transposase as seen in the case of *P*-elements and *hobo* elements (ENGELS *et al.* 1990; PERIQUET *et al.* 1990; PERIQUET *et al.* 1994).

During the invasive phase that follows the horizontal transfer of a transposable element into a new species, natural selection favors autonomous elements and this can be observed as a skewed ratio of synonymous and non-synonymous substitution rates (ROBERTSON and LAMPE 1995). In this study even though we saw evidence of purifying selection in the *Topi* transposase region, the neutral model of evolution was not rejected when compared to a discrete model. Only 14 of the total sample of 49 sequences that did not have any pre-mature stop codons were used in the analysis biasing towards the elements that are more conserved which might have led to the dN/dS ratios < 1. However, not being able to reject the neutral

model suggests that these might be molecular signals from the initial phase of selection that seem to persist in the genome for a long time.

In summary, this study shows that the *Topi* transposable element has been in the *An.gambiae* genome much longer than the *Herves* element. The insertion site-frequency distribution data indicates that the element is probably no longer active. The higher copy numbers observed in all locations does show that the activity of the element must have been much higher than the observed activity of *Herves*. It is also possible that the host regulation of these two elements was different. We found 14 different forms of the *Topi* transposase that were capable of producing the full-length protein. We saw no evidence of recent activity of *Topi* elements in these populations based on the insertion-site frequency distribution data, suggesting that these 14 forms are either not functional or are under the control of host regulation.

One of the striking observations from this study, as well as our earlier study of the *Herves* element are that Class II transposable elements in *An.gambiae* do not seem to evolve deleted forms as rapidly as observed with *P*-elements in *D.melanogaster* (O'HARE *et al.* 1992). They also seem to have at least one copy of the complete element for an extended period of time. This could be important for using Class II transposable elements as genetic drive agents as this fixed copy of the undeleted element would ensure the fixation of a copy of a refractory gene in the population. Because only one copy of a refractory gene is required to impair the development of the malaria parasite, *Plasmodium,* in the mosquito it could thereby result in reducing malaria transmission for an extended length of time.

Chapter 5: General Discussion

Global health burden due to vector-borne diseases is enormous; they collectively account for 1.5 million human deaths every year (HILL *et al.* 2005). Malaria, the most significant of the vector-borne diseases contributes to at least one million human deaths every year. Malaria is not just a disease that needs a cure, but is a complicated problem awaiting a solution. To combat this disease, health agencies have tried to minimize human contact with the vectors by getting rid of the mosquitoes using insecticides as well as by using pesticide-treated bed nets. Lack of adequate financial and political support for these vector-control programs in endemic countries hampers these efforts, and the insects are increasingly resistant to the insecticides that have been used discontinuously over a number of decades (GUBLER 1998). The parasites have also evolved, and are resistant to the widely used inexpensive anti-malarial drugs, such as chloroquine. The complex biology of the parasite *Plasmodium* has made development of an effective vaccine difficult to accomplish (GUBLER 1998).

To fight these harsh realities, a novel approach of genetically engineering transmission-incompetent mosquitoes and using them to replace the natural populations of mosquitoes is being explored. The success in generating transgenic mosquitoes with reduced vector-competence has raised hopes for using this strategy. In the last 10 years, tremendous progress has been achieved in identifying a number of effector genes and strategies to impair the development of the parasite in the mosquito (NIRMALA and JAMES 2003). We have successfully created at least three strains of *Anopheles* mosquitoes with impaired ability to transmit the malarial

parasite, *Plasmodium* (ITO *et al.* 2002; KIM *et al.* 2004; MOREIRA *et al.* 2002). Modern technology together with our understanding of the biology of the parasite as well as the mosquito has helped us to precisely express effector genes in the mosquito inhibiting the development of the parasites up to 87% (using Bee venom phospholipase A2). Realizing how quickly the parasites might evolve to overcome the barrier imposed by the effector molecules, there is a continuing search for more effectors and better targets to completely inhibit the parasite development.

Even after the identification of the effector gene, the success of this project depends on our ability to drive this transgene through natural populations of *Anopheles* mosquitoes. Class II transposable elements are a promising mechanism to drive transgenes because of their ability to move and rapidly increase in copy number under certain conditions (KIDWELL and RIBEIRO 1992). Compelling evidence from *P*-elements in *D.melanogaster* suggests that transposable elements are capable of efficiently spreading through large discontinuous populations in nature. *P*-elements have spread through the world's populations of *D.melanogaster* in the last 60 years after their introduction from a closely-related species, *Drosophila willistoni* (ANXOLABEHERE *et al.* 1988). However, we have limited understanding of the conditions in which such a rapid increase in frequency occurred. Our understanding of these conditions in *An.gambiae*, the target species for such a genetic control, will enable us to use Class II transposable elements as genetic drive agents in this species.

The projects described in this thesis have attempted to help better understand the behavior of Class II transposable elements in *An.gambiae*. We studied the behavior and evolution of two Class II transposable elements, *Herves* and *Topi*, in

natural populations of *An.gambiae* in Africa. We have used insertion-site frequency distributions to analyze the activity and copy number of these elements. We have used various analyses on the nucleotide sequences as well as the structure of these elements to understand their evolution in the natural populations of *An.gambiae*. We observed that *Topi* has been in *An.gambiae* genome longer than the *Herves* transposable element based on higher sequence diversity in *Topi* (average $\pi = 0.051$) compared to *Herves* elements (average $\pi = 0.0046$). The copy number of the *Topi* elements (10.2-33.8) was much higher than *Herves* (2.9-4.4) in all analyzed populations. There was no evidence of recent activity of *Topi* in *An.gambiae* as opposed to *Herves,* where a number of lines of evidence indicated activity. We saw evidence for *Herves* activity from the insertion-site frequency distribution in all locations studied, from DNA mobility assays in *Drosophila* as well as from the identification of functional forms of *Herves* transposase in natural populations of *An.gambiae*. However, we observed *Topi* transposase forms that are capable of producing full-length transposase; it is possible that these forms are functional and their inactivity is a result of the host repression system. The striking feature that is common to both elements is the presence of complete forms despite their long history in *An.gambiae*. Also, the activity of *Herves* together with the presence of *Topi* transposase forms that could be active suggests that these elements have been active for an extended period of time. Having a better understanding of the mutation rate in *An.gambiae* would have been helpful in deducing the time that these elements have stayed active. The structural integrity of these elements is an argument in favor of using Class II transposable elements as genetic drive agents. One of the concerns with

use of these elements as genetic drive agents is that they would lose the transgene long before they are fixed in natural populations of *An.gambiae*. This concern has arisen largely from the observation of rapid accumulation of deleted forms of *P*-elements in *D.melanogaster* (CARARETO *et al.* 1997; O'HARE *et al.* 1992). However, the results from this study are in contrast to what has been seen in *P*-elements; we observed that *Herves* and *Topi* had complete undeleted forms of the elements even though they have much longer histories in *An.gambiae* than *P*-elements in *D.melanogaster*. A single undeleted copy of the transposable element that we observed to be fixed in our studies, if it contained a copy of the refractory gene that is effective in inhibiting the development of *Plasmodium* would be enough to disrupt the transmission of malaria in natural populations of *An.gambiae*. These results are encouraging for the use of Class II transposable elements as genetic drive agents, at least in *An.gambiae*.

Future directions

Even though the strategy of genetically modified mosquitoes and population modification is promising, several challenges have to be met before it can be implemented to control vector-borne diseases. Transgenic mosquitoes have often been found to be less fit compared to mosquitoes not carrying the transgene. Genetically altered mosquitoes carrying the anti-parasite transgene would not have much impact in disrupting the disease-transmission in the wild, unless they are fit enough to compete with and eventually replace the natural populations of mosquitoes. Three of the five studies addressing the fitness of transgenic mosquitoes show that there was a reduction in fitness in mosquitoes carrying the transgenes (CATTERUCCIA

et al. 2003; IRVIN *et al.* 2004; MOREIRA *et al.* 2004); however, transgenic mosquitoes with one transgene, SM1, did not have any significant reduction in fitness compared to the non-transgenic mosquitoes (MOREIRA *et al.* 2004). Transgenic mosquitoes with SM1 (an anti-parasite gene shown to impair the development of the malaria parasite, *Plasmodium*) were also found to have a fitness advantage over the non-transgenic mosquitoes when both were fed with *Plasmodium*-infected blood (MARRELLI *et al.* 2007). Fitness of these genetically modified mosquitoes would also be reduced by the movement of the transposable elements used as genetic drive agents. This behavior, however, can not be avoided and is not as much of a concern as we have seen transposable elements (*P*-elements) are capable of spreading through populations in spite of the fitness cost associated with their mobility (ANXOLABEHERE *et al.* 1988). Efforts are, however, necessary to identify effector genes such as SM1 that cause less reduction in the fitness of the mosquitoes. Effector genes such as *cecA* (Cecropin A) that are part of the mosquito immune system may be less likely to impose a burden on the mosquito fitness. Exploring more immune effector genes may be helpful in identifying molecules that are effective in inhibiting the development of the parasite and are less disruptive to the mosquitoes.

Class II transposable elements have been shown to be capable of spreading in natural populations (*P*-elements) but their effectiveness as genetic drive agents is yet to be demonstrated in mosquitoes. A transposable element is yet to be identified that remobilizes at a sufficient rate in mosquitoes so that it can serve as a genetic drive agent. There have been two reports so far, both with *Mos1* elements in *Aedes aegypti*, that show the low germ-line remobilization rate of *Mos1* element in this mosquito

species. Wilson *et al* (2003) observed one new transposition event in 14,000 embryos that was screened in the next generation and Adelman *et al* (2007) observed two new insertions in three lines carrying a copy of *Mos1* elements. There have been no such studies in *An.gambiae*. Research towards identifying transposable elements that are capable of remobilizing in mosquitoes at a rate high enough to serve as genetic drive agents is absolutely critical. We have started addressing this deficiency, by understanding the remobilization potential of at least four Class II transposable elements, *piggyBac*, *Mos1*, *Hermes* and *Minos* in *An.gambiae*. Our results will show if any of these elements could serve as a genetic drive agent in this species. Similar efforts to identify transposable elements with higher remobilization rates, as well as developing methods to manipulate the existing transposable elements to increase their remobilization rates, are necessary so that they can serve as genetic drive agents in mosquitoes.

The population biology of *Anopheles* mosquitoes and malaria transmission in Africa is complex (FONTENILLE and SIMARD 2004). The mosquito-control programs employed are complicated by the presence of multiple vectors in the same area. Depending on the region, malaria can be transmitted by as many as five different species of *Anopheles*, *Anopheles gambiae, Anopheles arabiensis, Anopheles funestus, Anopheles nili* and *Anopheles moucheti* (FONTENILLE and SIMARD 2004). There is interspecies variation in the transmission of the disease according to season, ecological factors, urbanization, deforestation and agricultural practices (ANTONIO-NKONDJIO *et al.* 2002; ANTONIO-NKONDJIO *et al.* 2006; FONTENILLE and SIMARD 2004; HAY *et al.* 2005; MANGA *et al.* 1995). All of these species belong to different

groups of closely related species complexes that are morphologically indistinguishable. *An.gambiae* is a species complex, consisting of six different species, *An.gambiae s.s, An.arabiensis, An.merus, An.melas, An.quadrianulatus* and *An. bwambae*. *An.gambiae s.s,* the most efficient vector of the complex, has two molecular forms, M and S, which show differences in ecological tolerance and behavior. The M form has been observed to have a unique ability to breed in dry seasons. Even though there were no constraints in the mating of these two forms in the lab, and they were able to produce viable and fertile hybrids, a significant restriction in gene flow between these two forms has been observed in nature (DELLA TORRE *et al.* 2001; KRZYWINSKI and BESANSKY 2003; TAYLOR *et al.* 2001; TRIPET *et al.* 2001). Studies are in progress to understand the genetic basis for the reproductive isolation of these two forms in nature, which might shed light into the speciation process in *Anopheles* mosquitoes, as well as provide information for future malaria-control programs. Much still needs to be understood about the complexity and heterogeneity of malaria transmission, which is critical for effective malaria-control.

Besides the technical challenges, ethical and social challenges need to be addressed from the beginning. We can expect general public discomfort and anxiety when people realize the goal of this project is to make genetically modified mosquitoes that would outcompete the natural mosquitoes. In the past, there have been multiple occasions when novel mosquito-control programs of health agencies/governments have been falsely accused as being biological warfare. In the 1970's a WHO/Indian Council for Medical Research project of vector-control using the Sterile Insect Technique (SIT) had to be stopped after six years when a journalist

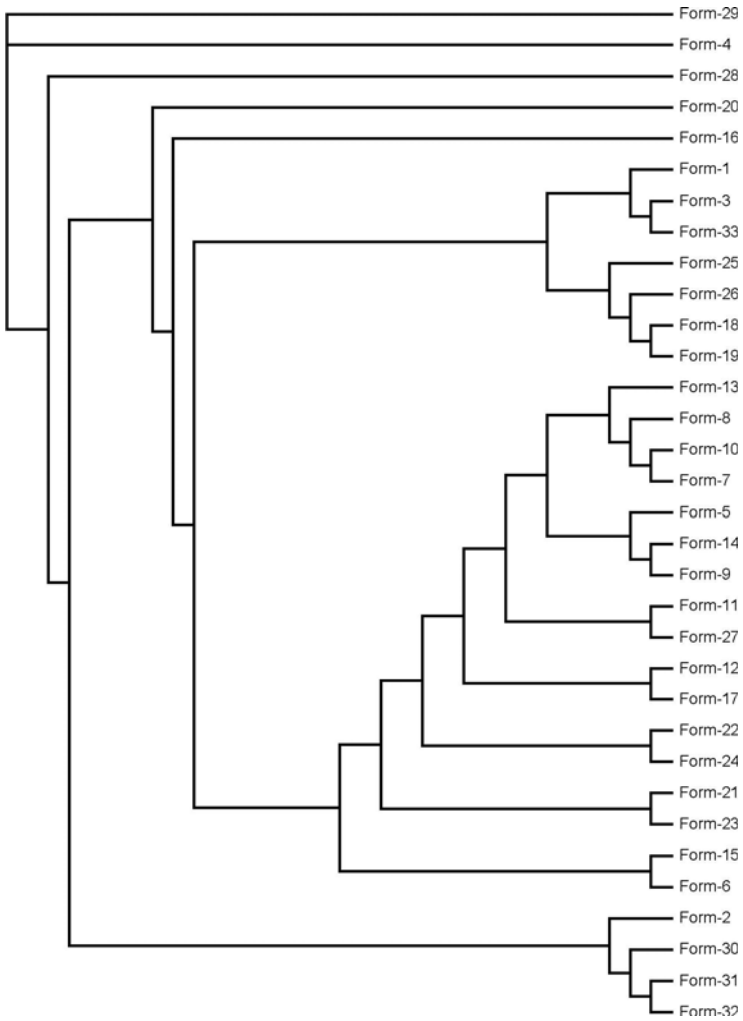
claimed that the intention of the project was not to research new methods of vector control but to engage in biological warfare (MACER 2005). Fears of biological warfare have increased after the anthrax scare in United States in 2001. Education before any intervention is critical to alleviate fears of using genetically modified mosquitoes. Scientists and health agencies need to educate the general public about the true biological properties of these genetically modified mosquitoes as well as the benefits of such an approach to human health. Careful evaluation of the risks involved - by gathering scientific data from field trials, accurately sharing the knowledge with the general public, and involving everyone in discussions - would help people to understand the potential of this approach and put to rest some of the fears that are a result of misinformation or lack of knowledge. Hopefully, people will realize that the goal of research like this is not to make dangerous mosquitoes but to improve human health by eliminating disease transmission.

The uncertainty, challenges and efforts being put into developing these new approaches has led some researchers to oppose these high-tech efforts. There are concerns that these novel high-tech efforts to tackle the disease take away resources from the already dwindling resources in the disease-control programs that use proven methods such as bed nets, insecticides and drugs. This is true and has to be avoided. The high-tech efforts should not grow at the expense of existing control programs.

Closing remarks

A better understanding of the behavior and evolution of Class II transposable elements in *An.gambiae* lead us to believe that Class II transposable elements still hold promise as a gene drive mechanism to spread refractory genes through natural

populations of *Anopheles gambiae.* Even though technical and social hurdles remain, undoubtedly genetically modified mosquitoes and population modification strategies will soon serve as a complementary strategy in our quest against malaria.



Supplementary Figure 1- 1: Neighbor Joining (NJ) tree of *Herves* forms
Neighbor Joining (NJ) tree of the thirty three different forms of *Herves* based on the first 528 bp of the 5' end of the transposase open reading frame.

Bibliography

ADELMAN, Z. N., N. JASINSKIENE and A. A. JAMES, 2002 Development and applications of transgenesis in the yellow fever mosquito, *Aedes aegypti.* Molecular and Biochemical Parasitology **121:** 1-10.

ADELMAN, Z. N., N. JASINSKIENE, S. ONAL, J. JUHN, A. ASHIKYAN *et al.*, 2007 *nanos* gene control DNA mediates developmentally regulated transposition in the yellow fever mosquito *Aedes aegypti*. Proceedings of the National Academy of Sciences of the United States of America **104:** 9970-9975.

AJIOKA, J. W., and W. F. EANES, 1989 The accumulation of *P*-elements on the tip of the X chromosomes of *Drosophila melanogaster.* Genet Res. **53:** 1-6.

ALPHEY, L., C. B. BEARD, P. BILLINGSLEY, M. COETZEE, A. CRISANTI *et al.*, 2002 Malaria control with genetically manipulated insect vectors. Science **298:** 119-121.

ANONYMOUS, 1991 Prospects for malaria control by genetic manipulation of its vectors (TDR/BCV/MAL-ENT/91.3). , World Health Organization, Geneva.

ANTONIO-NKONDJIO, C., P. AWONO-AMBENE, J. C. TOTO, J. Y. MEUNIER, S. ZEBAZE-KEMLEU *et al.*, 2002 High malaria transmission intensity in a village close to Yaounde, the capital city of Cameroon. Journal of Medical Entomology **39:** 350-355.

ANTONIO-NKONDJIO, C., C. H. KERAH, F. SIMARD, P. AWONO-AMBENE, M. CHOUAIBOU *et al.*, 2006 Complexity of the malaria vectorial system in Cameroon: Contribution of secondary vectors to malaria transmission. Journal of Medical Entomology **43:** 1215-1221.

ANXOLABEHERE, D., K. HU, D. NOUAUD and G. PERIQUET, 1990 PM system: A survey of *Drosophila melanogaster* strains from the People's Republic of China. Genet. Sel. Evol. **22:** 175-188.

ANXOLABEHERE, D., M. G. KIDWELL and G. PERIQUET, 1988 Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile *P* elements. Mol Biol Evol **5:** 252-269.

ARAVIND, L., 2000 The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. Trends in Biochemical Sciences **25:** 421-423.

ARENSBURGER, P., Y. J. KIM, J. ORSETTI, C. ALUVIHARE, D. A. O'BROCHTA *et al.*, 2005 An active transposable element, *Herves*, from the African malaria mosquito *Anopheles gambiae*. Genetics **169:** 697-708.

ASON, B., and W. S. REZNIKOFF, 2002 Mutational analysis of the base flipping event found in *Tn5* transposition. Journal of Biological Chemistry **277:** 11284-11291.

ATKINSON, P. W., A. C. PINKERTON and D. A. O'BROCHTA, 2001 Genetic transformation systems in insects. Annu Rev Entomol **46:** 317-346.

AUGE-GOUILLOU, C., B. BRILLET, S. GERMON, M. H. HAMELIN and Y. BIGOT, 2005 *Mariner Mos1* transposase dimerizes prior to ITR binding. Journal of Molecular Biology **351:** 117-130.

AUGE-GOUILLOU, C., M. H. HAMELIN, M. V. DEMATTEI, G. PERIQUET and Y. BIGOT, 2001 The ITR binding domain of the *mariner Mos-1* transposase. Molecular Genetics and Genomics **265:** 58-65.

BAKER, B., J. SCHELL, H. LORZ and N. FEDOROFF, 1986 Transposition of the maize controlling element *Activator* in Tobacco. Proceedings of the National Academy of Sciences of the United States of America **83:** 4844-4848.

BALU, B., D. A. SHOUE, M. J. FRASER and J. H. ADAMS, 2005 High-efficiency transformation of *Plasmodium falciparum* by the lepidopteran transposable element *piggyBac*. Proceedings of the National Academy of Sciences of the United States of America **102:** 16391-16396.

BEARD, C. B., C. CORDON-ROSALES and R. V. DURVASULA, 2002 Bacterial symbionts of the triatominae and their potential use in control of Chagas disease transmission. Ann. Rev. Entomology **47:** 123-141.

BEARD, C. B., R. V. DURVASULA and F. F. RICHARDS, 1998 Bacterial symbiosis in Arthropods and the control of disease transmission. Emerging Infectious Diseases **4:** 581-591.

BERGHAMMER, A. J., M. KLINGLER and E. A. WIMMER, 1999 A universal marker for transgenic insects. Nature **402:** 370-371.

BESANSKY, N. J., T. LEHMANN, G. T. FAHEY, D. FONTENILLE, L. E. O. BRAACK *et al.*, 1997 Patterns of mitochondrial variation within and between African malaria vectors, *Anopheles gambiae* and *An-arabiensis*, suggest extensive gene flow. Genetics **147:** 1817-1828.

BESANSKY, N. J., O. MUKABAYIRE, J. A. BEDELL and H. LUSZ, 1996 *Pegasus*, a small terminal inverted repeat transposable element found in the white gene of *Anopheles gambiae*. Genetica **98:** 119-129.

BESSEREAU, J. L., A. WRIGHT, D. C. WILLIAMS, K. SCHUSKE, M. W. DAVIS *et al.*, 2001 Mobilization of a *Drosophila* transposon in the *Caenorhabditis elegans* germ line. Nature **413:** 70-74.

BIEDLER, J., Y. QI, D. HOLLIGAN, A. DELLA TORRE, S. WESSLER *et al.*, 2003 Transposable element (TE) display and rapid detection of TE insertion polymorphism in the *Anopheles gambiae* species complex. Insect Molec. Biol. **12:** 211-216.

BIEDLER, J., and Z. J. TU, 2003 Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: Unprecedented diversity and evidence of recent activity. Molecular Biology and Evolution **20:** 1811-1825.

BIEMONT, C., F. LEMEUNIER, M. P. GARCIA GUERREIRO, J. F. BROOKFIELD, C. GAUTIER *et al.*, 1994 Population dynamics of the *copia*, *mdg1, mdg3, gypsy* and *P* transposable elements in a natural population of *Drosophila melanogaster.* Genet. Res. **63:** 197-212.

BIESSMANN, H., M. F. WALTER, D. LE, S. CHUAN and J. G. YAO, 1999 *Moose*, a new family of LTR-retrotransposons in the mosquito *Anopheles gambiae*. Insect Molecular Biology **8:** 201-212.

BLACKMAN, R. K., M. MACY, D. KOEHLER, R. GRIMAILA and W. M. GELBART, 1989 Identification of a fully-functional *hobo* transposable element and its use for germ-line transformation of *Drosophila*. EMBO J. **8:** 211-217.

BLANDIN, S., S. H. SHIAO, L. F. MOITA, C. J. JANSE, A. P. WATERS *et al.*, 2004 Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. Cell **116:** 661-670.

BOETE, C., and J. C. KOELLA, 2002 A theoretical approach to predicting the success of genetic manipulation of malaria mosquitoes in malaria control. Malaria J. **25:** 1-3.

BOETE, C., and J. C. KOELLA, 2003 Evolutionary ideas about genetically manipulated mosquitoes and malaria control. Trends in Parasitology **19:** 32-38.

BRAIG, H. R., and G. YAN (Editors), 2001 The spread of genetic constructs in natural insect populations, pp. 251–314 in Genetically engineered organisms: Assessing Environmental and Human Health Effects, edited by D. K. Letourneau and B. E. Burrows. CRC Press, Boca Raton, FL.

BROOKFIELD, J. F., 1986 A model for DNA sequence evolution within transposable element families. Genetics **112:** 393-407.

BROOKFIELD, J. F., 2005 The ecology of the genome - mobile DNA elements and their hosts. Nat Rev Genet **6:** 128-136.

BUCHETON, A., C. VAURY, M. C. CHABOISSIER, P. ABAD, A. PELISSON *et al.*, 1992 *I* elements and the *Drosophila* genome. Genetica **86:** 175-190.

BURT, A., 2003 Site-specific selfish genes as tools for the control and genetic engineering of natural populations. Proc Biol Sci. **270:** 921-928.

CARARETO, C. M., W. KIM, M. F. WOJCIECHOWSKI, P. O'GRADY, A. V. PROKCHOROVA *et al.*, 1997 Testing transposable elements as genetic drive mechanisms using *Drosophila P* element constructs as a model system. Genetica **101:** 13-33.

CATTERUCCIA, F., H. C. J. GODFRAY and A. CRISANTI, 2003 Impact of genetic manipulation on the fitness of *Anopheles stephensi* mosquitoes. Science **299:** 1225-1227.

CATTERUCCIA, F., T. NOLAN, T. G. LOUKERIS, C. BLASS, C. SAVAKIS *et al.*, 2000 Stable germline transformation of the malaria mosquito *Anopheles stephensi*. Nature **405:** 959-962.

CHARLESWORTH, B., and D. CHARLESWORTH, 1983 The population dynamics of transposable elements. Genet. Res. **42:** 1-27.

CHARLESWORTH, B., and C. H. LANGLEY, 1989 The population genetics of *Drosophila* transposable elements. Annu Rev Genet **23:** 251-287.

CLEMENT, M., D. POSADA and K. A. CRANDALL, 2000 TCS: a computer program to estimate gene genealogies. Molec. Ecol. **9:** 1657-1679.

COOLEY, L., R. KELLEY and A. SPRADLING, 1988 Insertional Mutagenesis of the *Drosophila* Genome with Single *P*-Elements. Science **239:** 1121-1128.

CRAIG, G. B., 1963 Prospects for vector control through manipulation of populations. Bull. World Health Organization **29:** 89-97.

CRAIG, N. L., 1997 Target site selection in transposition. Annual Review of Biochemistry **66:** 437-474.

CRAIG, N. L., R. CRAIGIE, M. GELLERT and A. M. LAMBOWITZ (Editors), 2002 *Mobile DNA II*. ASM Press, Washington,D.C.

CURTIS, C. F., 2003 Measuring public-health outcomes of release of transgenic mosquitoes, pp. 223-234 in *Ecological aspects for application of genetically modified mosquitoes.*, edited by W. TAKKEN and T. W. SCOTT. Kluwer Academic Publisher, Dordrecht.

DAVIES, D. R., L. M. BRAAM, W. S. REZNIKOFF and I. RAYMENT, 1999 The three-dimensional structure of a *Tn5* transposase-related protein determined to 2.9-angstrom resolution. Journal of Biological Chemistry **274:** 11904-11913.

DE CARVALHO, M. O., J. C. SILVA and E. L. S. LORETO, 2004 Analyses of *P*-like transposable element sequences from the genome of *Anopheles gambiae*. Insect Molecular Biology **13:** 55-63.

DELLA TORRE, A., C. FANELLO, M. AKOGBETO, J. DOSSOU-YOVO, G. FAVIA *et al.*, 2001 Molecular evidence of incipient speciation within *Anopheles gambiae* s.s. in West Africa. Insect Molecular Biology **10:** 9-18.

DIMOPOULOS, G., H. M. MULLER, E. A. LEVASHINA and F. C. KAFATOS, 2001 Innate immune defense against malaria infection in the mosquito. Current Opinion in Immunology **13:** 79-88.

DING, S., X. H. WU, G. LI, M. HAN, Y. ZHUANG *et al.*, 2005 Efficient transposition of the *piggyBac* resource (PB) transposon in mammalian cells and mice. Cell **122:** 473-483.

DUFFY, J. B., 2002 GAL4 system in Drosophila: A fly geneticist's Swiss army knife. Genesis **34:** 1-15.

ENGELS, W. R., 1989 *P*- elements in *Drosophila melanogaster*, pp. 439-484 in *Mobile DNA*, edited by D. E. BERG and M. M. HOW. American Society for Microbiology, Washington D.C.

ENGELS, W. R., D. M. JOHNSON-SCHLITZ, W. B. EGGLESTON and J. SVED, 1990 High-frequency *P* element loss in *Drosophila* is homologue dependent. Cell **62:** 515-525.

FESCHOTTE, C., 2006 The *piggyBac* transposon holds promise for human gene therapy. Proceedings of the National Academy of Sciences of the United States of America **103:** 14981-14982.

FONTENILLE, D., and F. SIMARD, 2004 Unravelling complexities in human malaria transmission dynamics in Africa through a comprehensive knowledge of vector populations. Comparative Immunology Microbiology and Infectious Diseases **27:** 357-375.

GALINDO, M. I., V. LADEVEZE, F. LEMEUNIER, R. KALMES, G. PERIQUET *et al.*, 1995 Spread of the autonomous transposable element *hobo* in the genome of *Drosophila melanogaster*. Mol Biol Evol **12:** 723-734.

GORROCHOTEGUI-ESCALANTE, N., and W. C. BLACK, 2003 Amplifying whole insect genomes with multiple displacement amplification. Insect Molecular Biology **12:** 195-200.

GROSSMAN, G. L., A. J. CORNEL, C. S. RAFFERTY, H. M. ROBERTSON and F. H. COLLINS, 1999 *Tsessebe, Topi and Tiang*: three distinct *Tc1*-like transposable elements in the malaria vector, *Anopheles gambiae*. Genetica **105:** 69-80.

GROSSMAN, G. L., C. S. RAFFERTY, J. R. CLAYTON, T. K. STEVENS, O. MUKABAYIRE *et al.*, 2001 Germline transformation of the malaria vector, *Anopheles gambiae*, with the *piggyBac* transposable element. Insect Molecular Biology **10:** 597-604.

GUBLER, D. J., 1998 Resurgent Vector-Borne Diseases as a Global Health Problem Emerging Infectious Diseases **4:** 442-450.

GUIMOND, N., D. K. BIDESHI, A. C. PINKERTON, P. W. ATKINSON and D. A. O'BROCHTA, 2003 Patterns of *Hermes* Transposition in *Drosophila melanogaster.* Molec. Gen. Genet. **268:** 779-790.

HANIFORD, D. B., 2006 Transpososome dynamics and regulation in *Tn10* transposition. Critical Reviews in Biochemistry and Molecular Biology **41:** 407-424.

HARTL, D. L., A. R. LOHE and E. R. LOZOVSKAYA, 1997a Modern thoughts on an ancyent marinere: Function, evolution, regulation. Annual Review of Genetics **31:** 337-358.

HARTL, D. L., E. R. LOZOVSKAYA, D. I. NURMINSKY and A. R. LOHE, 1997b What restricts the activity of *mariner*-like transposable elements. Trends Genet. **13:** 197-201.

HAY, S. I., C. A. GUERRA, A. J. TATEM, P. M. ATKINSON and R. W. SNOW, 2005 Urbanization, malaria transmission and disease burden in Africa. Nature Reviews Microbiology **3:** 81-90.

HICKMAN, A. B., Z. N. PEREZ, L. Q. ZHOU, P. MUSINGARIMI, R. GHIRLANDO *et al.*, 2005 Molecular architecture of a eukaryotic DNA transposase. Nature Structural & Molecular Biology **12:** 715-721.

HILL, C. A., F. C. KAFATOS, S. K. STANSFIELD and F. H. COLLINS, 2005 Arthropod-borne diseases: vector control in the genomics era. Nature Reviews Microbiology **3:** 262-268.

HOLT, R. A., G. M. SUBRAMANIAN, A. HALPERN, G. G. SUTTON, R. CHARLAB *et al.*, 2002 The genome sequence of the malaria mosquito *Anopheles gambiae*. Science **298:** 129-149.

HORIE, K., A. KUROIWA, M. IKAWA, M. OKABE, G. KONDOH *et al.*, 2001 Efficient chromosomal transposition of a *Tc1/mariner*-like transposon *Sleeping Beauty* in mice. Proceedings of the National Academy of Sciences of the United States of America **98:** 9191-9196.

IRVIN, N., M. S. HODDLE, D. A. O'BROCHTA, B. CAREY and P. W. ATKINSON, 2004 Assessing fitness costs for transgenic *Aedes aegypti* expressing the GFP marker and transposase genes. Proceedings of the National Academy of Sciences of the United States of America **101:** 891-896.

ITO, J., A. GHOSH, A. L. MOREIRA, A. W. ERNST and M. JACOBS-LORENA, 2002 Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. Nature **417** 452-455.

IVICS, Z., P. B. HACKETT, R. H. PLASTERK and Z. IZSVAK, 1997 Molecular reconstruction of *Sleeping beauty*, a *Tc1*-like transposon from fish, and its transposition in human cells. Cell **91:** 501-510.

IZSVAK, Z., and Z. IVICS, 2004 *Sleeping Beauty* transposition: Biology and applications for molecular therapy. Molecular Therapy **9:** 147-156.

JAMES, A. A., 1992 Mosquito molecular genetics: The hands that feed bite back. Science **257:** 37-38.

JAMES, A. A., 2005 Gene drive systems in mosquitoes: Rules of the road. Trends in Parasitology **21:** 64-67.

JASINSKIENE, N., C. J. COATES, M. Q. BENEDICT, A. J. CORNEL, C. S. RAFFERTY *et al.*, 1998 Stable transformation of the yellow fever mosquito, *Aedes aegypti*, with the *Hermes* element from the housefly. Proc Natl Acad Sci U S A **95:** 3743-3747.

KAPITONOV, V. V., and J. JURKA, 2005 RAG1 core and V(D)J recombination signal sequences were derived from *Transib* transposons. Plos Biology **3:** 998-1011.

KAPLAN, N., and J. F. Y. BROOKFIELD, 1983 Transposable elements in Mendelian populations. III. Statistical results. Genetics **104:** 485-495.

KAWAKAMI, K., and T. NODA, 2004 Transposition of the *Tol2* element, an *Ac*-like element from the Japanese medaka fish *Oryzias latipes*, in mouse embryonic stem cells. Genetics **166:** 895-899.

KEMPKEN, F., and F. WINDHOFER, 2001 The *hAT* family: a versatile transposon group common to plants, fungi, animals, and man. Chromosoma **110:** 1-9.

KIDWELL, M. G., 2002 Transposable elements and the evolution of genome size in eukaryotes. Genetica **115:** 49-63.

KIDWELL, M. G., and D. R. LISCH, 2001 Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution Int J Org Evolution **55:** 1-24.

KIDWELL, M. G., and J. M. C. RIBEIRO, 1992 Can transposable elements be used to drive disease refractoriness genes into vector populations?. Parasitol. Today **8:** 325-329.

KIKUNO, K., K. TANAKA, M. ITOH, Y. TANAKA, I. A. BOUSSY *et al.*, 2006 Patterns of *hobo* elements and their effects in natural populations of *Drosophila melanogaster* in Japan. Heredity **96:** 426-433.

KIM, S. J., S. H. SUNG, J. Y. SEUNG, H. Y. NAM, K. C. KIM *et al.*, 2003 Production of transgenic silkworm (*Bombyx mori L.*) by *P* element-mediated integration of recombinant luciferase with silkworm fibroin promoter. Korean Journal of Genetics **25:** 57-62.

KIM, W., H. KOO, A. M. RICHMAN, D. SEELEY, J. VIZIOLI *et al.*, 2004 Ectopic expression of a *cecropin* Transgene in the human malaria vector mosquito *Anopheles gambiae* (Diptera: Culicidae):Effects on Susceptibility to Plasmodium. J. Med. Entomol **41:** 447-455.

KISZEWSKI, A. E., and A. SPIELMAN, 1998 Spatially explicit model of transposon-based genetic drive mechanisms for displacing fluctuating populations of anopheline vector mosquitoes. Journal of Medical Entomology **35:** 584-590.

KLECKNER, N., R. M. CHALMERS, D. KWON, J. SAKAI and S. BOLLAND, 1996 *Tn10* and *IS10* transposition and chromosome rearrangements: Mechanism and regulation in vivo and in vitro, pp. 49-82 in *Transposable Elements*.

KRZYWINSKI, J., and N. J. BESANSKY, 2003 Molecular systematics of *Anopheles*: From subgenera to subpopulations. Annual Review of Entomology **48:** 111-139.

KUMAR, A., M. SERINGHAUS, M. C. BIERY, R. J. SARNOVSKY, L. UMANSKY *et al.*, 2004a Large-scale mutagenesis of the yeast genome using a *Tn7*-derived multipurpose transposon. Genome Research **14:** 1975-1986.

KUMAR, S., K. TAMURA and M. NEI, 2004b MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics **5:** 150-163.

KUNZE, R., and C. F. WEIL, 2002 The *hAT* and *CACTA* superfamilies of plant transposons., pp. 565-610 in *Mobile DNA II*, edited by N. L. CRAIG, R. CRAIGE, M. GELLERT and A. M. LAMBOWITZ. American Society of Microbiology, Washington, D.C.

LAMPE, D. J., T. E. GRANT and H. M. ROBERTSON, 1998 Factors affecting transposition of the *Himar1 mariner* transposon *in vitro.* Genetics **149:** 179-187.

LANGLEY, C. H., J. F. Y. BROOKFIELD and N. KAPLAN, 1983 Transposable elements in mendelian populations. I. A theory. Genetics **104:** 457-471.

LARGAESPADA, D. A., 2003 Generating and manipulating transgenic animals using transposable elements. Reprod Biol Endocrinol **1:** 80.

LEE, H. H., J. Y. YOON, H. S. KIM, J. Y. KANG, K. H. KIM *et al.*, 2006 Crystal structure of a metal ion-bound *IS200* transposase. Journal of Biological Chemistry **281:** 4261-4266.

LEHMANN, T., N. J. BESANSKY, W. A. HAWLEY, T. G. FAHEY, L. KAMAU *et al.*, 1997 Microgeographic structure of *Anopheles gambiae* in wetern Kenya based on mtDNA and microsatellite loci. Mol. Ecol. **6:** 243-253.

LEHMANN, T., W. A. HAWLEY, H. GREBERT and F. H. COLLINS, 1998 The effective population size of *Anopheles gambiae* in Kenya: Implications for population structure. Mol Biol Evol **15:** 264-276.

LEHMANN, T., W. A. HAWLEY, L. KAMAU, D. FONTENILLE, F. SIMARD *et al.*, 1996 Genetic differentiation of *Anopheles gambiae* from East and West Africa: comparison of microsatellite and allozyme loci. Heredity **77:** 192-200.

LEHMANN, T., M. LICHT, N. ELISSA, B. T. MAEGA, J. M. CHIMUMBWA *et al.*, 2003 Population Structure of *Anopheles gambiae* in Africa. J Hered **94:** 133-147.

LEIGH-BROWN, A. J., and J. E. MOSS, 1987 Transposition of the *I* element and *copia* in natural populations of *Drosophila melanogaster*. Genet Res. **49:** 231-237.

LEVIS, R. W., R. GANESAN, K. HOUTCHENS, L. A. TOLAR and F. M. SHEEN, 1993 Transposons in-Place of Telomeric Repeats at a Drosophila telomere. Cell **75:** 1083-1093.

LORENZEN, M. D., T. KIMZEY, T. D. SHIPPY, S. J. BROWN, R. E. DENELL *et al.*, 2007 *piggyBac*-based insertional mutagenesis in *Tribolium castaneum* using donor/helper hybrids. Insect Molecular Biology **16:** 265-275.

MACER, D., 2005 Ethical, legal and social issues of genetically modifying insect vectors for public health. Insect Biochemistry and Molecular Biology **35:** 649-660.

MANGA, L., J. C. TOTO and P. CARNEVALE, 1995 Malaria vectors and Transmission in an Area Deforested for a new international airport in southern Cameroon. Annales De La Societe Belge De Medecine Tropicale **75:** 43-49.

MARCUS, J. M., D. M. RAMOS and A. MONTEIRO, 2004. Germline transformation of the butterfly *Bicyclus anynana*. Proc Biol Sci **271 Suppl 5:** S263-265.

MARRELLI, M. T., C. Y. LI, J. L. RASGON and M. JACOBS-LORENA, 2007 Transgenic malaria-resistant mosquitoes have a fitness advantage when feeding on *Plasmodium*-infected blood. Proceedings of the National Academy of Sciences of the United States of America **104:** 5580-5583.

MASIDE, X., C. BARTOLOME and B. CHARLESWORTH, 2002 *S*-element insertions are associated with the evolution of the *Hsp70* genes in *Drosophila melanogaster*. Current Biology **12:** 1686-1691.

MEERAUS, W. H., C. MAXWELL and D. A. O'BROCHTA, 2005 *Herves* transposable elements in Tanzanian *Anopheles gambiae* - potential uses in malaria control. Trans. Royal Soc. Trop. Med. Hyg. **99:** 942-956.

MICHEL, K., and P. W. ATKINSON, 2003 Nuclear localization of the *Hermes* transposase depends on basic amino acid residues at the N-terminus of the protein. Journal of Cellular Biochemistry **89:** 778-790.

MICHEL, K., D. A. O'BROCHTA and P. W. ATKINSON, 2002 Does the proposed DSE motif form the active center in the *Hermes* transposase? Gene **298:** 141-146.

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:** 959-970.

MICHEL, K., A. STAMENOVA, A. C. PINKERTON, G. FRANZ, A. S. ROBINSON *et al.*, 2001 *Hermes*-mediated germ-line transformation of the Mediterranean fruit fly *Ceratitis capitata*. Insect Mol Biol **10:** 155-162.

MILLER, L. H., 1992 The Challenge of Malaria. Science **257:** 36-37.

MISKEY, C., Z. IZSVAK, K. KAWAKAMI and Z. IVICS, 2005. DNA transposons in vertebrate functional genomics. Cellular and Molecular Life Sciences **62:** 629-641.

MISKEY, C., Z. IZSVAK, R. H. PLASTERK and Z. IVICS, 2003 The *Frog Prince*: a reconstructed transposon from *Rana pipiens* with high transpositional activity in vertebrate cells. Nucleic Acids Research **31:** 6873-6881.

MOREIRA, L. A., M. J. EDWARDS, F. ADHAMI, N. JASINSKIENE, A. AJAMES *et al.*, 2000 Robust gut-specific gene expression in transgenic *Aedes aegypti* mosquitoes. Proc Natl Acad Sci **97:** 10895-10898.

MOREIRA, L. A., J. ITO, A. GHOSH, M. DEVENPORT, H. ZIELER *et al.*, 2002 Bee Venom Phospholipase inhibits malaria Parasite Development in Transgenic Mosquitoes. J Biol Chem **277:** 40839-40843.

MOREIRA, L. A., J. WANG, F. H. COLLINS and M. JACOBS-LORENA, 2004 Fitness of anopheline mosquitoes expressing transgenes that inhibit *Plasmodium* development. Genetics **166:** 1337-1341.

MORLAIS, I., N. PONCON, F. SIMARD, A. COHUET and D. FONTENILLE, 2004 Intraspecific nucleotide variation in *Anopheles gambiae*: new insights into the biology of malaria vectors. Am. J. Trop. Med. Hyg. **71:** 795-802.

NEI, M., and S. KUMAR, 2000 *Molecular evolution and phylogenetics.* University Press, New York.

NEI, M., and W. H. LI, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76:** 5269-5273.

NENE, V., J. R. WORTMAN, D. LAWSON, B. HAAS, C. KODIRA *et al.*, 2007 Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science **316:** 1718-1723.

NIRMALA, X., and A. A. JAMES, 2003 Engineering *Plasmodium*-refractory phenotypes in mosquitoes. Trends in Parasitology **19:** 384-387.

O'BROCHTA, D. A., and P. W. ATKINSON, 1996 Transposable elements and gene transformation in non-drosophilid insects. Insect Biochem Mol Biol **26:** 739-753.

O'BROCHTA, D. A., P. W. ATKINSON and M. J. LEHANE, 2000 Transformation of *Stomoxys calcitrans* with a *Hermes* gene vector. Insect Mol Biol **9:** 531-538.

O'BROCHTA, D. A., R. A. SUBRAMANIAN, J. ORSETTI, E. PECKHAM, N. NOLAN *et al.*, 2006 *hAT* element population genetics in *Anopheles gambiae s.l.* in Mozambique. Genetica **127:** 185-198.

O'BROCHTA, D. A., W. D. WARREN, K. J. SAVILLE and P. W. ATKINSON, 1996 *Hermes* , a functional non-drosophilid insect gene vector from *Musca domestica*. Genetics. **142:** 907-914.

O'HARE, K., A. DRIVER, S. MCGRATH and D. M. JOHNSON-SCHLITZ, 1992 Distribution and structure of cloned *P* elements for the *Drosophila melanogaster P* strain π2. Genet Res. **60:** 33-41.

PARINOV, S., I. KONDRICHIN, V. KORZH and A. EMELYANOV, 2004 *Tol2* transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes *in vivo*. Developmental Dynamics **231:** 449-459.

PASCUAL, L., and G. PERIQUET, 1991 Distribution of *hobo* transposable elements in natural-populations of *Drosophila melanogaster*. Molecular Biology and Evolution **8:** 282-296.

PERIQUET, G., M. H. HAMELIN, Y. BIGOT and H. KAI, 1989a Presence of the deleted *hobo* element *Th* in Eurasian populations of *Drosophila melanogaster*. Genetics Selection Evolution **21:** 107-111.

PERIQUET, G., M. H. HAMELIN, Y. BIGOT and A. LEPISSIER, 1989b Geographical and historical patterns of distribution of *hobo* elements in *Drosophila melanogaster* populations. Journal of Evolutionary Biology **2:** 223-229.

PERIQUET, G., M. H. HAMELIN, R. KALMES and J. EEKEN, 1990 *hobo* elements and their deletion-derivative sequences in *Drosophila melanogaster* and its sibling species *Drosophila simulans*, *D. mauritiana* and *D. sechellia*. Genetics Selection Evolution **22:** 393-402.

PERIQUET, G., F. LEMEUNIER, Y. BIGOT, M. H. HAMELIN, C. BAZIN *et al.*, 1994 The Evolutionary genetics of the *hobo* transposable element in the *Drosophila melanogaster* complex. Genetica **93:** 79-90.

PETERS, J. E., and N. L. CRAIG, 2001 *Tn7*: Smarter than we thought. Nature Reviews Molecular Cell Biology **2:** 806-814.

QUESNEVILLE, H., D. NOUAUD and D. ANXOLABEHERE, 2003 Detection of new transposable element families in *Drosophila melanogaster* and *Anopheles gambiae* genomes. Journal of Molecular Evolution **57:** S50-S59.

RASGON, J. L., and F. GOULD, 2005 Transposable element insertion location bias and the dynamics of gene drive in mosquito populations. Insect Molec. Biol. **14:** 493-500.

RAY, D. A., H. J. PAGAN, M. L. THOMPSON and R. D. STEVENS, 2007 Bats with *hATs*: evidence for recent DNA transposon activity in genus *Myotis*. Mol Biol Evol **24:** 632-639.

REZNIKOFF, W. S., 2003 *Tn5* as a model for understanding DNA transposition. Molecular Microbiology **47:** 1199-1206.

RIBEIRO, J. M. C., and M. G. KIDWELL, 1994 Transposable elements as population drive mechanisms - Specification of critical parameter values. Journal of Medical Entomology **31:** 10-16.

RICHARDSON, J. M., A. DAWSON, N. O'HAGAN, P. TAYLOR, D. J. FINNEGAN *et al.*, 2006 Mechanism of *Mos1* transposition: insights from structural analysis. Embo Journal **25:** 1324-1334.

RIEHLE, M. A., P. SRINIVASAN, C. K. MOREIRA and M. JACOBS-LORENA, 2003 Towards genetic manipulation of wild mosquito populations to combat malaria:advances and challenges. J Exp Biol. **206:** 3809-3816.

RIO, D. C., 2002 *P* transposable elements in *Drosophila melanogaster*., pp. 1204 in *Mobile DNA II*, edited by N. L. CRAIG, R. CRAIGE, M. GELLERT and A. M. LAMBOWITZ. ASM Press, Washington, D. C.

ROBERTS, R. J., and X. D. CHENG, 1998 Base flipping. Annual Review of Biochemistry **67:** 181-198.

ROBERTSON, H. M., 2002 Evolution of DNA transposons in eukaryotes, pp. 1093-1110 in *Mobile DNA II*, edited by N. L. CRAIG, R. CRAIGE, M. GELLERT and A. M. LAMBOWITZ. ASM Press, Washington, D. C.

ROBERTSON, H. M., and D. J. LAMPE, 1995 Recent horizontal transfer of a *mariner* element between Diptera and Neuroptera. Mol Biol Evol **12:** 850-862.

ROZAS, J., and R. ROZAS, 1995 DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. Comput. Appl. Biosci. **11:** 621-625.

ROZAS, J., J. C. SANCHEZ-DELBARRIO, X. MESSEGUER and R. ROZAS, 2003 DnaSP, DNA polymorphism analysis by the coalescent and other methods. Bioinformatics **19:** 2496-2497.

RUBIN, E., G. LITHWICK and A. A. LEVY, 2001 Structure and evolution of the *hAT* transposon superfamily. Genetics **158:** 949-957.

RUBIN, G. M., and A. C. SPRADLING, 1982 Genetic transformation of *Drosophila* with transposable element vectors. Science. **218:** 348-353.

SANCHEZ-GRACIA, A., X. MASIDE and B. CHARLESWORTH, 2005 High rate of horizontal transfer of transposable elements in *Drosophila*. Trends in Genetics **21:** 200-203.

SASAKURA, Y., S. AWAZU, S. CHIBA, S. KANO and N. SATOH, 2003 Application of *Minos*, one of the *Tc1/mariner* superfamily transposable elements, to ascidian embryos as a tool for insertional mutagenesis. Gene **308:** 11-20.

SCHLENKE, T. A., and D. J. BEGUN, 2004 Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. Proceedings of the National Academy of Sciences of the United States of America **101:** 1626-1631.

SCOTT, J. A., W. G. BROGDON and F. H. COLLINS, 1993 Identification of single specimens of the *Anopheles gambiae* complex by polymerase chain reaction. Am. J. Trop. Med. Hyg. **49:** 520-529.

SILVA, J. C., and M. G. KIDWELL, 2000 Horizontal transfer and selection in the evolution of *P* elements. Molec. Biol. Evol. **17:** 1542-1547.

SILVA, J. C., and M. G. KIDWELL, 2004 Evolution of *P* elements in natural populations of *Drosophila willistoni* and *D. sturtevanti*. Genetics **168:** 1323-1335.

SILVA, J. C., E. L. LORETO and J. B. CLARK, 2004 Factors that affect the horizontal transfer of transposable elements. Current Issues in Molecular Biology **6:** 57-71.

SIMMONS, G., 1992 Horizontal Transfer of *hobo* transposable elements within the *Drosophila melanogaster* species Complex: Evidence from DNA sequencing. Mol. Biol. Evol. **9:** 1050 - 1060.

SINDEN, R. E., 2002 Molecular interactions between *Plasmodium* and its insect vectors. Cellular Microbiology **4:** 713-724.

SPRADLING, A. C., 1986 *P* element-mediated transformation, pp. 175-198 in *Drosophila. A practical approach*, edited by D. B. ROBERTS. IRL Press, Oxford and Washington, D. C.

SUBRAMANIAN, R. A., P. ARENSBURGER, P. W. ATKINSON and A. O'BROCHTA D, 2007 Transposable element dynamics of the *hAT* element *Herves* in the human malaria vector *Anopheles gambiae* s.s. Genetics **176:** 2477-2487.

TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585-595.

TAMURA, T., C. THILBERT, C. ROYER, T. KANDA, E. ABRAHAM *et al.*, 2000 Germline transformation of the silkworm *Bombyx mori L*-using a *piggyBac* transposon-derived vector. Nature Biotechnology **18:** 81-84.

TAYLOR, C., Y. T. TOURE, J. CARNAHAN, D. E. NORRIS, G. DOLO *et al.*, 2001 Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, west Africa. Genetics **157:** 743-750.

TEMPLETON, A. R., K. A. CRANDALL and C. F. SING, 1992 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram Estimation. Genetics **132:** 619-633.

TREN, R., and R. BATE, 2001 When politics kills-Malaria and the DDT story. Institute of Economic Affairs.

TRIPET, F., Y. T. TOURE, C. E. TAYLOR, D. E. NORRIS, G. DOLO *et al.*, 2001 DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae*. Molecular Ecology **10:** 1725-1732.

TU, Z., and C. COATES, 2004 Mosquito transposable elements. Insect Biochemistry and Molecular Biology **34:** 631-644.

TU, Z. J., 2001 Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. Proceedings of the National Academy of Sciences of the United States of America **98:** 1699-1704.

UCHINO, K., M. IMAMURA, K. SHIMIZU, T. KANDA and T. TAMURA, 2007 Germ line transformation of the silkworm, *Bombyx mori*, using the transposable element *Minos*. Molecular Genetics and Genomics **277:** 213-220.

VANLUENEN, H., S. D. COLLOMS and R. H. A. PLASTERK, 1994 The mechanism of transposition of *Tc3* in *C.elegans*. Cell **79:** 293-301.

VANPOUDEROYEN, G., R. F. KETTING, A. PERRAKIS, R. H. A. PLASTERK and T. K. SIXMA, 1997 Crystal structure of the specific DNA-binding domain of *Tc3* transposase of *C.elegans* in complex with transposon DNA. Embo Journal **16:** 6044-6054.

VANSLUYS, M. A., J. TEMPE and N. FEDOROFF, 1987 Studies on the introduction and mobility of the maize *Activator* element in *Arabidopsis thaliana* and *Daucus carota*. Embo Journal **6:** 3881-3889.

VOS, P., R. HOGERS, M. BLEEKER, M. REIJANS, T. VAN DE LEE *et al.*, 1995 AFLP: a new technique for DNA fingerprinting. Nuc. Acid Res. **23:** 4407-4414.

VULULE, J. M., R. F. BEACH, F. K. ATIELE, J. M. ROBERTS, D. L. MOUNT *et al.*, 1994 Reduced susceptibility of *Anopheles gambiae* to permethrin associated wtih the use of permethrin-impregnated bednets and curtains in Kenya. Med. Vet. Entomol. **8:** 71-75.

WATERHOUSE, R. M., E. V. KRIVENTSEVA, S. MEISTER, Z. Y. XI, K. S. ALVAREZ *et al.*, 2007 Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. Science **316:** 1738-1743.

WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Pop. Biol. **7:** 256-276.

WHITTEN, M. J., 1985 The conceptual basis for genetic control,in Comprehensive Insect Physiology,Biochemistry and Pharmacolgy edited by C.A.Kerklot and L.I.Gilbert. 506-508.

WILSON, R., J. ORSETTI, A. D. KLOCKO, C. ALUVIHARE, E. PECKHAM *et al.*, 2003 Post-integration behavior of a *Mos1* gene vector in *Aedes aegypti*. Insect Biochem. Mol. Biol. **33:** 853-863.

WOOD, R. J., L. M. COOK, A. HAMILTON and A. WHITELAW, 1978 Transporting marker gene *Re* (Red Eye) into a laboratory cage population of *Aedes aegypti* (Diptera Culicidae), Using meiotic drive at Md Locus. Journal of Medical Entomology **14:** 461-464.

WRIGHT, S., 1951 The genetic structure of populations. Ann. Eugenics **15:** 323-354.

WRIGHT, S. I., Q. H. LE, D. J. SCHOEN and T. E. BUREAU, 2001 Population dynamics of an *Ac*-like transposable element in self- and cross-pollinating *Arabidopsis*. Genetics **158:** 1279-1288.

YAMASHITA, S., T. TAKANO-SHIMIZU, K. KITAMURA, T. MIKAMI and Y. KISHIMA, 1999 Resistance to gap repair of the transposon *Tam3* in *Antirrhinum majus*: a role of the end regions. Genetics **153:** 1899-1908.

YANG, Z., 1997 PAML: A program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:** 555-556.

YANG, Z., R. NIELSEN, N. GOLDMAN and A.-M. K. PEDERSEN, 2000 Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155:** 431-449.

YANT, S. R., L. MEUSE, W. CHIU, Z. IVICS, Z. IZSVAK *et al.*, 2000 Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. Nature Genetics **25:** 35-41.

ZHANG, P., and A. C. SPRADLING, 1994 Insertional mutagenesis of *Drosophila* heterochromatin with single *P*-elements. Proceedings of the National Academy of Sciences of the United States of America **91:** 3539-3543.

ZHOU, L., R. MITRA, P. W. ATKINSON, A. B. HICKMAN, F. DYDA *et al.*, 2004 Transposition of *hAT* elements links transposable elements and V(D)J recombination. Nature **432:** 995-1001.