DISSERTATION

POPULATION GENETICS AND VECTOR COMPETENCE OF *AEDES AEGYPTI* IN WEST AFRICA

Submitted by

Laura B. Dickson

Department of Microbiology, Immunology, and Pathology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2014

Doctoral Committee:

Advisor: William C. Black IV

Barry Miller
Brian Foy
Carol Blair
Kate Huyvaert

ABSTRACT

POPULATION GENETICS AND VECTOR COMPETENCE OF *AEDES AEGYPTI* IN WEST AFRICA

The mosquito, *Ae. aegypti* is the primary vector for all four serotypes of dengue (DENV 1-4) and yellow fever viruses worldwide. Dengue remains an important public health problem with an estimated 390 million cases per year and yellow fever outbreaks are still continuously reported even though a safe and effective vaccine exists. *Aedes aegypti (L)* is found globally in tropical and sub-tropical climates and exists as two subspecies: *Ae. aegypti aegypti (*Aaa*)* and *Ae. aegypti formosus*(Aaf) which have previously been defined by the presence or absence of scales on the first abdominal tergite. This classification system was developed in East Africa and is contradictory in West Africa where this mosquito is genetically diverse and exhibits great variation in susceptibility to DENV and YFV, which is a quantitative genetic trait. Understanding the population genetics, the vector competence, and the way in which genetic diversity contribute to vector competence of *Ae. aegypti* can improve our general understanding of mosquito/virus interactions and lead to potential ways to control these mosquitoes in nature.

In this dissertation, the distribution of F_{ST} values calculated from deep sequencing data between mosquitoes from two diverse locations in Senegal, one location in Mexico, and one location in Thailand across the entire genome were compared to identify the degree of genomic divergence as well as identify genes involved in speciation between the various populations. The distribution of F_{ST} values were also compared in different gene regions and mutation types to identify which parts of the genome provide the greatest resolution of subspecies population

structure. Genetic crossing experiments and deep sequencing of the sex determining locus of *Ae. aegypti* from Senegal and Thailand was used to demonstrate discrete genetic differences between the subspecies. The vector competence for a local sylvatic isolate of DENV-2 and two genetically diverse YFV isolates was compared between various collections of *Ae. aegypti* throughout Senegal to demonstrate that vector competence in these mosquitoes is dependent on the viral isolate. The genetic diversity of an important immune sensing gene, Dcr2 of the exo-siRNA pathway, was determined from mosquitoes with various vector competence phenotypes to test how the genetic diversity of this gene in individuals and populations of mosquitoes contribute to vector competence.

Overall, data in this dissertation suggest 1) allopatric speciation between *Ae. aegypti* from Senegal and Mexico or Thailand, and sympatric speciation within Senegal based on the distributions of F_{ST} values and variations in the sex determining locus between populations, 2) vector competence of *Ae. aegypti* from Senegal is dependent on the flavivirus species and viral genotype, and 3) increased genetic diversity of Dcr2 in individual *Ae. aegypti*, but not from populations of *Ae. aegypti*, from Senegal is correlated with the ability of *Ae. aegypti* to control DENV-2 infection. These results aid in our general understanding of the role of genomic divergence in speciation, as well as our understanding of genetics, vector competence, and the genetics of vector competence of West African *Ae. aegypti*, which could provide insight into the way we identify subspecies and make predictions about vector competence in this region.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisor, Dr. William Black IV, for providing guidance and support throughout the process of completing my PhD. Specifically, for the countless hours he tolerated me in his office teaching me statistics and population genetics, his amazing sense of humor, and his endless concern for the wellbeing of Chaz. The freedom granted by Bill to pursue my own interests, has helped shape me into an independent scientist. I would also like to thank the members of my committee, Drs. Brian Foy, Barry Miller, Carol Blair, and Kate Huyvaert. Dr. Brian Foy for his advice in writing this dissertation, as well as providing very useful career guidance. Dr. Barry Miller for allowing me to be trained in his lab on how to handle yellow fever virus and providing project advice and support. Dr. Kate Huyvaert for encouragement while writing the dissertation and very thoughtful and useful questions. Dr. Carol Blair for assistance in writing this dissertation and answering my many virology questions. I am very thankful and fortunate to have been able to train under such influential members of the vector biology field.

I would also like to thank my amazing family for the endless encouragement and love. The most gratitude goes to my husband, Nathan Goss, for being extremely supportive throughout completing my PhD. His encouragement to take breaks, belief in equal partnership, ability to tolerate my stress, and willingness and enthusiasm to follow me wherever my dreams take me have been instrumental in my ability to complete a PhD. I would not have achieved any of my success without the love and support of my parents, Cindy and Ron Dickson. My parents have provided me with every opportunity possible to succeed and have never stopped encouraging me to pursue my dreams with true passion, even if those dreams have taken me on some wild

adventures. My best friends and sisters, Karen and Ellie Dickson, have been my allies through life and have helped keep me sane and happy. Arguably most importantly, none of this would have been possible without the joy and balance that my two dogs, Chaz and Jake, bring me.

I would also like to thank everyone at AIDL and in the Black Lab for all the useful comments and suggestions along the way. Especially, Doug Brackney, Abhi Prasad, Ben Krajacich, Corey Campbell, Haoues Alout, and Nathan Grubaugh for the scientific discussions and countless hours of fun.

"If you have knowledge, let others light their candle in it" -Margaret Fuller

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Arboviruses

Infectious agents transmitted by arthropods that may cause disease (vector-borne diseases) are a persistent public health issue for humans. Arthropod-borne viruses, or arboviruses, are a group of viruses transmitted by arthropods between vertebrate hosts. Arboviruses are not a monophyletic group, instead there are members in seven known families of viruses. Under the Baltimore classification system [1], viruses are divided into seven groups based on the mechanism of mRNA transcription, which depends on the type of viral genome. The seven groups are I: double-stranded DNA, II: single-stranded DNA, III: double-stranded RNA, IV: positive-sense single-stranded RNA, V: negative-sense single-stranded RNA, VI: positive-sense single-stranded RNA that replicates through a DNA intermediate, VII: double-stranded DNA that replicates through a single-stranded RNA intermediate. Arboviruses in six major families (*Orthomyxoviridae*, *Rhabdoviridae*, *Bunyaviridae*, *Flaviviridae*, *Reoviridae*, and *Togaviridae*) have RNA genomes and African swine fever arboviruses (Family: *Asfarviridae*) have a double-stranded DNA genome. Within the family *Flaviviridae*, the genus *Flavivirus* contains many arboviruses and insect-specific viruses including medically important viruses such as yellow fever virus (YFV), dengue virus (DENV), West Nile virus (WNV), and tick-borne encephalitis virus (TBEV). This dissertation will focus on DENV and YFV.

DENV

Among the many medically important arboviruses that infect humans and cause disease are dengue virus (DENV) and yellow fever virus (YFV). DENV exists as four antigenically distinct serotypes (DENV-1 to DENV-4). Recently, evidence suggesting a putative 5th serotype

of DENV has been presented (N. Vasilakis, unpublished data). This remains contentious however, as a strict definition of how antigenically distinct two viruses must be to be considered different serotypes is lacking. Infection with DENV can cause dengue fever, which remains an important public health problem with an estimated 390 million infections per year globally [2]. Previously, dengue cases were defined as mild or asymptomatic (dengue fever, DF), moderately sever dengue hemorrhagic fever (DHF), and very severe dengue shock syndrome (DSS). In 2009, the World Health Organization (WHO) described new dengue case definitions. Severity is now classified as a) dengue without warning signs, b) dengue with warning signs, and c) severe dengue [3]. Patients with dengue without warning signs present with a fever and two of the following; nausea/vomiting, rash, aches/pains, leukopenia, or a positive tourniquet test. Patients with dengue with warning signs present with abdominal pain or tenderness, persistent vomiting, fluid accumulation, mucosal bleeding, lethargy, and liver enlargement. Patients with severe dengue present with fluid accumulation leading to shock (dengue shock syndrome, DSS), respiratory distress, or severe bleeding. Most DENV infections are subclinical or result in dengue without warning signs. According to the WHO, about 2.5% of people affected with severe dengue die. DENV is currently considered endemic in 11 African countries, and the WHO regularly reports new DENV outbreaks in previously unreported geographic areas [4]. Furthermore, it was recently estimated that in 2007 alone, 656 million fevers of all causes occurred in African children under the age of five [5]. Only 78 million of these cases were likely the result of *Plasmodium falciparum* infection, implicating other causative agents of febrile diseases. Despite the enormous number of acute non-malarial febrile illnesses in sub-Saharan Africa, their etiologies are poorly defined [5].

Antibody dependent enhancement

During primary DENV infection in humans, antibodies are produced that are specific to the serotype of the primary DENV infection. Upon a secondary DENV infection with a heterologous serotype, pre-existing plasma cells are triggered to rapidly produce antibodies; however, these are mostly directed against the initial DENV serotype. This process is referred to as "original antigenic sin" [6-8]. Most of the antibodies produced will bind to the heterologous serotype, but they are more likely to have non-neutralizing properties against the new serotype compared to the original serotype [9-11]. These cross-reacting antibodies have been found to enhance infection, a phenomenon known as antibody dependent enhancement (ADE) [7, 12-14]. Enhancement of infection is thought to be facilitated through efficient interaction of the virus-antibody complex with FcɣR receptors [15, 16]. FcɣR receptors are present on monocytes, macrophages, and dendritic cells, key cells involved in DENV replication. FcɣR receptors bind to antibodies and allow cells to phagocytize the pathogen [17], facilitating cell entry of DENV. The effects of ADE are cell type specific and are influenced by host genetics. Specifically, certain alleles in the interleukin-10 (IL-10) promoter region were associated with levels of IL-10, an important cytokine that controls the severity of ADE [18].

Yellow fever

YFV is another medically important arbovirus in Africa. Clinical symptoms of YF develop in only 15% of those infected with YFV and can be mild or severe. There are two main phases of the disease. In the first phase, the patient is viremic and symptoms include fever, headache, muscle pain, backache, weakness, red eyes, nausea, and vomiting. Symptoms in the second phase, or the toxic phase, include high fever, vomiting, epigastric pains, jaundice, hemorrhagic diathesis, coma, and death [19]. Mortality rates of those who reach the toxic phase

range from 20-50%, although the rates can be higher [20]. Surviving individuals are immune for life. Despite an effective vaccine, YF remains an important public health concern, with 200,000 cases of YF causing 30,000 deaths worldwide each year [21]. YFV is far more virulent than DENV in Africa; the case fatality rate for YF is 25%, whereas it is less than 10% for dengue [22].

Flavivirus **Genome**

Flaviviruses are enveloped viruses with a single-stranded positive-sense RNA genome that is approximately 11 kilobases in length. The flavivirus particle consists of a nucleocapsid protein and RNA surrounded by a host cell-derived lipid bilayer containing the E and prM/M proteins [23]. The flavivirus genome encodes three structural proteins at the 5' end and seven nonstructural proteins at the 3' end of the genome (Figure 1.1). While the exact function of each of the nonstructural proteins is not fully understood, NS3 and NS5 have been the most studied. The NS3 protein acts as a serine protease through the use of its cofactor NS2B protein [24, 25]. NS3 also contains an RNA helicase domain that is required for viral RNA synthesis [26]. The NS5 protein is a methyl transferase [27] and RNA-dependent RNA polymerase (RdRP) [28, 29]. Two characteristics of flavivirus RNA are that they have a single open reading frame and that proteolytic cleavage of the polyprotein, which is mediated by both virus-encoded and host proteases, occurs both co- and post-translationally.

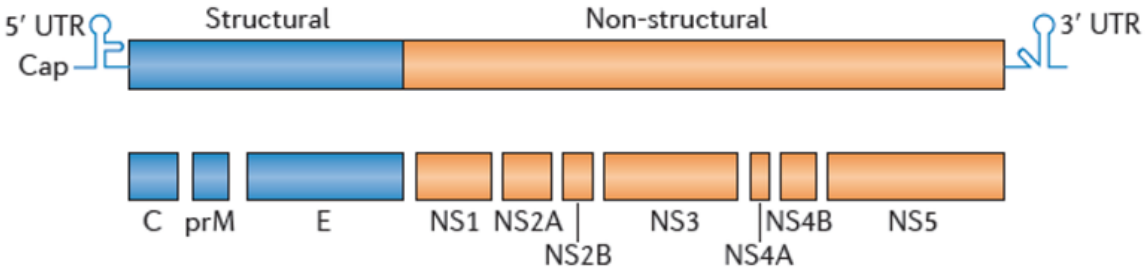


Figure 1.1: Flavivirus Genome
Beck, C., et al., 2013, *Flaviviruses in Europe: complex circulation patterns and their consequences for the diagnosis and control of West Nile disease* [30]. The flavivirus genome is approximately 11 kilobases and contains a single open reading frame flanked by non-coding regions. The 3'non-coding region consists of conserved structural elements that are involved in viral RNA replication, regulation of viral gene expression, and interactions with viral and cellular proteins. The genome encodes three structural proteins (capsid (C), membrane (M), and envelope (E) proteins), as well as seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5).

Flavivirus Cell Entry

Mammalian Cells

Upon infection through the bite of an infected mosquito, Langerhans cells, which are a type of resident dendritic cell in the skin, are among the first cells to be infected [31]. The virus is then transported to the draining lymph nodes [32] that contain other target cells such as resident dendritic cells [33-35], monocytes [13, 36], and macrophages [36-38]. Hepatocyes in the liver are also sites of infection [37, 39, 40].

A number of cell surface molecules have been proposed for flavivirus attachment to different types of cultured mammalian cells. These include heparan sulfate[39, 41-46] , heat shock proteins Hsp70 and Hsp90 [47-51], GRP78/BIP [52], CD14 [53], laminin receptor [54], dendritic cell (DC)-specific intercellular adhesion molecule 3 – grabbing nonintegrin (DC-SIGN) [35, 55-60], mannose receptor (CD206) [38], C-type lectin domain family 5, member A (CLEC5A/MDL-1) [61-63], αvβ3 integrins [64-66], and liver/lymph node-specific ICAM-3

grabbing nonintegrin (L-SIGN) [56], scavenger receptor class B type 1 [67], claudin-1 [68, 69], and natural killer cell activating receptor (NKp44) [70]. These factors are known to serve as attachment receptors for flavivirus infection, but none has been identified as an entry receptor for mammalian cells.

Mosquito Cells

The mosquito vector becomes infected by ingesting a bloodmeal from an infected vertebrate host. The initial site of infection in mosquitoes is cells of the midgut epithelium [71-73]. Studies have shown that a small number of cells of the midgut epithelium can be initially infected with WNV [74], DENV-2 [75] and Venezuelan equine encephalitis virus (VEEV) [76]. DENV-2 reaches its maximum titer in the midgut seven to ten days post-infection [75]. In a similar fashion to mammalian cell entry, flaviviruses are presumed to enter through receptor-mediated endocytosis. In *Culex quinquefasciatus* and *Aedes aegypti* mosquitoes, mosquito galactose specific C-type lectin-1 (mosGCTL-1) has been shown to facilitate the interaction of the E protein of WNV and the cellular surface [77]. Paralogs to the mosGCTL-1 protein have been shown to enhance DENV-2 infection of *Ae. aegypti* [78]. Other cell receptors which have been implicated in DENV entry into either cultured mosquito cells or whole mosquitoes include a laminin-binding protein [79], a heat-shock protein [80], prohibitin [81], a tubulin-like protein [82], carbohydrate moieties [83], as well as multiple others in either the midgut or salivary glands [84-90].

Flavivirus Replication

Flaviviruses enter host cells through receptor-mediated endocytosis [91] that is facilitated by the viral E glycoprotein. Clathrin-mediated endocytosis is utilized for flavivirus entry [92-95], after which the virion is delivered to early endosomes [92, 93]. The acidic pH inside the

endosome triggers an irreversible conformational change of the E protein from a homodimeric to a trimeric state resulting in exposure of the fusion loop. The fusion loop facilitates the fusion of the viral and cell membranes that results in the release of virus RNA into the cytosol [91, 96-101]. However, the final steps of fusion only occur in the late endosome which has the required repertoire of anionic lipids necessary for completion of fusion [92, 100].

Replication of the flavivirus genome occurs within virus-induced invaginations of the endoplasmic reticulum (ER) membrane into the lumen [102-105]. Similar to DENV replication in mammalian cells, DENV also replicates in intracellular membrane structures within the ER in the mosquito cell line, C6/36 [106]. During DENV replication, reorganization of ER membrane is induced by NS3 [107] and NS4A [108]. Viral replication begins with translation of the viral genome. Next, synthesis of full length negative-strand RNA occurs, which serves as a template for synthesis of the positive-strand RNA genome through the use of the newly translated RdRP encoded by NS5. During the process of viral RNA replication in membrane enclosed vesicles, dsRNA is generated, which serves as a molecular trigger for the innate immune response discussed in the "Mosquito Innate Immunity" section. Following viral RNA replication, immature flavivirus particles derive their lipid bilayer envelope by budding through the ER [109] and are then transported to the slightly acidic Golgi apparatus. The pH inside the Golgi apparatus triggers an irreversible conformational change of the E protein allowing for furin-mediated cleavage of the prM protein into the pr peptide and M protein that is necessary for the formation of mature infectious particles [110-112]. Mature virions exit the cell through exocytosis [23]. Flavivirus infected cells release a mixture of immature, mature, and partially-mature particles, indicating that the maturation process is not efficient [113]. Approximately 45% DENV particles are released from the mosquito cell line, C6/36 cells are immature [114].

DENV and YFV in West Africa

In Africa, DENV is maintained in two ecologically and evolutionarily distinct transmission cycles: an urban or human cycle and a sylvatic cycle [115] (Figure 1.2). Isolates of each of the four serotypes of DENV can be classified as endemic/epidemic (urban) or sylvatic and, in West Africa, a majority of sylvatic genotypes of DENV belong to the DENV-2 serotype [116]. Human illnesses and epidemics caused by DENV are generally a result of the urban cycle and endemic/epidemic viral isolates, though symptoms associated with sylvatic DENV and epidemic DENV are indistinguishable [117]. Infection with sylvatic DENV-2 has resulted in severe dengue in West Africa [118], and many cases of dengue assumed to be caused by endemic/epidemic isolates may in fact be sylvatic isolates. DENV is thought to have originated as a zoonotic virus in a sylvatic cycle, but sylvatic DENV-2 does not require adaptation to human hosts to replicate efficiently, suggesting the potential for re-emergence of sylvatic isolates into endemic cycles [119]. Transmission to humans of sylvatic DENV-2 in West Africa may be limited by the lack of competent *Ae. aegypti* near sylvatic foci, but the increasing urbanization of Africa could bring more efficient vectors to the sylvatic cycle.

Forest forms of *Aedes* mosquitoes (Figure 1.2) are responsible for spillover of sylvatic DENV-2 from non-human primates into the human population. Spillover of DENV-2 from the sylvatic cycle rarely results in outbreaks. In West Africa, the principal vectors in the sylvatic cycle appear to be *Ae. furcifer, Ae. taylori, and Ae. luteocephalus,* which overlap with the vector species of yellow fever virus, chikungunya virus, and Zika virus [116], although amplification of the viruses can occur in different physical locations [120, 121]. In West Africa, the non-human primate reservoirs are *Chlorocebus sabaeus, Papio papio, and Erythrocebus patas* [122]. In the

separate urban cycle, *Ae. aegypti* maintains epidemic genotypes of DENV at endemic levels and is responsible for epidemics.

Unlike DENV which is found in Africa, the Americas, and Asia, YFV is only found in Africa and the Americas despite a sufficient number of human hosts and mosquito vectors in Asia capable of infection and transmission. There are seven genotypes of YFV found worldwide, two of which (West African Genotypes I and II) are found in West Africa [123]. The differences in genome sequences among YFV isolates are specific to the geographic regions of Africa [124]. It is possible that different genotypes of YFV vary in virulence in humans as YF cases and epidemics are not uniform across the continent and correlate with YFV distribution. More outbreaks occur in West Africa compared to Central and East Africa [125] and West African genotype I is responsible for a majority of outbreaks and is genetically heterogeneous relative to other genotypes [123]. The difference in the number of cases and epidemics between West and East Africa could be the result of differences in the distribution of the mosquito vector. In West Africa, *Ae. aegypti* is responsible for urban YF outbreaks, but in East Africa, *Ae. africanus, Ae. keniensis,* and *Ae. simpsoni* are the main vectors [124, 126].

YFV has three transmission cycles in Africa: the urban cycle, the intermediate cycle, and the sylvatic cycle [127] (Figure 1.2). The urban cycle is maintained by humans and *Ae. aegypti*, and the sylvatic cycle is maintained by nonhuman primates and *Ae. africanus* [127]. The intermediate cycle bridges the urban cycle and the sylvatic cycle and is maintained by humans and numerous *Aedes* spp, such as *Ae. africanus, Ae. opok, Ae. metallicus, Ae. bromeliae, Ae. vittatus, Ae. luteocephalus, Ae. taylori, Ae. furcifer*, and members of the *Ae. simpsoni* complex [122, 128]. In Senegal, the only species of *Aedes* infected with YFV within human villages was *Ae. furcifer*, implicating it as the main mediator of human infection [129]. In this same study,

villages that contained YFV infected mosquitoes were significantly closer to large forests than villages with no infected mosquitoes [129]. In contrast to the distinct genotypes of DENV in the urban and sylvatic transmission cycles, the same genotypes of YFV that spillover from the sylvatic cycle move into human populations through the intermediate cycle known as the zone of emergence [130]. The non-human primates that act as reservoirs for YFV in Africa are *Alouatta* spp, *Colobus* spp, *Ceropitecus* spp, *and G. senegalensis* [122].

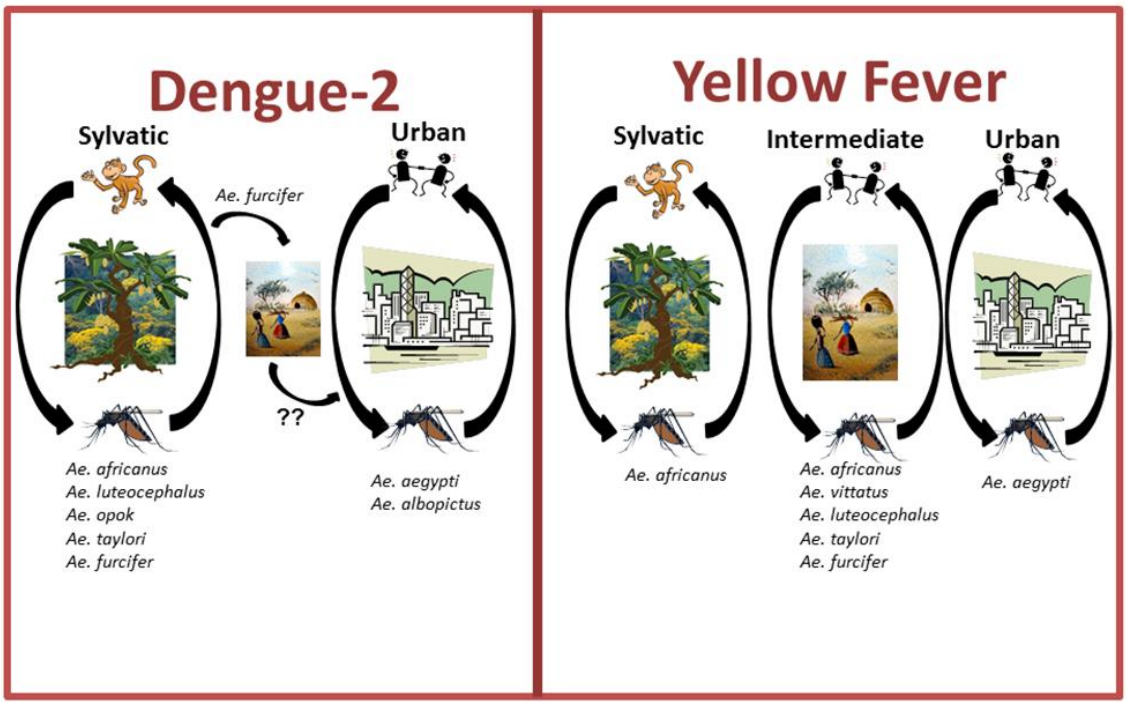


Figure 1.2: DENV and YFV transmission cycles in West Africa
DENV and YFV have different transmission cycles in West Africa. DENV circulates between non-human primates and forest forms of *Aedes* in the sylvatic cycle. Spillover from the sylvatic cycle occasionally results in human infection. DENV is endemic in urban environments, and epidemics occur through transmission of the virus between humans and *Ae. aegypti* or *Ae. albopictus*. There is potential for sylvatic isolates of DENV to emerge from the sylvatic cycle and be transmitted by *Aedes* mosquitoes [131]. YFV also has a sylvatic cycle and an urban cycle, but there is also an intermediate cycle that bridges the sylvatic and urban cycles. Epidemics of YF occur form spillover out of the sylvatic cycle into the urban cycle.

Aedes aegypti Morphology

Ae. aegypti is morphologically, ecologically, and genetically diverse [132-139]. The ecology and population biology of this species have been studied since the 1950s. Observations of *Ae. aegypti* in East Africa in the 1950s reported a higher frequency of paler forms near human dwellings and that darker forms predominate in the nearby bush [140, 141]. The correlation

between body color and behavior prompted Mattingly to revisit the biology and taxonomy of *Ae. aegypti* [138, 139] where he recognized a new subspecies, *Ae. aegypti formosus* (Walker). This form is restricted to sub-Saharan Africa, and is distinguished by the complete absence of pale scales on the first abdominal tergite. Mattingly also observed that the *formosus* subspecies most frequently breeds in natural containers such as tree holes and feeds on wild animals [138, 139]. In contrast, the type form is lighter in color, has pale scales on the first abdominal tergite, breeds in artificial containers, and has a feeding preference for humans. It was presumed that all *Ae. aegypti* from West Africa were *Ae. aegypti formosus* until McClelland performed a comprehensive study of the differences in scaling pattern in 69 different worldwide collections in 1974 [137]. He concluded that distinctions between the subspecies based on body color and behavior were not definitive in Africa, the only region in the world where both forms are found. In East Africa it was observed that in collections where the subspecies were mixed, there was a wide range of scaling. Later studies in West Africa found forms with no scales on the first abdominal tergite breeding domestically indoors in Senegal, Nigeria, and Cameroon in contradiction to Mattingly's early observations [142-145]. While the presence or absence of scales correlates with genetic differences in East Africa [135, 146], this is not the case in West Africa [142, 144, 145, 147] and will be discussed in the "Population Genetics in West Africa" section of the introduction.

***Aedes aegypti* Genetics**

Aedes aegypti is a diploid organism with three pairs of chromosomes. Early studies of *Ae. aegypti* genetics involved the construction of linkage maps based on meiotic recombination among isozymes and morphological mutant markers [148, 149]. Linkage maps are genetic maps that show the position of known genes or genetic markers relative to each other in terms of

recombination frequency of the markers instead of physical distances between the markers. Linkage maps can help identify the position of new markers and identify the location of genes associated with heritable disease. The first detailed DNA-marker based genetic map in *Ae. aegypti* was constructed using markers called restriction fragment length polymorphisms (RFLP) [150]. The RFLP markers for genes with identified function were derived from random cDNA clones, cDNA clones of known genes, and morphological mutant markers. Other early genetic mapping strategies used single strand conformation polymorphism (SSCP) analysis to identify single nucleotide polymorphisms (SNPs) in cDNA sequences [151, 152]. Severson et al.[153] used both RFLP markers and SSCP analysis in one mapping strategy and presented the expressed sequence tags (EST) for the cDNA markers. Amplified fragment length polymorphisms (AFLP) [154] and microsatellite strategies [155] have also been used. The linkage maps of *Ae. aegypti* have been used in quantitative trait loci (QTL) mapping studies to identify regions of the chromosomes involved in susceptibility to *Plasmodium gallinaceum[156]*, *Brugia malayi [157]*, and DENV-2[158-161].

In contrast to linkage maps where the position of a gene is described relative to genes around it, physical mapping determines the precise location of the gene on the chromosome, which can be useful for studying chromosomal rearrangements. Physical mapping with *in situ* hybridization has placed cDNA sequences on physical landmarks, such as regions of differential staining, on chromosomes. Physical mapping is achieved through fluorescent *in situ* hybridization (FISH) of cDNA or bacterial artificial chromosome (BAC) clones on *Ae. aegypti* chromosomes. The original linkage maps were confirmed by physical mapping with FISH [162]. More recently, 45% of the *Ae. aegypti* BAC clones were mapped to the genome using FISH [163]. The use of deep sequencing and restricted-site associated DNA (RAD) sequencing

revealed 14% of the markers had been previously misassembled and allowed for mapping of 60% of genome sequences [164].

While not fully assembled, publication of the whole genome sequence of *Ae. aegypti* has significantly advanced the field of mosquito genetics [165]. The genome sequence of *Ae. aegypti* has revealed significant differences when compared with other mosquito vectors like *An. gambiae* and *Cx. quinquefasciatus.* For example, the *Ae. aegypti* genome (1376 million basepairs) is about 5 times larger than *An. gambiae* and 2 times larger than *Cx. quinquefasciatus.* These large differences in genome size are attributed to the increased content of transposable elements in the *Ae. aegypti* genome [165]. The number of genes and average exon length are similar among the three mosquitoes, but the average intron length is much greater in *Ae. aegypti* [166].

Aedes aegypti population genetics

Early population genetic studies of *Ae. aegypti* by Tabachnick, Powell, Munstermann and Wallis demonstrated that collections fell into two clades [133-135, 167-173]. One clade contained *Ae. aegypti aegypti* from East Africa, South America, and the Caribbean suggesting that the New World populations originated from East Africa. The second clade contained *Ae. aegypti aegypti* populations from Asia and Southern USA with a basal *Ae. aegypti formosus* from both East and West Africa [135]. This finding suggested two independent introductions into the New World, one from East Africa and one from West Africa. The mitochondrial NADH subunit-4 (ND4) gene has also been used to examine gene flow between and among collections of *Ae. aegypti* outside Africa [174-176]. The addition of ND4 sequences from 10 collections in Senegal revealed that populations of *Ae. aegypti* outside Africa arose from two different clades, a basal West African clade and a second East African clade that arises from the first [177]. Further

studies have confirmed that *Ae. aegypti* originated from Africa where a domestic form arose through a single subspeciation event and spread throughout the rest of the subtropical world through human movement and trade [178].

A recent study examined 12 microsatellite loci in 24 worldwide collections of *Ae. aegypti* and demonstrated two distinct clusters: domestic populations outside Africa and both domestic and forest populations within Africa [146]. Specifically, *Ae. aegypti formosus* from Kenya and mosquitoes from Senegal, regardless of urban or sylvatic origin, were predicted to be a single population separate from *Ae. aegypti aegypti* from Kenya and other *Ae. aegypti aegypti* worldwide. In the same study, when the population structure was predicted within Africa, *Ae. aegypti* from Senegal appeared to be different from the homogeneous *Ae. aegypti formosus* from Kenya and Uganda, suggesting that *Ae. aegypti formosus* is not monophyletic in Africa [146].

Population genetics in West Africa

The correlation of genetic data and the presence of white scales on the first abdominal tergite in East Africa [135, 146, 179] has prompted multiple similar studies in West Africa. Comparisons of domestic and sylvatic populations in Cameroon revealed great genetic diversity in the sylvatic collection and reduced genetic diversity in the domestic collection [142] . A similar increase in genetic diversity in sylvatic collections compared to domestic collections was also observed in Senegal [144]. Huber et al. [144] analyzed the genetic diversity of four domestic and one sylvan collection in Senegal using allozymes. A lower level of genetic differentiation and genetic diversity was observed in the domestic collections compared to the sylvatic collection. Importantly, this study demonstrated genetic diversity between domestic and sylvan collections. Sylla et al. [145] performed a similar study but also included the McClelland scaling pattern [137] on the first abdominal tergite in the analysis. In this study, 19

collections from across Senegal that varied in subspecies composition, geographic location, distribution of vegetation types, and habitat were analyzed using 11 nuclear SNP loci. Regardless of collection site, the variance in allele frequencies was not associated with the presence of scales on the first abdominal tergite. In both studies, the variation arose from comparisons among collections within a subspecies rather than between subspecies, indicating the geographic or ecological habitat had more influence on genetic differentiation than did the scaling pattern. Paupy et al. [147] confirmed these results again by examining mosquitoes from one domestic collection site in western Senegal where the mosquitoes were scored for presence of white scales using the McClelland scale. Using microsatellite analysis, it was confirmed that there is no genetic distinctness based on the scaling pattern on the first abdominal tergite. These studies, along with Brown et al. [146], demonstrate that *Ae. aegypti* in Senegal may not be monophyletic, and the greatest amount of genetic diversity arises between geographic and ecological habitats. While multiple population genetics studies have been performed on closely and distantly related populations of *Ae. aegypti*, studies on gene flow and genomic differentiation between populations as it relates to speciation are lacking.

Speciation

The study of speciation can be divided into two branches, forces that caused the speciation event and forces that maintain reduced gene flow [180]. It is difficult to determine the forces that caused speciation due to the antiquity of the event. This has resulted in the current focus on the factors that are maintaining reduced gene flow. Studying the forces maintaining gene flow both aids in the understanding of how species are maintained in nature and may possibly give insights into the origin of the speciation event.

Biological species concepts define species based upon reproductive isolation [181]. Reproductive isolation can occur through pleiotrophy, whereby genes that evolve within a group have the side effect of creating isolating barriers between groups; or epistatis, whereby evolution of genes in one group creates isolating barriers by interacting with evolving genes in the other group. Reproductive isolating barriers can be prezygotic (e.g., behavioral) or postzygotic (e.g., hybrid sterility and inviability). Some consider pre-zygotic isolation the more important barrier to gene flow [182, 183] , while others argue that post-zygotic isolation is the more significant barrier [180]. Wright suggested that genetic drift is important to help incipient species cross an "adaptive valley" [184]. Coyne and Orr argue that speciation can occur without any fitness loss or genetic drift [180].

Before the era of modern genetics, Darwin first proposed that new species could arise without geographic isolation [185], resulting in years of debate and the idea of "divergence with gene flow" [186, 187]. It is now accepted that speciation can occur through allopatric speciation (geographic isolation and no gene exchange between populations), parapatric speciation (populations can exchange genes to a limited extent), or sympatric speciation (populations can exchange genes freely). Allopatric and parapatric speciation are the obvious and most accepted ways in which speciation can occur. Physical distance between the two populations can lead to genetic drift and is likely to result in reduced gene flow. The idea that sympatric speciation, or divergence with gene flow, might be an important cause of speciation is relatively recent. There are two different models of sympatric speciation: the continuous distribution of resources model and the discrete-habitat model [180]. In the continuous distribution of resources model, co-evolution of ecological traits and assortative mating divides a population into two sympatric groups that use different parts of the resource distribution [180]. In the discrete-habitat model,

sympatric species seek out different niches and preferably mate with their own species. This requires the evolution of three behaviors: niche preference, niche adaptation, and assortative mating.

There are limited examples of divergence with gene flow in natural populations, leaving few opportunities to study this process. An alternate path of research has been to study taxa that exist in partial reproductive isolation to elucidate the mechanism and maintenance of divergence [186]. During the process of reproductive isolation, diverging taxa should have relatively homogeneous genomes except in regions involved in ecological and reproductive isolation [187, 188]. The process of incipient speciation of *Anopheles gambiae* in West Africa is a good example of divergence with gene flow. *Anopheles gambiae* sensu stricto consists of two morphologically indistinguishable molecular forms called the M (Mopti) and S (Savannah) forms. The two forms are differentiated based on fixed differences at an X-linked ribosomal DNA marker [189, 190]. The M and S forms can be sympatric, but very few M-S hybrids have been detected [191-193]. Recently, the M and S forms have been proposed to be separate species within the *An. gambiae* species complex and are now called *An. coluzzii* (M form) and *An. gambiae* (S form) [194]. Spatial segregation of mating swarms [195] and assortative mating [193] appear to reduce hybridization of the two forms. Also, niche differentiation has been observed, suggesting that ecologically based divergent selection may be driving evolution of the two forms [192, 196, 197]. Recently, it has been shown that there is asymmetric introgression of the two forms and the extent of reproductive isolation varies across West Africa [198].

Vector Competence

Vector competence is the intrinsic ability of an arthropod vector for infection, replication, and transmission of a pathogen. In the case of viral infection, there are four main barriers to

transmission by the arthropod vector (Figure 1.3). First, the virus must infect and replicate within the cells of the midgut epithelium. Failure of the virus to do this is called a midgut infection barrier (MIB). Second, the virus must infect and replicate in tissues outside the midgut, particularly in hemocytes and the fat body, resulting in a disseminated infection. Failure of the virus to do this results in a midgut escape barrier (MEB). Finally, the virus must infect and replicate within the salivary glands where it is shed into the saliva and injected into the next vertebrate host upon which the arthropod feeds. Failure to infect the salivary glands constitutes salivary gland infection barrier (SIB), and failure of the virus to escape into the saliva represents a salivary gland escape barrier (SEB). The time from ingestion of the virus to the time when the mosquito is able to transmit the virus is called the extrinsic incubation period (EIP). Anatomical barriers to infection, such as the midgut and salivary glands, are important because they function as population bottlenecks for viruses and have implications for emergence of viral genotypes with epidemic potential [199]. These infection barriers, along with multiple arthropod, virus, and environmental factors, influence the vector competence of the arthropod.

The study of vector competence depends on which infection barrier phenotype is being measured. For example, studies may either focus on the arthropod's susceptibility to infection by measuring the MIB, or focus on rates of dissemination and transmission by measuring the MEB, SIB, and SEB. Great phenotypic variation has been found in susceptibility to infection between populations of mosquitoes within species in every mosquito-virus system examined [169, 200-217].

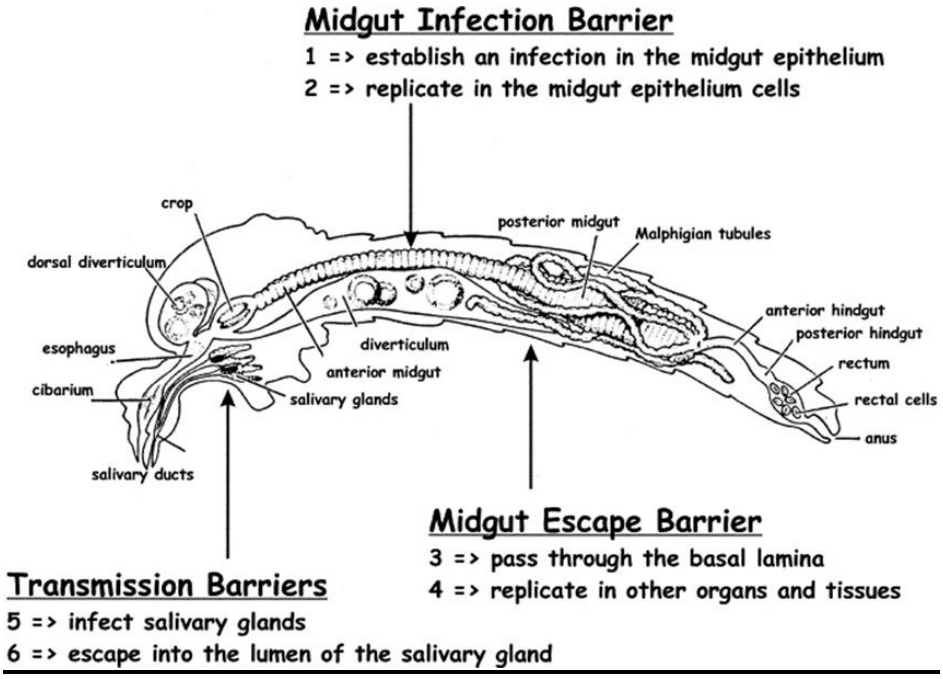


Figure 1.3: Barrier to arbovirus infection
Cross section of a female mosquito showing organs involved in arbovirus infection. The numbered steps underneath various barriers to infection represent the steps for viral infection, replication, and dissemination. *Black, W.C., et al., 2002. Flavivirus susceptibility in Aedes aegypti. [218]*

Genetics of Vector Competence

Differences in vector competence are detected in the laboratory under uniform environmental conditions, which highlights the mosquito's genetic contribution to vector competence. Genetic variation in vector competence within a mosquito species is widespread and has been broadly documented. Variation in vector competence has been examined in *Ae. aegypti* [144-146, 174, 176, 177, 211, 219-223], *Ae. albopictus* [224-227], *Cx. tarsalis* [228-237], and *Cx. pipiens* [238-243] mosquitoes. Significant progress has been made in finding ways to identify specific genes that contribute to vector competence [166, 244]. The ability to select for resistant/susceptible individuals provided the first evidence of a genetic component to vector competence. Subsequent quantitative genetics studies confirmed that vector competence phenotypes can be inherited [170, 245-248]. Susceptibility of *Ae. aegypti* to DENV is a

quantitative genetic trait [249]. A single genetic locus was proposed in the biting midge, *Culicoides sonorensis*, for its susceptibility to bluetongue virus [250]. In contrast, quantitative inheritance approaches using *Ae. aegypti* and DENV have identified several genetic loci involved in susceptibility [218]. These quantitative trait loci (QTL) studies have identified multiple regions on all three chromosomes that contribute to variation in MIB and MEB [158-161]. The identified QTL vary based on the geographic origin of the mosquitoes, indicating potentially different mechanisms of vector competence within a single species.

Many genomic studies have been performed to identify candidate genes involved in vector competence. These include proteomic and transcriptomic approaches in both *Aedes* and *Culex* [251-254]. A genome-wide transcriptome profile was generated in *Ae. aegypti* strains that differed in susceptibility to DENV [252]. In this study [252], many of the same candidate genes implicated in previous studies were identified that differed in expression levels between the susceptibility phenotypes in response to infection with DENV. These genes included important genes in the innate immune response and midgut enzymes.

Recently, the interaction of the mosquito genome with the virus genome has been shown to be an important determinant of vector competence and is changing the way mosquito genetics are studied in the context of viral infection [255]. Vector competence of *Ae. aegypti* for DENV is governed by interactions between mosquito genotype and viral genotype in natural collections [256], causing the genetics of vector competence field to move towards specific mosquito/virus systems. The importance of viral genetics in vector competence is discussed in the "Viral Determinants of Vector Competence" section below.

Viral Determinants of Vector Competence

It is important to note that, while mosquito genetics, or factors directly related to the mosquito, are important determinants of vector competence, viral factors play an important role as well. Genetic differences between viruses [257-272], the dose of virus that the mosquito feeds upon [265, 273-276], and the way in which the virus is prepared [258] can influence vector competence.

Emergence or re-emergence of an arbovirus into new amplification or human hosts can occur through mutations in the viral genome that confer greater fitness. It is thought that adaptive evolution of arboviruses is constrained by their requirement to replicate in divergent hosts like mosquitoes and humans [277]. Despite these theoretical constraints, simple point mutations have resulted in increased infection and replication in the mosquito vector. A single point mutation in Venezuelan equine encephalitis virus (VEEV) resulted in increased viremia in the equine population and increased infection of the mosquito vector [278]. Mutations in the Chikungunya virus (CHIKV) genome resulted in adaptation from *Ae. aegypti* to the more invasive *Ae. albopictus* [279, 280], resulting in major global expansion and a major epidemic [281]. The mutation in the E protein of CHIKV allowed for increased midgut infectivity, dissemination, and transmission in *Ae. albopictus*, but had no effect in *Ae. aegypti*. [279, 282]. Since the introduction of WNV (NY99) into New York City in 1999, a new genotype of WNV (WN02) has displaced the original NY99 genotype. WN02 differs from NY99 by a single point mutation in the E protein [283]. The WN02 genotype is thought to have shorter extrinsic incubation period (EIP) in *Cx. pipiens* and *Cx. tarsalis* resulting in more rapid transmission [267, 284], but this was not replicated in a separate study [285] leaving the impact of the WN02 genotype on EIP inconclusive. Although the origin of DENV is debated [116], DENV is thought

to have originated as an enzootic virus in Asia that was transmitted between non-human primates and arboreal *Aedes* spp. Emergence of DENV into the human population was most likely facilitated by mosquito host switching from arboreal *Aedes* spp. to urban *Aedes* spp. The genetic mechanisms explaining DENV emergence have not been defined.

Environmental Determinants of Vector Competence

Environmental factors, such as temperature, can influence viral replication in the mosquito leading to differences in the EIP [237, 286-300]. Temperature, nutrition, and competition during the larval stage can influence vector competence in adult females [203, 231, 298, 301-306]. Colonization, or multiple breeding generations in the lab, can also cause variation in the vector competence phenotype [173]. *Aedes aegypti* immunity to DENV can be altered by infection with the bacteria *Wolbachia* [307, 308], which is the most common endosymbiont of insects [309]. *Aedes aegypti* that were stably transfected with *Wolbachia* showed reduced infection and transmission of DENV [310-312]. *Wolbachia* are able to propagate through a population through cytoplasmic incompatibility [313, 314]. When infected males mate with uninfected females, the offspring are inviable. However the offspring are viable if both the male and female are infected. Therefore infected females are selected for in the population, which, as mentioned above, show reduced infection and transmission. A reduction in DENV replication has been shown in field collected *Ae. aegypti* infected with *Wolbachia* [315].

Vector Competence in West Africa

Previous studies on the vector competence of West African *Ae. aegypti* for YFV [245, 316] suggested that West African *Ae. aegypti* are more refractory to YFV infection than *Ae. aegypti* from the Americas and Asia. Tabachnick et al. [316] showed that two *Ae. aegypti*

collections from western Senegal, when infected with the Asibi isolate of YFV, were more refractory than collections from the Americas or Asia. Although the Asibi isolate is from Ghana, it had been passaged many times and may not have been representative of isolates involved in natural transmission cycles. Similar to the western Senegal collections, collections close to Dakar were more refractory than other collections throughout the country. Miller and Mitchell [245] showed that an *Ae. aegypti* collection from Nigeria was much more refractory (10% DI) when compared with collections from the Americas (90% DI) to a YFV isolate from Peru and was completely refractory (0% DI) with YFV BA-55, the same isolate to be used in this dissertation. The discrepancy in vector competence between the two populations with the same viral isolate demonstrates the specificity of the virus/mosquito interaction for vector competence.

A number of studies of *Ae. aegypti* from Senegal have examined vector competence for DENV-2 [131, 145, 216, 249, 316] When compared with *Ae. aegypti* from the Americas or Asia, West African *Ae. aegypti* were generally less susceptible to DENV-2 [249, 316]. In studies directly comparing collections within Senegal [131, 145, 216], there was variation in susceptibility, but the sylvatic *Ae. aegypti* were more refractory than domestic *Ae. aegypti*. However, the study by Sylla et al.[145] only examined the highly passaged DENV-2-Jam1409 isolate. Measuring vector competence in *Ae. aegypti* with a viral isolate collected in proximity may be the most informative approach to predicting the vector competence of that mosquito population [256]. Diallo et al. [131, 216] examined the vector competence of *Ae. aegypti* from Senegal with multiple local isolates of DENV-2. Diallo et al. [131] reported that sylvatic *Ae. aegypti* collections have lower infection rates than other sylvatic species of *Aedes*, but some sylvatic *Ae. aegypti* mosquitos developed a DI. Diallo et al. [216] reported low levels of midgut infection (0.0 - 26.3 %) and variable disseminated infection (0 – 100 %) in six collections from

Senegal regardless of geographic location. Importantly, both studies demonstrated variability in infection rates based on the isolate of DENV-2 and the collection site. Although some groups of *Ae. aegypti* are more refractory, they can still contribute to epidemics. This was the case with a large yellow fever outbreak in Nigeria where an incompetent vector was implicated in maintaining transmission [317].

Mosquito Innate Immunity

The mosquito innate immune pathways are an important component of *Ae. aegypti* vector competence to arboviruses. The innate immune response of vectors in response to arbovirus infection is studied because of the possible benefits of an increased understanding of mosquito-virus interactions and ways to manipulate the immune response in the vector as a way to reduce the infection in human hosts. *Drosophila melanogaster* is the primary model organism to study invertebrate innate immunity. In contrast to vertebrate innate immunity, invertebrates do not have the interferon (α/β or type I) pathway, which is the major pathway that responds to viral infection in vertebrates. Instead, the primary pathway in invertebrates is the RNA interference (RNAi) response [318] although other immune pathways (discussed below in the "Transcriptionally activated pathways" section) can be activated following viral infection. Also, unlike vertebrates, the innate immune response to viral infection in invertebrates does not lead to a protein-based adaptive immune response although there is growing evidence of immune memory in *Anopheles* mosquitoes infected with *Plasmodium* [319-321]. Although the two systems rely on different pathways, both invertebrates and vertebrates utilize recognition of pathogen-associated molecular patterns (PAMPs), in particular double-stranded RNA (dsRNA), by pathogen-recognition receptors (PRRs) in the host cell.

RNAi was first observed in plants as a defense mechanism against virus infection [322, 323] in the 1990s, but dsRNA acting as the molecular trigger for gene silencing was first discovered in 1998 in *Caenorhabditis elegans* [324] and *Drosophila melanogaster* [325]. During replication of viral RNA genomes, dsRNA intermediates accumulate in the cytoplasm and are though to serve as PAMPs for the RNAi response. In the context of viral infection, the exogenous small-interfering RNA (exo-siRNA) pathway is initiated by the recognition of cytoplasmic dsRNA generated during viral replication by the RNaseIII enzyme Dicer-2 (Dcr2). Flaviviruses such as DENV have positive-sense single-stranded RNA (+ssRNA) genomes. During replication of positive sense viral genomes in membrane bound vesicles, dsRNA replication intermediates accumulate in cytoplasmic vesicles [326], and serve as a PAMP that is recognized by Dcr2. Dcr2 then cleaves the long dsRNA into siRNA duplexes which are usually 21 bp in length and have 2 nucleotide 3' overhangs. In association with Dcr2 and another dsRNA binding protein, R2D2, siRNA duplexes are loaded into the RNA-induced silencing complex (RISC), that contains Argonaute-2 (Ago2) [327-329], the major catalytic component of the RISC [329, 330]. After unwinding of the siRNA duplex occurs within the RISC, one of the siRNA strands, known as the passenger strand, is degraded, while the other strand, known as the guide strand, is retained and leads the RISC to viral mRNA via sequence complementarity. This results in cleavage of the target transcript via the slicer endonuclease activity of Ago2 [329]. Thus, the exo-siRNA response inhibits viral replication by degrading viral mRNA. See the "RNAi response to viral infection" section below for a discussion of the RNAi response to viral infection in *Ae. aegypti*.

Transcriptionally activated pathways

Orthologs for genes in three other evolutionarily conserved *Drosophila* innate immune pathways (Toll, Imd, and JAK-STAT) have been identified in mosquito genomes [331]. Each of these three pathways works through the activation of a key transcription factor. One of the first descriptions of these pathways in mosquitoes following arbovirus infection was done with *Ae. aegypti* and Sindbis virus [332]. Expression of immune-related genes has been characterized in *Ae. aegypti* in response to DENV infection [333-335]. Through microarray analysis, these studies demonstrated transcriptional activation of genes linked to the Toll pathway and JAK-STAT pathway as well as two to three fold increases in viral load [334] when key genes in these pathways were silenced by RNAi. It was also observed that suppression of the JAK-STAT pathway led to increased susceptibility of *Ae. aegypti* to DENV [335]. In these studies, neither the PAMP nor the PRR were identified and impairment of the pathways led to only two or three fold change in the amount of DENV replication. In contrast, the PAMP and PRR are known in the RNAi pathway and impairment of the RNAi pathways results in more than tenfold changes to DENV replication, suggesting that RNAi is the main innate immune pathway controlling arbovirus infection [318].

RNAi Response to Viral Infection

Although replication of the DENV and YFV genomes occurs in ER-derived vesicles, replication intermediates (dsRNA) can still accumulate in the cytoplasm during infection [326]. During DENV infection, dsRNA is thought to serve as the PAMP to induce the RNAi response [336, 337], but both dsRNA and structured ssRNA may serve as PAMPs during WNV infection [338, 339]. It is likely that ssRNA may also serve as PAMPs during DENV infection similar to WNV infection, but this has yet to be shown. The role of the RNAi response against viral

infection was characterized in *Drosophila* [340-344]. Early studies also included expression of DENV RNA by a Sindbis virus transducing system in *Ae. aegypti* to generate antisense RNA targets to the premembrane coding region of DENV-2 that resulted in reduced DENV-2 replication in the mosquito [345]. The RNAi response is also important for controlling alphavirus and flavivirus replication in *Ae. albopictus* and *Ae. aegypti* cell cultures and *An. gambiae* and *Ae. aegypti* mosquitoes [345-352]. The importance of the RNAi pathway was confirmed by silencing Dcr2, Ago2, and R2D2 [346, 347, 350, 353] and observing increased viral load and a shortened extrinsic incubation period.

Dcr2 Processing of dsRNA

Dcr2 is responsible for recognition and processing of dsRNA into siRNAs [354, 355]. Dcr2 is a large protein that includes a putative dsRNA binding domain, two helicase domains (DExD/H-box and HELICc), a PAZ domain, DUF283 (an alternative dsRNA-binding fold), and two catalytic RNase III domains[356]. These diverse domains allow Dcr2 to function in recognition, cleavage, and processing of dsRNA. In *Drosophila*, the helicase domain is required for biogenesis of siRNAs from long dsRNA *in vitro* and *in vivo* [354, 357]. Interestingly, dsRNA with a 5' overhang or blunt termini require the helicase domain for efficient processing while dsRNA with a 3' overhang do not [358]. The DExD/H-box helicase domain of Dcr2 is homologous to the DExD/H-box helicase domain in the vertebrate RIG-I/MDA5 proteins, which are PRRs. In vertebrates the PRR, RIG-I, recognizes viral dsRNA PAMPs and triggers the innate immune response. Crystal structures of RIG-I-like DExD/H helicase domains have shown the helicase domain nonspecifically recognizes the phosphodiester backbone and 2'hydroxyl groups of dsRNA [359-363]. It remains unknown if the DExD/H helicase domain of Dcr2 also

recognizes the phosphodiester backbone and 2' hydroxyl groups of flavivirus derived dsRNA in mosquitoes.

Targeting of viral dsRNA by Dcr2 can degrade viral RNA by the direct cleavage of viral long dsRNA into siRNAs, targeting of viral mRNA by the RISC complex, and activation of other factors involved in controlling viral infection. In *Drosophila*, it was suggested that Dcr2 induces the gene *vago* and controls viral replication independently of AGO2 and R2D2 [364]. It has been proposed that in *Culex* mosquitoes, WNV infection is restricted by the Dcr2-dependent activation of *Culex* Vago (*Cx*Vago), which activates the transcription of the *Culex* ortholog of the virus-inducible *Drosophil*a anti-viral gene, *vir-1,* through the JAK-STAT pathway [365]. More recently, it was suggested that Rel2 plays an important role in activating *Cx*Vago through the TRAF-Rel2 signaling pathway in a Dcr2-dependent manner [366].

RNAi Pathway/Viral Co-evolution

Flaviviruses, like other RNA viruses, exist as a quasispecies, a group of genetically diverse virus genomes that are descendants from a common ancestral genotype within a single host or host cell. Quasispecies can be attributed to the virus-encoded error prone RNA polymerase used in viral replication [367]. Along with other RNA viruses, the population dynamics and genetic diversity of quasispecies have been studied in mostly WNV. WNV genomic RNA is more diverse in *Cx. quinquefasciatus* than in the bird host, and the genetic diversity is maintained through the extrinsic incubation period [338]. In contrast, in *Cx. pipiens*, WNV diversity decreased during midgut infection and through the EIP [368]. In fact, RNAi targeting of the West Nile virus genome in *Cx. quinquefasciatus* midguts is associated with increased viral RNA genetic diversity in "hot spots" of genome targeting [369]. WNV populations that were more genetically diverse were more fit and more likely to infect and

replicate in mosquitoes [370]. Bernhardt et al. [371] reported that refractoriness to DENV-2 infection was correlated with genetic diversity of Dcr2 in natural populations of *Ae. aegypti*, suggesting genetic diversity of both the virus and genes in the mosquito RNAi pathway might play a role in vector competence.

Many insect pathogenic viruses encode suppressors of the RNAi response [372], resulting in an "arms race" where the virus and host exert reciprocal selective pressure on each other and drive co-evolution of both the virus-encoded protein suppressor of RNAi (VSR) and the host genome encoding RNAi pathway genes. Obbard et al. [373] demonstrated a higher rate of diversifying selection in the siRNA pathway genes compared to the microRNA (miRNA) pathway genes in *Drosophila*. In contrast to *Drosophila*, both the siRNA and miRNA pathways are subject to diversifying selection in *Ae. aegypti* [371]. This study suggests that something other than arbovirus infection is driving the evolution of the siRNA and miRNA pathway genes in *Ae. aegypti*. Although arboviruses can reduce the survival of the mosquito vector [374], few mosquitoes in endemic areas are infected at a given time [375-377], making it unlikely that arboviruses are driving the evolution of siRNA pathway genes in mosquitoes.

Most insect pathogenic viruses and plant viruses express a protein VSR [378], but a VSR expressed during mosquito infection has not been identified [379-381], although a potential role for DENV NS4b as a viral suppressor of RNAi in mammalian cells has been proposed [382]. Recently it was shown that the noncoding RNA derived from the 3' untranslated region of flaviviruses, known as the subgenomic flavivirus RNA (sfRNA), is unable to be degraded by cellular ribonucleases and is essential for viral pathogenicity [383, 384]. During infection with DENV or Kunjin viruses, the sfRNA inhibits the activity of the exonuclease, XRN1, which is responsible for removing mRNA decay intermediates and de-stabilizes host cell mRNA [385].

Recently, sfRNA has been proposed to inhibit the miRNA and siRNA pathways in both mammalian and insect cells by inhibiting the cleavage of dsRNA by human Dicer, but conclusive evidence of inhibition of Dicer-dependent dsRNA cleavage is lacking [384].

Summary of Dissertation

Specific Aim 1: Define the genetic differences between global populations of *Aedes aegypti aegypti* and *Aedes aegypti formosus* populations in Senegal

The current understanding of the population genetics of global collections of *Ae. aegypti* is based on a limited number of isozyme, microsatellite, or mitochondrial markers. Using a limited number of markers can result in ascertainment bias. The current system of subspecies classification developed in East Africa is not informative in West Africa, highlighting the need to further examine genetic differences between sylvatic and urban forms of *Ae. aegypti* in West Africa.

For this dissertation, deep sequencing of the exome was performed on natural populations of *Ae. aegypti* from Mexico and Thailand, and two locations in Senegal (PK10 and Kaolack), to provide insight into the regions of the genome and types of mutations that are involved in the genetic divergence of global populations of *Ae. aegypti*. Differences in the distributions of allele frequencies represented as F_{ST} values per gene, and the ratio of replacement to synonymous mutations were calculated to compare the degree of positive selection between populations and identify genes and regions of genes involved in speciation. (Chapter 2)

Along with a population genetic approach, genetic differences in the sex determination locus were serendipitously identified between the subspecies. This was achieved by genetic crossing experiments and deep sequencing of genes involved in sex determination in natural

populations from PK10, Senegal, and Pai Lom, Thailand to identify sex-specific and subspecies-specific differences. (Chapter 3)

Specific Aim 2: Determine the distribution of vector competence for yellow fever virus and dengue virus type-2 in *Aedes aegypti* collections from Senegal.

Vector competence of *Aedes aegypti* mosquitoes is a quantitative genetic trait that varies among geographic locations and among different flavivirus species and genotypes within species. The subspecies *Ae. aegypti formosus*, found in sub-Saharan Africa, is considered to be refractory to both dengue (DENV) and yellow fever viruses (YFV) compared to the more globally distributed *Ae. aegypti aegypti*. Within Senegal, vector competence varies with collection site and DENV-2 viral isolate, but knowledge about the interaction of West African *Ae. aegypti* with other flaviviruses is lacking.

Research reported in this dissertation utilized low passage isolates of DENV-2 (DENV-2-75505 sylvatic genotype) and YFV (YFV BA-55 -West African Genotype I, or YFV DAK 1279-West African Genotype II) from West Africa and field derived *Ae. aegypti* collected from throughout Senegal to determine if vector competence is flavivirus or virus genotype dependent. (Chapter 4)

Specific Aim 3: Examine genetic diversity of Dcr2 and its correlation with vector competence in natural populations of *Aedes aegypti* from Senegal.

In the mosquito *Ae. aegypti*, susceptibility to dengue virus (DENV) infection is a complex and quantitative genetic trait. The major innate immune response to viral infection in mosquitoes is the RNA interference (RNAi) pathway, specifically the exo-siRNA pathway. Although genes in the exo-siRNA pathway, such as Dcr2, which plays a role in recognition of

viral dsRNA generated during replication, has been implicated in susceptibility to DENV; the genetic structure of Dcr2 in natural populations remains unknown.

For this dissertation, genetic diversity across the entire Dcr2 gene in populations from Mexico, Thailand, and Senegal were compared. A domain of Dcr2 involved in binding viral dsRNA was sequenced from individual mosquitoes from three locations in Senegal with known vector competence phenotypes to determine if increased genetic diversity of this domain of Dcr2 is associated with increased refractoriness to DENV (Chapter 5).

CHAPTER 2: EVIDENCE OF ALLOPATRIC AND SYMPATRIC SPECIATION IN *AEDES AEGYPTI*

Introduction

The mosquito *Ae. aegypti* is the primary vector for all four serotypes of dengue (DENV 1-4) and yellow fever viruses worldwide. Dengue remains an important public health problem with an estimated 390 million infections per year [2]. Yellow fever outbreaks are still continuously reported although a safe and effective vaccine exists. The species *Ae. aegypti (L)* is morphologically, ecologically, and genetically diverse [132-139]. *Aedes aegypti (L)* is found globally in tropical and sub-tropical climates and exists as two subspecies: *Ae. aegypti aegypti (*Aaa*)* and *Ae. aegypti formosus*(Aaf) [137, 138]. Early population genetic studies of *Ae. aegypti* by Tabachnick, Powell, Munstermann and Wallis demonstrated that collections fell into two clades [133-135, 167-173] based on a limited number of isozyme markers. One clade contained *Ae. aegypti aegypti* from East Africa, South America, and the Caribbean. The second clade contained *Ae. aegypti aegypti* populations from Asia and the Southwestern USA with a basal *Ae. aegypti formosus* from both East and West Africa. A more recent study analyzed 12 microsatellite loci in 24 worldwide collections of *Ae. aegypti* and demonstrated two distinct clusters; domestic populations outside Africa and both domestic and forest populations within Africa [146]. The mitochondrial NADH dehydrogenase subunit 4 (ND4) gene has also been used to examine gene flow between and among collections of *Ae. aegypti* outside Africa [174-176]. The addition of sequences from 10 collections from Senegal representing Aaa and Aaf revealed that populations of *Ae. aegypti* outside Africa arose from two different clades, a basal West African clade and a second East African clade that arises from the first [177].

Characteristics that distinguish the two subspecies were developed from data collected primarily in East Africa but are contradictory and confusing when identifying *Ae. aegypti* collected in West Africa. The current definition of the subspecies is based on the number or degree of white scales on the first abdominal tergite as defined by Mattingly and McClelland [137, 138]. *Aedes aegypti aegypti* has scales on the first abdominal tergite, has a light tan cuticle, is globally distributed, tends to be endophilic, and has a feeding preference for humans. In contrast, *Ae. aegypti formosus* has no white scales on the first abdominal tergite, has a dark or black cuticle, is found mostly in Sub-Saharan Africa in sylvatic environments, tends to be exophilic, and has a feeding preference for wild animals [139-141]. However, these distinctions become less clear in West Africa where *Ae. aegypti* with a dark black cuticle and scales (albeit usually few) are frequently detected near human habitats. In East Africa, the scaling pattern and behavior are also correlated with discrete genetic differences in allozyme and microsatellite markers [135, 146]. But in West Africa, the scaling pattern does not correlate with behavioral differences or genetic markers [144, 145] and leads to confusion in subspecies identification. Instead, genetic differences arise between domestic and sylvatic groups [142]. Furthermore, East African *Ae. aegypti aegypti* and *Ae. aegypti formosus* are genetically distinct from the monophyletic West African *Ae. aegypti* [135, 146].

Great variation in vector competence for DENV is seen among diverse mosquito collections [169, 245, 249], among geographically proximate collections of *Ae. aegypti* [207-209, 257], and depending upon the isolate of DENV-2 [131, 216]. *Aedes aegypti* vector competence for DENV-2 is a quantitative genetic trait [249] and multiple genetic loci confer resistance to DENV infection [158-161]. Understanding the population structure of *Ae. aegypti* worldwide, and being

able to distinguish between the subspecies, will allow us to better understand DENV or YFV transmission and potentially assist in identifying novel control methods.

The fixation index (F_{ST}) is an F-statistic that is derived from the inbreeding coefficient, F. F_{ST} is a measure of the variation in allele frequencies among subpopulations (S) relative to allele frequencies in the total population (T) (Wright 1951). The values of F_{ST} are continuous between 0 and 1. Large differences in allele frequencies between subpopulations result in large F_{ST} values, while small differences in allele frequencies between subpopulations yield small F_{ST} values. F_{ST} is frequently used to summarize genetic relationships among a group of subpopulations or between pairs of populations.

Previous population genetics studies of *Ae. aegypti* have been based upon isozyme markers or RFLP markers representing a small number of genes, or biased towards microsatellites which can lead to size homoplasy. In most applications, the mean F_{ST} has been used as a summary statistic representing a limited sample of the F_{ST} distribution. In reality, because of the many factors affecting F_{ST}, it should be treated as a distribution between populations [386]. With the advent of next generation sequencing [387] we first become able to examine the distribution of F_{ST} across an entire genome. Herein we report on the distributions of F_{ST} values among four disparate populations of the mosquito *Ae. aegypti* to test the hypothesis that sympatric speciation is occurring among *Ae. aegypti* in Senegal and allopatric speciation is occurring between *Ae. aegypti* subspecies within and outside of Senegal. This study also provides insight into the most informative genetic markers with which to study the population structure of *Ae. aegypti* subspecies.

Materials and Methods

Field Collection of Mosquitoes

Deep sequencing libraries were constructed from pools of F_1 individuals collected from the field. Mosquitoes were collected from PK10, Senegal (all Aaf) based on absence of scales on the first abdominal tergite [388]), Kaolack, Senegal (mixture of scales, Aaf), Merida , Mexico, and Pai Lom, Thailand. Mosquitoes from Mexico and Thailand represent *Ae. aegypti aegypti* (Aaa). A total of twelve libraries were constructed: PK10 Female Replicates 1 and 2, PK10 Male Replicates 1 and 2, Thailand Female Replicates 1 and 2, Thailand Male Replicates 1 and 2 (each with 12 individuals), Kaolack Replicates 1 and 2 (14 individuals each), and Mexico Replicates 1 and 2 (22 individuals each). Thus each of the four populations was represented by either 48 individuals, 44 individuals, or 28 individuals. DNA in individual mosquitoes was quantified using Pico Green (Life Technologies, catalog #P11496) and equal amounts of DNA per individual were pooled into each replicate.

Sequencing

We present results from libraries made from an exon-capture enrichment library. This enrichment is necessary because 60% of the *Ae. aegypti* genome consists of transposable elements and other forms of repetitive DNA [165]. For the exome capture procedure, TruSeq libraries (Illumina Inc., San Diego, CA) were prepared using 1 µg of genomic DNA per library, and DNA was sheared to 300-500 basepairs via sonication prior to library preparation. Exome capture was then performed to enrich for coding sequences using custom SeqCap EZ Developer probes (Nimblegen, Roche Life Sciences). Non-repetitive probes tiling coding sequences in the AaegL1.3 gene annotations were designed by Nimblegen. In total, 18,769 genes representing 26.7 Mb of the genome (20%) was targeted for enrichment. TruSeq libraries were hybridized to

the probes twice, unbound DNA was washed away, and the targeted DNA was eluted and amplified before being sequenced with 1 lane of 100 bp HiSeq2000 paired-end sequencing run. TruSeq library preparation, exome capture and sequencing were performed following manufacturer's instructions by the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (Oxford, UK).

Construction of Reference Genome

A series of FORTRAN programs were written to create a reference file used for the detailed analysis of each of the 18,769 genes in the AaegL1.4 reference genome on VectorBase. The coordinates of the features of each gene appearing in Aedes-aegypti-Liverpool_BASEFEATURES_Aaegl1.4.GTF were used to extract 600 bp upstream from the start of the 5' untranslated region (UTR) (called the 5'nontranscribed region, 5'NTR) , all of the 5'UTR, all exons, all introns, all of the 3'UTR, and 600 bp downstream of the 3'UTR.

Next, a file of the physical locations of 294 supercontigs [389] were used to identify the locations of 70 supercontigs containing 1,330 genes on chromosome 1, 142 mapped supercontigs containing 1,489 genes on chromosome 2, and 82 supercontigs containing 1,258 genes on chromosome 3. The remaining 14,692 genes have yet to be mapped. This file, called "All3U" contained information on all 301,348,667 nucleotides corresponding to 45% of the *Ae. aegypti* genome [389]. Sequences were assembled as a single string to act as the sequence index file for subsequent analyses.

Sequencing Analysis

The FASTQ files from each library were aligned to the All3U sequence index using GSNAP [390]. GSNAP was chosen because *Ae. aegypti* contains on average 1 SNP every 10-15 nucleotides [391] and other programs (e.g., Bowtie) reject reads with more than 2 SNPs per 100

nucleotide read. The SAM file produced by GSNAP was converted to a BAM file and then sorted. A Pileup file was created from the SortedBAM file using SAMtools [392]. The "readcounts" command in Varscan2 analyzed the Pileup file to report SNPs with a minimum of 15x coverage in each library.

A FORTRAN program *Line Reader 1* read the output from Varscan2 and produced a flat file containing the numbers of each nucleotide and the numbers and sequences of insertions and deletions. A second program *Line Reader 2* reduced the flat file for each of the polymorphic sites to a single line containing the SNP ID number followed by the counts of each nucleotide, insertion, and deletion. Even though several types of insertions and deletions were detected, they were all combined into a single category in the current analysis. The program *2x2* identified SNP ID numbers in common between the two replicate libraries to produce a single file containing nucleotide IDs that appeared in both libraries. A program *combine* added together males and females for rep 1 and rep 2 to form 4 files, PK10 Reps 1 and 2 and Thailand Reps 1 and 2. A program *2x2x2* combined all possible pairs of the four *2x2* tables (PK10 vs Kaolack, PK10 vs Mexico, PK10 vs Thailand, Kaolack vs Mexico, Kaolack vs Thailand, and Mexico vs Thailand). The program *tables* assembled a 4 x 6 contingency table with each of the 4 rows corresponding to a different library. A program *OneData* read in the contingency table file and produced a file containing the SNP ID number, the count of each of the four nucleotides and insertions and deletions and total coverage in a population. In addition, each SNP was coded as '1' if it was in an exon and encoded the first codon position, a'2' if it encoded the second codon position or '3' if it encoded the second codon position. SNPs in the 5' untranslated region (UTR) were coded as'4', in the 5'nontranscribed (NTS) as '5', in the 3' UTR as'6', in the 3'NTS as '7'. All SNPs in introns were encoded as '8.' Only SNPs with 2 alternate nucleotides, were

encoded as a 'transition,' 'transversion,' or 'indel.' SNPs in exons were scored as either 'silent'or 'replacement'

F_{ST} was estimated using the equation in Fumagalli et al. [393]. Genetic variance between populations (a_S) was calculated by:

$$a_s = \frac{4ni(P'_{i,s} - P'_s)^2 + 4nj(P'_{j,s} - P'_s)^2 - bs}{2(2ninj/(ni+nj))} \quad \text{(eq. 1)}$$

Within population variance (b_S) was calculated by:

$$b_s = \frac{n_i \alpha_{i,s} + nj\alpha_{j,s}}{n_i + nj - 1} \quad \text{(eq. 2)}$$

Where p'_i and p'_j are allele frequencies based in subpopulations *i* and j at SNPs, and n_i and n_j are the coverages in subpopulations *i* and j. $\alpha_{i,s}$ is the expected heterozysity at SNPs in subpopulation *i*:

$$\alpha_{i,s} = 2P'_{i,s}(1 - P'_{i,s}) \quad \text{(eq. 3)}$$

F_{ST} over all SNPs in a gene was calculated as:

$$F_{ST}^{(locus)} = \frac{\sum_{s=1}^{m} a_s}{\sum_{s=1}^{m}(a_s + b_s)} \quad \text{(eq. 4)}$$

A program *OneData Annotate* read each line of output from *OneData* and annotated each with a SNP ID number from the All3U reference file, the VectorBase name, codon, and position type (number 1-8). A program *6Stats* indicated if the mutation was a transition, transversion, or indel and encoded it as a silent or replacement substitution if it was in an exon.

A program *6stat-gene* calculated F_{ST} and expected heterozygosity for each gene and placed all genes on one of the three chromosomes or listed it as "unmapped." *6stat-mutation* calculated F_{ST} and expected heterozygosity for each SNP that was either a transition, transversion or InDel. *6stat-w* calculated F_{ST} and expected heterozygosity for each SNP that was either a silent or replacement substitution. Thus the F_{ST} distribution for all genes in a pairwise comparison were subdivided into 8 categories according to position in a gene or into three categories depending upon the type of mutation or into two categories depending upon whether they caused a silent or replacement substitutions.

The 'hist' routine in R 3.1.0 [394] with a bin size set to 0.01 and a range from 0 to 1 was used to place all genes into 100 bins. These counts were converted to frequencies as counts/sum (counts) and then plotted using ggplot2 in R 3.1.0. The beta (a,b) family of distributions are commonly used for continuously distributed random variables that are distributed from 0 to 1. The beta probability density function is:

$$f(x: a, b) = kx^{(a-1)}(1-x)^{(b-1)} \text{ for values of x between 0 and 1} \qquad \text{[eq. 5]}$$

Where $k = \Gamma(a + b)/\Gamma(a)\Gamma(b)$ and $\Gamma(a)=(a-1)!$. The component $x^{(a-1)}(1-x)^{(b-1)}$ determines the shape of the distribution while k is a constant needed to yield f (x;a,b) a probability density function (pdf). The mean of a beta distribution pdf is $E(x) = a/(a+b)$ and the variance is $Var(x) = ab/(a + b)^2(a + b+1)$. Figure 2.1 shows three examples of beta distributions likely to be assumed by F_{ST} and a fourth distribution that is very unlikely to appear among empirical distributions. Curve A would represent an empirical distribution of F_{ST} values arising among subpopulations sharing alleles through migration or mutation. There are very few genes that differ greatly in frequency. Alleles in the two subpopulations might differ through genetic drift or through limited local adaptation. Curve B represents an empirical F_{ST} distribution arising among subpopulations with

low migration or mutation. Genetic drift is an important factor for some genes while local adaptation is also a factor. Curve C represent an empirical F_{ST} distribution expected among subpopulations with very limited migration and mutation. Many genes throughout the genome are subject to genetic drift while local adaptation may also lead to large F_{ST} values. Curve D represents a very unlikely F_{ST} distribution among subpopulations with very limited migration or mutation. Notice that almost no genes have low F_{ST} values. This is unlikely because highly conserved genes will always be shared even among valid, reproductively isolated species. Thus as populations become more and more isolated, a shoulder covering low F_{ST} values should appear. Estimating parameters of a beta distribution allows confidence intervals of the average F_{ST} values to be determined and permits comparisons between the characteristics of both individual genes and subpopulations. Regions with a lack of fit to a beta distribution can also provide insights in mechanisms affecting genes with excessively high or low F_{ST} values. The distribution of F_{ST} values for a pair of populations was fit to beta distribution with the fitdistr package in the MASS package in R3.1.10. Expected values, variances, and 5% and 95% limits were calculated using the quantile option in the rbeta package in R3.1.10.

Next we wished to examine the distribution of F_{ST} values according to whether they belong to one of the eight regions in a gene, or whether they are a transition, transversion or InDel in a gene or whether they encode a silent or replacement substitution. For this we needed to estimate the conditional probability of sampling an F_{ST} values in a bin with a lower limit of *i* and an upper limit of *j* given that the gene region was r where r = 1-8 or

$$P\ (i\ \leq F_{ST} \leq\ j \mid r) \qquad \text{[eq. 6]}$$

This can be calculated as a posterior probability using Bayes' rule because:

$$P\ (i\ < F_{ST} <\ j \mid r) \quad = \quad \frac{P\ (r \mid i\ < F_{ST} <\ j)\ x\ P\ (i\ < F_{ST} <\ j)}{\sum_s^{100} P(r|i < FST(s) < j)\ x\ P(i < FST(s) < j)} \qquad \text{[eq. 7]}$$

Where the likelihood P (r | i < F_{ST} < j) is the frequency of F_{ST} in bin (i, j) for gene region r and the prior P (i < F_{ST} < j) is the overall frequency of F_{ST} in bin (i, j) over all r positions. The denominator is this product summed over all 100 bins. Equation 7 was used to calculate the posterior probability distribution of F_{ST} values for the 8 different regions of a gene, among transitions, transversions, and InDels and for silent and replacement amino acid substitutions.

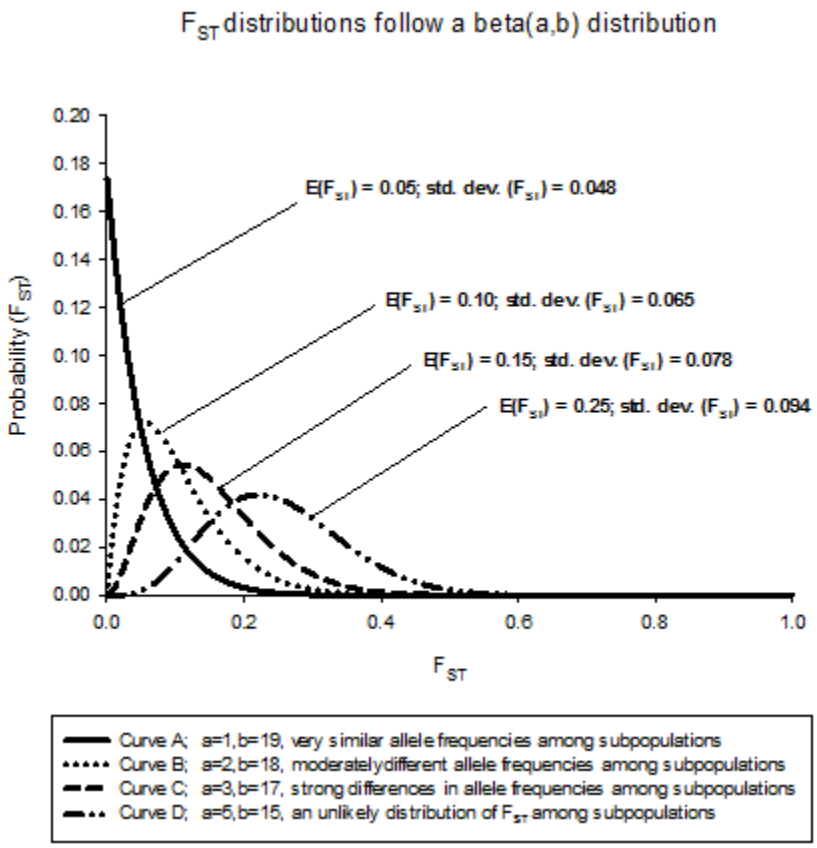


Figure 2.1: Representative examples of beta distributions likely to be assumed by F_{ST} values

Three probability distributions of F_{ST} that are likely to be encountered in natural populations, and a probability distribution that is unlikely (curve D). Curve A would represent an empirical distribution of F_{ST} values arising among subpopulations sharing alleles. Curve B represents an empirical F_{ST} distribution arising among subpopulations with low migration or mutation. Curve C represents an empirical F_{ST} distribution expected among subpopulations with very limited migration and mutation. Curve D represents a very unlikely F_{ST} distribution among subpopulations with very limited migration or mutation.

Results

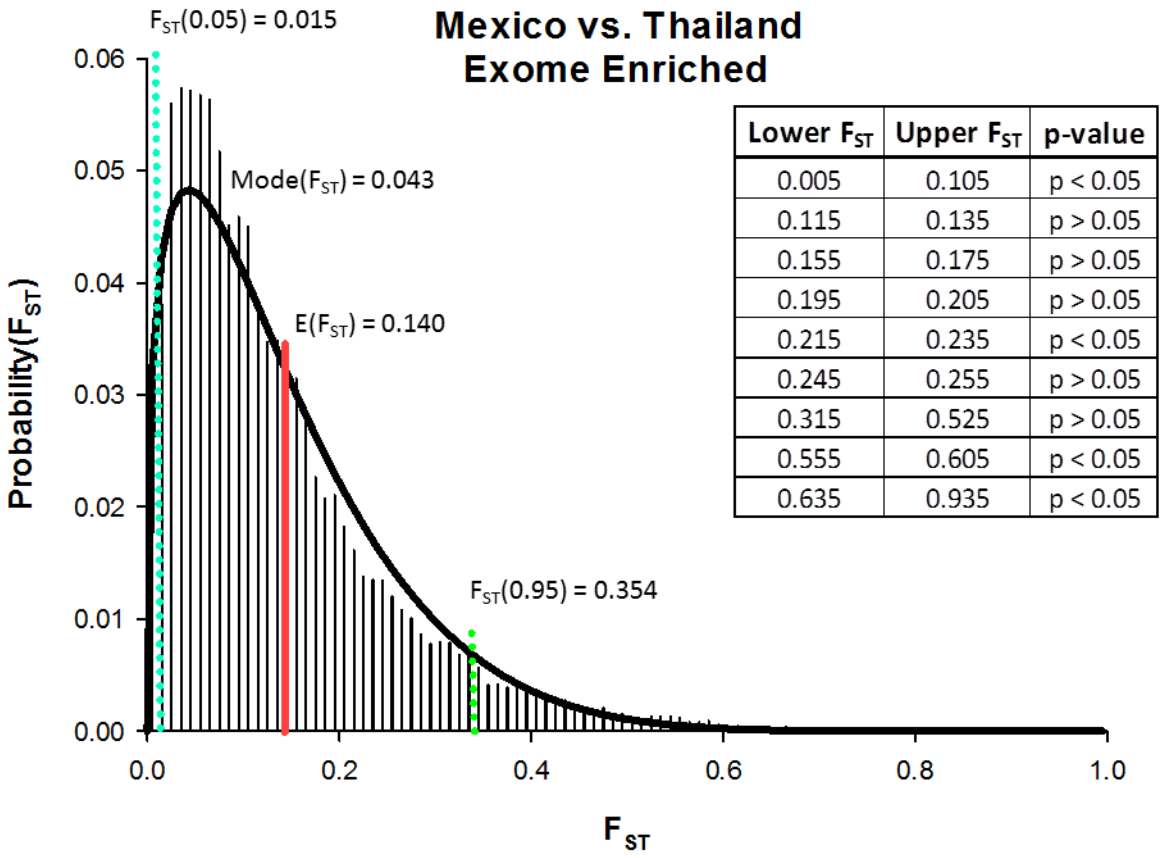
In each of the four libraries an average of 15,852 genes were represented out of the 18,769 genes in the current annotation of the *Ae. aegypti* genome on Vectorbase (Table 2.1). An average of 82% of all nucleotide sites were fixed and 18% were polymorphic in the four populations, however *Ae. aegypti aegypti* from Mexico or Thailand had 10% fewer polymorphic sites than *Ae. aegypti formosus* from Senegal. For all libraries, 24% - 43% of polymorphic sites were in protein coding sequences (Table 2.1).

Table 2.1: Summary of Deep Sequencing Libraries

	PK10	Kaolack	Mexico	Thailand
Number of Genes	15,947	16,080	15,730	15,652
Percentage of fixed sites	77.5%	78.1%	87.2%	85.5%
Percentage of polymorphic sites	22.5%	21.9%	12.8%	14.5%
Percent transitions	44.9%	42.3%	44.7%	45.7%
Percent transversions	27.6%	28.9%	27.7%	27.1%
Percent synonymous mutations	14.8%	11.2%	18.2%	19.9%
Percent replacement mutations	19.1%	12.8%	18.3%	24.0%

The F_{ST} values in each of the six pairwise comparisons of the four populations were fit to a beta distribution to determine if the comparison of beta distributions of F_{ST} values would be an informative approach to identify genetic differences between subpopulations. Each of the six pairwise comparisons fit a beta distribution (Figures 2.2-2.7) and the shape of the distribution varied among the six comparisons. The within subspecies comparisons (Figures 2.2 and 2.3) had lower mean F_{ST} values (F_{ST} = 0.140 – 0.144) than the between subspecies comparisons (F_{ST} = 0.193 – 0.230) (Figures 2.4 to 2.7). The characteristics of the distribution of F_{ST} values from each pairwise comparison are reported in Table 2.2. The beta distribution of the F_{ST} values from the within subspecies comparisons (Figures 2.2 and 2.3) were more narrow and had a high probability of low F_{ST} values. In the between Mexico and Thailand comparison (Figure 2.2), by

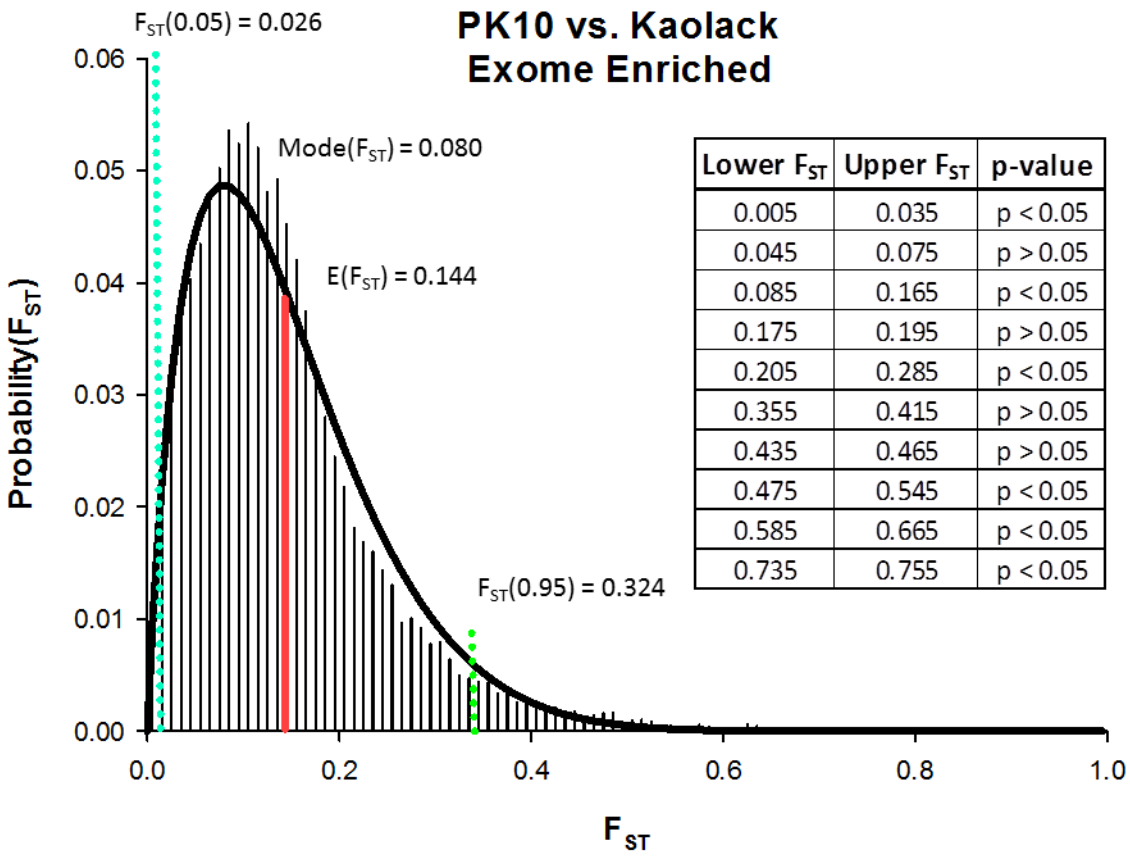
χ^2 analysis, there was a significantly significant excess of genes with low F_{ST} values and a deficiency of genes with high F_{ST} values as indicated by the observed F_{ST} values exceeding or receding from the expected beta distribution of F_{ST} values, but these deviations from the beta distribution were not as prominent in the within Senegal comparison (Figure 2.3). In contrast, the beta distributions of the F_{ST} values between subspecies (Figures 2.4 – 2.7) were more broad and were shifted to the right indicating a greater probability of higher F_{ST} values. Between Kaolack and Aaa comparisons (Figures 2.4 and 2.5), there were no larges differences in the observed F_{ST} values compared to the beta distribution, but between PK10 and Aaa comparisons (Figures 2.6 and 2.7), there was an excess of mid-range F_{ST} values and deficiencies of both low and high F_{ST} values as compared to the curve of the beta distribution. This excess or deficiency of high and low F_{ST} values was statistically significant by χ^2 analysis and shown on each individual graph.



Lower F_{ST}	Upper F_{ST}	p-value
0.005	0.105	p < 0.05
0.115	0.135	p > 0.05
0.155	0.175	p > 0.05
0.195	0.205	p > 0.05
0.215	0.235	p < 0.05
0.245	0.255	p > 0.05
0.315	0.525	p > 0.05
0.555	0.605	p < 0.05
0.635	0.935	p < 0.05

Figure 2.2: Probability distribution of F_{ST} values across the genome

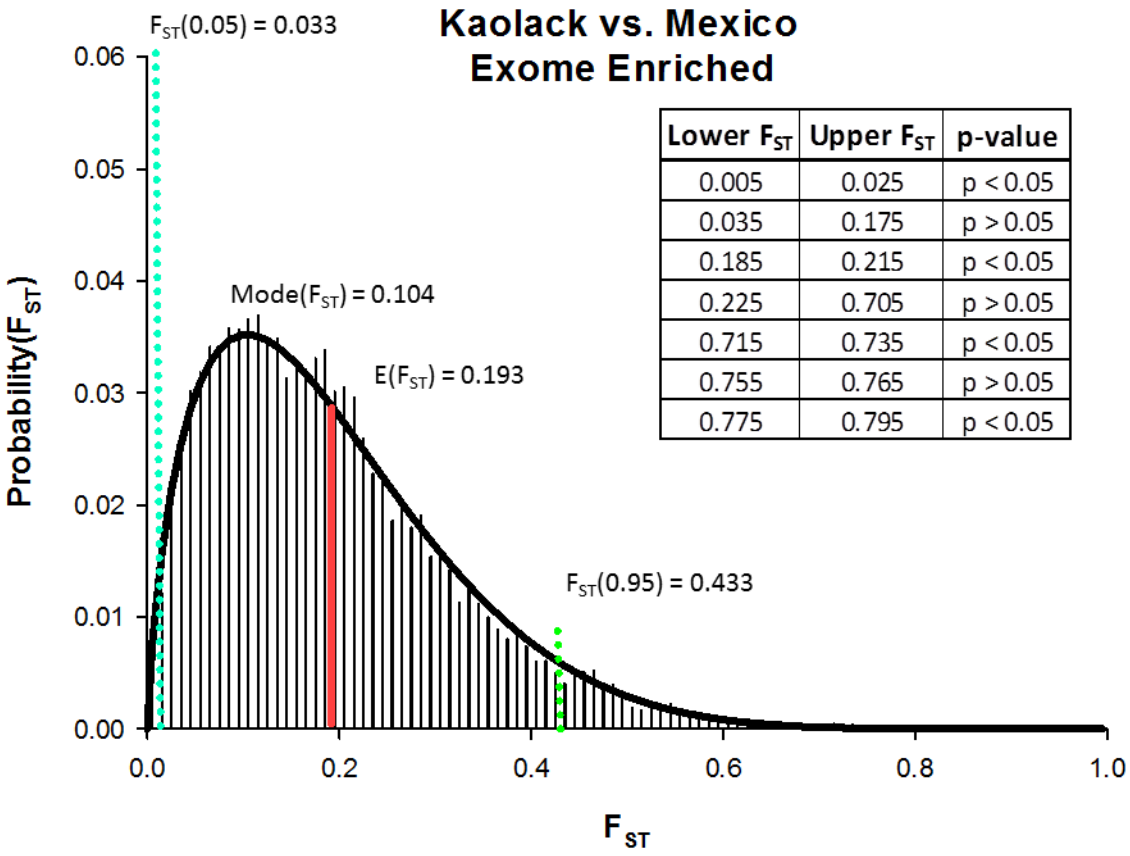
F_{ST} values were calculated in individual genes representing 83% of all annotated genes (vertical lines) and fit to a beta distribution ($\alpha = 1.316$, $\beta = 8.063$) represented with the smooth line. The mean of the beta distribution, $E(F_{ST})$ is represented by the red line. The upper 5% threshold F_{ST} value (green line) and lower 5% threshold F_{ST} value (blue line) of the distribution are also shown. The inset table shows the level of statistical significance calculated by χ^2 analysis of the observed F_{ST} values and the predicted beta distribution across regions representing a range of at least .02 F_{ST} across the entire distribution. P-values could not be calculated for the right tail of the distribution due to values of zero.



Lower F_{ST}	Upper F_{ST}	p-value
0.005	0.035	p < 0.05
0.045	0.075	p > 0.05
0.085	0.165	p < 0.05
0.175	0.195	p > 0.05
0.205	0.285	p < 0.05
0.355	0.415	p > 0.05
0.435	0.465	p > 0.05
0.475	0.545	p < 0.05
0.585	0.665	p < 0.05
0.735	0.755	p < 0.05

Figure 2.3: Probability distribution of F_{ST} values across the genome

F_{ST} values were calculated in individual genes representing 83% of all annotated genes (vertical lines) and fit to a beta distribution ($\alpha = 1.938$, $\beta = 8.796$) represented with the smooth line. The mean of the beta distribution, $E(F_{ST})$ is represented by the red line. The upper 5% threshold F_{ST} value (green line) and lower 5% threshold F_{ST} value (blue line) of the distribution are also shown. The inset table shows the level of statistical significance calculated by χ^2 analysis of the observed F_{ST} values and the predicted beta distribution across regions representing a range of at least .02 F_{ST} across the entire distribution. P-values could not be calculated for the right tail of the distribution due to values of zero.



Lower F_{ST}	Upper F_{ST}	p-value
0.005	0.025	$p < 0.05$
0.035	0.175	$p > 0.05$
0.185	0.215	$p < 0.05$
0.225	0.705	$p > 0.05$
0.715	0.735	$p < 0.05$
0.755	0.765	$p > 0.05$
0.775	0.795	$p < 0.05$

Figure 2.4: Probability distribution of F_{ST} values across the genome

F_{ST} values were calculated in individual genes representing 83% of all annotated genes (vertical lines) and fit to a beta distribution ($\alpha = 1.723$, $\beta = 7.226$) represented by the smooth line. The mean of the beta distribution, $E(F_{ST})$ is represented by the red line. The upper 5% threshold F_{ST} value (green line) and lower 5% threshold F_{ST} value (blue line) of the distribution are also shown. The inset table shows the level of statistical significance calculated by χ^2 analysis of the observed F_{ST} values and the predicted beta distribution across regions representing a range of at least .02 F_{ST} across the entire distribution. P-values could not be calculated for the right tail of the distribution due to values of zero.

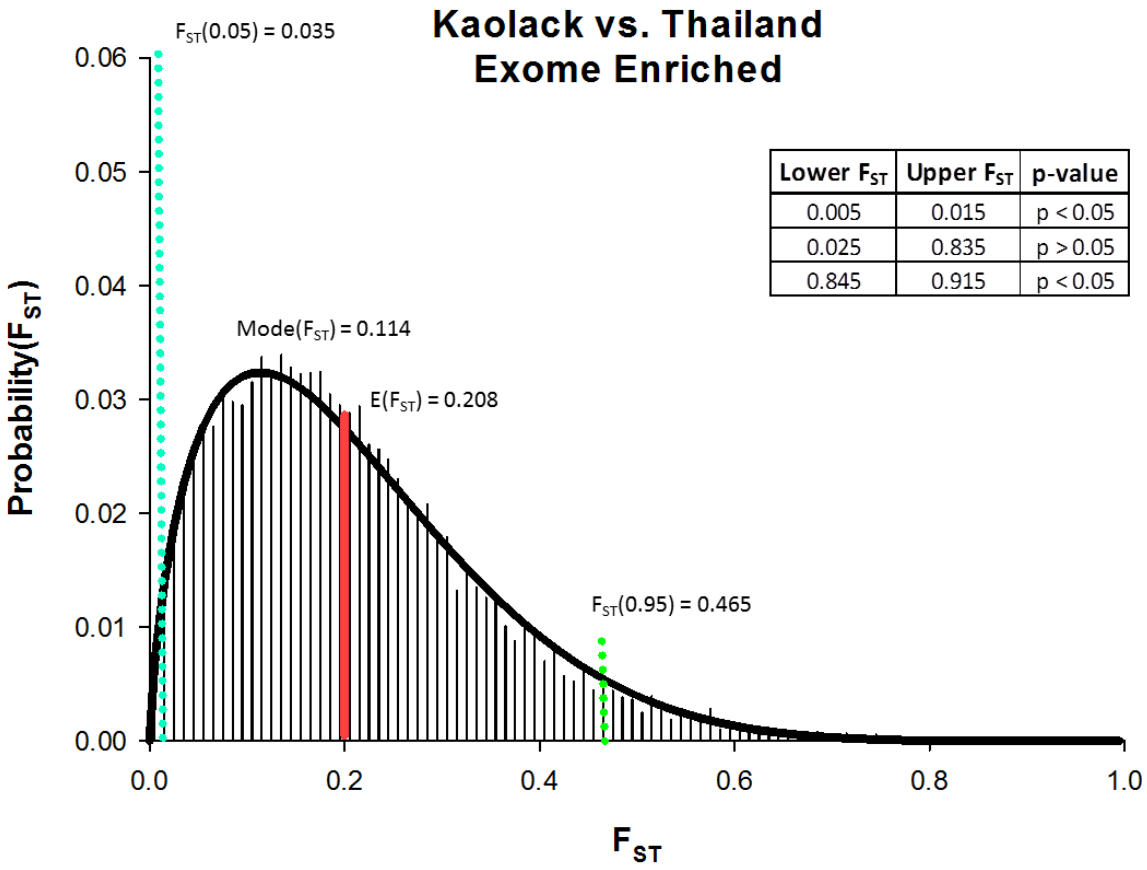
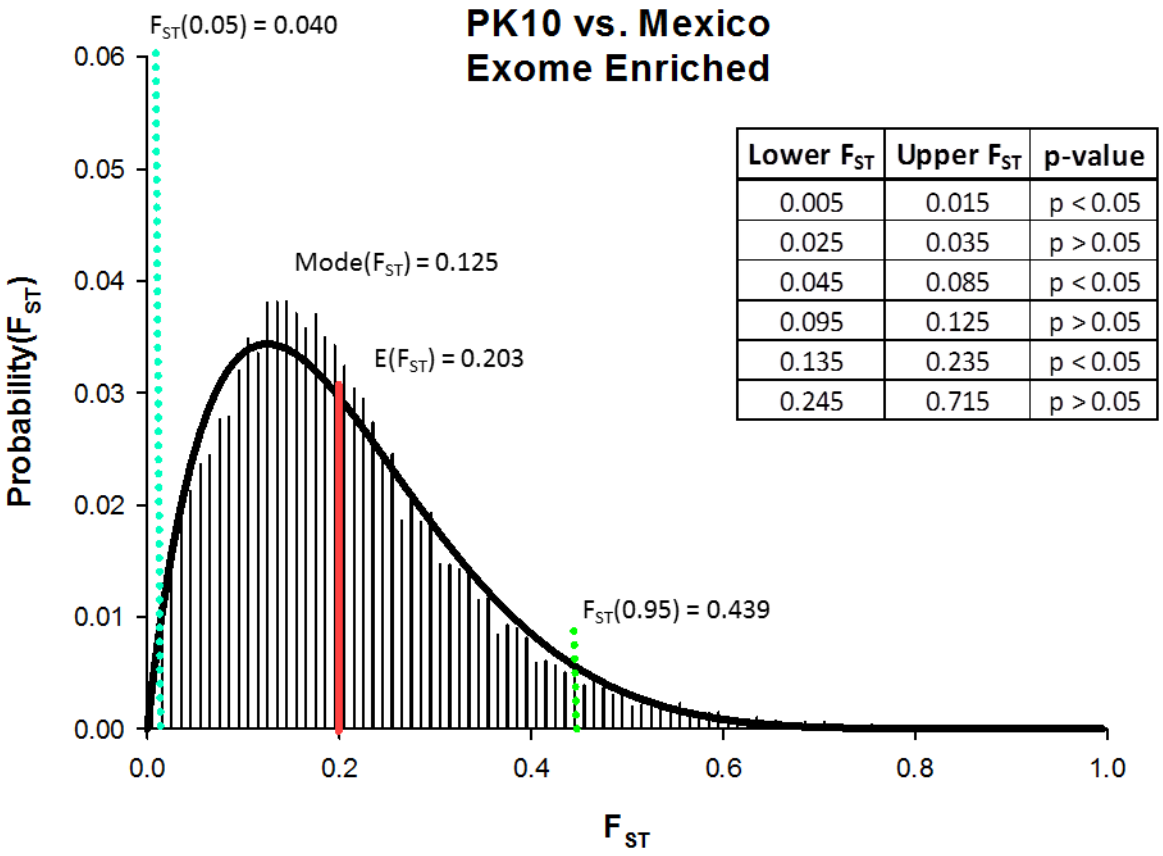


Figure 2.5: Probability distribution of F_{ST} values across the genome

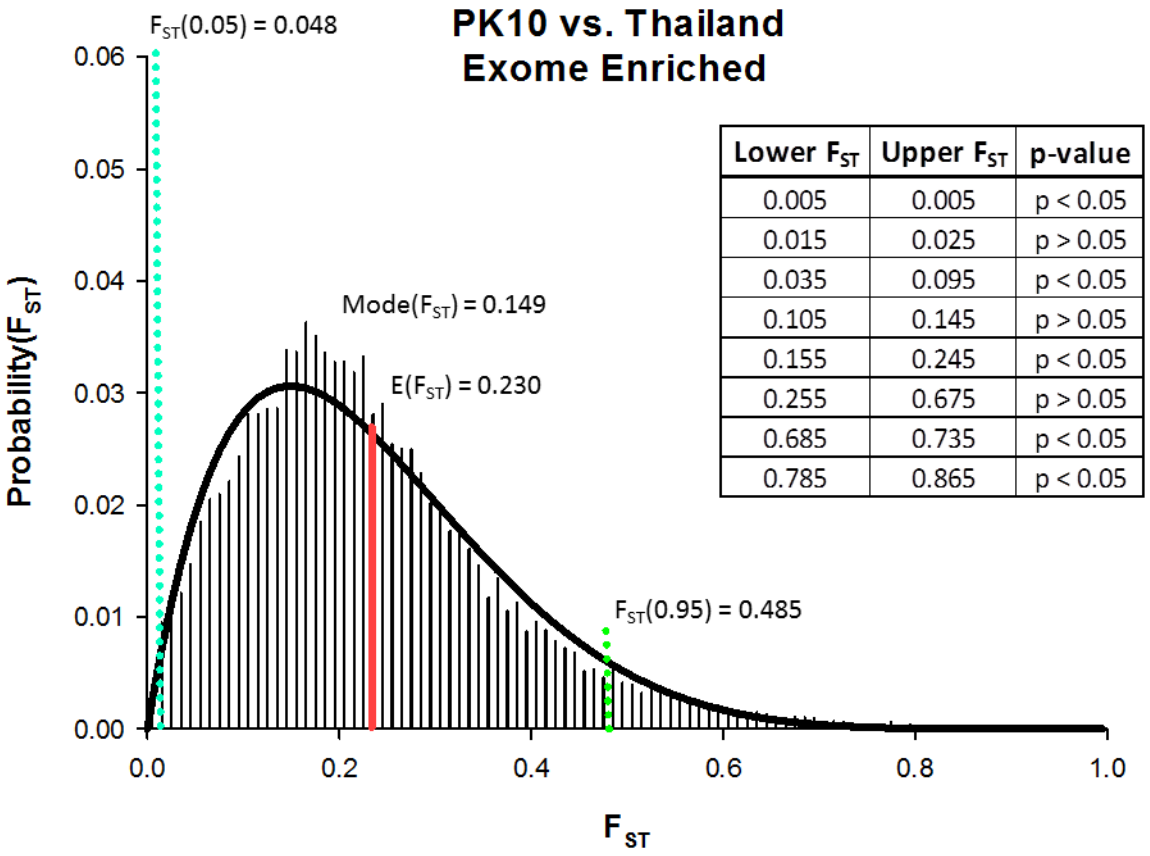
F_{ST} values were calculated in individual genes representing 83% of all annotated genes (vertical lines) and fit to a beta distribution ($\alpha = 1.703$, $\beta = 6.487$) represented by the smooth line. The mean of the beta distribution, E(F_{ST}) is represented by the red line. The upper 5% threshold F_{ST} value (green line) and lower 5% threshold F_{ST} value (blue line) of the distribution are also shown. The inset table shows the level of statistical significance calculated by χ^2 analysis of the observed F_{ST} values and the predicted beta distribution across regions representing a range of at least .02 F_{ST} across the entire distribution. P-values could not be calculated for the right tail of the distribution due to values of zero.



Lower F_{ST}	Upper F_{ST}	p-value
0.005	0.015	$p < 0.05$
0.025	0.035	$p > 0.05$
0.045	0.085	$p < 0.05$
0.095	0.125	$p > 0.05$
0.135	0.235	$p < 0.05$
0.245	0.715	$p > 0.05$

Figure 2.6: Probability distribution of F_{ST} values across the genome

F_{ST} values were calculated in individual genes representing 83% of all annotated genes (vertical lines) and fit to a beta distribution ($\alpha = 1.938$, $\beta = 7.596$) represented by the smooth line. The mean of the beta distribution, $E(F_{ST})$ is represented by the red line. The upper 5% threshold F_{ST} value (green line) and lower 5% threshold F_{ST} value (blue line) of the distribution are also shown. The inset table shows the level of statistical significance calculated by χ^2 analysis of the observed F_{ST} values and the predicted beta distribution across regions representing a range of at least .02 F_{ST} across the entire distribution. P-values could not be calculated for the right tail of the distribution due to values of zero.



Lower F_{ST}	Upper F_{ST}	p-value
0.005	0.005	p < 0.05
0.015	0.025	p > 0.05
0.035	0.095	p < 0.05
0.105	0.145	p > 0.05
0.155	0.245	p < 0.05
0.255	0.675	p > 0.05
0.685	0.735	p < 0.05
0.785	0.865	p < 0.05

Figure 2.7: Probability distribution of F_{ST} values across the genome

F_{ST} values were calculated in individual genes representing 83% of all annotated genes (vertical lines) and fit to a beta distribution ($\alpha = 1.996$, $\beta = 6.686$) represented by the smooth line. The mean of the beta distribution, $E(F_{ST})$ is represented by the red line. The upper 5% threshold F_{ST} value (green line) and lower 5% threshold F_{ST} value (blue line) of the distribution are also shown. The inset table shows the level of statistical significance calculated by χ^2 analysis of the observed F_{ST} values and the predicted beta distribution across regions representing a range of at least .02 F_{ST} across the entire distribution. P-values could not be calculated for the right tail of the distribution due to values of zero.

Table 2.2: Summary of the beta distribution of F_{ST} values across the entire gene

	# of genes	a	b	E(X)	FST_05	FST_95	Mode
MexThai	15,550	1.316	8.063	0.140	0.015	0.354	0.043
PK10Kao	15,815	1.886	11.208	0.144	0.026	0.324	0.080
PK10Mex	15,620	1.938	7.596	0.203	0.040	0.439	0.124
PK10Thai	15,574	1.996	6.686	0.230	0.048	0.485	0.149
KaoMex	15,630	1.723	7.226	0.193	0.033	0.433	0.104
KaoThai	15,572	1.703	6.487	0.208	0.035	0.465	0.114

Differences based on gene region

We used Bayes Rule to identify how F_{ST} values from specific gene regions (3'NTR, 3'UTR, codon position 1, codon position 2, codon position 3, intron, 5'NTR, and 5'UTR) contributed to the overall beta distribution of F_{ST} values. For polymorphic sites we compared the distribution of F_{ST} values among transitions (A to G or C to T substitution), transversions (purine to pyrimidine), and InDels (gap resulting from insertion or deletion). For polymorphic sites in exons, we compared the distribution of F_{ST} values between silent (synonymous) and replacement (non-synonymous) substitutions.

The probability of low F_{ST} values of SNPs in 5'NTR or 5'UTR was high ($p = 0.10 – 0.20$) (Figures 2.8 and 2.9). Although there was no difference between the population comparisons in the 5' region of genes, the 5'UTR appears to be more conserved between the populations than the 5'NTR as seen by the higher probability of very low F_{ST} values in the 5'UTR as compared to the 5' NTR.

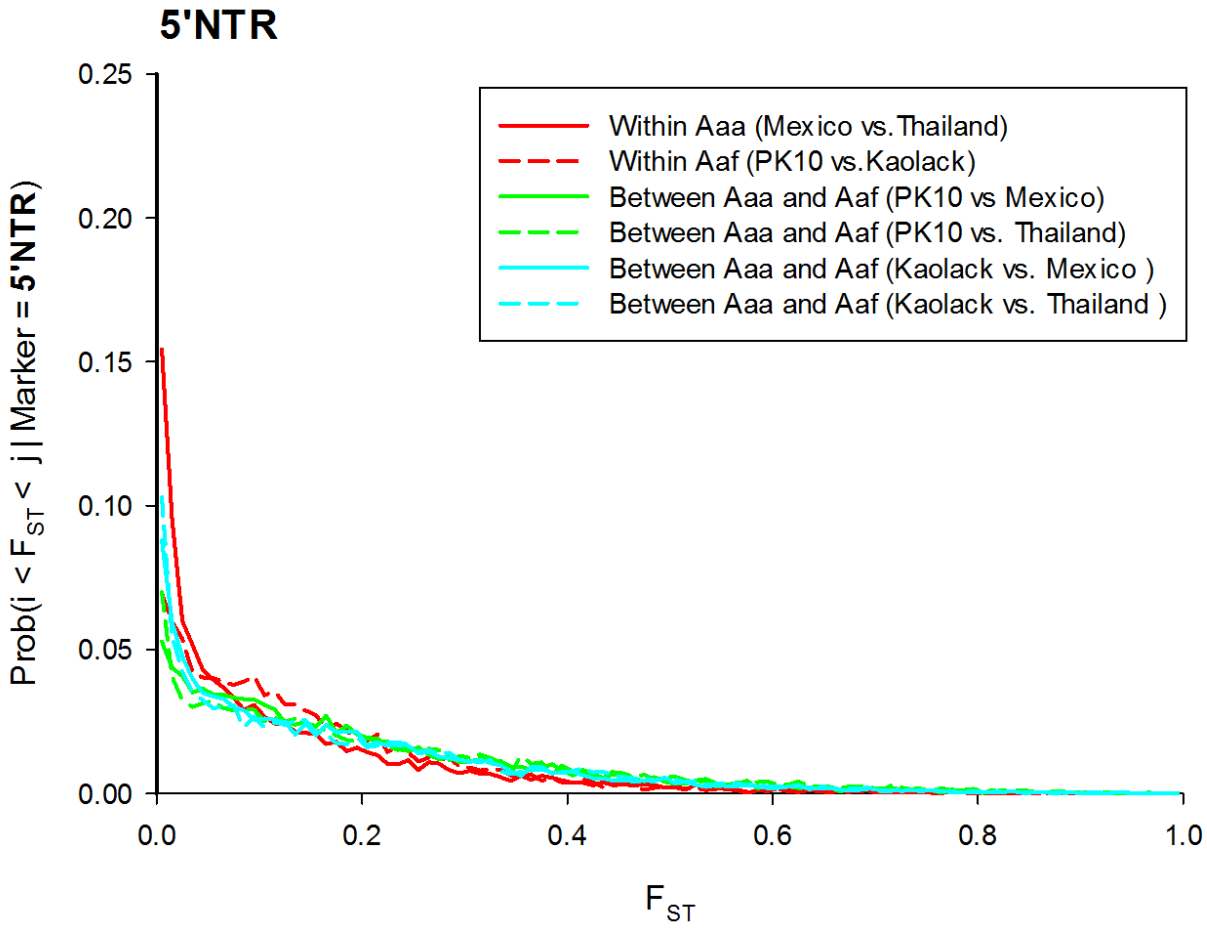


Figure 2.8: Probability distribution of F_{ST} values in 5'NTR

The probability distribution F_{ST} values was calculated using Bayes' rule. The probability of F_{ST} values (x-axis) in intervals from i to j in increments of 0.01 (y-axis) given that the allele was in the 5' nontranscribed region 600 basepairs upstream of the 5'UTR shown in the 6 pairwise comparisons. Within subspecies comparisons are in red. Between subspecies comparisons are in blue and green.

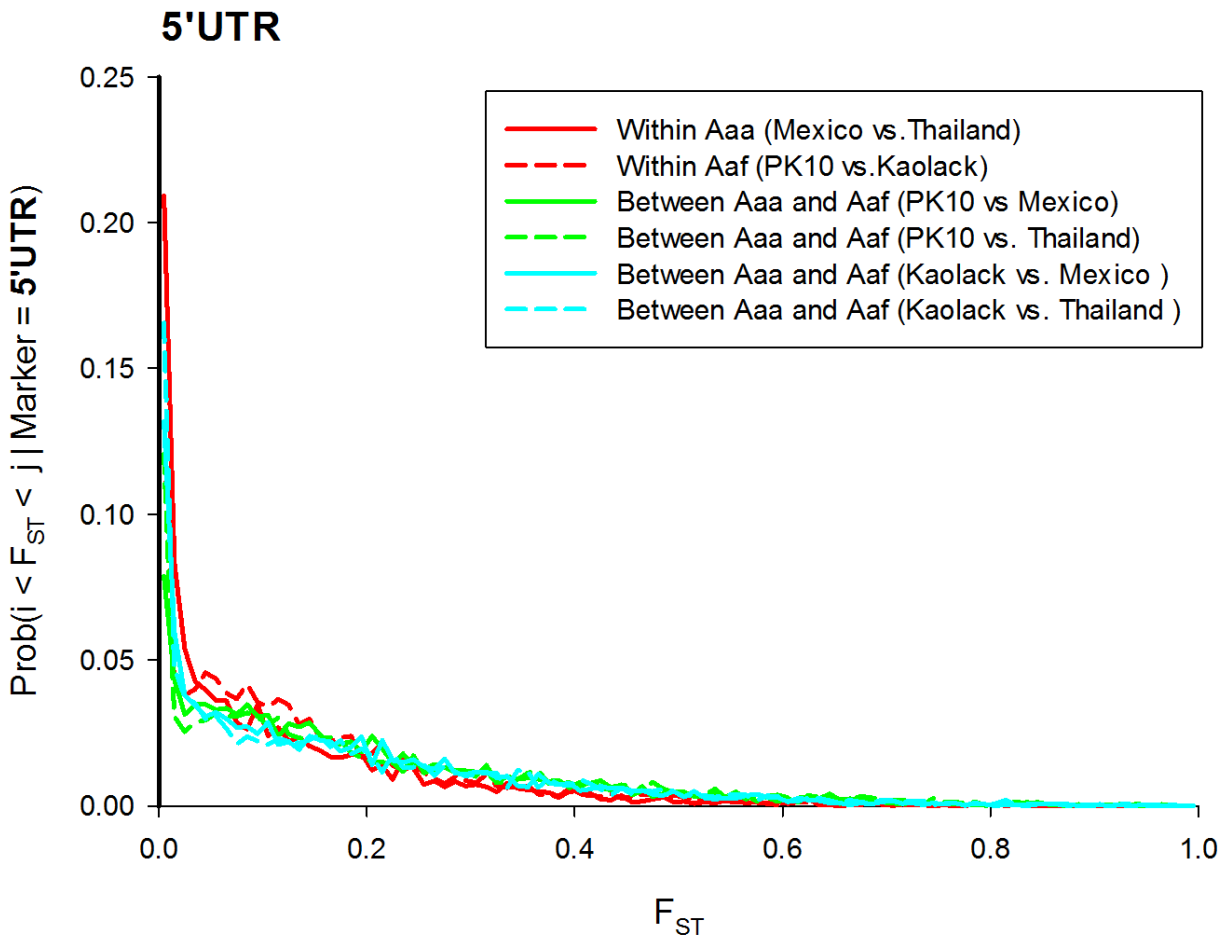


Figure 2.9: Probability distribution of F_{ST} values in 5'UTR
The probability distribution of F_{ST} values was calculated using Bayes' rule. The probability of F_{ST} values in intervals from i to j in increments of 0.01 given that the alternate allele was in the 5' untranslated region (5'UTR) for each of the 6 pairwise comparisons. Within subspecies comparisons are in red. Between subspecies comparisons are in blue and green.

The probability distributions of F_{ST} values among SNPs in codons varied among the three codon positions. When the polymorphic site was in the 1st or 2nd codon position, there was a high probability ($p = 0.1 - 0.3$) of very low F_{ST} values, or conversely a low probability ($p < 0.01$) of high F_{ST} values (Figure 2.10 and 2.11). In contrast, when the polymorphic site was in 3rd codon position, there was a much lower probability of low F_{ST} values ($p = 0.01 – 0.05$) (Figure 2.12). Also, the probability distribution of various F_{ST} values was much broader and there was a

higher probability of high F_{ST} values for polymorphic sites in codon position 3. The high probability of low F_{ST} values from polymorphic sites in codon position 1 or 2 indicates these are highly conserved nucleotides, while the shift in the probability distribution of F_{ST} values from polymorphic sites in codon position 3 indicates these nucleotides are less conserved. There are only slight differences in the probability distributions of F_{ST} values from polymorphic sites in codon position 1 or 2 in the within subspecies comparisons and between subspecies comparisons, but the probability distributions of F_{ST} values from polymorphic sites in codon position 3 show differences in the various subspecies comparisons indicating SNPs in codon position 3 could be informative markers within and among subspecies (Figure 2.12). FST values tended to be low in the within Aaa comparison, while the distribution of F_{ST} values from the within Aaf comparison has slightly higher F_{ST} values and the between Aaa and Aaf distributions have even higher F_{ST} values (Figures 2.10 – 2.12).

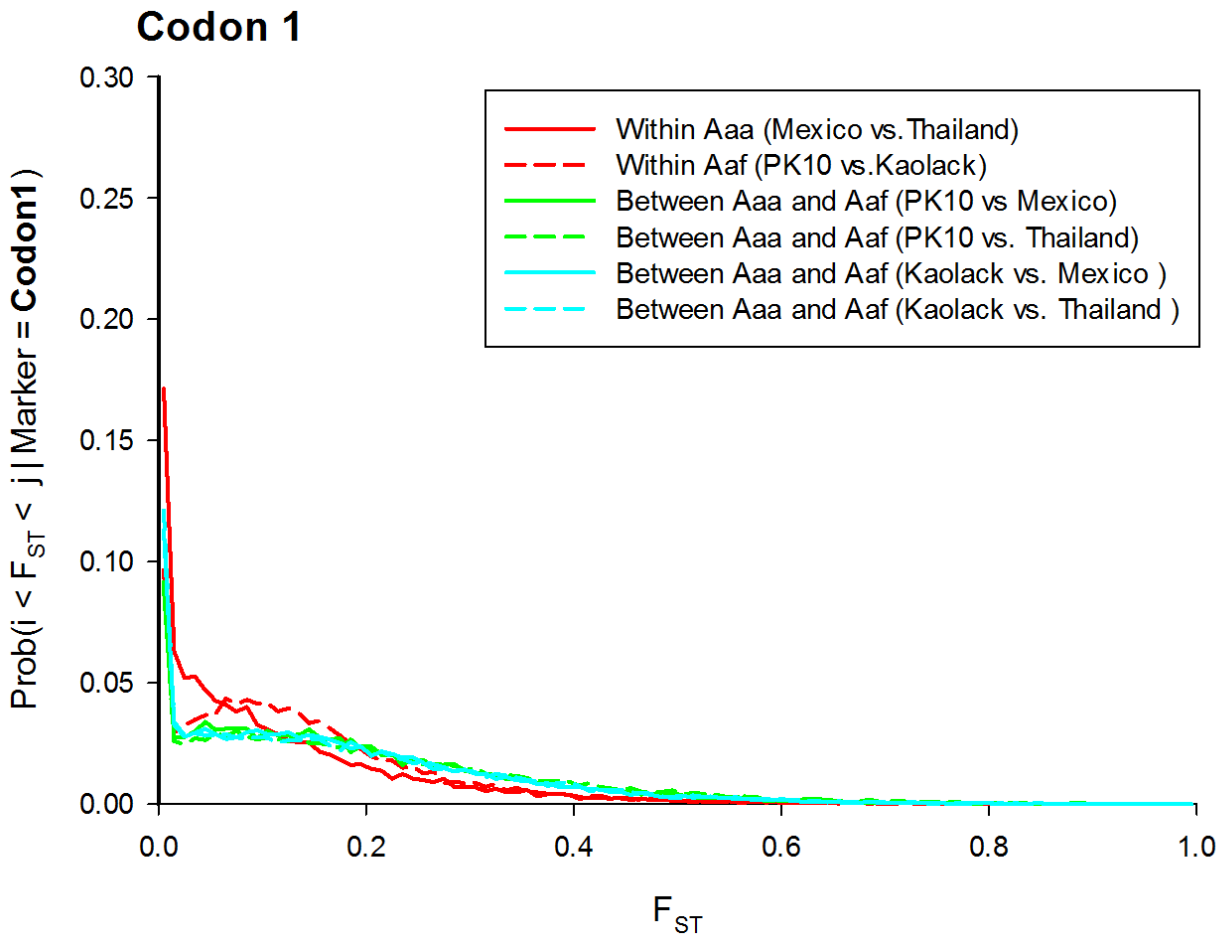


Figure 2.10: Probability distribution of F_{ST} values in 1st codon position

The probability distribution of F_{ST} values was calculated using Bayes' rule. The probability of F_{ST} values in intervals from i to j in increments of 0.01 given that the alternate allele was in the 1st codon position for each of the 6 pairwise comparisons. Within subspecies comparisons are in red. Between subspecies comparisons are in blue and green.

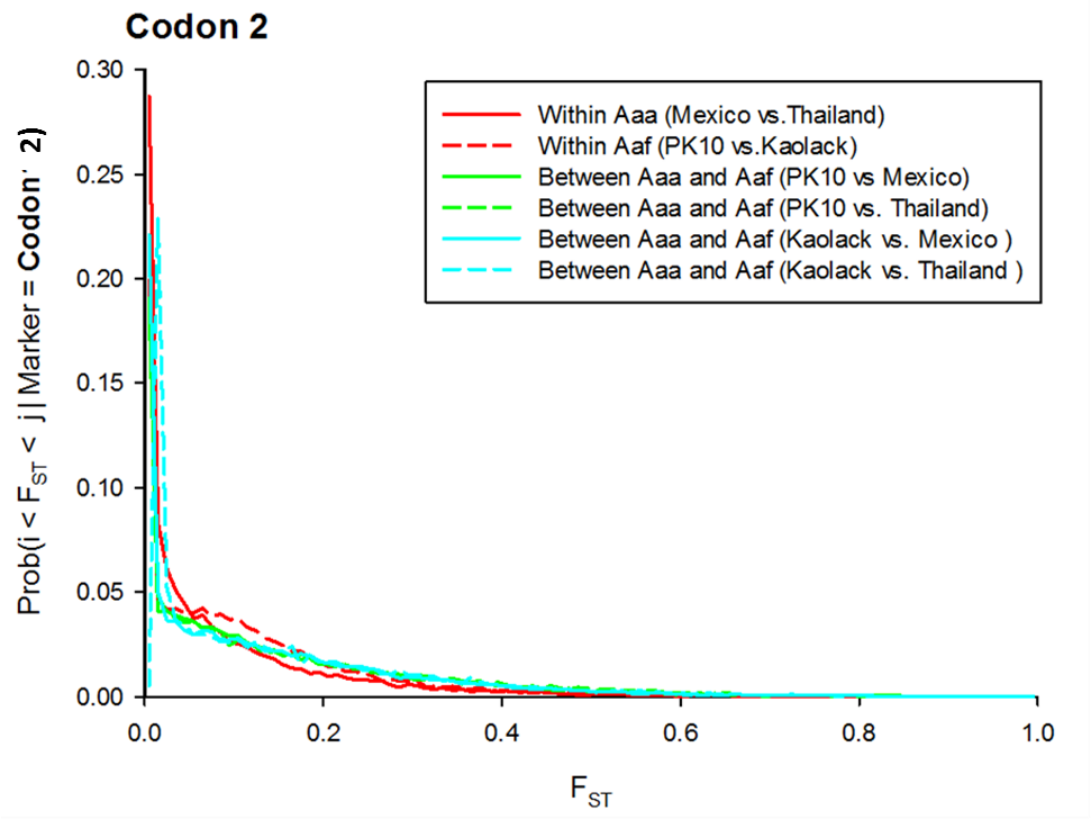


Figure 2.11: Probability distribution of F_{ST} values in 2nd codon position

The probability distribution of F_{ST} values was calculated using Bayes' rule. The probability of F_{ST} values in intervals from i to j in increments of 0.01 given that the alternate allele was in the 2nd codon position for each of the 6 pairwise comparisons. Within subspecies comparisons are in red. Between subspecies comparisons are in blue and green.

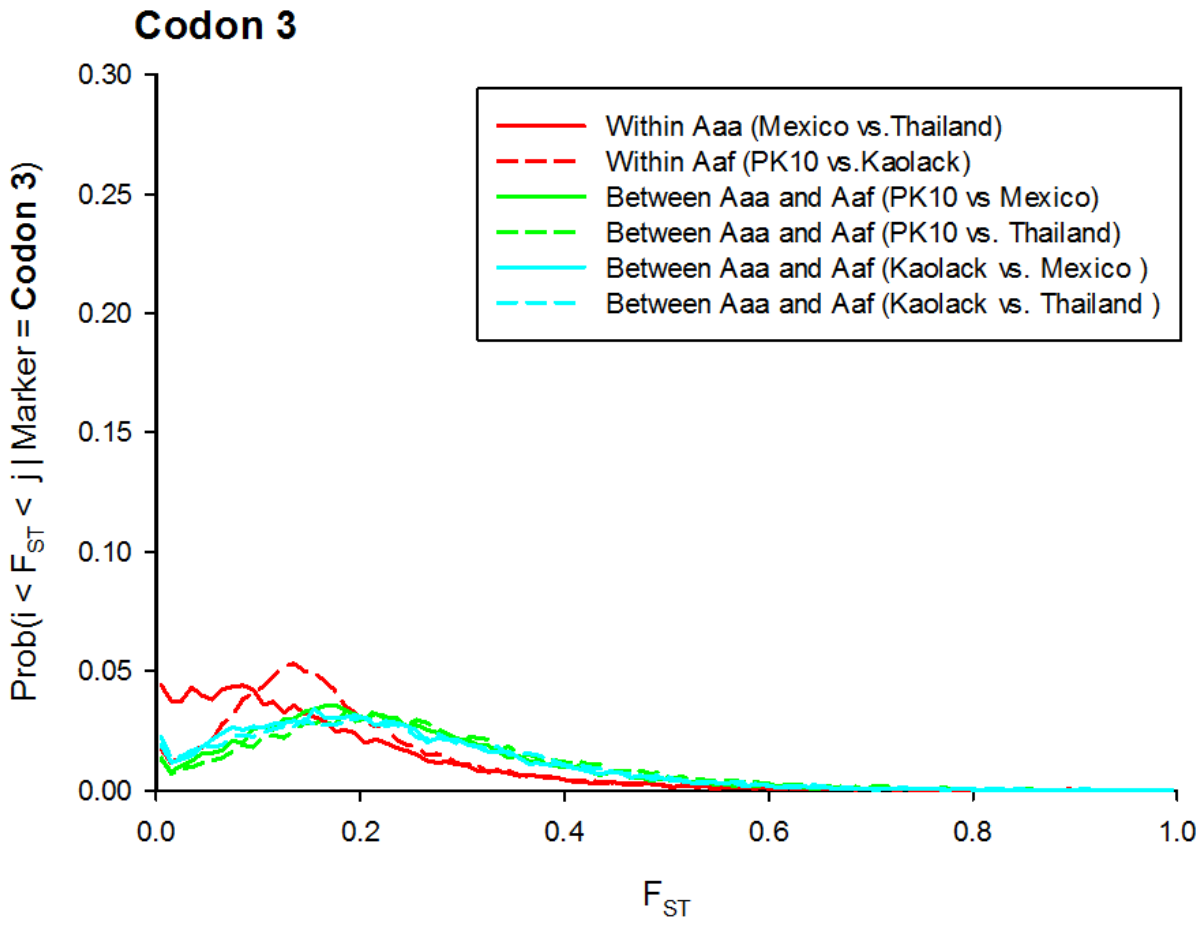


Figure 2.12: Probability distribution of F_{ST} values in 3rd codon position
The probability distribution of F_{ST} values was calculated using Bayes' rule. The probability of F_{ST} values in intervals from i to j in increments of 0.01 given that the alternate allele was in the 3rd codon position for each of the 6 pairwise comparisons. Within subspecies comparisons are in red. Between subspecies comparisons are in blue and green.

When a SNP was in an intron (Figure 2.13), the probability of low F_{ST} values ranged from 0.03-0.10. Unlike the F_{ST} values in the conserved 5'NTR, 5'UTR, and 1st and 2nd codon positions where there was no shoulder over low probability low F_{ST} values, in introns there is a shoulder with the maximum F_{ST} shifted towards higher F_{ST} values. Slight differences were detected in the different population comparisons. The within subspecies comparisons had a higher probability ($p = 0.10$) of low F_{ST} values, while the between subspecies comparisons had

lower probability (p = 0.03 - 0.05) of low F_{ST} values. Also, the probability of higher F_{ST} values was greater (p = 0.01) in the between subspecies comparisons and lower in the within subspecies comparison (p = 0.005). Although there are slight differences in the frequency of higher F_{ST} values in introns between the population comparisons, the mode of the distributions do not differ. As a result, introns do not appear to be as an informative marker, as compared to markers in the 3rd codon position, of subspecies population structure in *Ae. aegypti*.

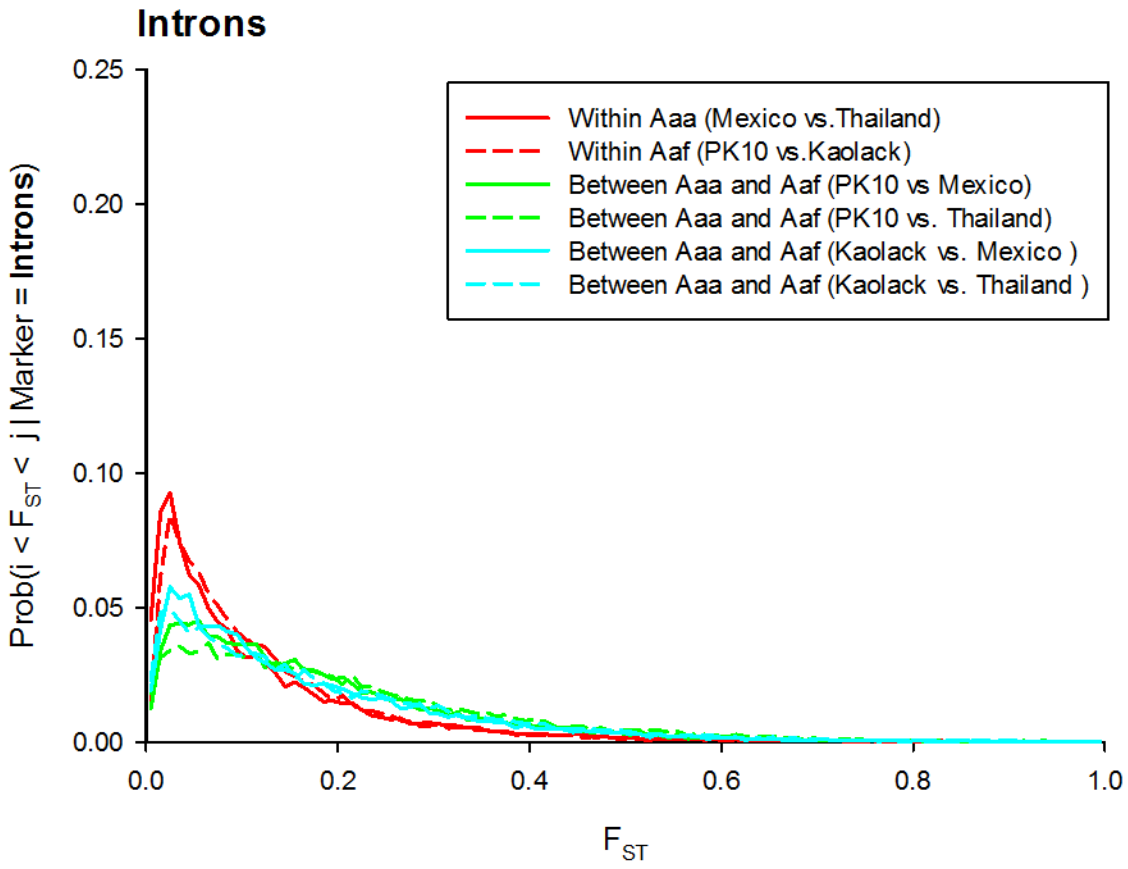


Figure 2.13: Probability distribution of F_{ST} values in introns
The probability distribution of F_{ST} values was calculated using Bayes' rule. The probability of F_{ST} values in intervals from i to j in increments of 0.01 given that the alternate allele was in an intron for each of the 6 pairwise comparisons. Within subspecies comparisons are in red. Between subspecies comparisons are in blue and green.

When the SNP was in a 3'NTR or 3'UTR, the probability of low F_{ST} values was high within Aaa comparisons (p = 0.12 -0.17) and low in the within Aaf comparison and between subspecies comparisons (p = 0.05 – 0.12) (Figures 2.14 and 2.15). The 5'UTR and 5' NTR appear to be more conserved than the 3'NTR and 3' UTR as seen by the higher probability of very low F_{ST} values in the 5' regions as compared to the 3' regions.

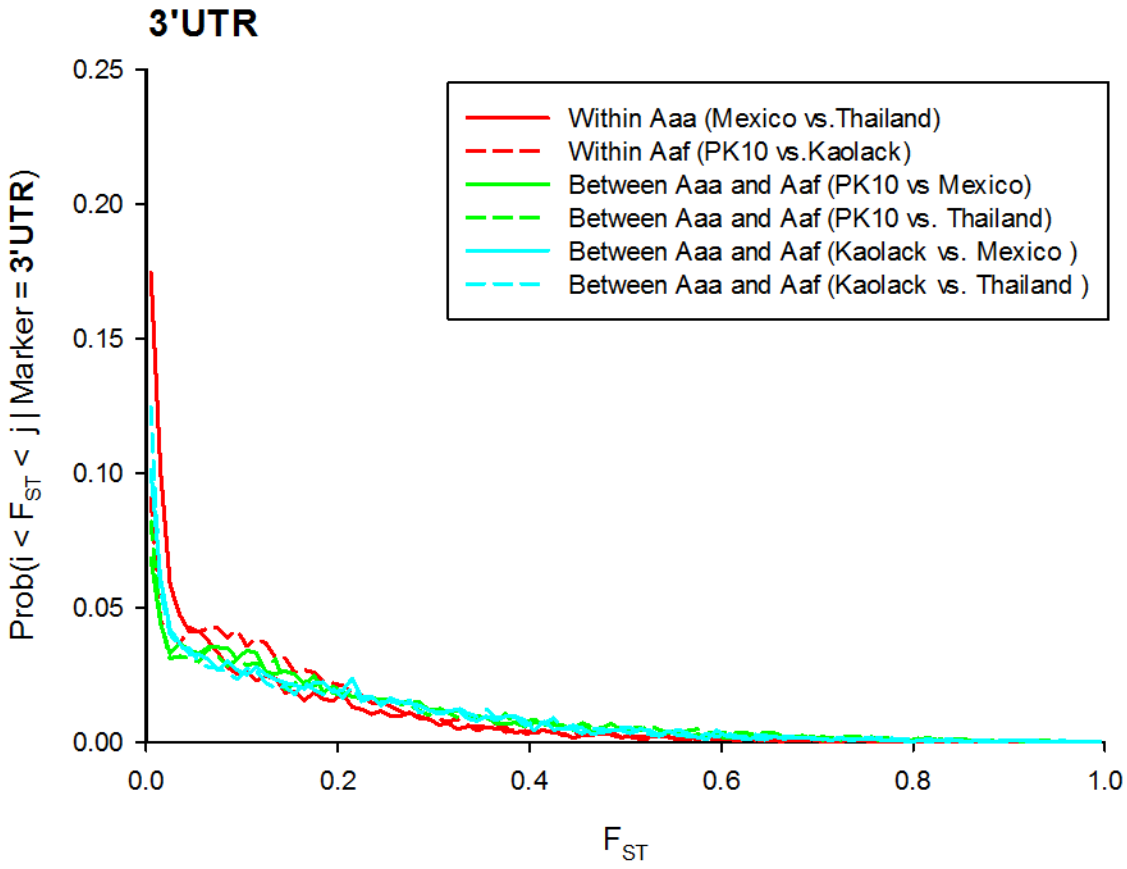


Figure 2.14: Probability distribution of F_{ST} values in 3'UTR

The probability distribution of F_{ST} values was calculated using Bayes' rule. The probability of F_{ST} values in intervals from i to j in increments of 0.01 given that the alternate allele was in the 3' untranslated region for each of the 6 pairwise comparisons. Within subspecies comparisons are in red. Between subspecies comparisons are in blue and green.

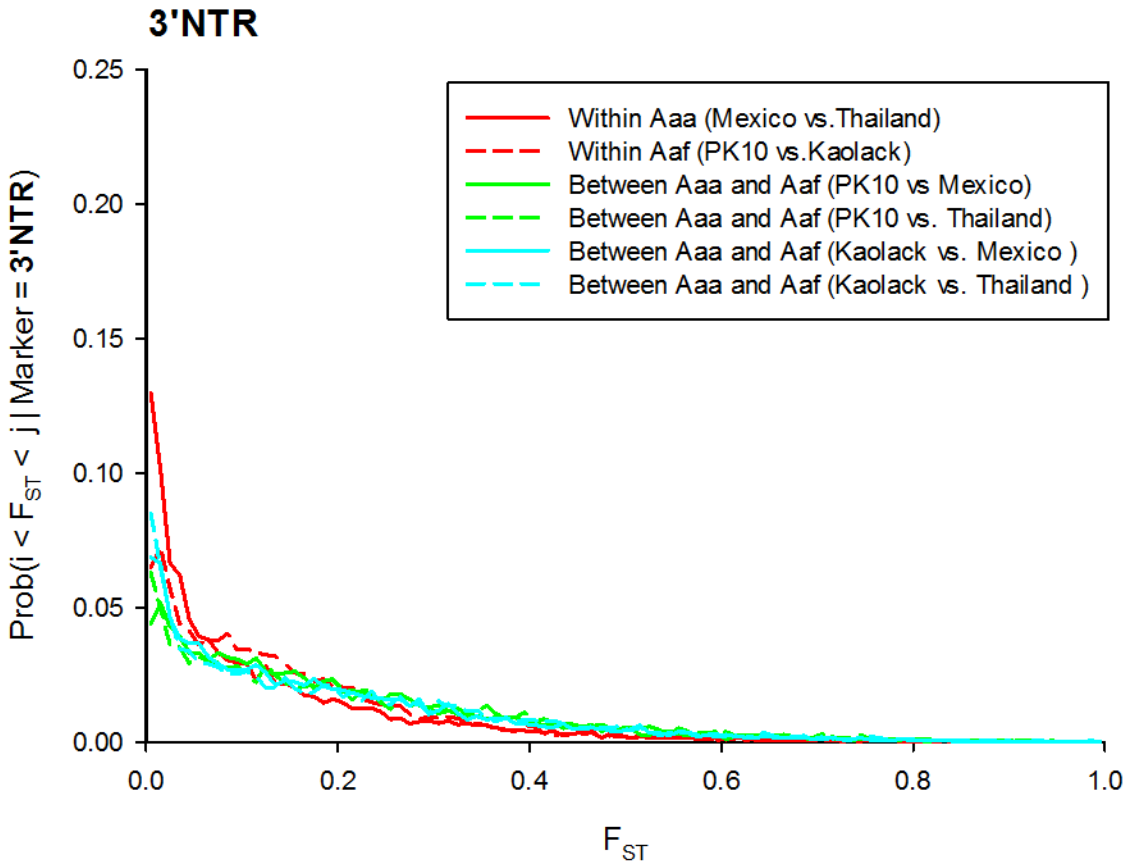


Figure 2.15: Probability distribution of F_{ST} values in 3'NTR
The probability distribution of F_{ST} values was calculated using Bayes' rule. The probability of F_{ST} values in intervals from i to j in increments of 0.01 given that the alternate allele was in the 3' nontranscribed region 600 basepairs downstream of the 3'UTR for each of the 6 pairwise comparisons. Within subspecies comparisons are in red. Between subspecies comparisons are in blue and green.

To confirm that F_{ST} values at the extreme ends of the distribution are not the result of a single or a few SNPs, the number of SNPs per gene and the average coverage per SNP was determined. Genes representing the 10 highest and the 10 lowest F_{ST} values in each of the six comparisons are shown in Table 2.3 as representative examples. The F_{ST} values in genes at either extreme of the F_{ST} distribution are a result of multiple SNPs (Table 2.3) that are distributed across the gene (Figure 2.16).

Table 2.3: Average coverage and # of SNPs in genes with the 10 highest and 10 lowest F_{ST} values. Gene annotations: hypothetical protein (HP), conserved hypothetical protein (CHP).

	Low F_{ST}					High F_{ST}				
	Accession Number	Gene	F_{ST}	Average Coverage /gene	#SNPs/ gene	Accession Number	Gene	F_{ST}	Average Coverage /gene	#SNPs/ gene
Mexico vs. Thailand	AAEL002140	transcription initiation factor IIA	0.000	260.89	293	AAEL005117	HP	0.865	293.93	822
	AAEL017618	28S_rRNA	0.000	127.47	93	AAEL007947	glutathione transferase	0.866	912.32	531
	AAEL008999	adult cuticle protein	0.000	414.42	243	AAEL013658	CHP	0.868	327.20	529
	AAEL010196	trypsin	0.001	704.80	547	AAEL002005	cytochrome P450	0.875	346.30	1318
	AAEL015628	glycine dehydrogenase	0.001	138.46	296	AAEL007336	ubiquitin-conjugating enzyme E2 q	0.876	256.89	212
	AAEL001620	CHP	0.001	216.50	129	AAEL013516	pupal cuticle protein	0.877	442.70	127
	AAEL002198	CHP	0.001	228.69	201	AAEL007962	glutathione transferase	0.885	983.50	931
	AAEL005193	HP	0.001	281.73	711	AAEL007948	glutathione transferase	0.913	1281.78	551
	AAEL002381	CHP	0.001	146.22	123	AAEL013696	CHP	0.914	291.39	195
	AAEL015376	CHP	0.002	418.19	280	AAEL007951	glutathione transferase	0.934	1411.71	1118
PK10 vs. Kaolack	Accession Number	Gene	F_{ST}	Average Coverage /gene	#SNPs/ gene	Accession Number	Gene	F_{ST}	Average Coverage /gene	#SNPs/ gene
	AAEL016404	tRNA-Val	0.002	102.23	74	AAEL012198	ADP,ATP carrier protein	0.735	395.43	292
	AAEL004446	HP	0.003	362.20	1491	AAEL002905	CHP	0.738	1171.79	766
	AAEL017670	no meta data	0.003	447.34	215	AAEL013846	HP	0.739	830.16	586
	AAEL007323	deoxyuridine 5'-triphosphate nucleotido-hydrolase	0.003	252.08	99	AAEL017055	no meta data	0.742	395.60	523
	AAEL015487	zinc finger protein, putative	0.003	642.53	874	AAEL007327	cornichon protein	0.746	1447.52	403
	AAEL016654	tRNA-Pro	0.004	112.16	74	AAEL011580	CHP	0.757	489.65	571
	AAEL017962	5.8S_rRNA	0.004	114.46	22	AAEL006661	histone acetyl-transferase	0.781	592.78	403
	AAEL015499	odorant binding protein OBP60	0.004	185.15	148	AAEL004203	CHP	0.801	464.07	660
	AAEL017413	no meta data	0.004	1665.47	181	AAEL012570	HP	0.811	83.73	37
	AAEL016401	tRNA-Pro	0.004	106.40	68	AAEL015481	neuronal calcium sensor	0.848	200.54	224

Table 2.3 Continued: Gene annotations: hypothetical protein (HP), conserved hypothetical protein (CHP).

	Low F_{ST}					High F_{ST}				
	Accession Number	Gene	F_{ST}	Average Coverage /gene	#SNPs/ gene	Accession Number	Gene	F_{ST}	Average Coverage /gene	#SNPs/ gene
Kaolack vs. Mexico	AAEL016160	tRNA-Ala	0.000	71.00	4	AAEL001960	cytochrome P450	0.790	252.04	1820
	AAEL017551	no meta data	0.001	94.54	78	AAEL001909	ribosomal protein	0.795	427.03	449
	AAEL002993	39S Ribosomal Protein L43	0.001	1376.79	631	AAEL008264	CHP	0.807	268.89	1076
	AAEL017926	no meta data	0.002	185.45	119	AAEL017489	no meta data	0.810	168.54	277
	AAEL016827	tRNA	0.003	94.67	3	AAEL013687	CHP	0.813	333.87	227
	AAEL015449	ribosome biogenesis protein brix	0.003	303.58	223	AAEL002005	cytochrome P450	0.853	336.15	1696
	AAEL016937	tRNA-Ala	0.004	69.25	4	AAEL012638	CHP	0.862	209.69	227
	AAEL000482	histone H3	0.004	183.48	105	AAEL002031	cytochrome P450	0.865	512.25	2160
	AAEL009684	HP	0.005	229.33	509	AAEL017006	no meta data	0.890	187.08	590
	AAEL017413	no meta data	0.005	1359.56	181	AAEL007336	ubiquitin-conjugating enzyme E2 q	0.955	609.56	212
Kaolack vs. Thailand	Accession Number	Gene	F_{ST}	Average Coverage /gene	#SNPs/ gene	Accession Number	Gene	F_{ST}	Average Coverage /gene	#SNPs/ gene
	AAEL008776	CHP	0.001	219.13	183	AAEL014049	predicted protein	0.844	514.96	1112
	AAEL017551	no meta data	0.001	95.41	39	AAEL008605	inosine triphosphatase	0.853	290.74	1057
	AAEL009699	HP	0.002	490.51	379	AAEL016997	GPCR Gastrin/ Bombesin Family	0.861	710.86	471
	AAEL017554	no meta data	0.002	184.38	314	AAEL007951	glutathione transferase	0.865	1195.64	964
	AAEL017784	18S_rRNA	0.002	792.48	131	AAEL017489	no meta data	0.868	371.83	282
	AAEL017057	no meta data	0.003	587.13	609	AAEL017006	no meta data	0.876	183.06	500
	AAEL017670	no meta data	0.003	336.65	189	AAEL015303	HP	0.892	379.93	259
	AAEL017030	no meta data	0.004	378.02	695	AAEL013516	pupal cuticle protein	0.900	612.38	127
	AAEL017413	no meta data	0.004	1005.67	181	AAEL005117	HP	0.911	413.09	822
	AAEL009683	HP	0.005	526.92	448	AAEL013696	CHP	0.928	336.15	289

Table 2.3 Continued: Gene annotations: hypothetical protein (HP), conserved hypothetical protein (CHP).

	Low F_{ST}					High F_{ST}				
	Accession Number	Gene	F_{ST}	Average Coverage /gene	#SNPs/ gene	Accession Number	Gene	F_{ST}	Average Coverage /gene	#SNPs/ gene
PK10 vs. Mexico	AAEL015628	glycine dehydrogenase	0.000	186.25	295	AAEL011545	cxyorf1	0.778	570.13	327
	AAEL017448	no meta data	0.001	527.47	171	AAEL017489	no meta data	0.788	178.48	101
	AAEL017354	no meta data	0.002	242.01	181	AAEL011972	actin binding protein	0.795	528.36	891
	AAEL017413	no meta data	0.002	1576.65	321	AAEL008264	CHP	0.800	295.35	1107
	AAEL009071	CHP	0.004	496.96	277	AAEL008967	adult cuticle protein	0.830	223.24	89
	AAEL017585	28S_rRNA	0.005	212.65	77	AAEL001911	CHP	0.847	495.07	565
	AAEL014223	vitellogenin receptor (VgR)	0.005	112.91	78	AAEL015481	neuronal calcium sensor	0.854	231.06	236
	AAEL017670	no meta data	0.005	375.46	186	AAEL007336	ubiquitin-conjugating enzyme E2 q	0.881	585.59	212
	AAEL008949	HP	0.005	258.97	299	AAEL009849	HP	0.883	84.94	17
	AAEL017730	no meta data	0.005	263.09	142	AAEL014817	serine-pyruvate amino-transferase	0.904	84.50	56
PK10 vs. Thailand	Accession Number	Gene	F_{ST}	Average Coverage /gene	#SNPs/ gene	Accession Number	Gene	F_{ST}	Average Coverage /gene	#SNPs/ gene
	AAEL017784	18S_rRNA	0.001	412.29	131	AAEL016997	GPCR Gastrin/ Bombesin Family	0.807	835.00	474
	AAEL015302	CHP	0.002	133.82	62	AAEL001672	multicopper oxidase	0.808	671.85	2850
	AAEL015521	CHP	0.002	266.47	315	AAEL017937	no meta data	0.824	265.49	96
	AAEL006785	60S ribosomal protein L18a	0.003	83.71	14	AAEL011793	aspartyl beta-hydroxylase	0.827	378.22	344
	AAEL010394	HP	0.003	199.86	95	AAEL017123	odorant receptor	0.838	517.91	1496
	AAEL008073	CHP	0.003	497.22	1372	AAEL017489	no meta data	0.841	386.98	104
	AAEL011618	CHP	0.003	140.33	9	AAEL017429	no meta data	0.853	277.35	429
	AAEL017683	28S_rRNA	0.003	438.72	129	AAEL004218	odorant receptor	0.858	319.67	1040
	AAEL000493	heparan sulphate 6-o-sulfotransferase	0.004	543.52	406	AAEL017221	odorant receptor	0.860	351.89	1317
	AAEL013865	CHP	0.004	239.77	328	AAEL001911	CHP	0.865	581.41	565

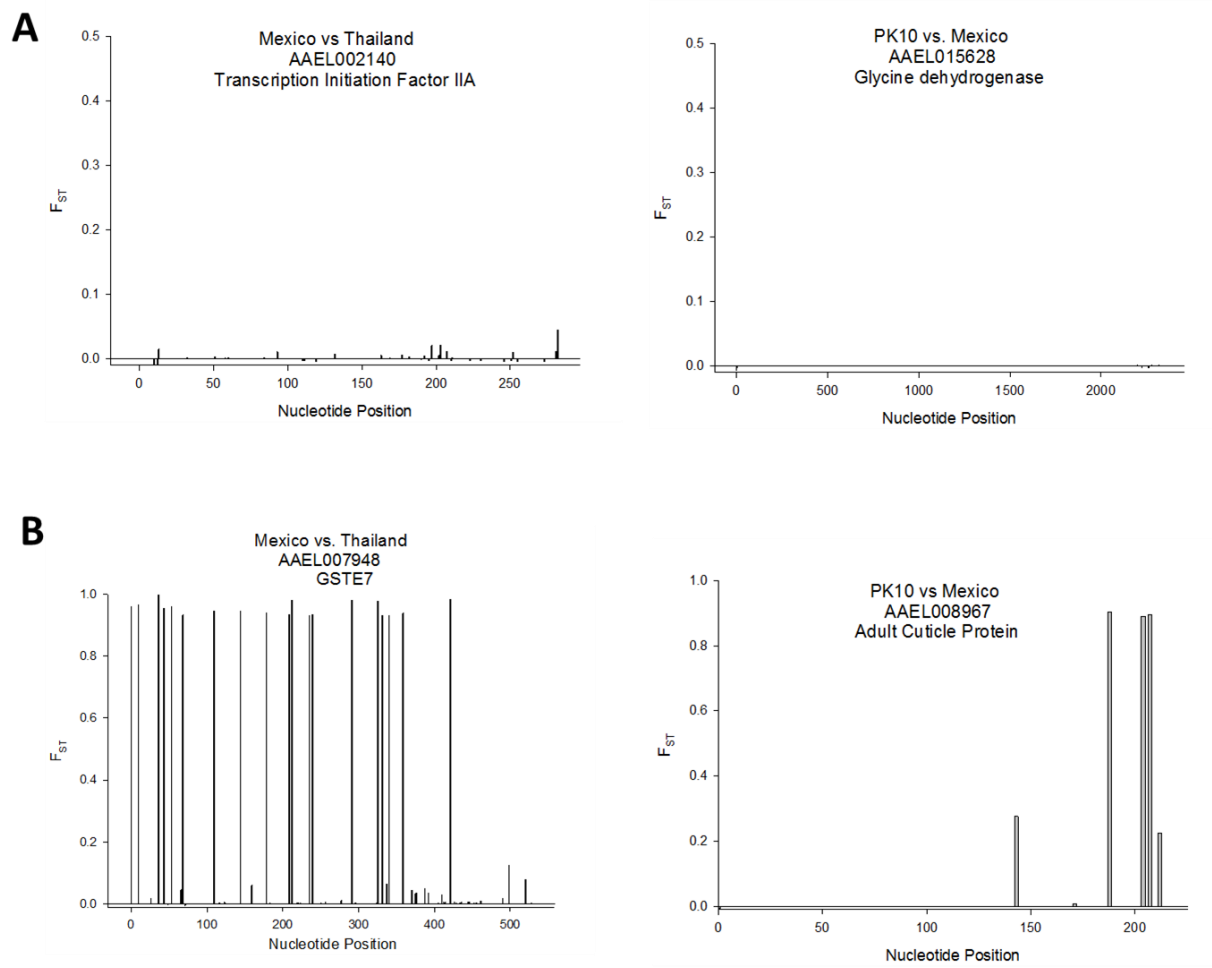


Figure 2.16: Distribution of F_{ST} values across a gene
F_{ST} values at individual nucleotides were mapped across 2 genes from the Mexico vs Thailand comparison (AAEL002140 and AAEL007948) and 2 genes from the PK10 vs Mexico comparison (AAEL015628 and AAEL008967) with low (A) and high (B) overall F_{ST} values across the gene.

Differences based on substitution type

When a polymorphic site was in an InDel, there was a high probability ($p = 0.15 – 0.30$) of very low F_{ST} values and a low probability ($p < 0.01$) of high F_{ST} values (Figure 2.17). The shape of the distribution between the population comparisons was very similar indicating SNPs in InDels are not an informative marker of subspecies population structure. In contrast, when the polymorphism resulted in a transition or transversion, there was a much lower probability (p

= 0.01 – 0.05) of low F_{ST} values (Figure 2.18 and 2.19). Similar to polymorphisms in the third codon position, the distribution of F_{ST} values was broad for polymorphisms resulting in transitions or transversions. The probability distributions of F_{ST} values from polymorphic sites resulting in transitions or transversions show differences in the various population comparisons indicating that transitions or transversions could be informative markers of subspecies population structure. Specifically, the distribution from the within Aaa comparison is shifted towards lower F_{ST} values, while the distribution of the within Aaf comparison is shifted towards higher F_{ST} values and between Aaa and Aaf distributions are shifted towards even higher F_{ST} values.

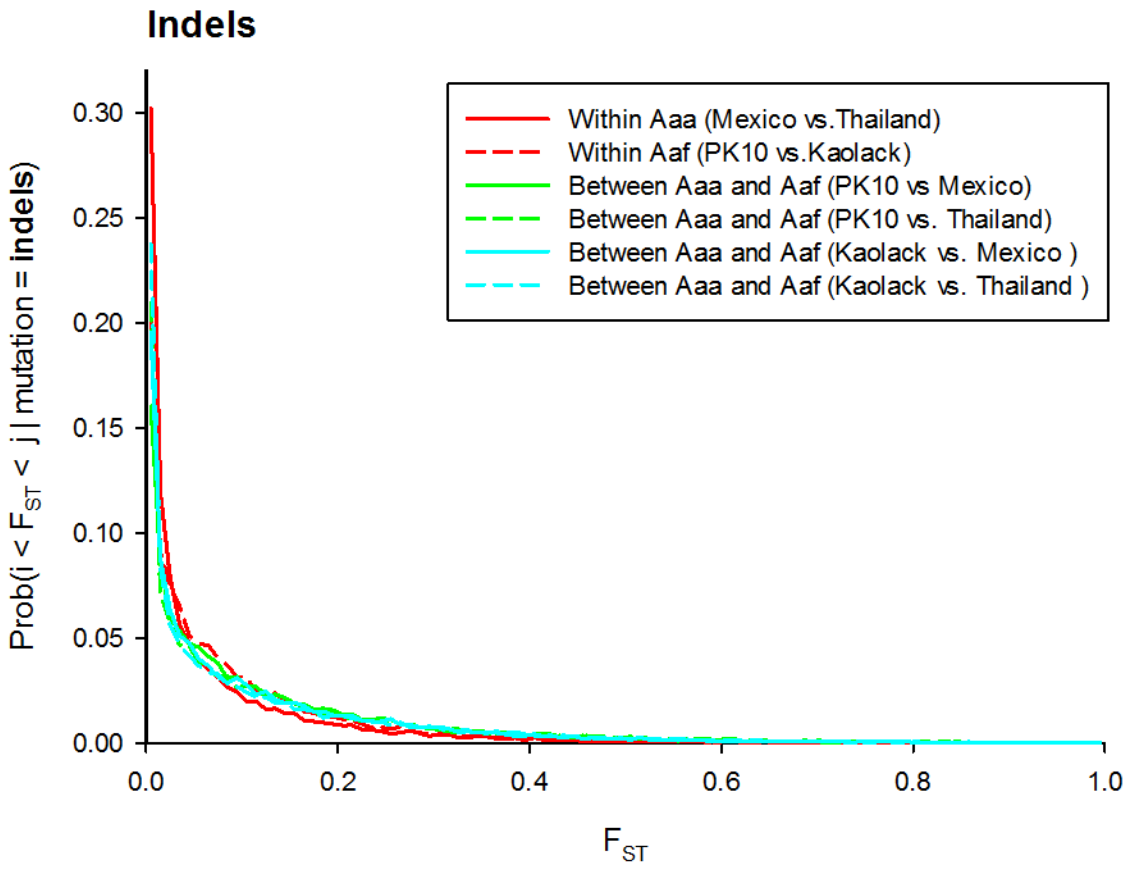


Figure 2.17: Probability distribution of F_{ST} values in InDel substitutions
The probability distribution of F_{ST} values was calculated using Bayes' rule. The probability of F_{ST} values in intervals from i to j in increments of 0.01 given that the alternate allele was an insertion or deletion for each of the 6 pairwise comparisons. Within subspecies comparisons are in red. Between subspecies comparisons are in blue and green.

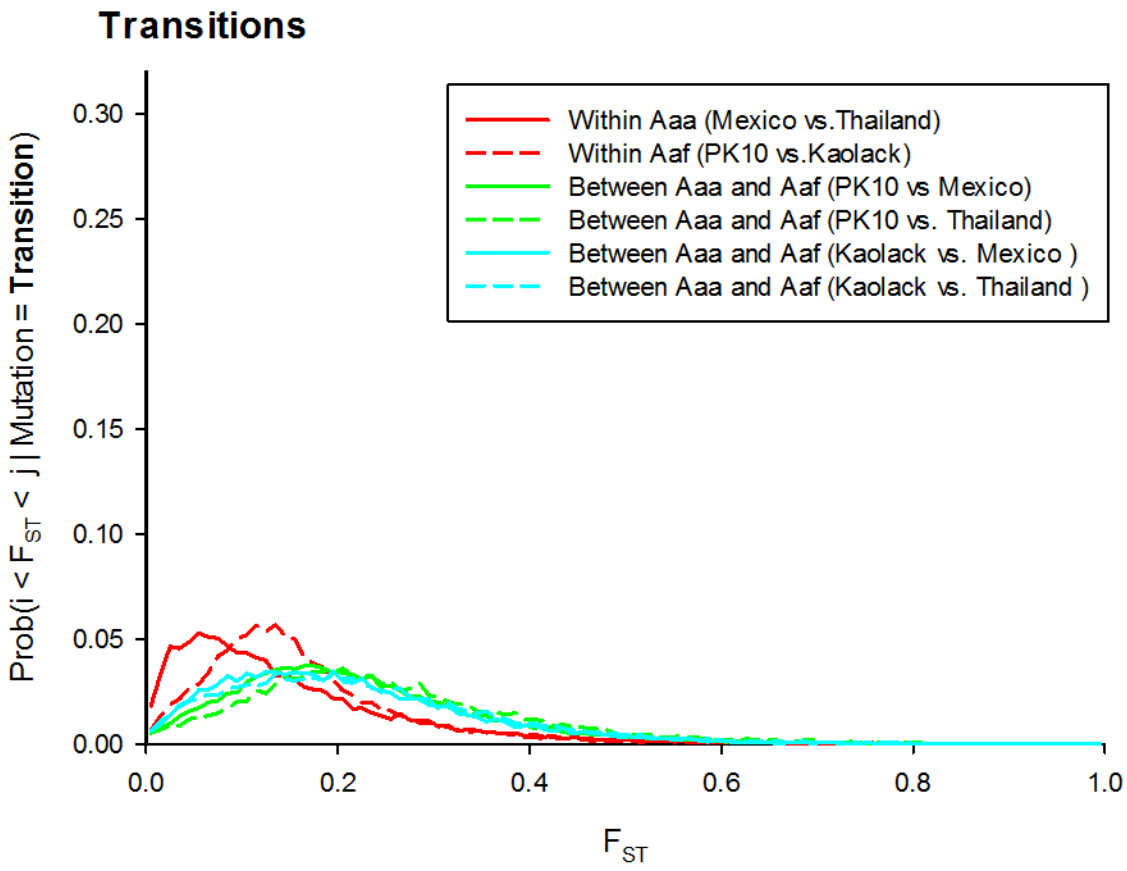


Figure 2.18: Probability distribution of F_{ST} values in transition substitutions

The probability distribution of F_{ST} values was calculated using Bayes' rule. The probability of F_{ST} values in intervals from i to j in increments of 0.01 given that the alternate allele resulted in a transition substitution for each of the 6 pairwise comparisons. Within subspecies comparisons are in red. Between subspecies comparisons are in blue and green.

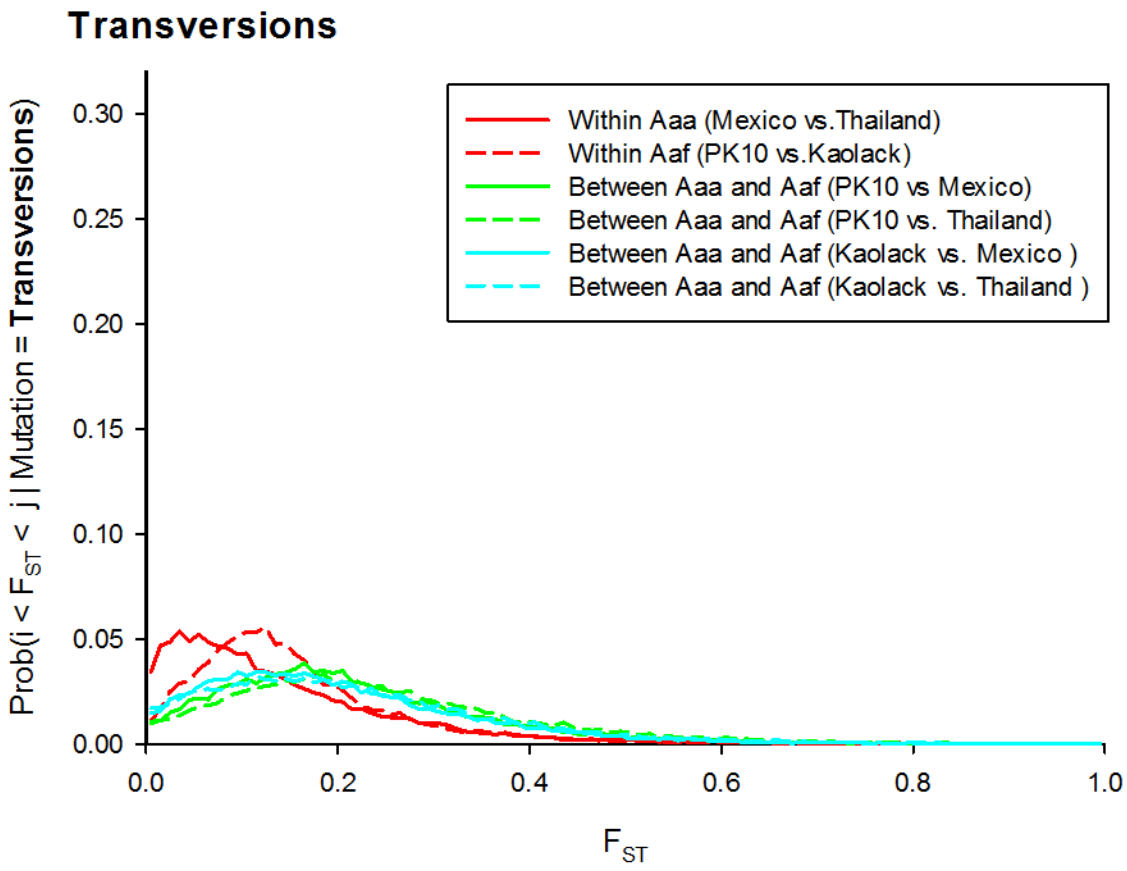


Figure 2.19: Probability distribution of F_{ST} values in transversion substitutions
The probability distribution of F_{ST} values was calculated using Bayes' rule. The probability of F_{ST} values in intervals from i to j in increments of 0.01 given that the alternate allele resulted in a transversion substitution for each of the 6 pairwise comparisons. Within subspecies comparisons are in red. Between subspecies comparisons are in blue and green.

When a SNP involved a replacement substitution, there was a high probability (p =0.12 – 0.18) of very low F_{ST} values and a low probability ($p < 0.01$) of high F_{ST} values (Figure 2.20). The shape of the distribution among the population comparisons were very similar indicating SNPs resulting in replacement substitutions are not informative markers of subspecies population structure. In contrast, when the SNP resulted in silent substitution, there was a much lower probability (p = 0.01 – 0.05) of low F_{ST} values (Figure 2.21). Similar to polymorphisms in codon position 3, transitions, and transversions, the distribution of F_{ST} values of SNPs resulting

in silent substitutions was broad. The probability distributions of F_{ST} values from polymorphic sites resulting in silent substitutions show differences in the various population comparisons indicating that SNPs resulting in silent substitutions could be informative markers of subspecies.

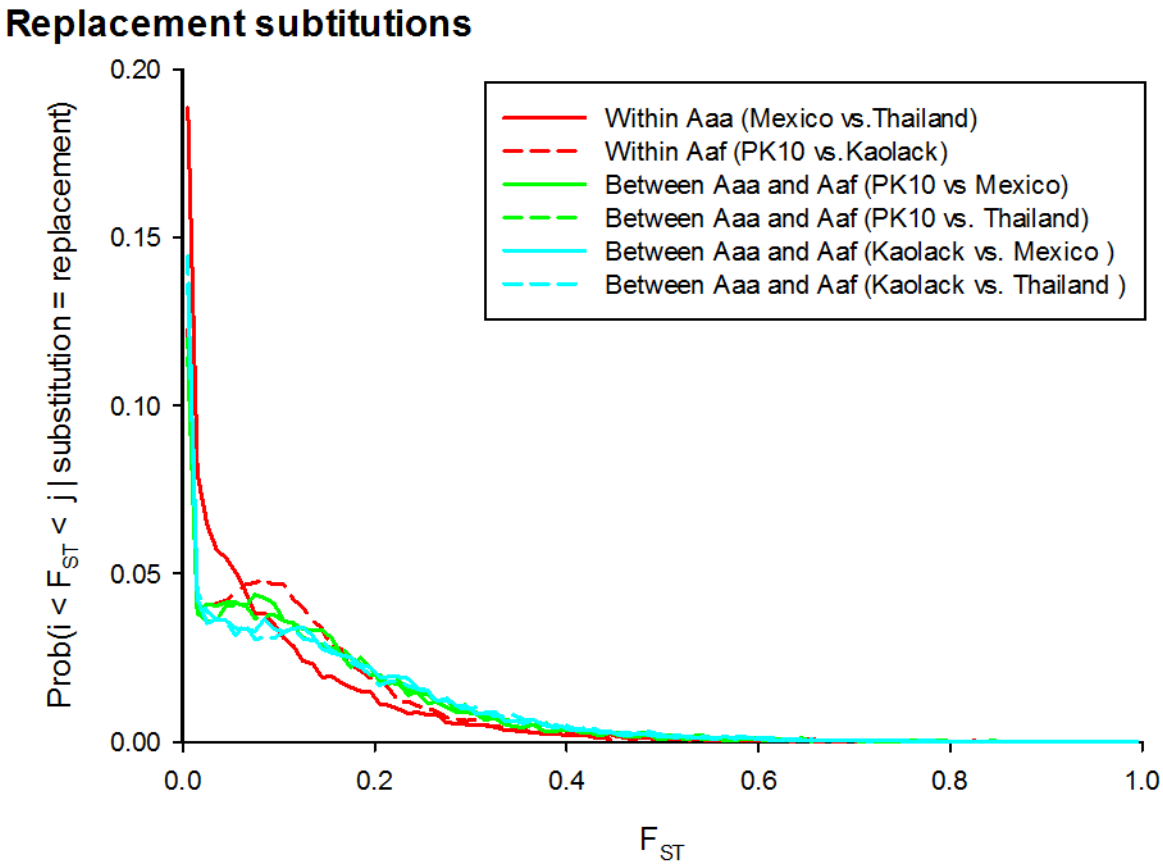


Figure 2.20: Probability distribution of F_{ST} values in replacement substitutions
The probability distribution of F_{ST} values was calculated using Bayes' rule. The probability of F_{ST} values in intervals from i to j in increments of 0.01 given that the alternate allele resulted in a replacement amino acid substitution for each of the 6 pairwise comparisons. Within subspecies comparisons are in red. Between subspecies comparisons are in blue and green.

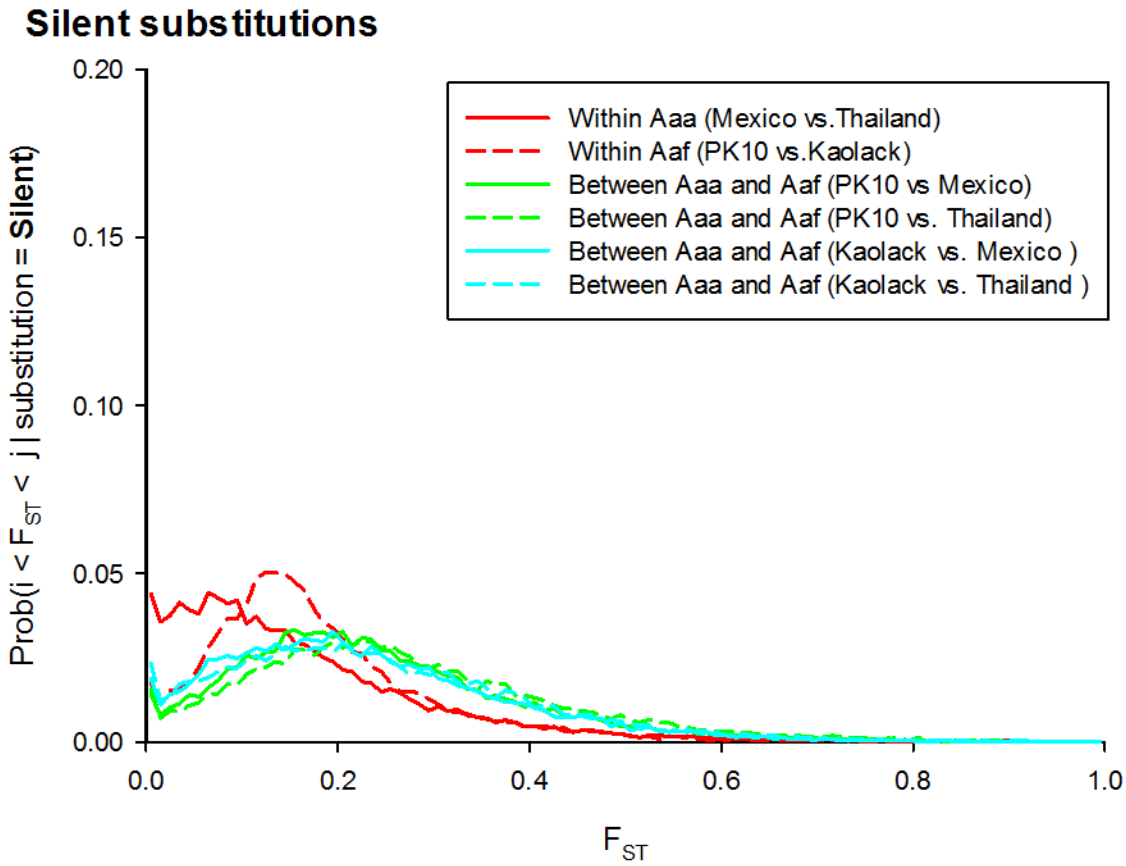


Figure 2.21: Probability distribution of F_{ST} values in silent substitutions
The probability distribution of F_{ST} values was calculated using Bayes' rule. The probability of F_{ST} values in intervals from i to j in increments of 0.01 given that the alternate allele resulted in a silent amino acid substitution for each of the 6 pairwise comparisons. Within subspecies comparisons are in red. Between subspecies comparisons are in blue and green.

F_{ST} values across the chromosomes

During the process of speciation, chromosomal regions with high F_{ST} values tend to spread to nearby regions through a process called divergent hitchhiking [395]. F_{ST} values were plotted against chromosome position along all three chromosomes to test for chromosomal regions of divergence. Although peaks and troughs in F_{ST} values are observed, no obvious regions of divergence were detected on chromosome 1(Figure 2.22), chromosome 2 (Figure 2.23), or chromosome 3 (Figure 2.24).

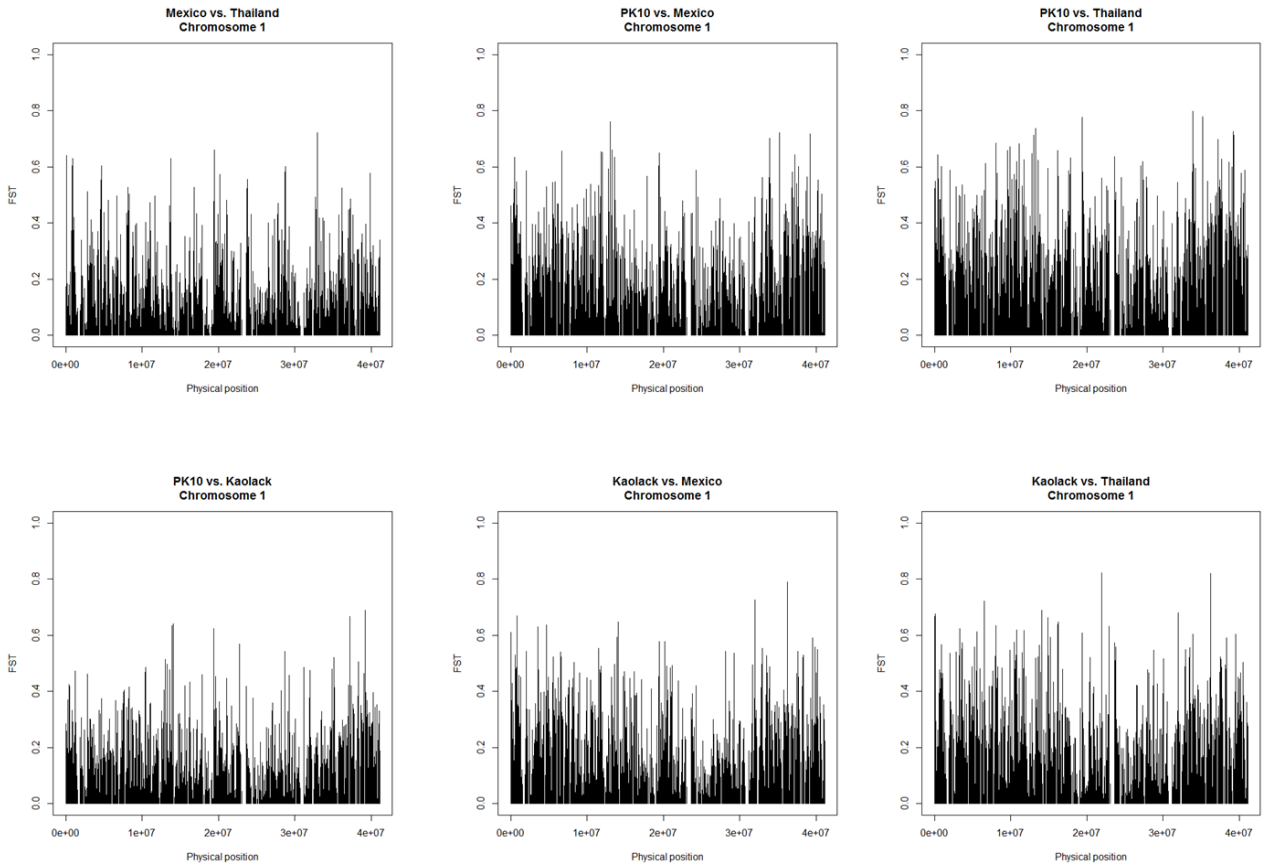


Figure 2.22: F_{ST} Values Across Chromosome I

F_{ST} values per gene were plotted against the average nucleotide position of the gene across chromosome I.

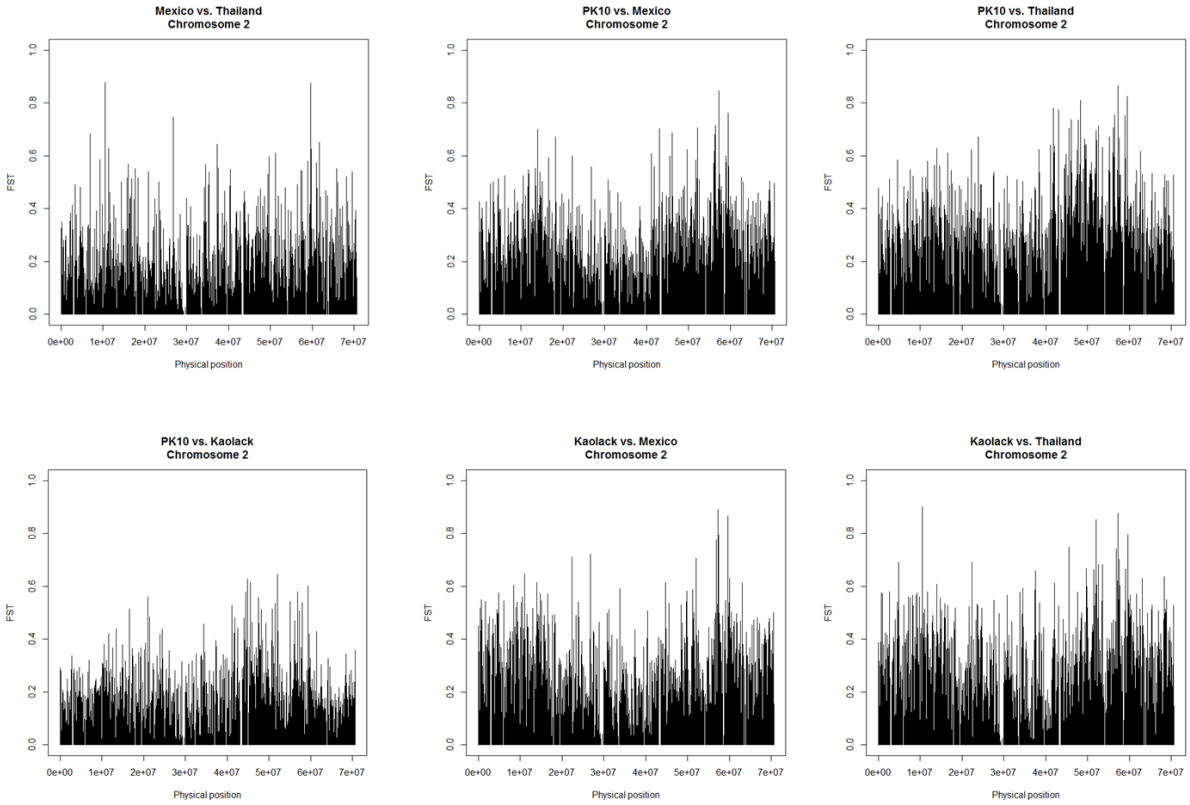


Figure 2.23: F_{ST} Values Across Chromosome II

F_{ST} values per gene were plotted against the average nucleotide position of the gene across chromosome II.

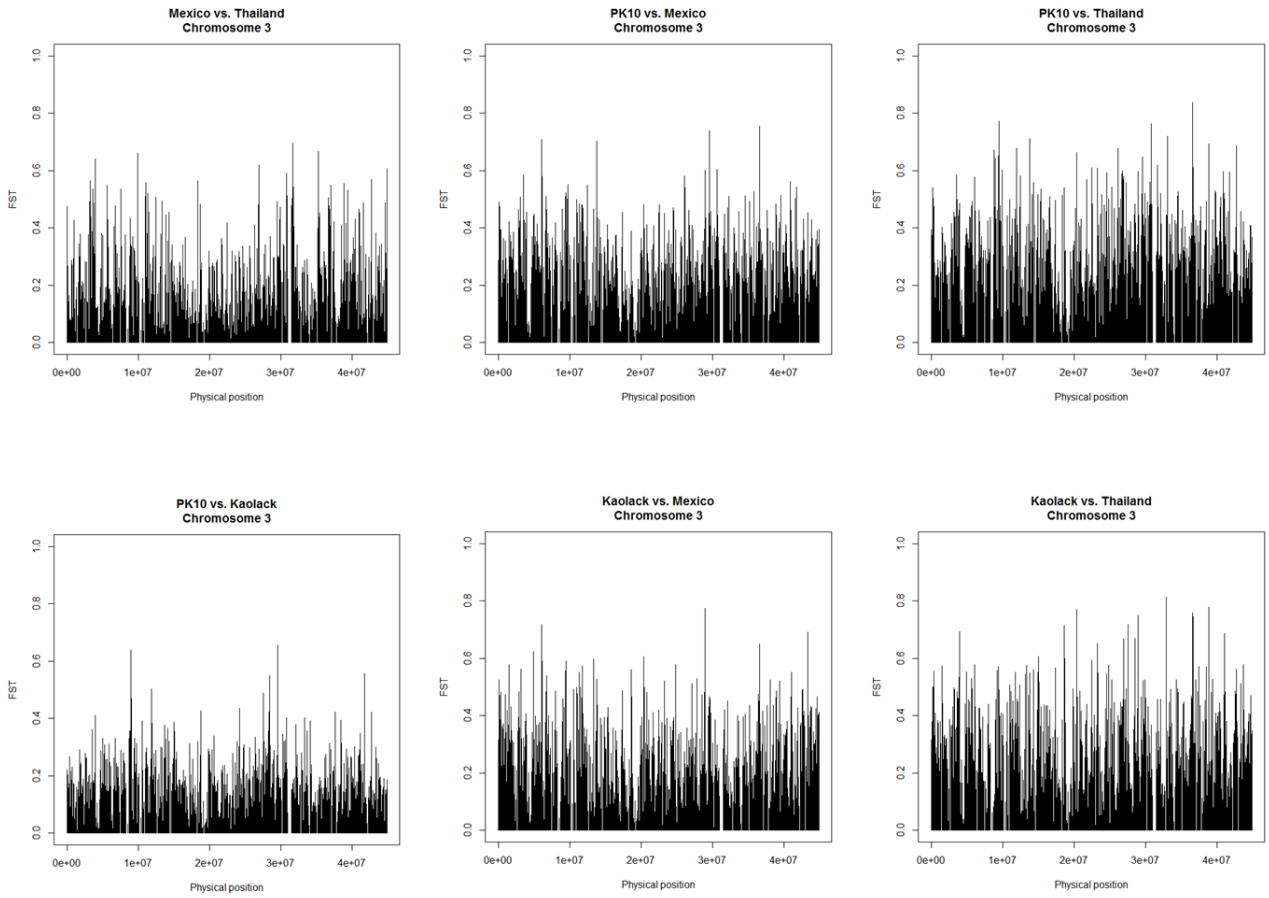


Figure 2.24: F_{ST} Values Across Chromosome III
F_{ST} values per gene were plotted against the average nucleotide position of the gene across chromosome III.

Classification of Genes

Using the beta distributions, we sought to determine what percentage of genes fell into 12 different categories defined by F_{ST}, within population diversity (b_s, eq.2), and the ratio of replacement to synonymous amino acid changes (ω) to gain insight into the mechanism and types of genes involved in genomic speciation. The 12 categories, along with the percent of genes in each comparison in each category, are presented in Table 2.4. The 12 categories consist of combinations of a F_{ST} value above or below the mean, a value of b_s above or below the mean, and a value of ω greater than, less than, or equal to one. Unlike the distributions of F_{ST} and b_s,

the values for ω fit a gamma distribution. Values of ω that were in the top 95% of the distribution were given the classification of $\omega > 1$. Values of ω that were in the bottom 5% were give the classification of $\omega < 1$. Most of the ω values were centered on 1 (data not shown).

Table 2.4: Twelve categories of predicted divergence mechanisms

Genes were grouped based on whether their F_{ST} value and within population diversity was above or below the mean and whether the ratio of replacement to synonymous mutations was above the 95% interval or below the 0.05% interval. Predictions were made about the relevant genetic forces giving rise to each category and the percentage of genes in each population comparison that fit that category

					Aaa vs Aaa	**Aaa vs Aaf**	**Aaa vs Aaa**	**Aaa vs Aaa**	**Aaf vs Aaa**	**Aaf vs Aaa**
Category					Mex_Thai	Kao_PK10	Kao_Mex	Kao_Thai	PK10_Mex	PK10_Thai
1	FST > Mean	Hexp > Mean	ω < 1	opposing forces, expected to be rare	0.2%	0.3%	0.3%	0.3%	0.4%	0.3%
2	FST > Mean	Hexp > Mean	ω = 1	Genetic drift	13.0%	24.1%	28.0%	27.6%	43.2%	37.6%
3	FST > Mean	Hexp > Mean	ω >1	Local adaptation	0.3%	0.5%	0.5%	0.4%	0.5%	0.6%
4	FST > Mean	Hexp < Mean	ω < 1	Genetic drift and population bottlenecks	0.1%	0.1%	0.2%	0.1%	0.2%	0.2%
5	FST > Mean	Hexp < Mean	ω = 1	Genetic drift and population bottlenecks	9.5%	18.2%	17.2%	18.1%	21.8%	34.4%
6	FST > Mean	Hexp < Mean	ω >1	Local selective sweeps	0.4%	0.7%	0.6%	0.6%	0.7%	1.2%
7	FST < Mean	Hexp > Mean	ω < 1	Similarity maintained by selection	0.4%	0.7%	0.4%	0.4%	0.3%	0.3%
8	FST < Mean	Hexp > Mean	ω = 1	Similarity maintained by migration	24.5%	16.5%	13.0%	14.8%	9.1%	4.9%
9	FST < Mean	Hexp > Mean	ω >1	Shared adaptation or migration	0.6%	0.5%	0.4%	0.4%	0.4%	0.2%
10	FST < Mean	Hexp < Mean	ω < 1	Genes strongly constrained by purifying selection in both collections	0.8%	1.0%	1.0%	0.8%	0.7%	0.6%
11	FST < Mean	Hexp < Mean	ω = 1	Nucleotide evolution is constrained	47.8%	35.4%	36.2%	34.4%	21.0%	18.0%
12	FST < Mean	Hexp < Mean	ω >1	Parallel directional selection, low diversity arising from selective sweeps	2.5%	2.0%	2.2%	2.0%	1.7%	1.6%
					100%	100%	100%	100%	100%	100%

Predictions were made about what type of force or selective force would result in each of the 12 categories (Table 2.4). For example, category one is defined by opposing forces and predicted to be rare. Category three is made up of genes with a F_{ST} value > the mean (allele frequencies different between the two populations), a b_s value > the mean (large nucleotide diversity within each population), and $\omega > 1$ (diversifying selection). Taken together, this pattern suggests genes that fit these criteria are involved in local adaptation and could be involved in speciation.

Comparing the percentage of genes that fell into each category provided insight into the type of divergence and the forces giving rise to the divergence in the six different population comparisons. Most of the divergent genes between populations were predicted to have arisen through genetic drift (categories 2 and 5), lack of migration (category 8), and constrained nucleotide evolution (category 11). Far fewer genes appear to be under selection regardless of being diversifying selection (categories 3, 6, 9, 12) or purifying selection (categories 7, 10). In categories 2, 3, 5, and 6, where F_{ST} is greater than the mean, the percentage of genes in the PK10_Kao comparison were more similar to the between subspecies comparisons (Aaa vs. Aaf) than the within subspecies comparison (Mexico vs Thailand) further suggesting that *Ae. aegypti* in Senegal does not represent a homogeneous subspecies.

Discussion

In an effort to identify an informative approach to study the process of subspeciation in *Ae. aegypti*, F_{ST} values from genes, gene regions, and substitution types across the entire genome were fit to a beta distribution to allow for comparisons of natural populations from Thailand, Mexico, and two from Senegal using a deep sequencing approach. Small differences in allele frequencies across genes were detected between populations. However, large differences in

allele frequencies between populations became apparent when gene regions and substitution types were looked at individually. Major conclusions from this study include 1) measuring F_{ST} differences across the whole genome instead of a limited number of markers can be more informative, 2) comparing the distribution of F_{ST} values instead of point estimates of F_{ST} provides more information about population differences as well as provides a tool for identifying highly conserved or divergent genes between populations, and 3) the distribution of F_{ST} values changes based on gene region and substitution type, indicating that some regions of the genome are not as informative as other regions.

Overall, polymorphisms in the third codon position and polymorphisms that resulted in transitions, transversions, and silent amino acid substitutions showed large differences between populations, but highly conserved regions did not show resolution of the population structure. No obvious regions in the three chromosomes were identified that were highly conserved or divergent between the populations. We were able to use the beta distributions of F_{ST} values to identify genes that have undergone genetic drift, migration, and selection. Overall, these data show that comparing beta distributions of F_{ST} values from genes, gene regions, and substitution types between populations can provide insight into genome regions associated with allopatric and sympatric speciation within and between populations of *Ae. aegypti*.

Differences in F_{ST} values can be explained by multiple forces. Low F_{ST} values can result from purifying selection or balancing selection in the entire population, migration and mutation, or parallel or convergent selection and adaptation among subpopulations High F_{ST} values can result from a catastrophic reduction in subpopulation size, local directional selection and adaptation, and mating systems that promote inbreeding or forced inbreeding because of spatial clustering around a limited resource. Highly conserved regions of the gene that are under

purifying selection such as 5'NTR, 5'UTR, 1st codon position, 2nd codon position, 3'UTR, and 3'NTR showed low F_{ST} values among all population comparisons indicating alleles in these regions are not involved in speciation. The allele conservation in these regions can be explained by the fact that nucleotide sequences upstream of the coding region, such as the 5'NTR and 5'UTR, and downstream of the coding region, such as the 3' NTR and 3' UTR, contain important regulatory regions involved in regulation of gene expression. Also, changes to the nucleotide in the first or second codon position are likely to change the amino acid, while changes to the nucleotide in the third codon position are likely to have no effect on the amino acid. In this study, the low F_{ST} values observed in the nucleotides in the first and second codon position demonstrate the conservation of amino acid sequence between populations. Nucleotide changes in the third codon position and nucleotide changes that resulted in a silent substitution, transitions, or transversions (non-conserved regions) had the largest F_{ST} values and provided resolution of subspecies suggesting that measuring differences in these types of nucleotide changes could be the most informative markers to study speciation of *Ae. aegypti*.

When non-conserved regions were used as markers, the within Aaa comparison was different than the within Aaf comparison and both were different than the between Aaa and Aaf comparisons, indicating an intermediate subspeciation event is occurring within Senegal. Alternatively, the difference between Kaolack and PK10 is greater than the difference between Mexico and Thailand, but not as different as the between Aaa and Aaf comparisons. One would expect the distribution of F_{ST} values in the PK10 vs Kaolack comparison to be similar to the distribution of the F_{ST} values in the Mexico vs Thailand comparison since PK10 and Kaolack are close geographically and represent a monophyletic group in previous studies [135, 146] using isozyme or microsatellite markers. This is not what we found by comparing the distribution of

F_{ST} values across the entire genome. Instead it appears that PK10 and Kaolack represent two population structures that are different from a third population structure specific to Mexico or Thailand.

A problem in molecular evolution is elucidating the contribution of genetic drift and natural selection to patterns of genetic variation. According to the neutral theory of evolution, most mutations are neutral or deleterious and changes in allele frequency are the result of genetic drift [396-398]. This was seen in this study as well, where approximately 94% of genes fell into a category of neutral amino acid changes and/or genetic drift (Table 2.4). This problem can be avoided by measuring differences across the whole genome instead of a limited number of markers where genes that are involved in adaptation associated with speciation events would show signatures of positive selection [399]. In this study, the categories predicted to be involved in local adaptation were categories 3 and 6. Genes in both these categories had F_{ST} values greater than the mean and $\omega > 1$ indicating positive selection at the amino acid level. The Mexico vs Thailand comparison had fewer genes in categories 3 and 6 than in any other comparison suggesting more genes are under selection between Aaf and Aaa and within Aaf, compared to within Aaa.

Higher mean F_{ST} values and signatures of positive selection in the PK10 vs Kaolack comparison as well as the between Aaa and Aaf comparisons suggest sympatric speciation within Senegal and allopatric speciation between *Ae. aegypti* from Senegal and global populations. Darwin first proposed that new species could arise without geographic isolation [185] resulting in years of debate and idea of "divergence with gene flow" [186, 187]. It is now accepted that speciation can occur through allopatric speciation (geographic isolation and no gene exchange between populations), parapatric speciation (populations can exchange genes to a

limited extent), or sympatric speciation (populations can exchange genes freely). Allopatric and parapatric speciation are the obvious and most accepted ways in which speciation can occur. Physical distance between the two populations, or barriers to gene flow, and therefore reduced exchange of genetic material and the resulting genetic drift are bound to result in reduced gene flow. The idea that sympatric speciation, or divergence with gene flow, is an important cause of speciation is relatively recent. There are two different models of sympatric speciation: the continuous distribution of resources model and the discrete-habitat model [180]. In the continuous distribution of resources model, coevolution of ecological traits and assortative mating divides a population into two sympatric groups that use different parts of the resource distribution [180]. An example of assortative mating contributing species divergence is the stickleback fish in Iceland [400]. In the discrete-habitat model, sympatric species seek out different niches and mate with their own species, such as *Rhagoletis* flies that have a preference for host plant A or host plant B on which they mate [401]. This requires the evolution of three behaviors: niche preference, niche adaptation, and assortative mating. The current study does not provide evidence for either model in Senegal, but it is likely that the discrete- habitat model is involved in sympatric speciation of *Ae. aegypti* in West Africa. Kaolack is an urban environment, and mosquitoes here breed in artificial containers or tires and predominantly bite humans. PK10 is a sylvatic environment and these mosquitoes breed in treeholes or dried fruit husks and bite other animals besides humans.

Future directions for this project include providing more information on the function of the genes predicted to be undergoing positive selection as well looking more closely for islands of speciation and evidence of divergence hitchhiking. Although we did not detect any chromosomal regions with obvious signatures of divergent hitchhiking, scanning along the

chromosomes in finer scale might yield informative results. Evidence of chromosomal rearrangements, such as inversions, as well as evidence of assortative mating or discrete habitats between the two putative groups in

CHAPTER 3: DIVERGENT EVOLUTION OF THE SEX-DETERMINING LOCUS IN *AEDES AEGYPTI* SUBSPECIES

Introduction

Genes that determine sexual differentiation are highly variable among the fruitfly *Drosophila melanogaster*, the nematode *Caenorhabditis elegans*, and the yeast *Saccharomyces cerevisiae*, and mammals. Comparison of the four systems reveals divergent as well as conserved genetic elements [402]. In most *Diptera*, there is a primary signal, either genetic or environmental [403], that results in the differentiation of males and females; this signal initiates the process in the differential expression of a key effector gene. The effector gene, in turn, regulates expression of other regulatory and sex-determining genes, resulting in male or female phenotypes. A comparison of *Drosophila* and the housefly *Musca* spp. revealed differences in both the primary signal (X:A chromosome ratio in *Drosophila* and the M gene in *Musca*), as well as the effector gene (*sexlethal* in *Drosophila* and the F gene in *Musca*). These different effector genes both alter expression of *doublesex* (*dsx*), a transcription factor which controls male and female development [404] . Although there is great diversity in the upstream effector genes among insects, most converge on the regulation of the alternative splicing of *dsx* [402].

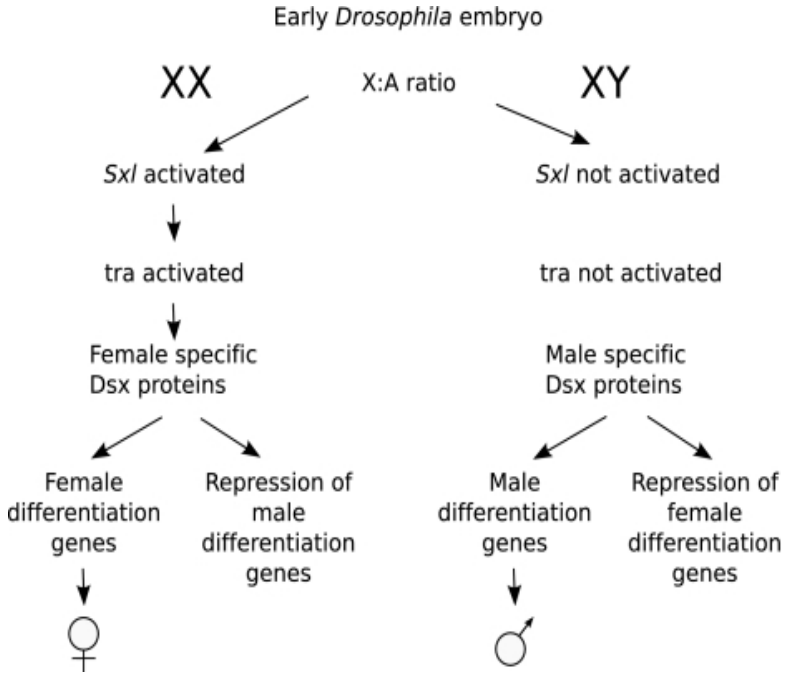


Figure 3.1: Sex-determination pathway in Drosophila.
The ratio of the expression of the X chromosome regulates the alternative splicing of *sexlethal* (*sxl*) which results in the active form of *sxl*. *Sxl* then regulates the alternative splicing of *transformer* (*tra*) resulting in the active form of *tra*. *Tra* then regulates the alternative splicing of the transcription factor, *doublesex* (*dsx*), which then regulates other genes involved in female development and represses genes involved in male development.

In most dipterans, *tra* is a key splicing regulator of the female *dsx* isoform [405-412]. Gene duplications and allelic variants in *tra* or *tra*-like genes have led to diverse mechanisms in the upstream regulation of *dsx* [406, 409, 413-415]. In *Drosophila*, *sexlethal*, a regulator of alternative splicing, controls expression of the transformer (*tra*) gene, which, in turn, is responsible for the alternative splicing of *dsx* mRNA into female- and male-specific isoforms [416-418] . The upstream regulators in *Drosophila*, *sexlethal* and *tra*, as well as the transcription factor, *dsx*, are expressed in a male- and female-dependent manner through alternative splicing of sex-specific isoforms (Figure 3.1).

In Diptera, sex determination mechanisms have been characterized in the more evolutionarily derived genera of the order, such as *Anastrepha*, *Musca*, and *Drosophila*, but less

is known about the more ancestral members, such as those in the suborder Nematocera, such as *Aedes aegypti*. *Aedes aegypti* does not have heteromorphic sex chromosomes and sex is controlled by an autosomal locus that contains an M gene that acts as a dominant male determiner [149, 419-421]. Genes linked to the sex determining locus (GLS) has been mapped to chromosome 1 and segregates with the gene LF284 (AAEL009994) on supercontig 1.446 on the p arm of chromosome 1[153, 163]. Availability of the complete sequence of the *Ae. aegypti* genome has led to further genetic characterization of sex- determining genes [165]. A male-biased gene tightly linked to GLS, myo-sex, was recently characterized and is only present in the male genome [422]. Recently, orthologs of the *dsx* gene and another important gene in sex determination, *fruitless,* have been identified in the genome of *Ae. aegypti* and are expressed in a sex-specific manner throughout development [423, 424]. *Aedes aegypti* fruitless and *dsx* share a cis-regulatory element that does not utilize the splicing factor *tra*, and they lack an obvious ortholog for *tra*. However, they do possess multiple transformer-2 (*tra-2*) orthologs, an accessory factor which is responsible for the alternative splicing of *tra* in *Anastrepha* species [425]. The mechanism for the regulation of the alternative splicing of *dsx* in *Ae. aegypti* remains unknown.

The mosquito, *Ae. aegypti*, is the primary vector for the four serotypes of dengue virus, as well as yellow fever virus worldwide. *Aedes aegypti (L)* is found globally in tropical and sub-tropical climates and exists as two subspecies: *Ae. aegypti aegypti* and *Ae. aegypti formosus* [137, 138]. The current definition of the subspecies is based on the number or degree of white scales on the first abdominal tergite as defined by Mattingly and McClelland [137, 138]. *Aedes a. aegypti* has scales on the first abdominal tergite, has a light tan cuticle, and is globally distributed. In contrast, *Ae. aegypti formosus* has no white scales on the first abdominal tergite,

has a dark or black cuticle, and is found mostly in Sub-Saharan Africa in sylvatic environments [139-141]. However, these distinctions become less clear in West Africa where *Ae. aegypti* with a dark black cuticle and scales (albeit usually few) are frequently detected near human habitats [144, 145, 147]. In East Africa, the scaling pattern and behavior are also correlated with discrete genetic differences in allozymes and microsatellites [135, 146]. In West Africa, the scaling pattern does not correlate with these genetic markers [144, 145] or with behavioral differences between the subspecies, and leads to confusion in identification. Therefore, discrete genetic differences remain poorly defined between West African *Ae. aegypti* and global populations.

The difficulties in differentiating West African *Ae. aegypti* subspecies using morphological markers led us to compare genes in linked to the sex-determination locus of *Ae. aegypti aegypti* from Thailand and *Ae. aegypti formosus* from Senegal to elucidate discrete genetic differences between the subspecies. Difference in the sex-determination pathway is of particular relevance between the two sub-species because of the potential for generating reproductive isolation. We hypothesized that sex-specific features are evolving between *Ae. aegypti* subspecies and contributing to a separate sub-speciation even in West Africa. To test this, we studied the genetic linkage pattern of the sex-determining locus between subspecies and performed genomic DNA deep sequencing on replicate pools from Thailand and Senegal to characterize genetic differences in genes linked to the sex-determining locus. Validation relied on RT-PCR and haplotype analysis to build a more detailed picture of the genes linked to the sex-determination locus in *Aedes* spp.

Materials and Methods

Crossing Design

To determine the linkage position of the sex-determining locus in various populations of *Ae. aegypti*, populations of interest were crossed with a well characterized mutant strain of *Ae. aegypti*, Higgs white-eye (HWE). The HWE colony was started by Stephen Higgs by isolating individuals from a wild-type Rexville D colony that exhibited the spontaneous mutation of white eyes. The white-eye locus is a single locus mapped to a linkage position at 20 centimorgans (cM) on chromosome 1. The sex-determining locus is also a single locus mapped to a linkage position at 34 cM on chromosome 1 [149]. The two loci therefore 14 cM apart. The crossing design was as follows (Figure 3.2): males from a population of interest (n=30) were crossed to a virgin HWE female to produce F1 individuals heterozygous for the white-eye locus and the sex-determining locus. Eggs from this cross were collected and raised to adults. Male F1 individuals (n=10) were selected and backcrossed to virgin Higgs white-eye females. F2 adults were scored for sex and eye color. There were four possible phenotypes. Two of the phenotypes represented the parental genotype (white-eye females and black-eye males), and two of the phenotypes represented the recombinant genotype (white-eye males and black eye females).

The cM distance between the white-eye and sex-determining loci was the number of white-eye males or black-eye females divided by the total number of individuals and multiplied by 100. Three different models were considered and evaluated using chi-square goodness-of-fit. The highest probability represented the most likely scenario. The three scenarios were that the white-eye locus and sex locus are 1) 14 cM apart, 2) completely linked (0 cM apart), or 3) unlinked (50 cM apart).

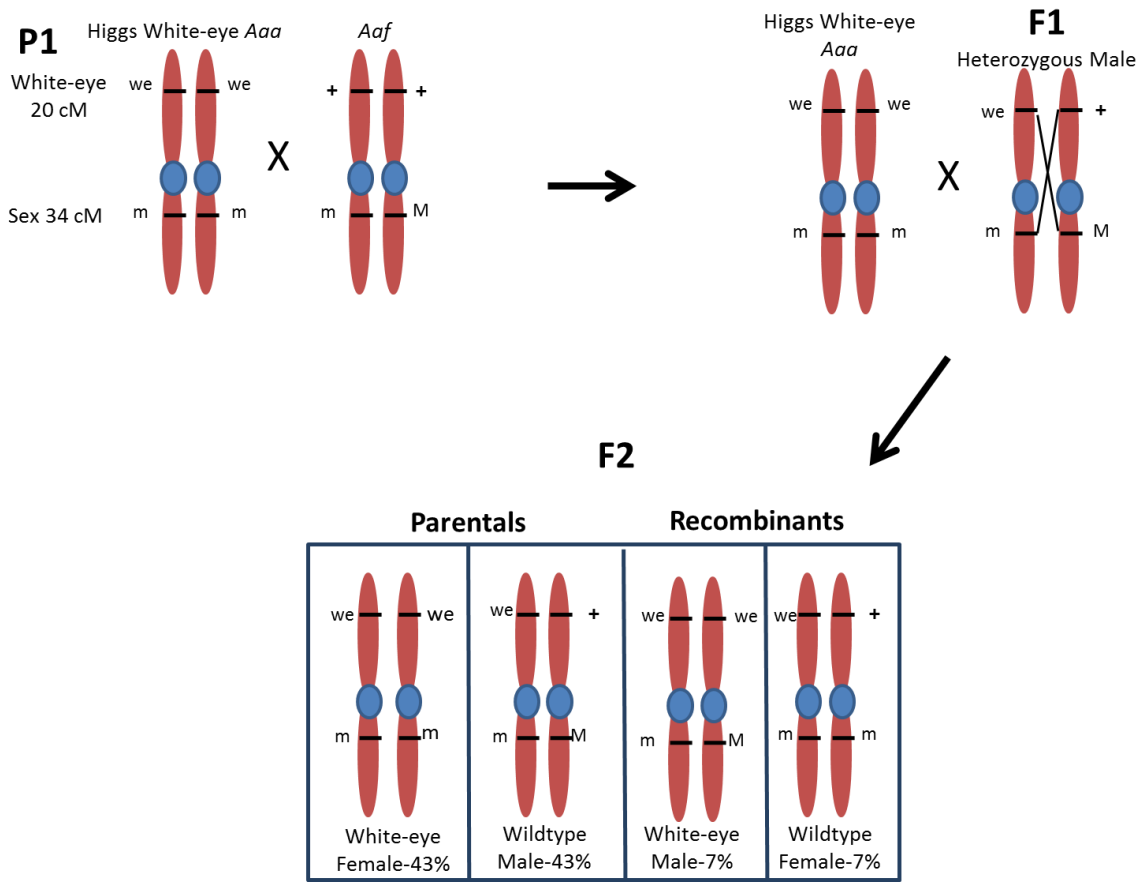


Figure 3.2: Crossing design to measure recombination between the white-locus and sex-determining locus.

A male (M/m) *Ae. aegypti formosus* (Aaf) from the population of interest (n=30) that were wildtype (+) for eye color (+/+, M/m) was crossed to a virgin Higgs white-eye female (we/we, m/m) *Ae. aegypti aegypti* (Aaa) to produce F1 individuals heterozygous for the white-eye locus (we/+) and the sex-determining locus (M/m). Eggs from this cross were collected and raised to adults. Male F1 individuals (n=10) were selected and backcrossed to a virgin Higgs white-eye females. Eggs were collected again, and the F2 individuals were scored for sex and eye color. There were four possible phenotypes. Two of the phenotypes represented the parental genotype (white-eye females and black-eye males), and two of the phenotypes represented the recombinant genotype (white-eye males and black eye females). If the two loci are the expected 14 cM apart, 86% of individuals should have a phenotype representing the parental genotype and 14% of individuals should have a phenotype representing the recombinant genotype.

Doublesex expression

Total RNA was extracted using Trizol from 1-2 day old adult male and female mosquitoes from Martinez de la Torre, Mexico and in the forested PK10 area of Senegal. A total of 10 males and 10 females from each location were tested. The samples were treated with DNase to eliminate genomic DNA. Equal amounts of RNA were used in the RT-PCR reaction. Expression of doublesex was measured using the Qiagen One-Step RT-PCR kit and primers described in Salvemini et al.[424] (dsxF: 5'- GATACTGAAAGGCGCCGACG-3', dsxR: 5'-GCAGAATATGGGACTGGTGC-3'). Expression of *dsx* was confirmed by visualizing the PCR product by agarose gel electrophoresis.

Deep Sequencing

Deep sequencing libraries were made from pools of F1 individuals collected from PK10, Senegal and Pai Lom, Thailand. Mosquitoes from PK10 represent *Ae. aegypti formosus* [388] and mosquitoes from Thailand represent *Ae. aegypti aegypti* [223]. A total of eight libraries were constructed: PK10 Female Replicates 1 and 2, PK10 Male Replicates 1 and 2, Thailand Female Replicates 1 and 2, and Thailand Male Replicates 1 and 2. Each replicate was made up of 12 individuals. DNA was quantified using picogreen and equal amounts of DNA per individual were pooled

DNA sequencing libraries were prepared using TruSeq DNA library preparation kits (Illumina, Somewhere CA) with 1 ug of genomic DNA per library. Genomic DNA was sheared via sonication prior to library preparation. Exome capture was then performed to enrich for 15,722 coding sequences using custom SeqCap EZ Developer probes (Nimblegen, Roche Life Sciences). Non-repetitive probes tiling coding sequence in the AaegL1.3 gene annotations were designed by Nimblegen. In total, 26.7Mb of the genome (20%) was targeted for

enrichment. TruSeq libraries were hybridized to the probes, unbound DNA was washed away, and the targeted DNA was eluted and amplified before being sequenced with 1 lane of 100 bp HiSeq2000 paired-end sequencing. TruSeq library preparation, exome capture and sequencing were performed following manufacturer's instructions by the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (Oxford, UK).

Construction of Reference Genome

A series of FORTRAN programs were written to create a reference file used for the detailed analysis of each of the 18,769 genes in the AaegL1.4 reference genome on Vectorbase. The coordinates of the features of each gene appearing in Aedes-aegypti-Liverpool_BASEFEATURES_Aaegl1.3.GTF were used to extract 1) 600 bp upstream from the start of the 5'UTR, 2) all of the 5'UTR, 3) all exons, 4) all introns, 5) all of the 3'UTR, and 6) 600 bp downstream of the 3'UTR. Next, a file of the physical locations of 294 supercontigs (kindly provided by Vladamir Timoshevskiy and Maria Sharakova) 1,330 genes on chromosome 1, 1,489 genes on chromosome 2, 1,258 genes on chromosome 3. The remaining 14,692 genes have yet to be mapped to chromosomes. This hash file, called "All3U" contained information on all 10 features for 301,348,667 nucleotides corresponding to 45% of the *Ae. aegypti* genome [389]. The 10 features are: 1) nucleotide ID number (id = 1,…, 301,348,667), 2) the VectorBase ID number, 3) the supercontig number, 4) the physical location following Timoshevskiy et al 2014, 5) the gene region (e.g. 5` nontranscribed, intron, exon etc.), and 6) the VectorBase reference nucleotide. For nucleotides in exons, each line also contained the 7) three letter codon, 8) the position in the codon (1, 2, or 3), 9) the amino acid encoded by the codon and 10) classification of the gene as mRNA for protein coding genes, tRNA, miRNA, snRNA, snoRNA,

misc_RNA or rRNA. Sequences were assembled as a single string to act as the sequence index file for subsequent analyses.

Sequencing Analysis

The FASTQ files from each library were aligned to the reference All3U genome assembly described earlier using GSNAP [390]. GSNAP was chosen because *Ae. aegypti* contains on average 1 SNP every 10-15 nucleotides [391] and other programs (eg. Bowtie) only allow 2 SNPs per 100-nucleotide read. The SAM file produced by GSNAP was converted to a BAM file and then sorted. A Pileup file was created from the SortedBAM file using SAMtools [392]. The "readcounts" command in Varscan2 analyzed the Pileup file to report SNPs with a minimum of 15x coverage in each library. The FORTRAN program, *Line Reader 1*, read the output from Varscan2 [426] and produced a flat file containing the numbers of each nucleotide and the numbers of insertions and deletions. The program *2x2* was used to identify SNP ID numbers in common between the two replicates to produce a single file containing nucleotide IDs that appear in both libraries. The program *2x2x2* compared SNP IDs between the PK10 male 2x2 file and the PK10 female 2x2 file to identify SNP IDs common to both. The program *tables* assembled a 4 x 6 contingency table with each of the 4 rows corresponding to a different library. The program *EasyChi* performed a contingency Chi-square analysis of each SNP ID table, calculated the probability with 3(n-1) degrees of freedom where n is the number of nucleotides segregating at that site. This probability was converted to a Log of odd ratio (LOD) score as $-\log_{10}(\text{prob})$. The file containing LOD values was annotated using the All3U reference file described above.

Haplotype Prediction

The Trinity *de novo* assembly package [427] was used to assemble genomic haplotypes. Trinity was developed for RNA-Seq analysis, but is useful for haplotype analysis in the context used here (personal communication, Brian Haas, The Broad Institute). For each population and sex of interest, raw FASTQ reads from the target-capture deep sequencing runs were assembled using default parameters, which were expected to produce consensus, not exhaustive, haplotypes for each population. The reason for this is that at one step of the Butterfly phase of the Trinity assembly, groups of sequence variants that are more than 95% similar are collapsed into a single contig. Each sex-specific group of haplotypes was searched using stand-alone BLASTn against the *Aedes* transcriptome release 1.4. Specific haplotypes of putative sex determinant genes were aligned with the Geneious version 7.1.3 (Biomatters, Ltd.) sequence analysis package to identify replacement sex-specific and subspecies-specific differences.

Copy Number Variation

Sequences were aligned to the reference genome using the GSNAP program. The "copynumber" command in Varscan2 [426] was used to compare the coverage of 100 base pair fragments between two libraries. PK10 Males were compared to Thai Males and PK10 Females were compared to Thai Females. The log2 ratio scores from VarScan were then analyzed with the DNAcopy package [428] in R. The DNAcopy package uses circular binary segmentation [429] to translate noisy intensity measurements into regions of equal copy number. Regions identified as having variable copy number between the two populations were required to be two standard deviations away from the equal copy number value.

qPCR

Genomic DNA was extracted from individual mosquitoes using the salt extraction method [430]. The actin gene (AAEL001928) and LF284 gene (AAEL009994, segregates with sex-determining loci) were amplified by PCR and their concentrations were determined and converted to copies/ul. 10-fold serial dilutions of the PCR products were then used to generate a standard curve for both actin and LF284. The same primers were used to amplify both genes in the individual of interest using qPCR. The number of copies of the LF284 in both males and females was determined by taking the copies/ul of LF284 generated from comparison to the standard curve and dividing by copies/ul of actin. This number was then normalized as the mean copies/ul of individual mosquitoes from Mexico to get a fold-increase of the copies/ul of LF284 in *Ae. aegypti formosus* compared *Ae. aegypti aegypti*. The assay was validated using individual mosquitoes from our crossing experiment that showed 14 cM linkage of the sex locus and white-eye locus. The primer sequences were:

Actin F : 5'- CTCTATCTACCTTCCAGGTAT - 3'

Actin R: 5'- GACGCTGACAAGTATCACAA - 3'

LF284 F: 5'- CATCTACGCCAAGGACGAG -3'

LF284 R: 5' –CAGGACCTTACGGATTTCAT-3'

Results

Variable linkage of the SDL

To measure the linkage patterns of the previously described sex-determining locus (SDL), recombination rates between the mutant white-eye locus and the SDL were measured in males from four different locations in Senegal and one location in Mexico. Four collections from Senegal varied in the proportion of individuals in which the two loci were significantly

more than the expected 14 cM apart (Table 3.1, Figures 3.3 and 3.4), indicating *Ae. aegypti* collections from Senegal have variable linkage of the SDL and white-eye. In contrast, no individuals from Vergel, Mexico exhibited this phenotype, suggesting that the variable linkage of the SDL and the white-eye locus is specific to *Ae. aegypti* collections from Senegal.

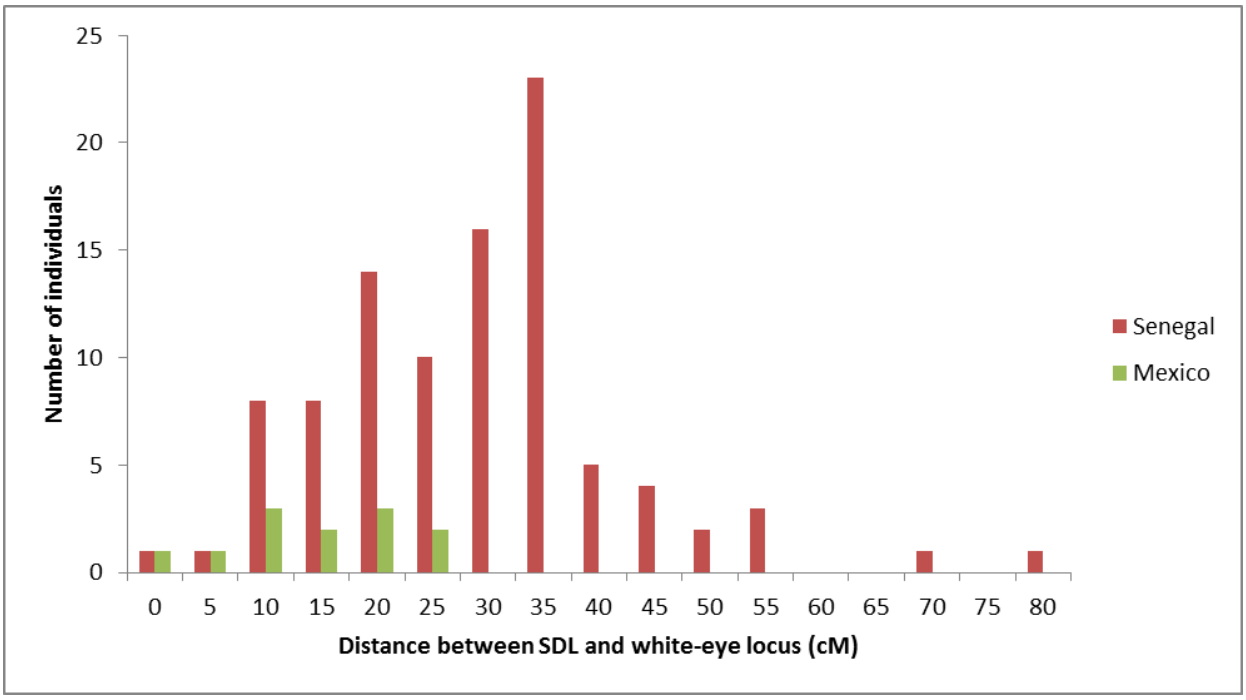


Figure 3.3: Distribution of linkage distance between SDL and white-eye locus
Histogram showing the number of individuals from Mexico and Senegal with various linkage distances between the SDL and white-eye. The cM distance was measured between the white-eye locus and SDL in 5 collections from Senegal and 1 collection from Mexico.

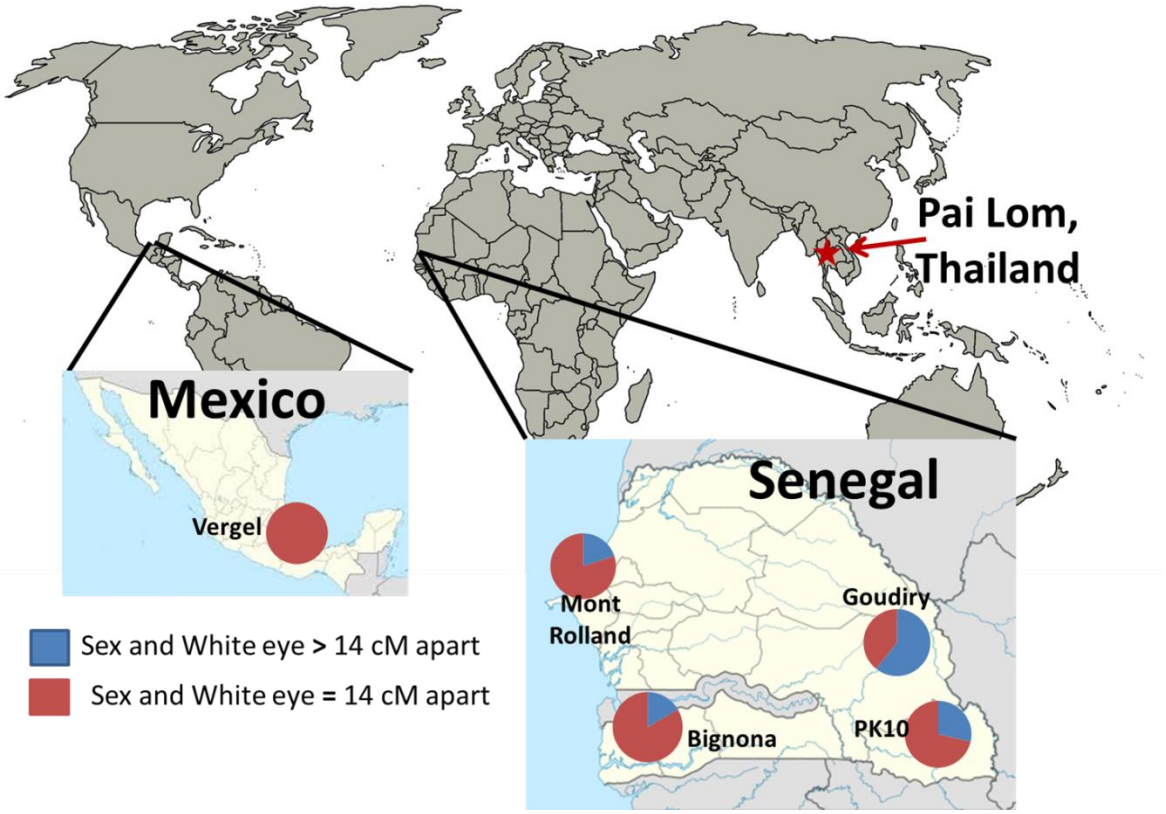


Figure 3.4: Linkage of the white-eye and sex locus is variable in 4 collections from Senegal
The linkage of the white-eye locus and the sex-determining locus was determined by crossing 25-30 individual males from five different collection sites in Senegal with white-eye females from the HWE colony and measuring the recombination rate between the two loci. In each collection: 27 families were analyzed in PK10, 27 families in Goudiry, 35 families in Mont Rolland, 5 families in Bignona, and 11 families in Vergel.

Table 3.1: Crossing Families

		Parental	Parental	Recombinant	Recombinant					
		White-Eye Female (n)	Wildtype Male (n)	Wildtype Female (n)	White-eye Male (n)	total	χ^2(r=0.5 cM)	χ2(r=0.0 cM)	χ2(r=0.14 cM)	Hypothesis Supported
	P51	0	34	0	9	43	2.08E-17	9.66E-40	9.49E-12	wild-type
	P17	0	54	2	7	63	3.94E-29	1.17E-31	5.18E-14	wild-type
	P32	1	35	0	16	52	3.43E-15	1.66E-112	1.20E-17	Unlinked
	P50	1	27	1	2	31	1.35E-15	1.44E-07	3.30E-07	wild-type
	P22	4	65	5	19	93	7.33E-25	9.88E-170	1.81E-17	wild-type
	P26	4	34	5	9	52	1.01E-11	1.05E-46	5.49E-08	wild-type
	P56	6	40	4	13	63	4.05E-13	3.15E-80	3.56E-10	wild-type
	P46	6	48	2	12	68	8.62E-19	9.68E-67	5.15E-11	wild-type
	P19	6	25	7	7	45	2.14E-06	2.87E-41	6.72E-06	wild-type
	P56	6	14	0	12	32	1.08E-04	6.98E-61	2.38E-12	Unlinked
	P44	8	48	4	18	78	6.09E-15	2.58E-146	1.29E-13	wild-type
	P31	8	23	10	3	44	8.52E-06	9.44E-46	1.86E-06	Unlinked
	P38	9	25	5	4	43	2.65E-07	1.57E-17	3.01E-03	wild-type
PK10	P31	13	40	3	9	65	2.09E-12	6.42E-39	2.13E-05	wild-type
	P54	13	20	8	4	45	3.68E-04	1.72E-32	1.74E-03	wild-type
	P27	15	15	8	10	48	7.52E-02	5.05E-67	1.98E-06	Unlinked
	P9	19	20	7	7	53	5.83E-04	2.36E-39	9.08E-03	wild-type
	P47	19	18	2	5	44	4.82E-06	1.27E-11	2.03E-01	wild-type
	P48	20	44	14	11	89	3.83E-08	4.36E-132	1.69E-06	wild-type
	P2	21	28	7	7	63	4.59E-06	2.09E-39	3.51E-02	wild-type
	P25	23	24	2	7	56	2.36E-07	2.17E-21	6.48E-02	wild-type
	P28	25	63	8	12	108	6.39E-17	1.19E-87	1.76E-05	wild-type
	P3	28	54	11	29	122	2.78E-08	0.00E+00	5.22E-15	Unlinked
	P21	38	26	23	8	95	1.16E-05	4.69E-249	1.14E-11	Unlinked
	P8	39	35	16	16	106	3.85E-05	5.12E-213	1.42E-06	Unlinked
	P30	45	44	13	15	117	1.63E-08	8.54E-163	1.71E-03	wild-type
	P24	47	33	21	7	108	1.32E-08	4.76E-205	1.32E-07	wild-type
	G11	39	24	17	14	94	6.78E-05	2.55E-202	1.91E-08	Unlinked
	G12	10	27	0	2	39	9.47E-12	3.79E-04	5.62E-04	wild-type
	G13	6	12	4	4	26	1.01E-02	4.53E-13	5.58E-03	Unlinked
	G18	85	39	39	23	186	1.08E-11	0.00E+00	2.49E-19	Unlinked
	G19	13	21	5	12	51	1.48E-03	1.08E-69	8.89E-07	Unlinked
	G1A	19	8	10	0	37	8.80E-06	1.75E-42	1.03E-07	Unlinked
	G22	73	44	18	17	152	1.03E-13	1.80E-256	4.23E-05	wild-type
	G23	0	22	0	2	24	3.68E-14	3.26E-07	4.96E-07	wild-type
	G24	19	40	4	12	75	6.65E-10	8.98E-67	5.46E-05	wild-type
	G25	1	13	0	3	17	5.39E-07	2.39E-06	2.01E-04	wild-type
	G26	34	22	14	11	81	7.65E-05	3.38E-131	2.73E-06	Unlinked
	G30	21	25	20	14	80	7.83E-02	1.11E-249	3.53E-14	Unlinked
	G31	25	22	15	8	70	1.67E-03	1.75E-119	3.74E-07	Unlinked
Goudiry	G34	2	36	0	3	41	1.26E-20	9.13E-11	2.11E-09	wild-type
	G42	37	28	23	12	100	3.05E-04	2.17E-282	1.06E-11	Unlinked
	G45	2	47	0	5	54	4.00E-26	4.80E-19	4.44E-12	wild-type
	G46	9	7	18	19	53	3.53E-03	1.87E-289	1.02E-31	Unlinked
	G47	32	36	19	23	110	9.49E-03	0.00E+00	1.45E-13	Unlinked
	G48	32	39	9	9	89	1.10E-08	1.20E-65	6.12E-02	wild-type
	G49	22	27	5	6	60	5.93E-07	3.16E-24	2.15E-01	wild-type
	G4B	23	12	11	6	52	5.78E-04	1.11E-64	4.07E-06	Unlinked
	G51	15	14	8	5	42	1.04E-02	6.46E-36	6.65E-04	Unlinked
	G54	51	45	10	7	113	7.15E-14	3.77E-60	3.07E-01	wild-type
	G55	28	17	9	2	56	2.36E-07	3.80E-35	1.37E-03	wild-type
	G56	23	38	13	18	92	9.58E-05	7.41E-206	4.43E-09	Unlinked
	G57	24	20	11	9	64	1.92E-03	1.14E-82	4.73E-05	Unlinked
	G5B	13	1	11	2	27	4.37E-05	2.82E-53	2.24E-13	Unlinked

Table 3.1 Continued: Crossing Families

		Parental	Parental	Recombinant	Recombinant					
		White-Eye Female (n)	Wildtype Male (n)	Wildtype Female (n)	White-eye Male (n)	total	χ^2(r=0.5 cM)	χ2(r=0.0 cM)	χ2(r=0.14 cM)	Hypothesis Supported
Mont Rolland	MR10	18	23	2	3	46	6.19E-08	2.31E-05	2.81E-01	wild-type
	MR11	12	17	2	1	32	1.85E-06	1.48E-02	1.92E-01	wild-type
	MR13	32	17	17	6	72	1.31E-05	9.21E-136	3.48E-09	Unlinked
	MR14	9	16	1	2	28	4.95E-06	8.32E-03	1.11E-01	wild-type
	MR15	21	0	11	1	33	2.91E-09	8.09E-54	9.76E-13	Unlinked
	MR16	22	22	2	2	48	7.76E-09	2.43E-03	2.58E-01	wild-type
	MR17	42	21	6	2	71	9.45E-14	2.15E-17	2.33E-03	wild-type
	MR18	34	38	10	5	87	6.11E-10	1.59E-50	8.20E-02	wild-type
	MR19	7	11	1	1	20	1.48E-04	1.69E-01	2.74E-01	wild-type
	MR1A	15	30	11	8	64	2.36E-05	2.70E-76	2.02E-05	Unlinked
	MR20	42	39	4	12	97	1.91E-11	2.42E-65	2.11E-02	wild-type
	MR22	11	19	3	1	34	1.02E-06	9.61E-05	7.50E-02	wild-type
	MR24	13	29	1	6	49	1.55E-09	2.60E-16	1.82E-03	wild-type
	MR26	6	9	0	0	15	5.70E-05	1.90E-01	7.64E-02	linked
	MR27	43	46	7	6	102	4.77E-14	1.00E-33	5.80E-01	wild-type
	MR29	39	45	15	11	110	1.97E-08	9.72E-143	1.65E-03	wild-type
	MR30	13	6	1	1	21	1.76E-05	6.14E-02	8.01E-02	wild-type
	MR31	33	47	5	8	93	3.93E-13	5.97E-36	7.63E-02	wild-type
	MR32	2	11	2	0	15	1.06E-05	9.84E-04	4.21E-03	wild-type
	MR33	15	14	6	1	36	1.14E-04	3.34E-15	1.53E-02	wild-type
	MR34	16	17	9	9	51	3.49E-02	4.47E-66	1.16E-05	Unlinked
	MR35	10	6	1	0	17	9.49E-05	1.57E-01	1.18E-01	linked
	MR36	16	21	2	3	42	4.16E-07	2.21E-05	3.14E-01	wild-type
	MR37	21	45	1	3	70	2.79E-17	3.58E-06	1.86E-04	wild-type
	MR39	7	9	0	0	16	4.87E-05	2.48E-01	8.88E-02	linked
	MR4	23	12	3	2	40	8.90E-08	5.92E-06	5.22E-02	wild-type
	MR40	21	38	11	8	78	1.12E-07	3.94E-76	5.08E-04	wild-type
	MR41	5	5	4	1	15	9.04E-02	3.17E-07	2.79E-03	Unlinked
	MR42	4	7	2	5	18	8.92E-02	5.10E-12	2.50E-04	Unlinked
	MR43	6	16	4	3	29	1.24E-04	9.82E-11	9.47E-03	wild-type
	MR44	23	29	1	0	53	1.21E-12	1.93E-01	6.58E-03	linked
	MR6	37	35	10	8	90	1.15E-08	2.12E-66	8.02E-02	wild-type
	MR7	11	15	3	3	32	2.39E-04	2.89E-07	2.77E-01	wild-type
	MR8	35	51	6	12	104	1.48E-12	8.31E-74	1.22E-02	wild-type
	MR9	6	4	1	3	14	5.39E-02	1.83E-04	2.78E-02	Unlinked
Bignona	B1	13	44	6	32	95	6.18E-10	0.00E+00	4.93E-27	Unlinked
	B7	24	9	1	3	37	3.12E-09	6.33E-06	4.29E-03	linked
	B8	9	11	2	3	25	2.17E-03	4.52E-06	2.70E-01	linked
	B3	36	42	7	12	97	1.10E-09	1.20E-93	2.87E-02	linked
	B5	9	25	2	4	40	1.13E-08	8.92E-10	4.21E-03	linked
Vergel	V15	21	46	4	9	80	2.98E-19	9.16E-41	6.41E-04	wild-type
	V16	22	47	5	11	85	5.06E-18	6.47E-61	2.77E-04	wild-type
	V17	23	18	3	3	47	9.96E-07	3.26E-07	4.10E-01	wild-type
	V31	7	11	2	1	21	9.72E-05	1.43E-02	2.68E-01	wild-type
	V32	11	16	0	1	28	1.24E-08	1.65E-01	5.06E-02	linked
	V34	8	19	1	1	29	4.13E-11	2.06E-02	1.38E-02	linked
	V41	12	26	1	3	42	1.50E-13	2.21E-05	9.09E-03	wild-type
	V42	6	12	2	4	24	1.28E-04	1.84E-08	2.08E-02	wild-type
	V5	31	33	2	4	70	6.80E-13	5.83E-08	1.39E-01	wild-type
	V7A	11	7	4	2	24	2.30E-02	2.89E-08	3.64E-02	wild-type
	V9	21	35	3	9	68	1.14E-12	4.74E-37	4.99E-03	wild-type

Deep sequencing

To identify sex-specific and sub-species specific genetic differences, the genomes of *Ae. aegypti aegypti* and *Ae. aegypti formosus* were sequenced. Sex-specific and subspecies-specific single nucleotide polymorphisms (SNPs) were identified for the *Ae. aegypti aegypti* and *Ae. aegypti formosus* sequencing datasets. Allele frequencies were compared using chi-square analysis and log of odd ratios (LODs) were reported. More SNPs and more significant sex-specific SNP LODs were detected in *Ae. aegypti formosus* than in *Ae. aegypti aegypti* across all three chromosomes (Figure 3.5). Chromosome 1, which contains GLS and *dsx*, had the most SNPs and the most significant SNPs in both subspecies (Figure 3.5). Also, regions of Chromosome 1 contain highly significant groups of sex-specific SNPs in *Ae. aegypti formosus* that differ from *Ae. aegypti aegypti* (Figure 3.5). Due to the large number of SNPs across the genomes, we decided to focus on SNPs in GLS for this analysis.

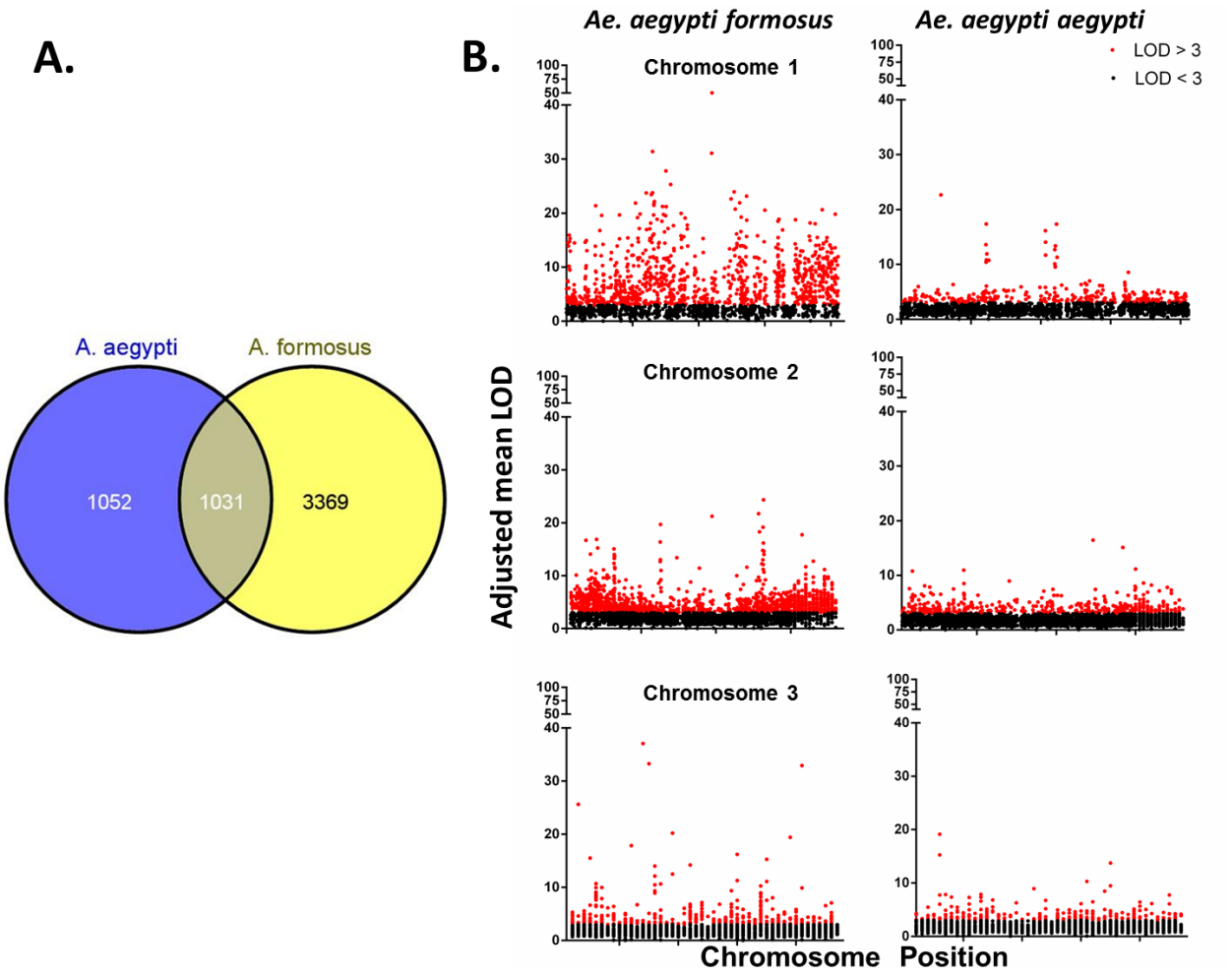


Figure 3.5: Common genes among male vs female comparisons

A) Venn diagram showing the number of genes with sex-specific differences (adjusted mean LOD >3) found only in *Ae. aegypti aegypti*(1052), only in *Ae. aegypti formosus*(3369), or shared between the two subspecies (1031). B) Adjusted mean LOD per gene plotted against chromosome position in *Ae. aegypti formosus* and *Ae. aegypti aegypti*. Adjusted mean LOD values for each gene were calculated by taking the average of the 10 highest LOD scores for each gene. If 10 SNPs were not detected in a single gene, the sum of the LOD scores was still divided by 10. The *Ae. aegypti aegypti* libraries had 9x coverage and *Ae. aegypti formosus* had 15x coverage.

Sex-determining locus genotypes and haplotypes

To provide further evidence of subspecies-specific differences in GLS, haplotypes were assembled from the combined replicate deep sequencing FASTQ files for each sex and subspecies. In the *Ae. aegypti* genome, SDL segregates with supercontig 1.446 [153, 163], therefore this genomic location was examined more closely. Supercontig 1.446 contains 20 genes, but SNPs were only detected in 13 of these genes. Of the 13 genes, three genes with the annotation of conserved hypothetical protein (AAEL010004, AAEL010005, and AAEL009997) contained sub-species specific and sex-specific SNPs and haplotypes. In AAEL010004 and AAEL010005, three amino acid replacement sex-specific substitutions with large LOD scores were common to both subspecies (Table 3.2). These SNPs are also detected in haplotype predictions (Figure 3.6). All three SNPs were heterozygous for serine in *Ae. aegypti formosus* males, but had low heterozygosity for serine in *Ae. aegypti aegypti* males. AAEL010005 and AAEL009997 bear seven replacement substitutions found only in *Ae. aegypti formosus* (Table 3.3) and six others exclusive to *Ae. aegypti aegypti* (Table 3.4). Interestingly, 61.5 % of replacement substitutions in GLS of *Ae. aegypti formosus* (LOD score > 3) were heterozygous in males and had low heterozygosity in females (Tables 3.2 and 3.3), but only 9.0% of sex-specific coding changes in *Ae. aegypti aegypti* followed this pattern (Tables 3.2 and 3.4). Instead, females were heterozygous and males had low heteroygosity in 45.5% of replacement SNPs in *Ae. aegypti aegypti*.

Table 3.2: Replacement substitutions in GLS shared between *Ae. aegypti aegypti* and *Ae. aegypti formosus*

Three replacement substitutions in the GLS with an LOD score greater than 3 were shared between the subspecies. The VectorBase accession number, SNP ID, frequency of each amino acid in either subspecies, and LOD score are reported.

		Ae. aegypti formosus			*Ae. aegypti aegypti*		
Gene	SNP ID	Amino Acid Frequency in Males	Amino Acid Frequency in Females	MvF LOD	Amino Acid Frequency in Males	Amino Acid Frequency in Females	MvF LOD
AAEL010004	26375401	0.435 (G), 0.565 (S)	0.986(G)	24.16	0.932 (G)	0.676(G), 0.324 (S)	4.48
AAEL010004	26375402	0.425 (T), 0.575 (S)	0.965(T)	22.99	0.932 (T)	0.671(T), 0.329 (S)	4.63
AAEL010005	26429142	0.526 (S), 0.474 (N)	0.991(S)	15.17	0.877 (S)	0.642(N), 0.358(S)	2.97

Table 3.3: Replacement substitutions in *Ae. aegypti formosus*

Seven replacement substitutions in the SDL with an LOD score greater than 3 were specific to *Ae. aegypti formosus*. The VectorBase accession number, SNP ID, frequency of the alternate amino acid in either subspecies, and LOD score are reported. SNPs highlighted in red are heterozygous in males and homozygous in females. SNPS highlighted in green are heterozygous in females and homozygous in males. SNPs highlighted in blue have similar alternate amino acid frequencies between the sexes.

Gene	SNPID	Frequency of alt. allele in Females	Frequency of alt. allele in Males	FvM LOD	Amino acid, Alternate amino acid
AAEL009997	26426605	0.85	0.59	15.64	M, T
AAEL009997	26426707	0.46	0.62	3.54	G, D
AAEL009997	26426847	0.87	0.60	4.1	A, T
AAEL010005	26429000	0.98	0.58	22.64	T, S
AAEL010005	26429313	0.53	0.81	3.83	E, G
AAEL010005	26428937	0.970	0.312	10.49	G, S
AAEL010005	26429133	0.992	0.566	14.41	S, N

Table 3.4: Replacement substitutions in *Ae. aegypti aegypti*

Six replacement substitutions in the SDL with an LOD score greater than 3 only in *Ae. aegypti aegypti*. The VectorBase accession number, SNP ID, frequency of the alternate amino acid in either subspecies, and LOD score are reported. SNPs highlighted in red are heterozygous in males and homozygous in females. SNPS highlighted in green are heterozygous in females and homozygous in males. SNPs that are highlighted in blue have similar alternate amino acid frequencies between the sexes.

Gene	SNPID	Frequency of alt. allele in Females	Frequency of alt. allele in Males	FvM LOD	Amino acid, Alternate amino acid
AAEL009997	26426971	0.89	0.79	3.96	P, Q
AAEL009997	26426668	0.86	0.70	3.58	P, L
AAEL009997	26426707	0.87	0.65	4.11	G, D
AAEL010004	26374966	0.53	0.83	5.06	P, S
AAEL010005	26430096	0.59	0.89	4.68	A, V
AAEL010005	26429798	0.69	0.37	6.53	A, T

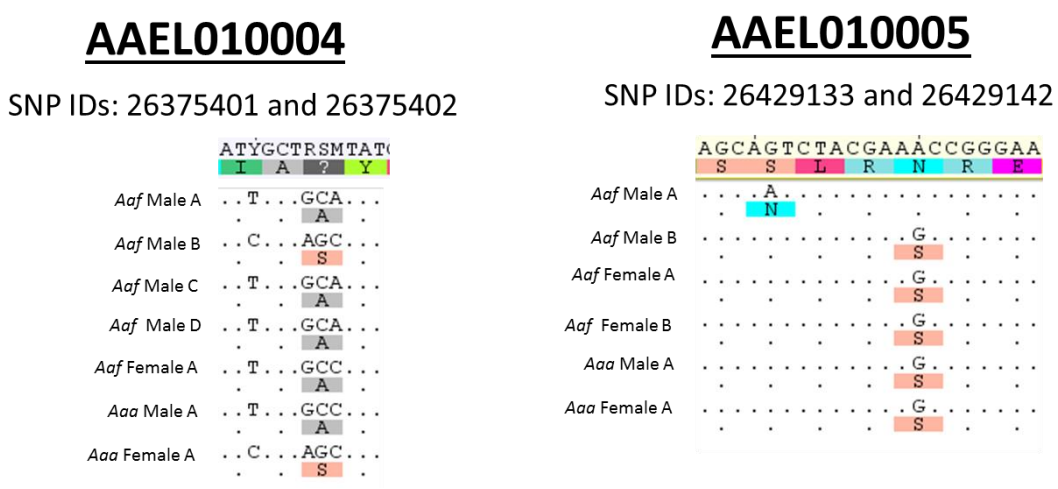
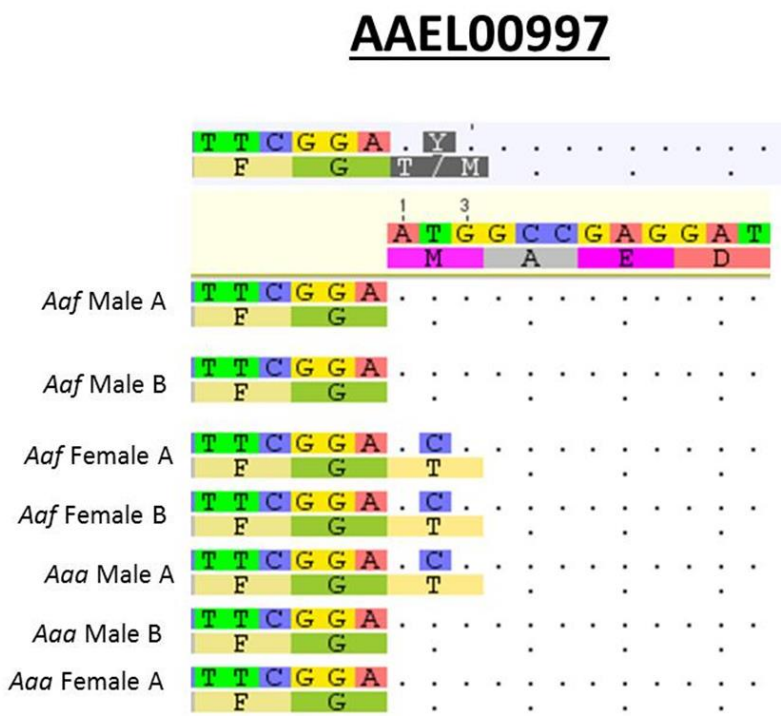


Figure 3.6: Haplotypes for 3 of the SNPs in Table 3.2
Haplotype predictions for *Ae. aegypti aegypti* (Aaa) and *Ae. aegypti formosus* (Aaf). Unique sex-specific haplotypes (ie., A, B, C, D) represent consensus sequences which were generated in a reference-independent assembly of raw *.fastq files. For each individual haplotype, the nucleotide sequence is on top, and the amino acid sequence is on the bottom. The consensus sequence is represented as dots and variable nucleotides and corresponding amino acids are highlighted. The consensus nucleotide and amino acid sequence is shown on the top line. The consensus nucleotide sequence includes the IUPAC ambiguity codes (Y,R,S,M) representative of heteroygosity of that allele.

An interesting SNP was identified in the start codon of AAEL009997 (Figure 3.7). *Aedes aegypti formosus* males are heterozygous for a methionine, while *Ae. aegypti formosus* females had low heterozygosity for a threonine. This SNP was not represented in the *Ae. aegypti aegypti* female exome capture dataset. The presence of a methionine start codon in *Ae. aegypti aegypti* females from Thailand was later confirmed by amplifying and sequencing genomic DNA from the same population used for the deep sequencing. Although the presence of a methionine start codon was confirmed through Sanger sequencing , the frequency of methionine in *Ae. aegypti aegypti* females remains unknown since it is not present in the deep sequencing data set. Overall, the SNP frequency and haplotype data in AAEL009997, AAEL010004, and

AAEL010005 indicate there are genetic differences in GLS between *Ae. aegypti aegypti* and *Ae. aegypti formosus*.



SNPID	Aaf Female	Aaf Male	Aaf LOD	Aaa Female	Aaa Male	Aaa LOD
26426605	0.847(T)	0.593(T) 0.407(M)	15.64	N/A	N/A	N/A

Figure 3.7: Haplotype prediction in the start codon of AAEL009997
Haplotype predictions for *Ae. aegypti aegypti* (Aaa) and *Ae. aegypti formosus* (Aaf). For each individual haplotype, the nucleotide sequence is on top, and the amino acid sequence is on the bottom. The consensus sequence is represented as dots and variable nucleotides and corresponding amino acids are highlighted. The consensus nucleotide and amino acid sequence is shown on the top line. The amino acid frequencies in Aaf males and females are also shown. This SNP was not represented in the Aaa female deep sequencing dataset (represented as N/A in table), but the sequence was confirmed by PCR amplification and Sanger sequencing in a female from Thailand. The frequencies of a threonine (T) or methionine (M) in Aaf are reported in the table.

Copy number variation

The precedent of gene duplications leading to the evolution of *tra* and *tra*-like genes in *Apis mellifera* [431] led us to look for variations in copy number of GLS as an explanation for the variable linkage of the SDL in mosquitoes from Senegal. The probability that the number of sequencing reads of a continuous sequence of at least 400 basepairs was at least 2 standard deviations away from the number of sequencing reads of surrounding sequences was calculated using circular binary segmentation with the DNAcopy program in R [428]. Examination of the number of sequencing reads in supercontigs 1.446 ad 4 surrounding supercontigs, revealed that loci within supercontig 1.446 are were represented in higher copy number in *Ae. aegypti formosus* males compared to *Ae. aegypti aegypti* males (Figure 3.8). In females, this locus is represented in equal abundance between the subspecies. Specifically, genes at the 5' end of supercontig 1.446 (AAEL010007, AAEL010004, AAEL009997, and AAEL010012) are duplicated in *Ae. aegypti formosus* males (Table 3.5).

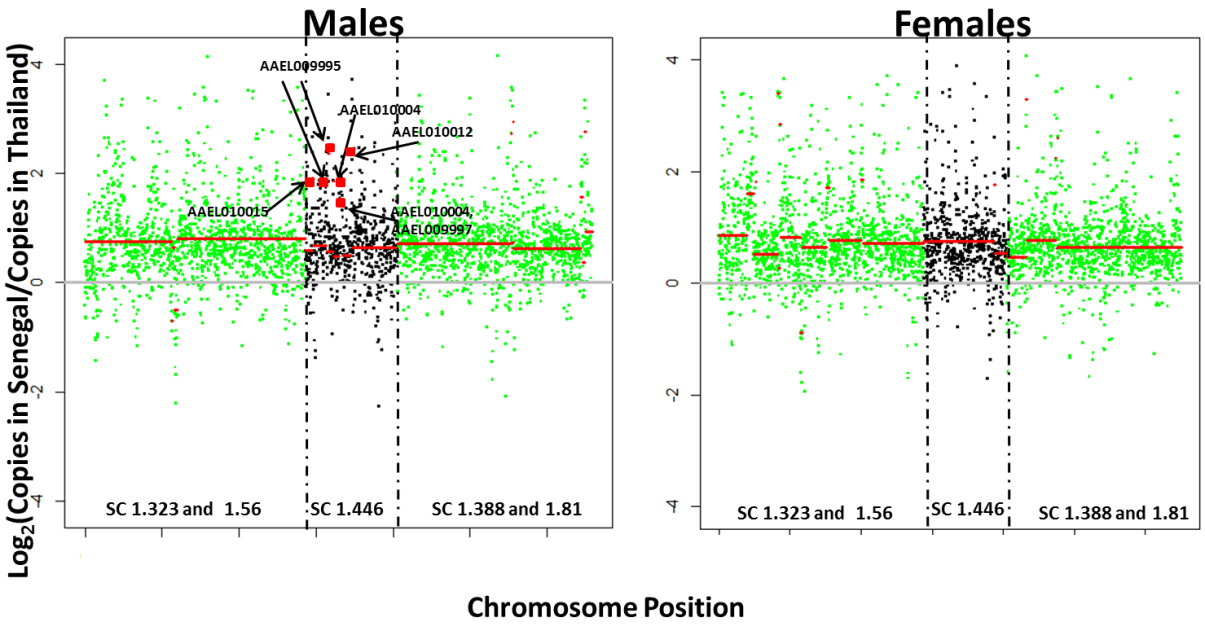


Figure 3.8: The sex-determining locus has variable copy numbers in males from Senegal
Black dots represent the $\log_2$(copies in Senegal/ copies in Thailand) for 100 basepair fragments across supercontig 1.446 and green dots represent the $\log_2$(copies in Senegal/ copies in Thailand) for 100 basepair fragments across 2 supercontigs upstream and 2 supercontigs downstream of supercontigs 1.446. The red lines indicate regions of the same copy number calculated using circular binary segmentation with DNAcopy R program [428]. For a gene segment (minimum 400 basepairs) to be classified as having variable copy number, it had to be 2 standard deviations from the surrounding 100 basepair reads A) Senegal males vs Thailand Males. B) Senegal females vs Thailand females. The red oval contains red lines with a higher $\log_2$(ratio), indicating gene regions in supercontigs 1.446 that have increased copy number in males from Senegal

Table 3.5: Predicted Segments with Copy Number Variation.

supercontig	Number of Markers in segment	Number of basepairs in segment	Segment Mean $\log_2$(copies Senegal /copies Thailand)	Accession Number
supercontig 1.323/1.56	564	453,786	0.7541	
supercontig 1.323/1.56	3	42	-0.7054	
supercontig 1.323/1.56	15	8,736	0.6416	
supercontig 1.323/1.56	18	1,829	-0.5083	
supercontig 1.323/1.56	832	919,673	0.8085	
supercontig 1.446	43	13,283	0.5858	
supercontig 1.446	4	5,572	1.913	AAEL010015
supercontig 1.446	84	29,618	0.6705	
supercontig 1.446	9	7,112	1.863	AAEL009995
supercontig 1.446	5	1,132	0.5739	
supercontig 1.446	4	302	2.3535	AAEL009995
supercontig 1.446	32	8,133	0.5734	
supercontig 1.446	3	454	1.8512	AAEL010004
supercontig 1.446	35	13,501	0.4846	
supercontig 1.446	23	36,034	1.4432	AAEL010004/AAEL009997
supercontig 1.446	54	10,869	0.4877	
supercontig 1.446	9	9,334	2.3368	AAEL010012
supercontig 1.446	287	89,692	0.6343	
supercontig 1.388/1.81	739	437,122	0.7218	
supercontig 1.388/1.81	5	275	2.7202	
supercontig 1.388/1.81	9	8,753	0.7263	
supercontig 1.388/1.81	3	1,830	2.9362	
supercontig 1.388/1.81	437	136,499	0.623	
supercontig 1.388/1.81	12	4,925	1.5672	
supercontig 1.388/1.81	13	2,458	0.3704	
supercontig 1.388/1.81	4	1,189	2.76	
supercontig 1.388/1.81	46	47,539	0.9228	
	3,292	2,249,692		

Copy number variation (CNV) of GLS between *Ae. aegypti aegypti* and *Ae. aegypti formosus* was validated with quantitative PCR. The gene LF284 (AAEL009994) was used as the marker due to its co-segregation with the sex phenotype [153, 163]. The distribution of the number of copies of LF284 in individual males from Mexico and Thailand ranged from 1 to 3

copies, while the number of copies in individual males from Senegal had a much wider distribution varying from 2 to 7 copies (Figure 3.9a). Interestingly, this phenomenon was not seen in females. Individual females from Mexico, Thailand, and Senegal had a much smaller number of copies of LF284 and ranged from 1 to 3 copies (Figure 3.9B).

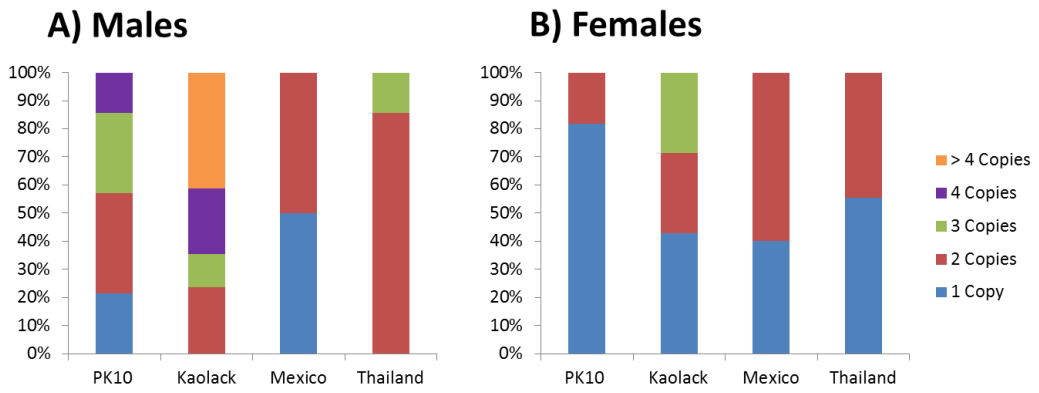


Figure 3.9: The sex-determining locus has variable copy numbers in males from Senegal, but not Thailand or Mexico

qPCR was used to validate the copy number data from the crossing experiment and deep sequencing results. The number of copies of the gene AAEL009994 in the SDL was measured in males (A) and females (B). Each collection site represents 10-20 individuals. The percent of individuals in each collection site that had 1, 2, 3, 4, or greater than 4 copies of AAEL009994 is shown for both sexes.

To determine if the genetic differences in GLS between subspecies resulted in subspecies-specific differences in *dsx*, a downstream gene in the sex determination pathway, expression of *dsx* was measured in adults from each subspecies. In one day old adults in both subspecies, the two expected isoforms were expressed in females and one isoform was expressed in males (Figure 3.10), as previously reported [424]. Both *Ae. aegypti formosus* and *Ae. aegypti aegypti* expressed *dsx* sex-specific isoforms in a similar manner, indicating the sex determination pathway converges on dsx regardless of the subspecies-specific genetic differences and duplication of the SDL.

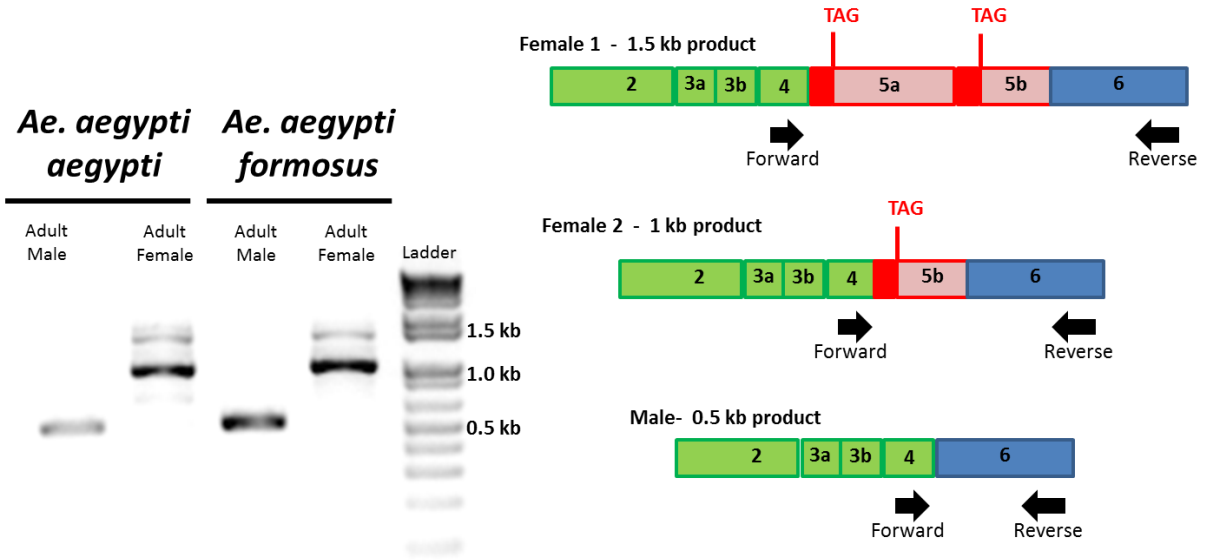


Figure 3.10: Sex-Specific expression of *doublesex* isoforms
Total RNA was extracted from 1-2 day old adult mosquitoes and treated with DNase to remove genomic DNA. Sex-specific expression of the three *dsx* isoforms was determined by RT-PCR in each subspecies. In the left panel, an agarose gel of the RT-PCR product is shown. Females from both subspecies express both of the female isoforms, and males from both subspecies express the male isoform. In the right panel, cartoons of the three predicted isoforms including the exons are shown along with the position of the primers and the expected product size. Pink exons are female-specific, blue exons are male-specific, and green exon are not sex-specific. In females, retention of exon 5a and 5b or just exon 5b results in a cDNA fragment of 1.5 or 1.0 kb respectively. In males, the absences of exon 5 results in a cDNA fragment of 0.5 kb.

Discussion

In an effort to identify *Ae. aegypti* subspecies-specific genetic differences, variation in genes linked to the sex determining locus (GLS) was analyzed. Variable linkage of the sex determining locus (SDL) on chromosome 1 in *Ae. aegypti formosus* males was shown through a classical crossing experiment. A pattern of sex-specific heterozygosity was identified, which would be predicted based on early studies demonstrating a dominant/recessive aspect to sex determination in *Ae. aegypti* [149, 421]. Importantly, the pattern of sex-specific heterozygosity in GLS was reciprocal in *Ae. aegypti formosus* males and *Ae. aegypti aegypti* females. Close

examination of SNPs in GLS revealed that 3 out of 3 replacement sex-specific SNPs are heterozygous in *Ae. aegypti formosus* males and had low heterozygosity in *Ae. aegypti aegypti* males. The SNP differences also correlate with haplotype signatures at these loci. The variable linkage appears to be the result of variations in the copy number of GLS, as demonstrated by analysis of the deep sequencing data and validated by qPCR. Taken together, these data suggest the Senegalese collection represents a diverse group of separate subspecies, each of which have the SDL in a different location and that *Ae. aegypti aegypti* represents a more genetically homogeneous population. Earlier work with *dsx* and another important gene in sexual development, *fruitless*, in *Ae. aegypti* suggest that *Ae. aegypti* utilizes upstream regulators other than *tra* [424]. Instead of using *tra* to regulate *dsx* splicing events, Salvemini *et al.* [424] proposed a model in which a male-specific M factor leads to male-specific splicing of *dsx* mRNA. In females, this M factor is not present and results in female-specific splicing of *dsx*. The identity of this M factor is not known, but it could be a gene or groups of genes in the SDL. The presence of male-biased gene in the same linkage position as the SDL [422] further supports the importance of this genomic region in sex determination. The current study builds on the story of *Ae. aegypti* sex determination by demonstrating duplication of GLS in some *Ae. aegypti formosus* in Senegal and a pattern of sex-specific heterozygosity in GLS.

Divergent sex determination systems within a species or closely related species is not novel. The housefly *Musca domestica* displays multiple different sex determination systems [432-435] among populations. These systems range from a dominant male-determiner on the Y chromosome providing a masculinizing activity, or the dominant male-determiner on any of the other five chromosomes, or on the X chromosome. In another *Musca domestica* system, a dominant female determiner directs sexual development [409].

One of the most interesting findings of this study is that within Senegal, there is a mixed frequency of males in which the white-eye locus and SDL are unlinked. In contrast, in all individuals from Mexico the white-eye locus and SDL were the expected 14 cM apart. This supports the hypothesis that Ae. aegypti formosus in Senegal is not a single subspecies as evident by the variation in the linkage of the white-locus and the SDL in the collections from Senegal as a result of these collections being a mixture of different subspecies. Greater genetic diversity, as evident by multiple SDL in Senegal, makes sense since *Ae. aegypti formosus* are the ancestral form of *Ae. aegypti* [146, 177]. Evidence of *Ae. aegypti formosus* being the ancestral form was also supported by greater genetic diversity, as evident by more SNPs in *Ae. aegypti formosus*. In the lab, colonies from these collection sites interbreed, but assortative mating within each collection has not been tested. This study provides evidence of discrete genetic differences between subspecies that have an important functional aspect.

Dominant alleles in the SDL result in male development, while female development is recessive [421, 436, 437]. The dominant/recessive nature of sex determination in *Ae. aegypti* predicts that alleles in the SDL will be heterozygous in males and homozygous in females. SNPs resulting in replacement substitutions in GLS were heterozygous in males and had low heterozygosity in females in *Ae. aegypti formosus*. Interestingly, the reverse was not true in *Ae. aegypti aegypti*. Multiple SNPs in GLS in *Ae. aegypti aegypti* were heterozygous in females and low heterozygosity in males. The pattern of heterozygous females in *Ae. aegypti aegypti* was not as consistent as the pattern of heterozygous males in *Ae. aegypti formosus*, further supporting the hypothesis that subspecies differences are substantial

The observation that sex and white-eye are unlinked in some individuals is consistent with three alternate hypotheses: 1) multiple different sex-determining loci in *Ae. aegypti* from

Senegal, 2) variable numbers of the white-eye locus in *Ae. aegypti* from Senegal, 3) duplication of the SDL in *Ae. aegypti* from Senegal. Gene duplication is a major force in generating novel genetic variants [438]. In honeybees (*Apis mellifera*), the complementary sex determination gene (csd) evolved through a duplication of the *fem*/*tra* gene [406]. In honey bees sex is determined by heterozygosity at a single locus (the Sex Determination Locus) that contains the csd gene. Female bees are heterozygous at this locus, while males are hemizygous (haploid individuals). Interestingly the heterozygosity of *csd* gene is only required for the induction of the female pathway through a positive feedback loop, while the fem gene remains active throughout development and controls the splicing of *dsx* [405, 406]. In this system, fem is similar to *tra* in that it is alternatively spliced in a sex specific manner and then regulates the sex specific splicing of *dsx*. The heterozygosity of the *csd* gene activates *fem*, a duplicated form of *csd*. A hypothesis could be that in the individuals from Senegal that show the white-eye locus and SDL being unlinked, the duplicated copy of the SDL regulates the original SDL at 34 CM which is involved in downstream regulation of *dsx* alternative splicing through an unknown mechanism. But in *Ae. aegypti aegypti*, the single copy SDL is regulated by a different unknown gene. It is likely then that the dominant alleles would be in the duplicated copy of the SDL in *Ae. aegypti formosus*. A gene other than the duplicated SDL that regulates the original SDL in *Ae. aegypti aegypti* would explain the lack of heterozygosity of the SNPs in the single copy SDL of *Ae. aegypti aegypti* males. We would expect the gene involved in regulating the SDL in *Ae. aegypti aegypti* males to be heterozygous.

While these data suggest that *Ae. aegypti aegypti* and *Ae. aegypti formosus* have evolved different mechanisms for sex determination, the characterization of the duplicated copy of the SDL has yet to be done. Functional studies that involve silencing genes in the SDL could

provide supportive evidence for the involvement of the SDL in the regulation of alternative splicing of *dsx*. It is also possible that the white-eye locus is duplicated in the populations from Senegal or that white-eye mutants are common in our colonies from Senegal, but neither of these hypotheses have been tested.

CHAPTER 4: VECTOR COMPETENCE IN WEST AFRICAN *AEDES AEGYTPI* IS FLAVIVIRUS AND GENOTYPE DEPENDENT

Introduction

Aedes aegypti is the primary epidemic vector of yellow fever virus (YFV) and all four serotypes of dengue viruses (DENV1-4), as well as being a known vector of chikungunya virus. Dengue remains an important public health problem with an expected 390 million infections per year [2]. Although there are fewer dengue cases in Africa compared to other regions, there are 11 African countries endemic for DENV, and the World Health Organization (WHO) continually reports DENV outbreaks in previously unreported geographic areas. Furthermore, it was recently estimated that in 2007 alone, 656 million fevers occurred in African children under the age of five [5]. However, in only 78 million of these cases was it likely that the child was infected with *Plasmodium falciparum.* Despite the enormous number of acute non-malarial febrile illnesses in sub-Saharan Africa, their etiologies are poorly defined [5].

Despite an effective vaccine, YF outbreaks still occur. Outbreaks are occasionally reported in South America (11 endemic countries), but a majority of the cases are in Africa (33 endemic countries), and most of the outbreaks are in West Africa [124]. According to the WHO, 200,000 cases of YF cause 30,000 deaths each year [439].

Both YFV and DENV belong to the family *Flaviviridae* and have a single-stranded positive sense RNA genome. There are four antigenically distinct serotypes of DENV (1-4) and all four serotypes are currently found in Africa. In West Africa, DENV-2 is an important serotype because it includes the sylvatic genotype with a high potential for emergence [116, 119, 440]. Sylvatic genotypes are transmitted between monkeys in forested areas, while cosmopolitan genotypes are transmitted between humans in urban areas. The sylvatic genotype

has been isolated from mosquitoes, monkeys, and humans [441-443] in West Africa, but is genetically distinct from epidemic isolates [115, 444]. Different genotypes as well as different lineages within genotypes can result in differences in both vector competence and the severity of human disease [445]. Genetic differences in YFV isolates are also an important predictor of vector competence. There are seven genotypes of YFV found worldwide, two of which (West African Genotypes I and II) are endemic in West Africa. The genetic differences among YFV isolates are geographically associated in Africa. West Africa genotype I is responsible for a majority of outbreaks and is genetically heterogeneous relative to other genotypes [123].

The vector for both viruses, *Aedes aegypti (L)*, exists as two subspecies: *Ae. aegypti aegypti* and *Ae. aegypti formosus* [137, 138]. Characters that distinguish the two subspecies were developed primarily in East Africa but are contradictory and confusing when identifying *Ae. aegypti* forms collected in West Africa. The current definition of the subspecies is based on the number or degree of white scales on the first abdominal tergite as defined by Mattingly and McClelland [137, 138]. *Aedes aegypti aegypti* has scales on the first abdominal tergite, has a light tan cuticle, is globally distributed, tends to be endophilic, and has a feeding preference for humans. In contrast, *Ae. aegypti formosus* has no white scales on the first abdominal tergite, has a dark or black cuticle, is found mostly in Sub-Saharan Africa in sylvatic environments, tends to be exophilic, and has a feeding preference for wild animals [139-141]. However, these distinctions become unclear in West Africa where *Ae. aegypti* with a dark black cuticle and white scales (albeit usually few) are frequently detected. Furthermore, *Ae. aegypti* without scales frequently breed near human habitats and bite humans. In East Africa, the scaling pattern and behavior are also correlated with discrete genetic differences in allozymes and microsatellites [135, 146]. But in West Africa, the scaling pattern does not correlate with these genetic markers

[144, 145] or behavioral differences and leads to confusion in subspecies identification. Further, these genetic studies reveal that East African *Ae. aegypti aegypti* and *Ae. aegypti formosus* are genetically distinct from the monophyletic West African *Ae. aegypti* [135, 146]. Due to this ambiguity the present study will hereafter refer to all *Ae. aegypti* collections based on their breeding site, habitat, and phytogeographic region of the collection site in Senegal.

Previous vector competence studies on *Ae. aegypti* from West Africa have shown these mosquitoes to be more refractory for both DENV [249, 316] and YFV[245, 316] compared to *Ae. aegypti* collected worldwide. Studies that have specifically examined collections across Senegal showed wide variation in vector competence for both high-passage [145] and low passage field isolates of DENV-2 [131, 216] . In particular, sylvatic collections from southeastern Senegal were more refractory than other collections from throughout Senegal.

Great variation in vector competence is seen within closely related collections of *Ae. aegypti* [207-209, 257] and different isolates of DENV-2 [131, 216] or YFV [217]. Furthermore, geographically distinct collections of *Ae. aegypti* from Senegal are genetically diverse[144, 145]. It has been demonstrated that vector competence of *Ae. aegypti* for DENV is governed by interactions between mosquito strain and virus genotype in natural collections [256] . This underscores the importance of using viruses and vectors that are geographically proximate and genetically diverse to draw conclusions about vector competence among collections Mosquito strain by flavivirus genotype interactions have yet to be examined in West African *Ae. aegypti*. Therefore, the objective of the current study was to examine these interactions by quantifying the vector competence of West African *Ae. aegypti* populations to DENV-2 and YFV field isolates.

Materials and Methods

Mosquito Collections

Aedes aegypti were collected as larvae in 8 locations in Senegal (Table 4.1). The sylvatic collections from the southeast (PK10 [388], and Kedougou,) were made in 2011, all others were made in 2010. The larvae were transported to a temporary local laboratory in Kedougou or Theis, reared to adults, and given a bloodmeal to generate eggs to start a colony. Eggs from each collection were brought back to Colorado State University and maintained for approximately 10-15 generations before being challenged with an artificial bloodmeal containing virus as described below. It was intended to use collections with fewer generations, but we were not able to get F_1 mosquitoes collected from the field to survive, mate, or bloodfeed sufficiently to perform vector competence assays. Adult mosquitoes were kept in incubators maintained at 28^{o}C with 70-80% relative humidity and a 12:12 hour photocycle. Eggs were collected on filter papers and stored for up to 5 months in a high humidity chamber. The collection sites consist of domestic sites around huts in urban environments and rural villages, tires in urban environments, and forests (sylvatic) where mosquito larvae were collected from treeholes and the discarded fruit husks of *Saba senegalensis* [388] (Table 4.1).

Table 4.1: Mosquito Collections

	Date Collected	Breeding Site	Habitat	Phytogeographic Region	Latitude	Longitude
Fatick	August 2010	Tire	Urban	Acacia-Savanna	14˚20'21.5"N	16˚24'38.8"W
Bignona	September 2011	Water Jar	Urban	Marsh/Swamp	12˚48'6.1"N	16˚13'37.8"W
Richard Toll	October 2010	Tire	Urban	Acacia-Savanna	16˚27'54.1"N	15˚41'1.8"W
Goudiry	October 2010	Tire	Urban	Acacia-Savanna	14˚11'2.2"N	12˚42'57.3"W
Kedougou	September 2011	Tire	Rural village	Deciduous Forest/Shrub	12˚33'29.4"N	12˚11'19.0"W
PK10	September 2011	Treehole	Forest Gallery	Deciduous Forest/Shrub	12˚36'45.1"N	12˚14'51.2"W
Mont Rolland	August 2010	Water Jar	Urban	Acacia-Savanna	14˚55'18.1"N	16˚59'37.3"W
Rufisque	October 2010	Tire	Urban	Acacia-Savanna	14˚42'56.8"N	17˚16'15.0"W

Viruses

Two low passage genotypes of YFV from West Africa were used in these studies. The YFV BA-55 from Nigeria is representative of West African Genotype I and YFV DAK 1279 from Senegal is a West African Genotype II [123]. YFV BA-55 was isolated from a human during an outbreak in Nigeria [317] and has been used in previous vector competence studies [217, 245]. The DENV-2 isolate was from *Ae. luteocephalus* in Kedougou, Senegal (Table 4.2). YFV BA-55 and DAK 1279 were inoculated into Vero cells and DENV-2-75505 was inoculated into C6/36 cells to generate a virus stock. C6/36 cells were used for DENV-2-75505 to insure virus recovery because this isolate came from a mosquito and there were no data on the titer of the initial stock. The supernatants from these infections were clarified and aliquoted and stored in minimum essential medium (MEM) with 20% fetal bovine serum (FBS) at -80^{o}C for all subsequent use. Growth curves were performed on all viruses to determine the optimal number

of days post-infection on which to harvest the virus for mosquito oral infection. For mosquito feeds, C6/36 cells were infected with DENV-2-75505 at a multiplicity of infection (MOI) of 0 .001 or YFV BA-55 and DAK 1279 at an MOI of 0.01 and grown to a titer of approximately 6.5-7 logs of plaque forming units (PFU)/ml before being mixed 1:1 with defibrinated sheep blood for a titer of approximately 6-6.5 logs of plaque forming units (PFU)/ml for the mosquito feeds. The medium was removed from the C6/36 cells infected with DENV-2-75505 six days post-infection and replaced with fresh medium. Final virus was harvested 12 days post-infection and fed directly to mosquitoes. Both isolates of YFV were grown in C6/36 cells for five days before being harvested and fed directly to mosquitoes.

Table 4.2: Viral Isolates

	Isolated from	Location	Year	Passage History
DENV-2-75505	*Aedes luteocephalus*	Kedougou, Senegal	1990	AP61 p6, C6/36 p4, Vero p1
DENV-2-JAM1409	Human	Jamaica	1983	C6/36 >25 times
YFV BA-55	Human	Nigeria	1986	Suckling Mice p2, Vero p1
YFV Dak1279	Mosquito	Diourbel, Senegal	1965	Suckling Mice p6

Mosquito Infections

Groups of approximately 30 five-to-seven-day-old female mosquitoes from each colony were exposed to a bloodmeal containing approximately six to seven logs of virus for 30 minutes. The bloodmeal titer was determined from blood removed before the feed. Almost all of the mosquitoes fed within 10 minutes, so the pre-feed blood represents the bloodmeal titer ingested. Females that were not completely engorged and males were removed from the study immediately following the bloodmeal. The fully engorged female mosquitoes were held for 14

days at 28°C, 70-80% relative humidity, 12:12 hour photocycle, and were fed water and raisins under BSL3 containment. After 14 days, legs, heads/thoraces, and midguts were separated into individual tubes. Each tissue was triturated in 100 µL (legs and midguts) or 200 µL (heads/thoraces) minimum essential medium (MEM) with 20% fetal bovine serum (FBS) and 1.5 µg/ml Fungizone. Virus titer in each sample was determined by plaque assay. The manipulation of DENV-2-JAM1409 is very similar and was previously described [347]

Plaque Assays

All plaque assays were performed on Vero cells in 12-well tissue culture plates. When Vero cells reached 95% confluency, each triturated sample was diluted and added directly to the cells and allowed to incubate for 1 hour at 37°C. After 1 hour, the first overlay (19.6% 10x Earle's Buffered Salt Solution, 63.4% water, 6.6% Yeast extract-Lactalbumin hydrolysate (YE-LAH, 0.02 g/ml Yeast and 0.11 g/ml Lactalbumin hydrolysate), 4% FBS, 2.2% sodium bicarbonate, 0.5 µg/ml Gentamycin, and 2.0 µg/ml Fungizone mixed 1:1 with 2% SeaKem LE agarose was applied. Four days post infection with either YFV isolate, or 7 days with DENV-2-75505, the second overlay (same as the first overlay plus 2.0 ml of 0.33% Neutral Red (Sigma-Aldrich, St Louis, MO, N2889) per 100ml overlay) was applied. Plaques were counted for 3 days following the addition of the second overlay. The plaque assays for DENV-2-JAM1409 were similar and have been previously described [347].

Data Analysis

For each collection, disseminated infection (DI), midgut escape barrier (MEB) rates, and midgut infection barrier (MIB) rates were calculated. The number of mosquitoes with a DI is the number of mosquitoes with virus in the head/thorax or legs divided by the total number of bloodfed mosquitoes. The MEB rate is the number of mosquitoes without virus in the head/thorax or legs

divided by the total number of midgut infected mosquitoes, and MIB is the number of mosquitoes without virus in the midgut divided by the total number of bloodfed mosquitoes. Proportions of infected mosquitoes were compared among collection sites (villages) and among viral isolates by calculating Bayesian 95% Highest Density Intervals (95% HDI) using WinBUGS 1.4 [446] and an analysis of contingency tables script (Box 6.13 in [447]). Mean virus titers were compared among isolates at each of four locations using a one way ANOVA script (Box 6.1 in [447]) run with WinBUGS and comparing 95% HDI among isolates. We used two-way ANOVA in R [model = summary(aov (Midgut ~ Virus*Village))] to test for significant virus by village interactions. Correlation analyses were also performed in R [cor.test (BloodmealTiter, ProportionDisseminated)].

Results

Vector competence was measured in 647 mosquitoes and each village-virus combination represented 14-30 mosquitoes (mean = 28 and median = 30). Due to the necessity of using freshly grown YFV [258] and DENV, it was difficult to obtain identical titers in all bloodmeals. Nevertheless, bloodmeal titers were not correlated with the midgut infection (MI) rate or the disseminated infection (DI) rate for YFV BA-55 (MI: Pearson Correlation Coefficient (r) = 0.008; DI: r = 0.18). Correlation analysis was not performed for the other two viruses because only one or two bloodmeal titers were used in the mosquito feeds. In some collections, legs were used instead of the head/thorax to measure disseminated infection with YFV BA-55. But the tissue analyzed did not affect either comparisons of infection frequencies or titers. The infection frequencies (r = 0.97, P = 0 .03) and YFV BA-55 titers (r = 0.87, P = 0.01) whether assaying legs or heads were strongly correlated. This correlation analysis was performed on infection data from head/thorax or legs of mosquitoes from the same collection sites infected with YFV BA-55.

Vector competence throughout Senegal

Vector competence varied among collections and viral isolates (Figure 4.1, Table 4.3). The proportion and distribution of mosquitoes with a DI, MEB, and MIB infected with YFV BA-55, YFV DAK1279, or DENV-2-75505 varied across Senegal (Figure 4.1). Collection sites in the forested area of southeast Senegal were more refractory to YFV BA-55 than the collection sites in western Senegal with the exception of collections in urban sites (Mont Rolland and Rufisque) close to or within the capital of Dakar. Other collections from western Senegal (Richard Toll, Fatick, and Bignona) had appreciable, albeit variable, DI due to a low frequency of both MIB and MEB. Mosquitoes from Fatick in particular had a 100% midgut infectin (MI). In contrast, when infected with YFV DAK1279 the same populations of mosquitoes were highly refractory. Only five mosquitoes from Richard Toll and one mosquito from Mont Rolland developed a DI.

Table 4.3: Proportions of *Aedes aegypti* with YFV or DENV-2 infected midguts (MI) or disseminated infections (DI).

Bloodmeal titer was determined by plaque assay at the time of the mosquito feeds. Midgut infection (MI) and disseminated infection (DI) was determined by plaque assay 14 days post infection. Disseminated infections were assessed in either the head/thorax or the legs. * Disseminated infection determined using legs.

	YFV BA-55			DENV-2-75505			YFV-DAK1279		
	Titer (log_{10}(PFU/ml)	MI #positive/N (%)	DI #positive/N (%)	Titer(log_{10}(PFU/ml)	MI #positive/N (%)	DI #positive/N (%)	Titer(log_{10}(PFU/ml)	MI #positive/N (%)	DI #positive/N (%)
Fatick	6.22	17/17(100)	10/17(59)*	6.02	21/23(91)	14/23(61)	5.90	3/18(17)	0/18(0)
Bignona	6.22	25/30(83)	4/30(13)*	6.02	7/14(50)	4/14(29)	7.79	10/30(33)	0/30(0)
Richard Toll	6.32	17/30(57)	3/30(10)*	6.02	21/30(70)	8/30(30)	7.79	17/30(57)	5/30(17)
Goudiry	6.04	16/30(53)	0/30(0)*	6.02	10/18(56)	7/18(39)	5.90	3/30(10)	0/30(0)
Kedougou	5.34	8/23(35)	0/20(0)	6.02	26/30(87)	18/30(60)	5.90	3/29(10)	0/29(0)
PK10	6.04	8/30(27)	1/30(3)*	6.02	26/30(87)	17/30(57)	5.90	6/27(22)	0/27(0)
Mont Rolland	6.2	8/30(27)	0/30(0)	7.00	25/30(83)	28/30(93)	5.90	6/30(20)	1/30(3)
Rufisque	6.13	5/30(17)	0/30(0)*	6.02	25/30(83)	10/30(33)	5.90	3/28(11)	0/28(0)

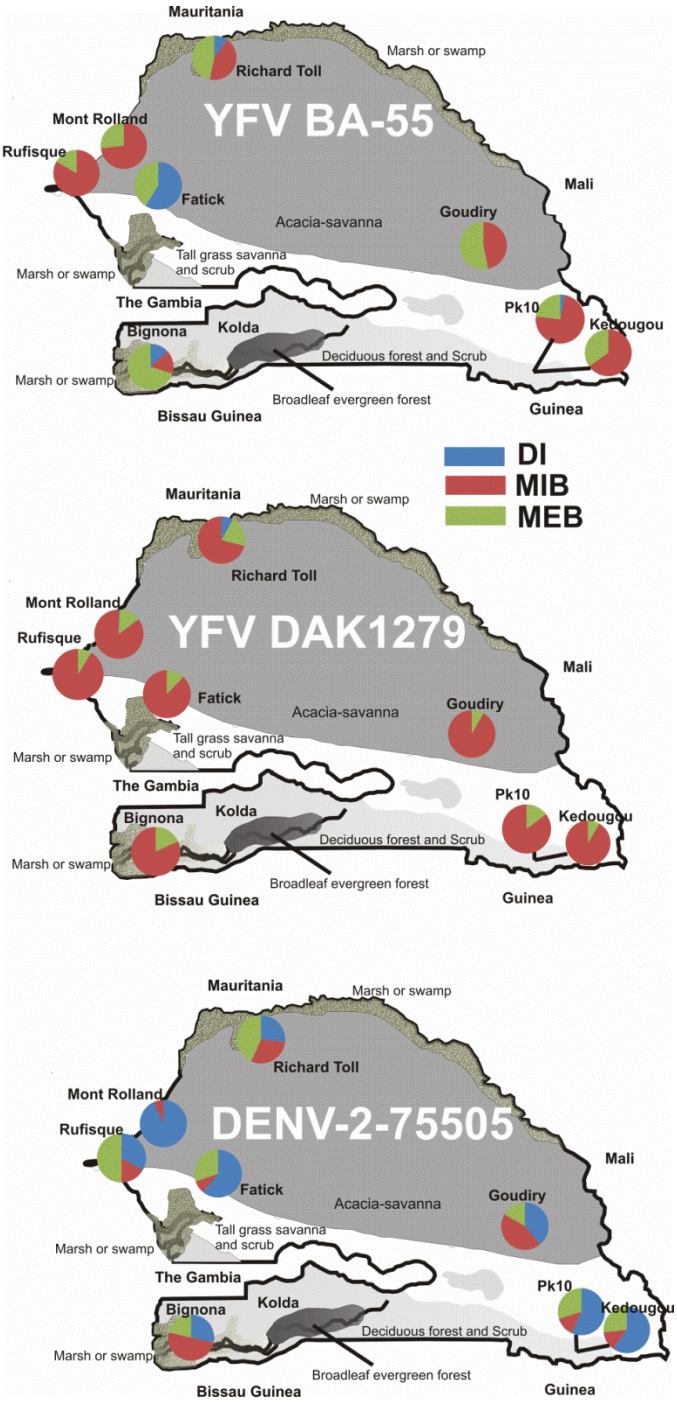


Figure 4.1: Vector Competence for DENV-2 and YFV throughout Senegal
Pie charts represent the proportion of individuals with a disseminated infection (DI), midgut escape barrier (MEB), and midgut infection barrier (MIB). DI is calculated by the number of mosquitoes with virus in the head/thorax or legs divided by the total number of mosquitoes that bloodfed. The MEB rate is the number of mosquitoes without virus in the head/thorax or legs divided by the total number of midgut infected mosquitoes, and MIB is the number of mosquitoes without virus in the midgut divided by the total number of bloodfed mosquitoes.

Infection rates with DENV-2-75505 were different than both isolates of YFV (Figure 4.1). In contrast to Sylla et al. [145], the collections from Kedougou and PK10 had high DI when infected with DENV-2-75505 rather than DENV-2-JAM1409 as previously reported. We confirmed similar infection rates with DENV-2 JAM1409 to those reported in Sylla et al. [145] that used younger generation mosquitoes, indicating high DI rates are not attributed to multiple generations in the lab (Table 4.4). The high DI rate is especially noteworthy because the Kedougou and PK10 collection sites are sylvatic and 84% and 82% respectively of the females collected had no scales on the first abdominal tergite, thereby classifying them as classical *Ae. aegypti formosus* based on the McClelland scale [388]. Notably, all collections that were tested with DENV-2-75505 developed a disseminated infection. There was no clear association with geographic location and susceptibility. Note that the Mont Rolland and Rufisque sites in the west and Kedougou and PK10 in the southeast that were refractory to YFV BA-55 and YFV DAK1279 had high number of individuals with a DI with DENV-2-75505.

Table 4.4: Proportions of *Aedes aegypti* with DENV-2-JAM1409 infected midguts (MI) or disseminated infections (DI).
Disseminated infections were assessed in head/thorax. Infection status was determined by plaque assay 14 days post infection.

	Titer (log_{10}(PFU/ml)	MI #positive/N (%)	DI #positive/N (%)
Mont Rolland	7.52	18/30 (60)	10/30 (33)
Richard Toll	7.52	18/30 (60)	11/30 (36)
Ngari	7.52	25/30 (83)	3/30 (10)
Kolda	7.52	21/30 (70)	15/30 (50)

Mosquito/Virus Specificity

To test for a mosquito strain by virus genotype interaction, the proportion of MIs and DIs were compared among the eight geographically distinct collection sites and three viruses. The proportion of MIs and DIs was dependent on both the virus and collection site (Figures 4.2a and 4.2b), indicating a mosquito strain by virus genotype interaction. Significant differences in the proportion of MIs and DIs were seen in all the collection sites, except for Richard Toll. There was no significant difference in MI rate in mosquitoes from Goudiry infected with DENV-2 75505 or YFV BA-55, but the DI rate was significantly different. Also, MI rates in Bignona were significantly different between YFV BA-55 and YFV DAK1279, but DI rates were not significantly different (Figures 4.2a and 4.2b).

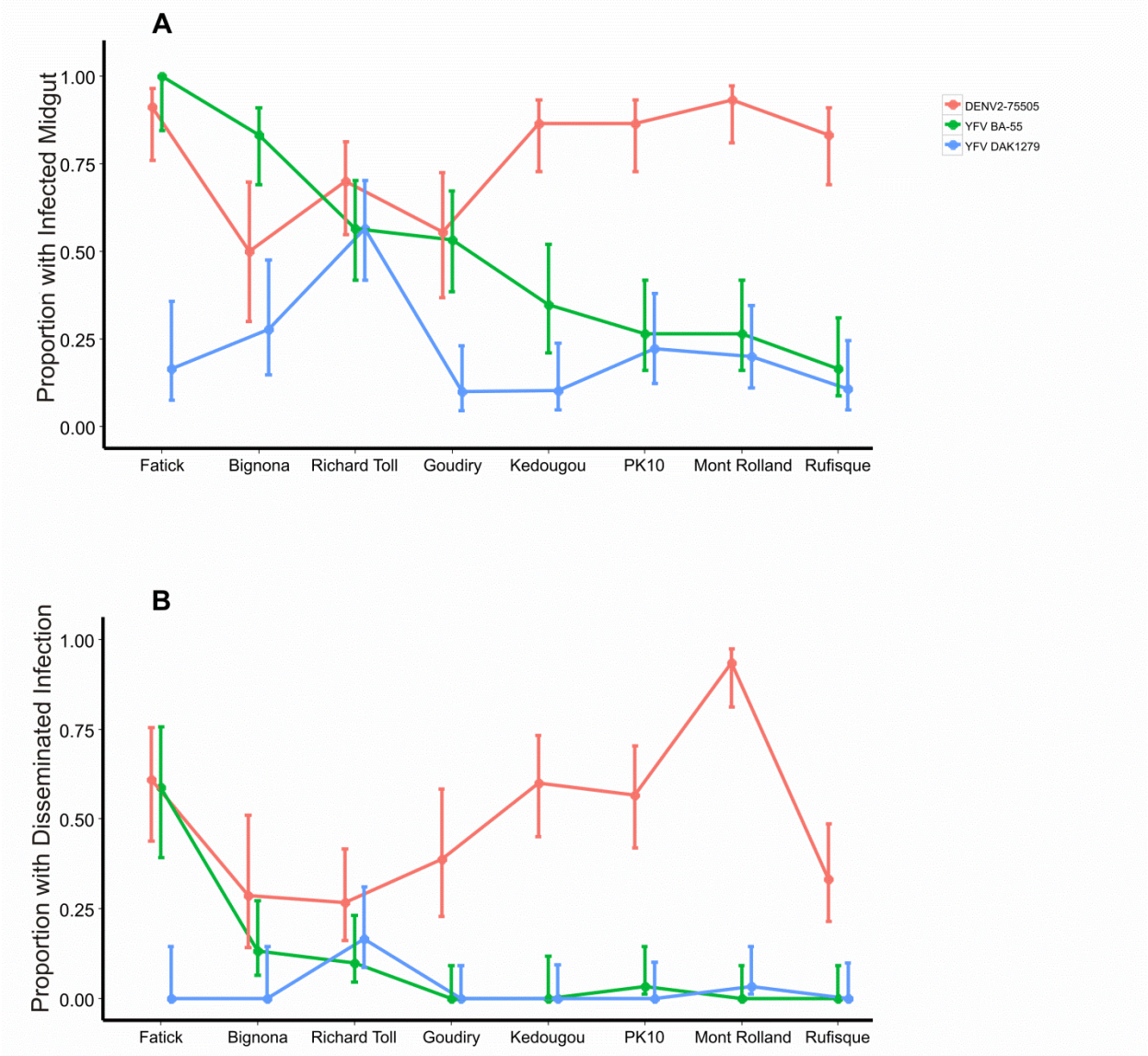


Figure 4.2: Proportion of midgut and disseminated infections
(A) The proportion of individuals with a midgut infection 14 DPI. (B) The proportion of individuals with a disseminated infection 14 DPI. Infection was determined by plaque assay. Error bars represent Bayesian 95% credible intervals and non-overlapping error bars indicate statistical significance.

Although the MI rate was variable among collections and viruses, the titer of virus in mosquitoes that developed a MI were similar among the eight collection sites and the three viruses (Figure 4.3a). The average titer of virus in the midgut was distributed around 3 logs per ml (Figure 4.3a) and ranged from 2.5 to 4.5 logs per ml (Figure 4.4) in individual mosquitoes. In contrast there was a broad distribution in DI (Figure 4.3b) extending from zero up to 5.5 logs per ml and ranging from 1.5 to 5.5 logs per ml (Figure4. 4). The virus titer was dependent on the virus, the collection site, and the virus by mosquito collection site interaction in the midgut (F-statistic = 3.41; P - value = $1.79x10^{-2}$) and in DI (F-statistic = 28.50; P - value = $4.91x10^{-4}$) (Figure 4.3). The virus titer in midguts was not correlated with the virus titer in disseminated infections regardless of viral isolate (Figure 4.4a) or collection site (Figure 4.4b). Only three collection sites (Richard Toll, Bignona, and Fatick) had individuals that developed a DI with all three viruses (Figure 4.4b). In general, as reported earlier [245, 249], the efficiency of viral replication in the midgut does not affect the efficiency of viral replication in other tissues after dissemination.

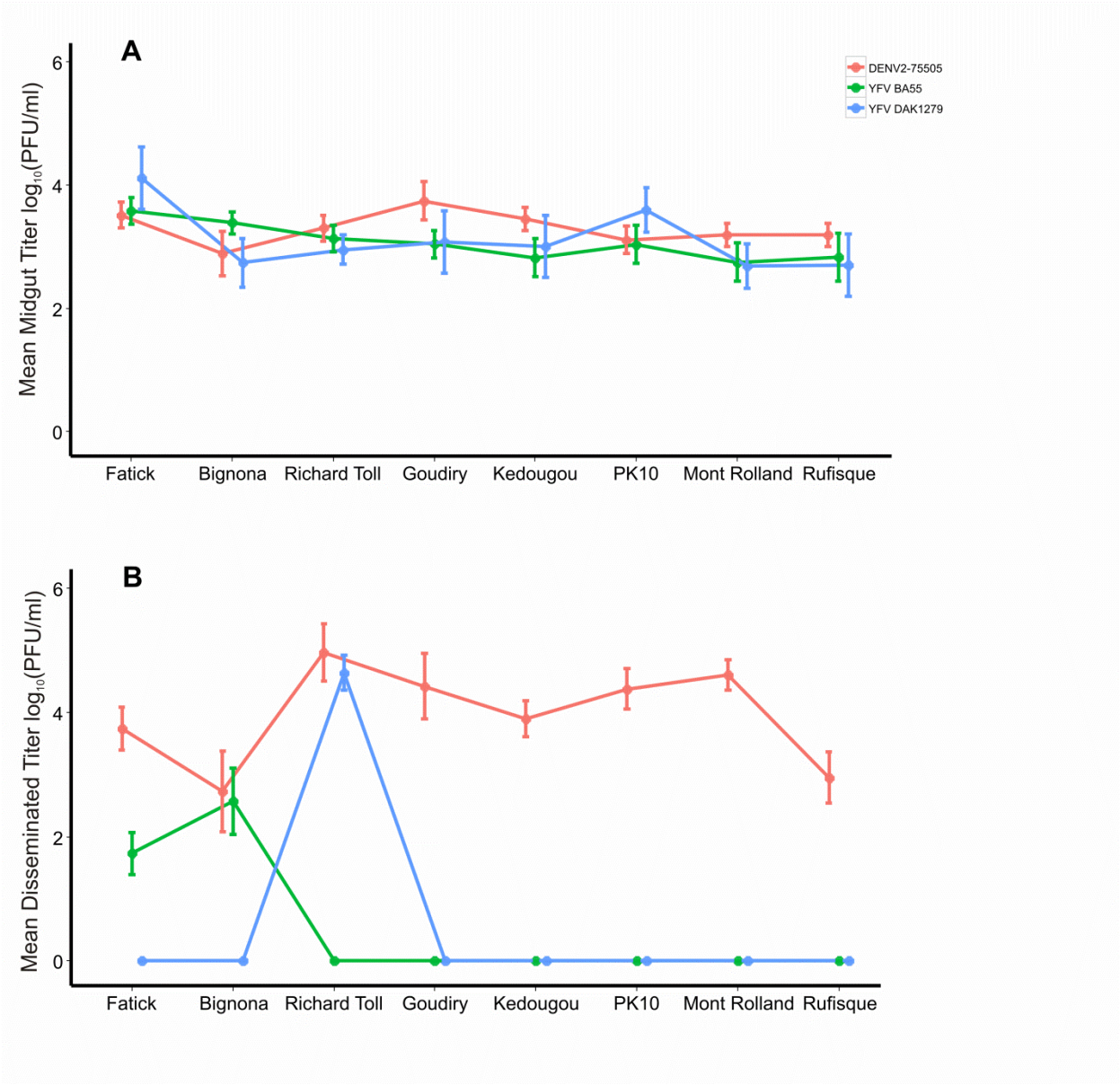


Figure 4.3: Mean virus titer in midgut and disseminated infections
(A) The mean viral titer in the midgut. (B) The mean viral titer in disseminated infections. Virus titers were determined 14 DPI by plaque assay. Error bars represent Bayesian 95% credible intervals and non-overlapping error bars indicate statistical significance.

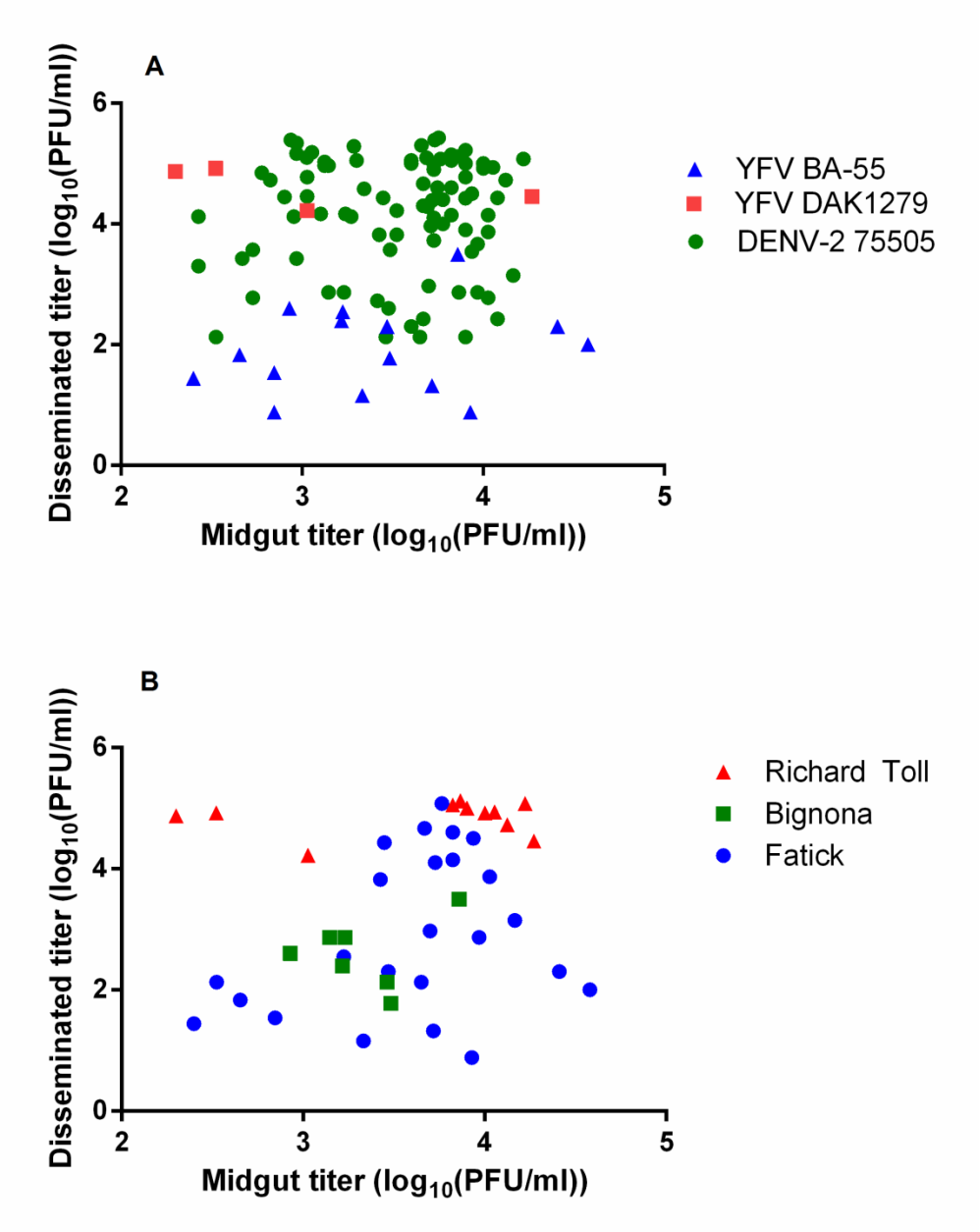


Figure 4.4: The virus titer in the midgut is not correlated with the virus titer in other tissues

Virus titer in midguts (log_{10}(PFU/ml)) was compared with virus titer in DIs (log_{10}(PFU/ml)) from the same individual . The analyses were grouped by viral isolate (A) or collection site (B). The collection sites shown in (B) are the only 3 that had individuals with a disseminated infection with all 3 viruses. Each point represents an individual. Pearson correlation coefficients: DENV-2-75505 (r = 0.041, P = 0.707), YFV DAK1279 (r = -0.607, P = 0.394), YFV BA-55 (r = 0.213, P = 0.446), Fatick (r = 0.309, P = 0.143), Richard Toll (r = 0.156, P = 0.647), Bignona (r = 0.231, P = 0.619).

The mean virus titer of in MIs was correlated with the proportion of individuals within a collection site that developed DIs with YFV BA-55 (r = 0.843; P = 0.0086), but not with YFV DAK1279 or DENV-2-75505 (Figure 4.5). No correlation existed between midgut titer and DI rates when analyzed by collection site. This indicates that MEB and DI rates are independent of the efficiency of viral replication in the midgut, but in a virus dependent manner.

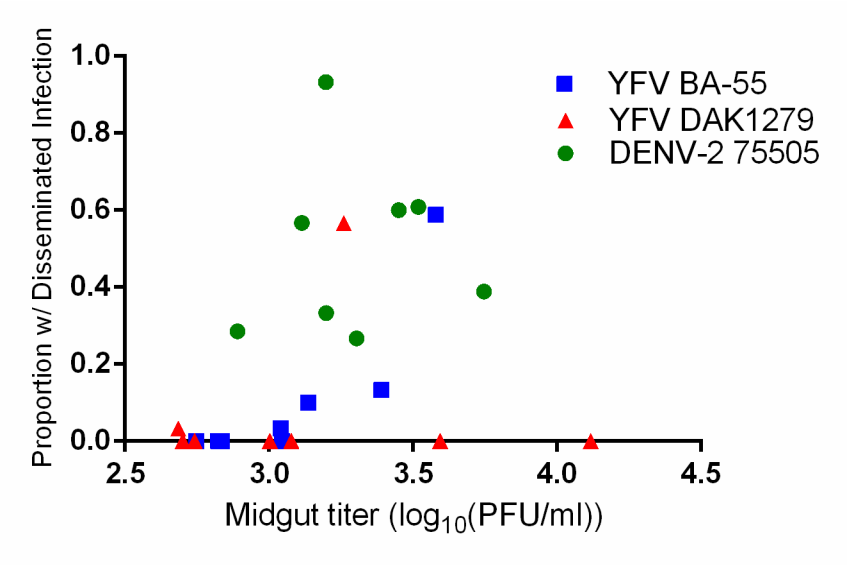


Figure 4.5: The titer of virus in the midgut is correlated with dissemination rate upon infection with YFV BA-55
Mean virus titers (log_{10}(PFU/ml)) in the midgut for each of the 8 collection sites were compared with the proportion of individuals in that collection that developed a disseminated infection using a Pearson correlation analysis. Mean viral titer in the midgut was correlated with DI rate upon infection with YFV BA-55(r = 0.843, P = 0.0086). There was no correlation in YFV DAK279 or DENV-2-75505.

Discussion

We document great variability in vector competence for both DENV-2 and YFV in collections of *Ae. aegypti* from across Senegal. The northwest-southeast decline in the susceptibility to YFV BA-55 is very similar to that seen with DENV-2-JAM1409 [145]. However, contrary to previous work with DENV-2 JAM-1409, the same collections from across Senegal, including sylvatic collections, developed a DI with DENV-2-75505, a sylvatic isolate of

DENV-2 from the same region. Comparison of infection rates and virus titer revealed a mosquito strain by virus genotype interaction in MI and DI. Although there was significant variability in MI rates, the virus titer in the midgut was similar among viruses and mosquito collections. The virus titer in DIs was much more variable among viruses and mosquito collections than MI titers. The efficiency of viral replication in the midgut was not correlated with the efficiency of viral replication in other tissues after DI. The proportion of individuals in each collection that developed a DI was correlated with the efficiency of YFV BA-55 replication in the midgut. In contrast, the proportion of individuals in each collection that developed a DI was not correlated with the efficiency of YFV DAK1279 or DENV-2-75505 replication in the midgut The current study builds on previous work by quantifying differences in infection rates and viral titers between two field isolates of YFV (YFV BA-55 and YFV DAK 1279) and a sylvatic field isolate of DENV-2 (DENV-2-75505) in the same mosquito stains with known genetic diversity [144, 145].

Previous studies on the vector competence of West African *Ae. aegypti* for YFV [245, 316] suggested that West African *Ae. aegypti* are more refractory to YFV infection than *Ae. aegypti* from the Americas and Asia. Tabachnick et al. [316] showed that two *Ae. aegypti* collections from western Senegal, when infected with the Asibi isolate of YFV, were more refractory than collections from the Americas or Asia. Although the Asibi isolate is from Ghana, it had been passaged many times and may not have been representative of isolates involved in natural transmission cycles. Similar to their western Senegal collections, collections close to Dakar were more refractory than other collections throughout the country. Miller and Mitchell [245] showed that an *Ae. aegypti* collection from Nigeria was much more refractory (10% DI) when compared with collections from the Americas (90% DI) to a YFV isolate from Peru and

was completely refractory (0% DI) with YFV BA-55, the same isolate used in the present study. The current study demonstrates that some collections from Senegal also had 0% DI when infected with YFV BA-55, but that other collections did develop DIs. Although it is difficult to draw an informative conclusion by comparing results obtained with mosquitoes from Nigeria with the results obtained here, the discrepancy in vector competence between the two populations with the same viral isolate further demonstrates specificity of the virus/mosquito interaction for vector competence.

A number of studies of *Ae. aegypti* from Senegal have examined vector competence for DENV-2 [131, 145, 216, 249, 316] Those results differ from those presented here. When compared with *Ae. aegypti* from the Americas or Asia, West African *Ae. aegypti* were less susceptible to DENV-2 JAM1409 [249, 316]. In studies directly comparing mosquito collections within Senegal [131, 145, 216], there was variation in susceptibility, but the sylvatic *Ae. aegypti* were more refractory than domestic *Ae. aegypti*. However, the study by Sylla et al.[145] only examined the highly passaged DENV-2-Jam1409 isolate. Measuring vector competence in *Ae. aegypti* with a viral isolate collected in proximity may be the most informative approach [256]. Diallo et al. [131, 216] did so by examining the vector competence of *Ae. aegypti* from Senegal with multiple local isolates of DENV-2. Diallo et al. [131] reported that sylvatic *Ae. aegypti* collections have lower infection rates than other sylvatic species of *Aedes (Ae. furcifer, Ae. luteocephalus, and Ae. vittatus),* but some sylvatic *Ae. aegypti* mosquitos developed a DI. Diallo et al. [216] reported low levels of midgut infection (0.0 - 26.3 %) and variable disseminated infection (0 – 100 %) in six collections from Senegal regardless of geographic location. Importantly, both studies demonstrated variability in infection rates based on the isolate of DENV-2 and the mosquito collection site.

The high rates of infection with DENV-2-75505 in sylvatic collections in the current study are not congruent with the low rates of infection in the previous studies [131, 145, 216]. An explanation for our differing results could be a result of multiple generations in the lab. Unlike Diallo et al.[216], we were unable to get F_1 mosquitoes collected from the field to survive, mate, or bloodfeed sufficiently to perform vector competence assays. High DI rates as a result of lab adaptation are unlikely because similar DI rates with DENV-2-JAM1409 confirmed that the older generation mosquito collections used in this study had similar vector competence as our younger generation mosquitoes previously reported [145]. Another explanation could be that this sylvatic DENV-2-75505 isolate is more infectious in these mosquito populations than isolates used by Diallo et al. [216], highlighting the importance of the viral isolate in vector competence assays.

Although sylvatic isolates of DENV-2 are not traditionally thought to have large public health implications, infection with sylvatic DENV-2 has resulted in DHF in West Africa [118], and symptoms associated with sylvatic DENV and epidemic DENV are indistinguishable [117]. The degree of human exposure to sylvatic DENV, as well as the role of humans in the amplification of sylvatic DENV is not well characterized. DENV is thought to have originated as a zoonotic virus emerging from a sylvatic cycle, but interestingly, sylvatic DENV-2 does not require adaptation to human hosts to replicate efficiently, suggesting that the re-emergence potential of sylvatic isolates into endemic cycles is high [119]. Sylvatic DENV-2-75505 used in this study was isolated from *Ae. luteocephalus* and its ability to infect humans has not been confirmed.

These populations were not screened for insect only viruses, therefore we cannot rule out the possibility they are present or if co-infection with insect only viruses is contributing to our

observed differences in vector competence. The role of insect only viruses in vector competence is currently unresolved in *Ae. aegypti*, however *Cx. pipiens* infected with Culex flavivirus had a significantly lower proportion of mosquitoes develop a disseminated infection with WNV (7 days post infection (DPI)), but not at 14 DPI [448].

Local adaptation between the virus and mosquito vector has been documented before and demonstrates the need to use a viral isolate circulating in the same geographic region when making assumptions about vector competence. Lambrechts et al. [256] demonstrated that differences in vector competence among three *Ae. aegypti* collections from Thailand infected with three genotypes of DENV-1 was a result of mosquito genotype by virus genotype interactions. The current study builds on this and demonstrates mosquito genotype by virus genotype interactions also occur with sylvatic DENV-2 and with YFV.

Lambrechts et al. [256] and Bosio et al. [249] showed that the virus titer in the midgut did not correlate with proportion of disseminated infections or the virus titer outside the midgut. We found a slight correlation between the efficiency of viral replication in the midgut and the proportion of DIs. But this correlation was dependent on the virus. It is possible that the different viruses interact differently with the mosquito innate immune response, or that different genes in the mosquito are involved in the immune response to YFV or DENV-2. The use of plaque assays in the current study to quantify virus may result in different results than quantitative RT-PCR [256] or $TCID_{50}$ [249] due to differences in how the number of infectious virus particles are being measured and counted in various cell types.

The lack of variation in titers in the midgut compared to more variation in other tissues observed here could provide interesting insights into differing mechanisms of vector competence in different tissues. These results could point to different genes or different mechanisms in the

mosquito being involved in viral defense inside the midgut as compared with the tissues outside the midgut.

An advantage of the current study is that we were able to quantify virus in the midgut and disseminated infections through the use of plaque assays while other studies have only compared infection rates. Looking at quantitative differences in viral titers creates a more complete picture about how different viruses are interacting with different mosquito collections. The mechanisms underlying *Ae. aegypti*/virus interactions remain unclear, but these results suggest virus specific mechanisms. Examining more collections throughout Senegal as well as infecting these collections with a non-sylvatic isolate of DENV-2 from Senegal may provide more insight into the mechanisms. Genetic association studies with different collections, different viruses, and different tissues might provide clues as to whether different genes in *Ae. aegypti* play a role in various vector competence phenotypes when infected with different viruses.

CHAPTER 5: Dicer2 DIVERSITY IN NATURAL POPULATIONS OF *AEDES AEGYPTI* AND ITS ROLE IN VECTOR COMPETENCE

Introduction

Aedes aegypti is the primary vector of yellow fever virus (YFV) and all four serotypes of dengue viruses (DENV1-4), as well as being a known vector of chikungunya virus. Dengue remains an important public health problem with an estimated 390 million infections per year [2]. Yellow fever outbreaks are still continuously reported even though a safe and effective vaccine exists. Great variation in vector competence is seen between diverse mosquito collections [169, 245, 249], within closely related collections of *Ae. aegypti* [207-209, 257], and different isolates of DENV-2 [131, 216]. *Aedes aegypti* vector competence for DENV-2 is a quantitative genetic trait [249] and multiple genetic loci confer resistance to DENV infection [158-161].

Invertebrates lack interferon - α/β or type I induced pathways, which are the major pathways that respond to viral infection in vertebrates. Instead, invertebrates primarily control viral infection through RNA interference (RNAi) [318] which results in the cleavage of viral mRNA. In the context of viral infection, the exogenous small-interfering RNA (exo-siRNA) pathway is initiated by the recognition of cytoplasmic dsRNA generated during viral replication by the RNaseIII enzyme Dicer-2 (Dcr2). Flaviviruses such as DENV have positive-sense single-stranded RNA (+ssRNA) genomes. During replication of the viral genome, dsRNA replication intermediates are formed in the cytoplasm [326], and serve as a pathogen associated molecular pattern (PAMP) that is recognized by Dcr2. Dcr2 then cleaves the long dsRNA into siRNA duplexes which are usually 21 bp in length and usually have 2 nucleotide 3' overhangs. In association with Dcr2 and another dsRNA binding protein, R2D2, siRNA duplexes are loaded

into the RNA-induced silencing complex (RISC), which contains Argonaute-2 (Ago2) [327-329], the major catalytic component of the RISC [329, 330]. After unwinding of the siRNA duplex within the RISC, one of the siRNA strands, known as the passenger strand, is degraded, while the other strand, known as the guide strand, is retained and leads the RISC to viral mRNA via sequence complementarity. This results in cleavage of the target transcript via the slicer endonuclease activity of Ago2 [329]. Thus, the exo-siRNA response inhibits viral replication by degrading viral mRNA. In *Ae. aegypti*, the importance of the RNAi pathway in controlling DENV infection has been established by dsRNA-silencing of Dcr2, Ago2, and R2D2 [346, 347, 350, 353] resulting in an increase in viral load and a shortened extrinsic incubation period.

Dcr2 is responsible for recognition and processing of dsRNA into siRNAs [354, 355]. Dcr2 is a large protein that includes a putative dsRNA binding domain, two helicase domains (DExD/H-box and HELICc), a PAZ domain, DUF283 (an alternative dsRNA-binding fold), and two catalytic RNase III domains[356]. These diverse domains allow Dcr2 to function in recognition, cleavage, and processing of dsRNA. In *Drosophila*, the helicase domain is required for biogenesis of siRNAs from long dsRNA *in vitro* and *in vivo* [354, 357]. Interestingly, dsRNA with a 5' overhang or blunt termini require the helicase domain for efficient processing, while dsRNAs with a 3' overhang do not [358].

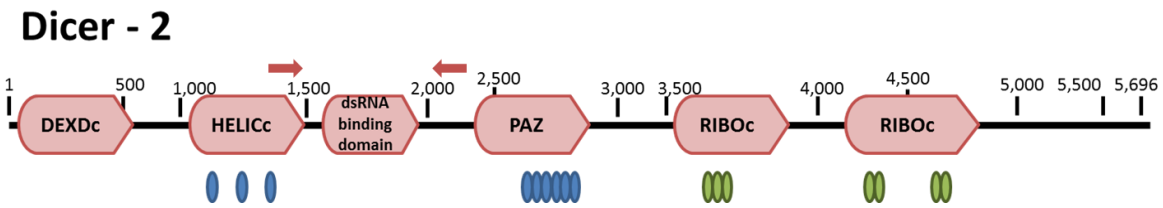


Figure 5.1: Major Protein Domains encoded by the Dcr2 gene
In *Ae. aegypti*, the Dcr2 protein has multiple helicase domains (DEXDc and HELICc) at the N-terminus and the 2 catalytic RNase III domains (RIBOc) at the C-terminus (adapted from [371]). Blue ovals represent nucleic acid binding interfaces, green ovals represent active sites, and numbers represent nucleotide position on cDNA. The location of the primers (red arrows) show the region of Dcr2 that was amplified. It spans the putative dsRNA binding domain (pfam03368) in exon 8 and two flanking exons (exons 6 and 7).

Genetic specificity of the putative dsRNA binding domain in exon 8 of Dcr2 in *Ae. aegypti* for different genotypes of DENV-1 has been demonstrated [449], highlighting the importance of both mosquito and viral genotypes in vector competence. Nucleotide diversity of the Dcr2 gene was correlated with refractoriness of DENV-2 midgut escape (midgut escape barrier) in a population of *Ae. aegypti* from Mexico [371], but the genetic diversity and vector competence assays were assessed in different groups of individuals.

Although evidence exists demonstrating the genetic specificity of the RNAi response in *Ae. aegypti*, as well as a suggested role of genetic diversity of Dcr2 in vector competence, there is limited knowledge about the structure of genes in the RNAi pathway in natural populations of *Ae. aegypti*, and their relationship to vector competence. In this study, we test the hypothesis that the degree of genetic diversity in Dcr2 varies among natural populations of *Ae. aegypti* and is associated with vector competence phenotypes. Genetic diversity of Dcr2 was measured in natural *Ae. aegypti* populations from Mexico, Thailand, and Senegal. The number of haplotypes and nucleotide differences were calculated from sequences of a putative dsRNA binding domain

of Dcr2 from individual *Ae. aegypti* from Senegal with a known vector competence phenotype to test for an association between Dcr2 diversity and vector competence.

Materials and Methods

Mosquitoes

Aedes aegypti were collected as larvae in 3 locations in Senegal: Kedougou (2011), Richard Toll (2010), and Fatick (2010). The larvae were transported to a temporary local laboratory in Kedougou or Theis, reared to adults, given a bloodmeal, and allowed to oviposit. Eggs from each collection were brought back to Colorado State University and maintained for approximately 10-15 generations before being challenged with an artificial bloodmeal containing virus as described below. Adult mosquitoes were kept in incubators maintained at 28^{o}C with 70-80% relative humidity and a 12:12 hour light:dark cycle. Eggs were collected on filter paper and stored for up to 5 months in a high humidity chamber. The collection sites consisted of domestic sites around huts and in tires in urban environments and rural villages or tires in urban environments. Information about the collection site and the nucleotide sequences of Dcr2 genes from Mexico, Senegal, and Thailand were previously published [371].

Viruses

The DENV-2-75505 used in this study was isolated from a sylvatic mosquito (*Aedes luteocephalus)* from Kedougou, Senegal in 1990. The passage history of DENV-2-75505 is given in Table 4.5 in Chapter 4 and was passaged once in C6/36 cells to generate a virus stock in our lab. The supernatant from this infection was clarified and aliquoted and stored in minimum essential medium (MEM) with 20% fetal bovine serum (FBS) at -80^{o}C for all subsequent use. Growth curves were performed to determine the optimal number of days post-infection on which to harvest virus for mosquito oral infection. For mosquito feeds, C6/36 cells were infected with

DENV-2-75505 at a multiplicity of infection (MOI) of 0.001and grown to a titer of 6.5-7.5 logs plaque forming units per milliliter (PFU/ml) before being mixed with defibrinated sheep blood to a titer of 6.0-7.0 logs PFU/ml for the mosquito feeds. The medium was removed from the C6/36 cells infected with DENV-2-75505 six days post-infection and replaced with fresh medium. Final virus was harvested 12 days post-infection, mixed with blood, and fed directly to mosquitoes.

Mosquito Infections

Groups of approximately 30 five-to-seven-day-old mosquitoes from each colony were exposed to a bloodmeal containing approximately 6.0 to 7.0 logs (log_{10}(PFU/ml)) of DENV for 30 minutes. The bloodmeal titer was determined before the feed. Almost all the mosquitoes fed within 10 minutes, so the pre-feed blood titer represented the bloodmeal titer ingested. Fully engorged female mosquitoes were held for 14 days at 28^{o}C, 70-80% relative humidity, 12:12 hour light:dark cycle, and were provided water and raisins under BSL-3 containment. After 14 days heads/thoraces and midguts were separately harvested into individual tubes. Each tissue was triturated in 100µL (midguts) or 200µL (heads/thoraxes) MEM with 20% FBS and 1.5 µg/ml Fungizone. Virus titer in each sample was determined by plaque assay.

Plaque Assays

All plaque assays were performed on Vero cells in 12-well tissue culture plates. When Vero cells reached 95% confluency, each triturated sample was diluted and added directly to the cells and allowed to incubate for 1 hour at 37^{o}C. After 1 hour, the first overlay (19.6% 10x Earle's Buffered Salt Solution, 63.4% water, 6.6% Yeast extract-Lactalbumin hydrolysate (YE-LAH, 0.02 g/ml Yeast and 0.11 g/ml Lactalbumin hydrolysate), 4% FBS, 2.2% sodium bicarbonate, 0.5 µg/ml Gentamycin, and 2.0 µg/ml Fungizone mixed 1:1 with 2% SeaKem LE

agarose was applied. Seven days post infection, the second overlay (same as the first overlay plus 2.0ml of 0.33% neutral red (Sigma-Aldrich, St. Louis, MO) per 100ml overlay) was applied. Plaques were counted for 3 days following the addition of the second overlay.

Vector Competence Phenotypes

For each collection, disseminated infection (DI) rates, midgut escape barrier (MEB) rates, and midgut infection barrier (MIB) rates were calculated. The DI rate is the number of mosquitoes with infectious virus in the head/thorax or legs divided by the total number mosquitoes with infectious virus in the midgut. The MEB rate is the number of mosquitoes without virus in the head/thorax or legs divided by the total number of midgut infected mosquitoes, and MIB rate is the number of mosquitoes without detectable infectious virus in the midgut divided by the total number of bloodfed mosquitoes.

Sequencing

The entire Dcr2 gene (Vectorbase ID: AAEL006794) was sequenced in 20 *Ae. aegypti* individuals from Senegal, 40 individuals from Thailand, and 20 individuals from Mexico as reported in [371]. For the analysis of Dcr2 in infected individuals, total genomic DNA was isolated from individual mosquitoes with known vector competence phenotypes from Senegal (Richard Toll, Fatick, and PK10). Exons 6-8 of the Dicer-2 gene (Figure 5.1) were amplified from genomic DNA by PCR using the following primers:

Forward: 5’ - TTTGCGAAAGCCTGAGGCATCAG-3’,

Reverse: 5’ – GCATCAAAATCCTTCCAATGACGG – 3’

The presence of a single band at the anticipated size of 1154 basepairs in the PCR product was verified by electrophoresis on an agarose gel. The amplicon was purified from the PCR reaction using the Qiagen PCR Purification Kit. Amplicons were sequenced using Sanger sequencing at

Functional Biosciences in Madison WI. Sequences were trimmed and aligned using the Geneious Software program (Geneious version 7.1.7 created by Biomatters. Available from http://www.geneious.com/.

Sequence Analysis

The association between individual segregating sites and the MIB, MEB, and DI phenotypes was determined using the PGenome package (http://pgenome.sourceforge.net/). In this package, segregating sites are tested for association with a specific phenotype using the Doerge and Churchill Method [450] and the genotype at each segregating site is tested for association with a specific phenotype. For the prediction of haplotypes, the PHASE, (a group of SNPs that occur together in cis), of each sequence was determined using the DNAsp50 Software program [451]. DNAsp was used to estimate the most likely haplotypes, calculate the various measures of diversity (haplotype diversity and π [452], equations 8.4, 8.12 and 10.5), as well as the variance of haplotype diversity and π. π is the number of nucleotide differences divided by number of nucleotides in the sequence. Homozygous or heterozygous haplotypes were compared between viral titers in the midgut or disseminated infection using ANOVA in R [394]. Odds ratios for homozygous or heterozygous haplotypes generated by DNAsp and MIB, MEB, or DI were determined by Bayes' analysis in WinBugs [446]. A correlation between pairwise measures of genetic diversity at the population level (haplotype diversity and π) and the proportion of individuals in each collection having a MIB, MEB, and DI was determined with the "cor.test" command in R [394].

Results

Genetic diversity in natural populations

Genetic diversity was measured across all ten exons of Dcr2 in 20 individual mosquitoes from Mexico, Thailand, and Senegal. The amount of genetic diversity varied among collections (Figure 5.2). Overall, *Ae. aegypti* from Mexico and Thailand had fewer nucleotide differences across the Dcr2 gene than *Ae. aegypti* from Senegal, indicating that Dcr2 is more diverse in Senegalese *Ae. aegypti*.

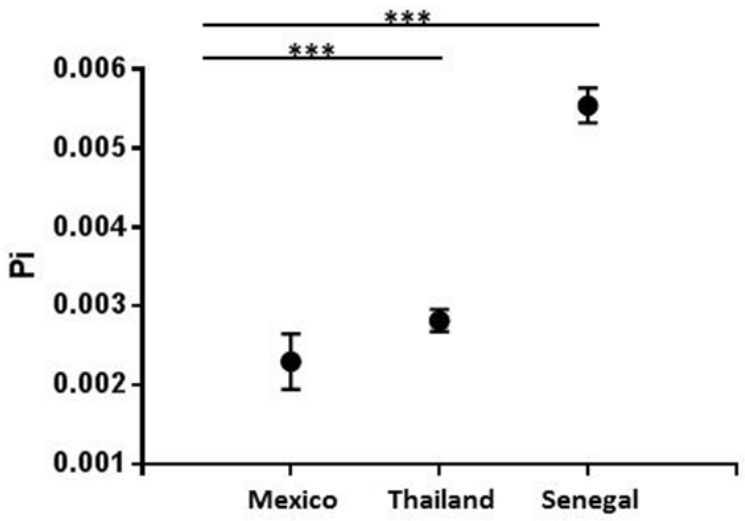


Figure 5.2: Nucleotide diversity (pi) of the Dcr2 gene varied among collection sites.
Nucleotide diversity across the entire Dcr2 gene was measured in 20 individual mosquitoes from each population from Mexico, Thailand, and Senegal. Error bars represent 95% confidence intervals and statistical significance was determined by two-tailed t-test. The sequences used in this analysis were previously published [371]

Genetic diversity in individuals and vector competence

To examine the role of Dcr2 genetic diversity in vector competence, exons 6-8 which contain a putative dsRNA binding domain of Dcr2 were sequenced in 57 individuals from three different collection sites in Senegal with known vector competence phenotypes. A total of 45 segregating sites were identified among the three different collections. The segregating sites

were distributed across all three exons (exons 6-8) with most of them in exon 6 and exon 8. None of the 45 SNPs was associated with MIB, MEB, or DI (Figure 5.3). Also, the genotype of each SNP was not associated with MIB, MEB or DI (Figure 5.4) indicating that SNPs in this region of Dcr2 are not correlated with vector competence for DENV-2.

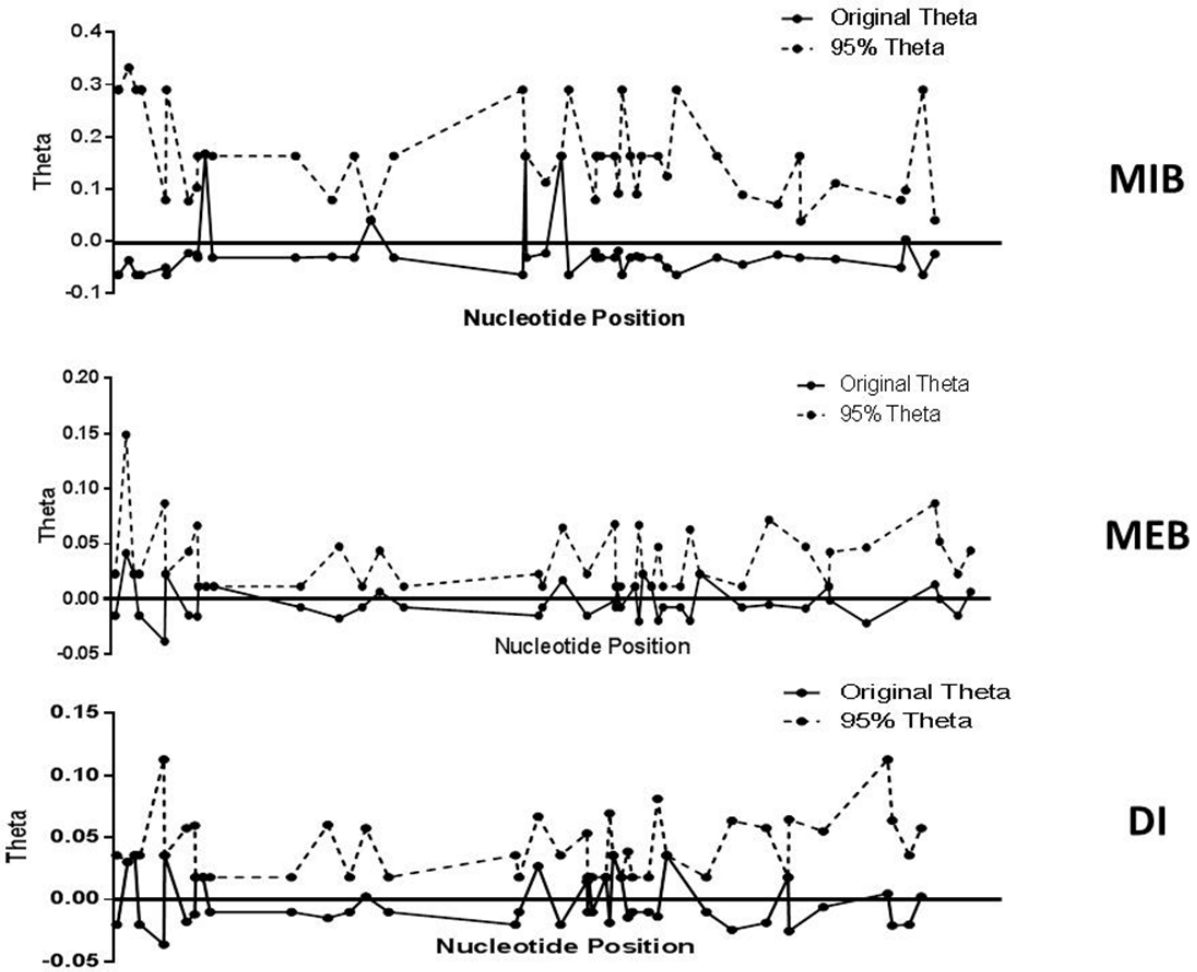


Figure 5.3: Association of SNPs with various vector competence phenotypes
Association mapping of MIB, MEB, and DI based upon nucleotide frequencies. Each segregating nucleotide position was plotted against theta (population mutation rate) [453] for each of the 45 SNPs across exons 6-8 of Dcr2. A putative QTN is identified at a segregating site when the original theta (solid line and circles) exceeded the 95% threshold (dotted line, empty circle). No segregating sites had an original theta value greater than the 95% threshold. Thresholds values for theta were computed using the Doerge and Churchill method [450].

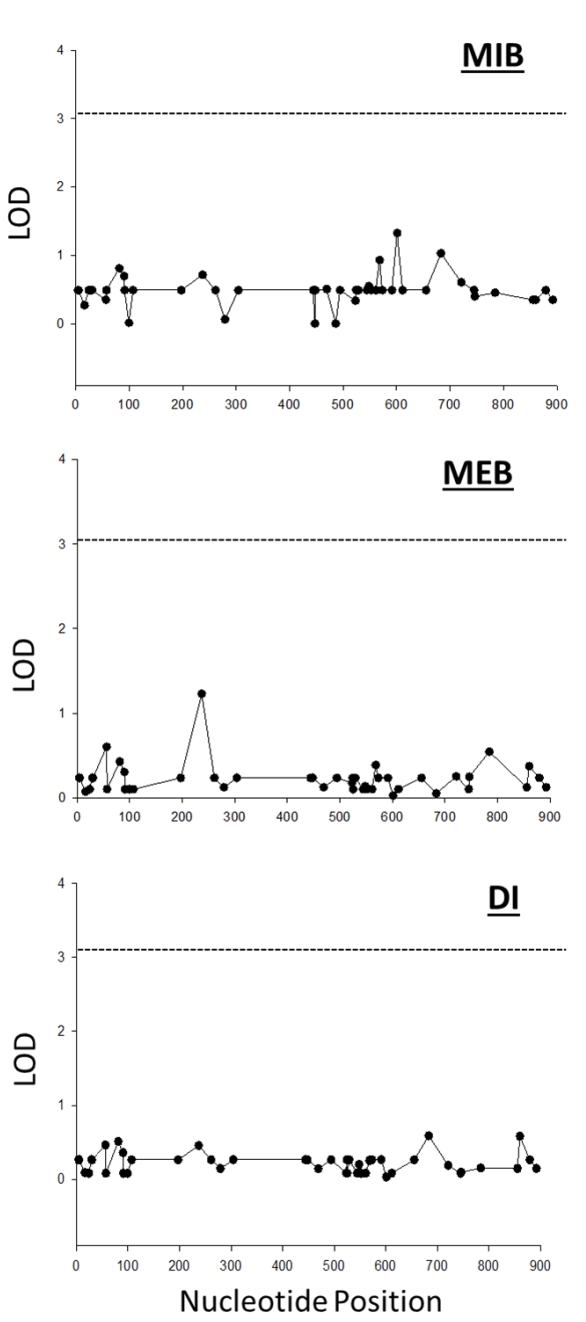


Figure 5.4: Association mapping of various vector competence phenotypes with genotype frequencies.

Contingency χ^2 tests of association between alternate genotypes and the number of mosquitoes with and without an MIB, MEB, or DI was performed. The LOD ($-\log_{10}(p)$ score where p = 1 - the cumulative probability associated with the test χ^2 statistic) was then plotted against nucleotide position at each of the 45 segregating nucleotide positions. The dotted line represents the average 95% threshold value.

The association of individuals with one or two haplotypes (as determined with PHASE in DNAsp, see materials and methods) in individual mosquitoes and vector competence was then examined. The 57 individuals contained 32 different haplotypes, and 61.4% of the individuals contained at least one of the four most common haplotypes. The number of mosquitoes with MEB or DI were not distributed independently of the homozygosity or heterozygosity of these individuals. Individuals that were heterozygous for a given haplotype were more likely to have a MEB resulting in no disseminated infection (Table 5.1), indicating individuals with a single haplotype of Dcr2 (homozygous) are more susceptible than individuals with two haplotypes (heterozygous). No difference in the viral titers in the midgut ($F = 1.778$, $p = 0.19$) or head ($F = 0.621$, $p = 0.44$) was detected between homozygous and heterozygous individuals.

Table 5.1: Odds ratio for heterozygous haplotypes and MIB, MEB, and DI

Haplotypes were predicted with PHASE in DNAsp, and odds ratios were calculated using Bayes analysis. The Bayes analysis shows that there is a 2:1 greater chance of a Dcr2 heterozygote having a midgut infection barrier (MIB+) compared to a Dcr2 homozygote but the 97.5% HDI for the ratio contains 1 so is not credibly different than 1. The Bayes analysis shows that there is a 8.6:1 greater chance of a Dcr2 heterozygote having a midgut escape barrier (MEB+) compared to a Dcr2 homozygote and the 97.5% HDI for the ratio does not contain 1 so is credible. The Bayes analysis shows that there is a 13.2:1 greater chance of a Dcr2 heterozygote having no dissemination infection (DI-) compared to a Dcr2 homozygote and the 97.5% HDI for the ratio does not contain 1 so is credible.

	Median odds ratio	2.5 % credible interval	97.5 % credible interval
Heterozygous: MIB +	2.0	0.2	30.1
Heterozygous: MEB+	8.6	1.2	118.8
Heterozygous: DI-	13.2	1.8	186.8

Genetic diversity in populations and vector competence

Genetic diversity of one putative dsRNA binding domain of Dcr2 in each collection was tested for a correlation with vector competence. π was calculated across all three exons (Figure 5.5) as a measure of nucleotide diversity within a collection. Nucleotide diversity was correlated with MIB and MEB, but not DI. To test for a correlation between diversity in the putative dsRNA binding domain (Exon 8, pfam03368) and vector competence, π was calculated in individual exons (Figure 5.6). Nucleotide diversity in exon 8 was correlated with MEB and DI, but no significant correlation was detected in exon 6 or exon 7. Haplotype diversity was not significantly correlated with any vector competence phenotype regardless whether exons 6-8 were treated as one sequence or analyzed as individual sequences (Figures 5.5 and 5.6).

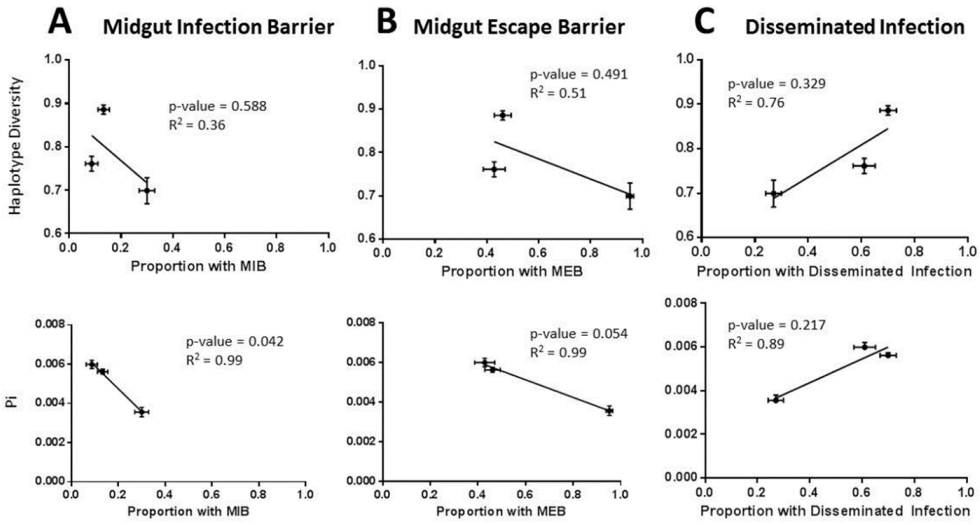


Figure 5.5: Nucleotide diversity across all three exons is correlated with MIB and MEB
Correlation of haplotype diversity and nucleotide diversity (pi) with A) Midgut infection Barrier (MIB), B) Midgut escape Barrier (MEB), and C) Disseminated Infection (DI). Each point represents a separate collection site (Kedougou, Richard Toll, or Fatick) consisting of an average of 20 individual mosquitoes. Error bars represent 95% confidence intervals for both the x and y-axis.

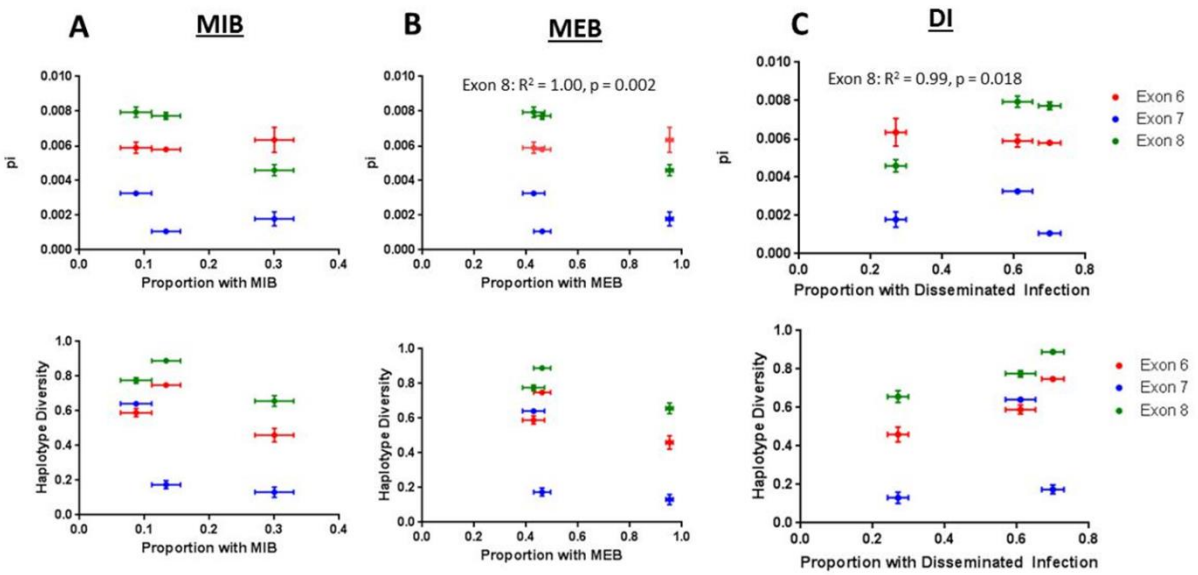


Figure 5.6: Nucleotide diversity in exon 8 is correlated with MEB and DI
Correlation of haplotype diversity and nucleotide diversity (pi) in each exon with A) Midgut infection Barrier (MIB), B) Midgut escape Barrier (MEB), and C) Disseminated Infection (DI). Each point represents a separate collection site (Fatick, Kedougou, or Richard Toll) consisting of an average of 20 individual mosquitoes. The R^2 value and p-values are only shown on graphs in which the p-value was significant. Error bars represent 95% confidence intervals for both the x and y-axis. For pi: Exon 6 (MIB: $R^2 = 0.86$, $p = 0.242$, MEB: $R^2 = 0.95$, $p = 0.145$, DI: $R^2 = 0.93$, $p = 0.165$). Exon 7 (MIB: $R^2 = 0.15$, $p = 0.743$, MEB: $R^2 = 0.062$, $p = 0.839$, DI: $R^2 = 0.00002$, $p = 0.997$). Exon 8 (MIB: $R^2 = 0.98$, $p = 0.095$, MEB: $R^2 = 1.00$, $p = 0.002$, DI: $R^2 = 0.99$, $p = 0.018$). For Haplotype diversity: Exon 6 (MIB: $R^2 = 0.49$, $p = 0.503$, MEB: $R^2 = 0.65$, $p = 0.406$, DI: $R^2 = 0.86$, $p = 0.243$). Exon 7 (MIB: $R^2 = 0.52$, $p = 0.484$, MEB: $R^2 = 0.37$, $p = 0.581$, DI: $R^2 = 0.15$, $p = 0.744$). Exon 8 (MIB: $R^2 = 0.46$, $p = 0.566$, MEB: $R^2 = 0.71$, $p = 0.361$, DI: $R^2 = 0.91$, $p = 0.198$).

Discussion

This study demonstrates that the amount of genetic diversity of the Dcr2 gene varies among natural populations of *Ae. aegypti* from Mexico, Thailand, and Senegal. Haplotype diversity of the putative dsRNA binding domain of Dcr2 in exon 8 in individual *Ae. aegypti* was positively correlated with DENV-2 infection in the midgut, and nucleotide diversity of one putative dsRNA binding domain of Dcr2 in populations of *Ae. aegypti* was negatively correlated with controlling DENV-2 infection in the midgut.

It was not surprising that specific alleles or the genotype of those alleles was not associated with vector competence. Bosio *et al.* [249] reported that 41% of variation in vector

competence was attributed to genes that act additively. The genetic specificity of the *Ae. aegypti*/DENV interaction [449] supports the importance of the mosquito genotype on vector competence, but it is unlikely that a specific genotype of Dcr2 alone confers resistance or susceptibility to flaviviruses. Instead, these data provide evidence for the importance of heterozygosity in individual mosquitoes. Greater genetic diversity in populations will generate greater genetic diversity at the level of individual mosquitoes.

Studies on the evolution of the RNAi genes in natural populations of *Ae. aegypti* have demonstrated that unlike *Drosophila*, both the exo-siRNA and miRNA pathways are undergoing rapid diversifying selection [371]. It is unlikely this is a function of arboviral infection because arboviral infections are rare in nature. Although arbovirus infection may not be driving evolution of the RNAi pathway genes in the mosquito, the genotype of RNAi pathway genes may alter its interactions with the viral population and therefore vector competence. RNAi targeting of the West Nile genome in the *Cx. quinquefasciatus* midgut is associated with increased viral genetic diversity [369] and the increased West Nile genomic genetic diversity results in fitness gains in the mosquito [370]. In contrast, in *Cx. pipiens*, WNV diversity decreased during midgut infection and through the EIP [368]. Increased WNV genetic diversity and subsequent fitness gains may be correlated with increased vector competence [370]. Genetic diversity of Dcr2 could alter its ability to target viral dsRNA and result in variable rates of evolution of the viral genome and subsequent effects on vector competence.

The exact mechanism of Dcr2 detection of DENV PAMPs generated during viral infection remains unknown, but replacement substitutions may affect the enzymatic properties of Dcr2. Lack of genetic diversity within an individual, such as homozygous haplotypes, could result in increased disseminated infection as a result of having a single genotype of Dcr2. One

genotype may not be sufficient to reduce viral infection. Two haplotypes in an individual may cause diversifying selection on the viral population, resulting in a more genetically diverse viral population. This could possibly lower the overall fitness of the viral population by reducing the frequency of the most fit viral genotype. In contrast, the presence of one haplotype in a homozygous individual would create less diversifying selection, allowing more fit viral genotypes to survive, possibly resulting in increased disseminated infection.

Although haplotype diversity in exons 6-8 of Dcr2 was important in individual mosquitoes, it was not correlated with vector competence in a whole population. Interestingly, only increased nucleotide diversity in exons 6-8 of the Dcr2 gene at the population level was correlated with increased DENV-2 infection and dissemination. At first, it seems counterintuitive that decreased genetic diversity in an individual results in increased viral dissemination, while increased genetic diversity in a population results in increased viral dissemination. However, increased nucleotide diversity in a population would increase individual diversity. Within an individual, a single haplotype allows increased disseminated infection, possibly as a result of decreased selective pressure on the viral population compared to the selective pressure applied by two haplotypes.

While these data provide evidence for a correlation between nucleotide diversity of the putative dsRNA binding domain of Dcr2 in exon 8 in a population and vector competence, measuring nucleotide diversity of this region of Dcr2 in a population should not be used to predict vector competence. Instead, we observed that nucleotide diversity was negatively correlated with midgut infection and midgut escape barriers, but positively correlated with disseminated infection. In this study, the mosquitoes from Kedougou had the highest nucleotide diversity in exon 8 of Dcr2 and the highest DI rate. However, mosquitoes from Kedougou have

a very low DI rate when infected with other isolates of DENV-2 [145] or YFV [454]. The vector competence of a group of mosquitoes varies based on the viral genotype [256, 454], indicating the best predictor of vector competence in a collection of mosquitoes is repeated testing with genetically diverse viral genotypes.

Although sylvatic isolates of DENV-2 are not thought to have large public health impacts, infection with sylvatic DENV-2 has resulted in DHF in West Africa [118], and symptoms associated with sylvatic DENV and epidemic DENV are indistinguishable [117]. The degree of human exposure to sylvatic DENV, as well as the role of humans in the amplification of sylvatic DENV is not well characterized. DENV is thought to have originated as a zoonotic virus emerging from a sylvatic cycle, but interestingly, sylvatic DENV-2 does not require adaptation to human hosts to replicate efficiently, suggesting that the re-emergence potential of sylvatic isolates into endemic cycles is high [119]. Sylvatic DENV-2-75505 used in this study was initially isolated from *Ae. luteocephalus* and its ability to infect humans has not been confirmed.

Sequencing the entire Dcr2 gene in more individuals with known vector competence phenotypes will increase the statistical power and provide a more comprehensive view of the role of Dcr2 diversity in vector competence. In the current study, only sequences from exons 6-8 were used because they contain a putative dsRNA binding domain and has been used in previous studies [449]. But, other domains may also play an important role in dsRNA recognition and cleavage, therefore no conclusions about the functional role of Dcr2 diversity in vector competence can be made until more regions of the gene are sequenced. Comparing the correlation of diversity among various functional domains with vector competence phenotypes may provide insights into functional implications of Dcr2 diversity. Also, comparing viral

genetic diversity in mosquitoes with homozygous or heterozygous haplotypes of the Dcr2 gene could provide insights into how diversity of Dcr2 influences viral dissemination. The current study only indicates that mosquitoes differ in dissemination rates when homozygosity or heterozygosity of haplotypes is compared. Comparing the degree to which haplotypes differ could provide more insights. A major deficiency of this study is the ability to draw strong conclusions about the role of Dcr2 diversity in vector competence based on the observed correlation. Further studies determining the transcription/translation efficiency of the different haplotypes as well as whether specific haplotypes confer more or less susceptibility need to be performed to strengthen the conclusion that haplotype diversity has a functional role in vector competence.

CHAPTER 6: CONCLUSIONS

Throughout the process of completing this dissertation, our methods and approaches changed drastically. At first, I approached my first aim of identifying genetic differences between *Ae. aegypti* subspecies by comparing the sequences of a limited number of markers through the use of Sanger sequencing. During the process, we shifted to deep sequencing of the entire genome and exome. Although much effort was spent deciding on and developing our analysis pipeline, it turned out to be very fruitful and informative in the end. The major conclusions from this dissertation follow.

Major conclusions from this study include 1) measuring F_{ST} differences across the whole genome instead of a limited number of markers can be more informative, 2) comparing the distribution of F_{ST} values instead of point estimates of F_{ST} provides more information about population differences as well as provides a tool for identifying highly conserved or divergent genes between populations, and 3) the distribution of F_{ST} values changes based on gene region and substitution type indicating that some regions of the genome are not as informative as other regions. Overall, these data suggest allopatric speciation between populations from Senegal and Mexico and Thailand, and sympatric speciation among populations within Senegal.

In an effort to identify discrete subspecies-specific genetic differences, variations in the SDL were analyzed. The most important finding of this study was the differences in the SDL in *Ae. aegypti* subspecies. Variable linkage of the SDL with the white-eye locus in *Ae. aegypti formosus* males was shown through a classical crossing experiment. A pattern of sex-specific heterozygosity was identified, which would be predicted based on early studies demonstrating a dominant/recessive aspect to sex determination in *Ae. aegypti* [149, 421]. Importantly, the pattern of sex-specific heterozygosity in the SDL was reciprocal in *Ae. aegypti formosus* males

and *Ae. aegypti aegypti* females. Close examination of SNPs in the SDL revealed that 3 out of 3 replacement sex-specific SNPs are heterozygous in *Ae. aegypti formosus* males and homozygous in *Ae. aegypti aegypti* males. The SNP differences also correlate with haplotype signatures at these loci. The variable linkage appears to be the result of variations in the copy number of the SDL, as demonstrated by analysis of the deep sequencing data and validated by qPCR

We documented great variability in vector competence for both DENV-2 and YFV in collections of *Ae. aegypti* from across Senegal. The northwest-southeast cline in the susceptibility to YFV BA-55 is very similar to that seen with DENV-2-JAM1409 [145]. However, contrary to previous work with DENV-2-JAM 1409, the same collections from across Senegal, including sylvatic collections, developed a DI with DENV-2-75505, a sylvatic isolate of DENV-2 from the same region. Comparison of infection rates and virus titer revealed a mosquito strain by virus genotype interaction in MI and DI. Although there was significant variability in MI rates, the titer of virus in the midgut was similar among YFV BA-55, YFV DAK1279, and DENV-2-75505 and mosquito collections. The virus titer in DIs was much more variable among YFV BA-55, YFV DAK1279, and DENV-2-75505 and mosquito collections than MI titers. The efficiency of viral replication in the midgut versus other tissues following DI was strongly dependent on the virus and the mosquito collection site. The current study builds on previous work by quantifying differences in infection rates and viral titers between two field isolates of YFV (YFV BA-55 and YFV DAK 1279) and a sylvatic field isolate of DENV-2 (DENV-2-75505) in the same mosquito stains with known genetic diversity [144, 145].

Nucleotide diversity in a putative dsRNA binding domain of the Dcr2 gene is greater in Senegal compared to Mexico and Thailand. Specific single nucleotide polymorphisms (SNPs) or the genotype of the SNP was not correlated with viral infection or dissemination and the number

of haplotypes of Dcr2 was not correlated with the titer of DENV-2 in the midgut or disseminated infection. The number of haplotypes of Dcr2 in individual mosquitoes and nucleotide diversity (pi) of one dsRNA binding domain of Dcr2 in populations of mosquitoes was correlated with escape of DENV-2 from the midgut resulting in a disseminated infection. Overall, these data demonstrate that diversity of Dcr2 in natural populations of *Ae. aegypti* from Senegal is associated with disseminated infection of DENV-2.

This dissertation demonstrates the genetic complexity of *Ae. aegypti* in West Africa and highlights the need to re-evaluate our current classification system in order to change the way we study and draw conclusions about the vector competence in natural populations of these mosquitoes. Future work should aim to address what types of genes are involved in the speciation of *Ae. aegypti*, elucidating the niche preference and assortative mating of groups of *Ae. aegypti* with variable linkage of the sex-determining loci, and the contribution of different genes in different sub-populations of *Ae. aegypti* to vector competence for different flaviviruses and different genotypes of the same flavivirus.

REFERENCES

1. Baltimore, D., *Expression of animal virus genomes.* Bacteriological reviews, 1971. **35**(3): p. 235-41.
2. Bhatt, S., et al., *The global distribution and burden of dengue.* Nature, 2013. **496**(7446): p. 504-7.
3. CDC. 2009; Available from: http://www.cdc.gov/dengue/clinicalLab/caseDef.html.
4. Organization, T.W.H. *Dengue and severe dengue*. 2014; Available from: http://www.who.int/mediacentre/factsheets/fs117/en/.
5. Gething, P.W., et al., *Estimating the Number of Paediatric Fevers Associated with Malaria Infection Presenting to Africa's Public Health Sector in 2007.* Plos Medicine, 2010. **7**(7).
6. Halstead, S.B., S. Rojanasuphot, and N. Sangkawibha, *Original antigenic sin in dengue.* The American journal of tropical medicine and hygiene, 1983. **32**(1): p. 154-6.
7. Midgley, C.M., et al., *An in-depth analysis of original antigenic sin in dengue virus infection.* Journal of virology, 2011. **85**(1): p. 410-21.
8. Zompi, S., et al., *Dominant cross-reactive B cell response during secondary acute dengue virus infection in humans.* PLoS neglected tropical diseases, 2012. **6**(3): p. e1568.
9. Beltramello, M., et al., *The human immune response to Dengue virus is dominated by highly cross-reactive antibodies endowed with neutralizing and enhancing activity.* Cell host & microbe, 2010. **8**(3): p. 271-83.
10. Schieffelin, J.S., et al., *Neutralizing and non-neutralizing monoclonal antibodies against dengue virus E protein derived from a naturally infected patient.* Virology journal, 2010. **7**: p. 28.
11. Dejnirattisai, W., et al., *Cross-reacting antibodies enhance dengue virus infection in humans.* Science, 2010. **328**(5979): p. 745-8.
12. Halstead, S.B. and E.J. O'Rourke, *Dengue viruses and mononuclear phagocytes. I. Infection enhancement by non-neutralizing antibody.* The Journal of experimental medicine, 1977. **146**(1): p. 201-17.
13. Halstead, S.B., J.S. Porterfield, and E.J. O'Rourke, *Enhancement of dengue virus infection in monocytes by flavivirus antisera.* The American journal of tropical medicine and hygiene, 1980. **29**(4): p. 638-42.
14. Halstead, S.B., et al., *Intrinsic antibody-dependent enhancement of microbial infection in macrophages: disease regulation by immune complexes.* The Lancet infectious diseases, 2010. **10**(10): p. 712-22.
15. Rodrigo, W.W., et al., *Differential enhancement of dengue virus immune complex infectivity mediated by signaling-competent and signaling-incompetent human Fcgamma RIA (CD64) or FcgammaRIIA (CD32).* Journal of virology, 2006. **80**(20): p. 10128-38.
16. Moi, M.L., et al., *Involvement of the Fc gamma receptor IIA cytoplasmic domain in antibody-dependent enhancement of dengue virus infection.* The Journal of general virology, 2010. **91**(Pt 1): p. 103-11.
17. Rossman, M.D., et al., *Fc gamma receptor recognition of IgG ligand by human monocytes and macrophages.* American journal of respiratory cell and molecular biology, 1989. **1**(3): p. 211-20.

18. Boonnak, K., et al., *Cell type specificity and host genetic polymorphisms influence antibody-dependent enhancement of dengue virus infection.* Journal of virology, 2011. **85**(4): p. 1671-83.
19. Monath, T.P., *Yellow fever: an update.* The Lancet infectious diseases, 2001. **1**(1): p. 11-20.
20. Tomori, O., *Impact of yellow fever on the developing world.* Advances in virus research, 1999. **53**: p. 5-34.
21. Organization, T.W.H.; Available from: http://www.who.int/mediacentre/factsheets/fs100/en/.
22. Barrett, A.D. and S. Higgs, *Yellow fever: a disease that has yet to be conquered.* Annual review of entomology, 2007. **52**: p. 209-29.
23. Mukhopadhyay, S., R.J. Kuhn, and M.G. Rossmann, *A structural perspective of the flavivirus life cycle.* Nature reviews. Microbiology, 2005. **3**(1): p. 13-22.
24. Arias, C.F., F. Preugschat, and J.H. Strauss, *Dengue 2 virus NS2B and NS3 form a stable complex that can cleave NS3 within the helicase domain.* Virology, 1993. **193**(2): p. 888-99.
25. Falgout, B., et al., *Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins.* Journal of virology, 1991. **65**(5): p. 2467-75.
26. Li, H., et al., *The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids.* Journal of virology, 1999. **73**(4): p. 3108-16.
27. Egloff, M.P., et al., *An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization.* The EMBO journal, 2002. **21**(11): p. 2757-68.
28. Nomaguchi, M., et al., *De novo synthesis of negative-strand RNA by Dengue virus RNA-dependent RNA polymerase in vitro: nucleotide, primer, and template parameters.* Journal of virology, 2003. **77**(16): p. 8831-42.
29. Tan, B.H., et al., *Recombinant dengue type 1 virus NS5 protein expressed in Escherichia coli exhibits RNA-dependent RNA polymerase activity.* Virology, 1996. **216**(2): p. 317-25.
30. Beck, C., et al., *Flaviviruses in Europe: complex circulation patterns and their consequences for the diagnosis and control of West Nile disease.* International journal of environmental research and public health, 2013. **10**(11): p. 6049-83.
31. Wu, S.J., et al., *Human skin Langerhans cells are targets of dengue virus infection.* Nature medicine, 2000. **6**(7): p. 816-20.
32. Johnston, L.J., G.M. Halliday, and N.J. King, *Langerhans cells migrate to local lymph nodes following cutaneous infection with an arbovirus.* The Journal of investigative dermatology, 2000. **114**(3): p. 560-8.
33. Marovich, M., et al., *Human dendritic cells as targets of dengue virus infection.* The journal of investigative dermatology. Symposium proceedings / the Society for Investigative Dermatology, Inc. [and] European Society for Dermatological Research, 2001. **6**(3): p. 219-24.
34. Boonnak, K., et al., *Role of dendritic cells in antibody-dependent enhancement of dengue virus infection.* Journal of virology, 2008. **82**(8): p. 3939-51.

35. Lozach, P.Y., et al., *Dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin (DC-SIGN)-mediated enhancement of dengue virus infection is independent of DC-SIGN internalization signals.* The Journal of biological chemistry, 2005. **280**(25): p. 23698-708.
36. Jessie, K., et al., *Localization of dengue virus in naturally infected human tissues, by immunohistochemistry and in situ hybridization.* The Journal of infectious diseases, 2004. **189**(8): p. 1411-8.
37. Kangwanpong, D., N. Bhamarapravati, and H.L. Lucia, *Diagnosing dengue virus infection in archived autopsy tissues by means of the in situ PCR method: a case report.* Clinical and diagnostic virology, 1995. **3**(2): p. 165-72.
38. Miller, J.L., et al., *The mannose receptor mediates dengue virus infection of macrophages.* PLoS pathogens, 2008. **4**(2): p. e17.
39. Hilgard, P. and R. Stockert, *Heparan sulfate proteoglycans initiate dengue virus infection of hepatocytes.* Hepatology, 2000. **32**(5): p. 1069-77.
40. Suksanpaisan, L., A. Cabrera-Hernandez, and D.R. Smith, *Infection of human primary hepatocytes with dengue virus serotype 2.* Journal of medical virology, 2007. **79**(3): p. 300-7.
41. Chen, Y., et al., *Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate.* Nature medicine, 1997. **3**(8): p. 866-71.
42. Germi, R., et al., *Heparan sulfate-mediated binding of infectious dengue virus type 2 and yellow fever virus.* Virology, 2002. **292**(1): p. 162-8.
43. Lin, Y.L., et al., *Heparin inhibits dengue-2 virus infection of five human liver cell lines.* Antiviral research, 2002. **56**(1): p. 93-6.
44. Kroschewski, H., et al., *Role of heparan sulfate for attachment and entry of tick-borne encephalitis virus.* Virology, 2003. **308**(1): p. 92-100.
45. Lee, E., et al., *Antiviral effect of the heparan sulfate mimetic, PI-88, against dengue and encephalitic flaviviruses.* Antiviral research, 2006. **69**(1): p. 31-8.
46. Okamoto, K., et al., *Dengue virus strain DEN2 16681 utilizes a specific glycochain of syndecan-2 proteoglycan as a receptor.* The Journal of general virology, 2012. **93**(Pt 4): p. 761-70.
47. Reyes-Del Valle, J., et al., *Heat shock protein 90 and heat shock protein 70 are components of dengue virus receptor complex in human cells.* Journal of virology, 2005. **79**(8): p. 4557-67.
48. Das, S., et al., *Heat shock protein 70 on Neuro2a cells is a putative receptor for Japanese encephalitis virus.* Virology, 2009. **385**(1): p. 47-57.
49. Zhu, Y.Z., et al., *Association of heat-shock protein 70 with lipid rafts is required for Japanese encephalitis virus infection in Huh7 cells.* The Journal of general virology, 2012. **93**(Pt 1): p. 61-71.
50. Thongtan, T., et al., *Characterization of putative Japanese encephalitis virus receptor molecules on microglial cells.* Journal of medical virology, 2012. **84**(4): p. 615-23.
51. Cabrera-Hernandez, A., et al., *Dengue virus entry into liver (HepG2) cells is independent of hsp90 and hsp70.* Journal of medical virology, 2007. **79**(4): p. 386-92.
52. Jindadamrongwech, S., C. Thepparit, and D.R. Smith, *Identification of GRP 78 (BiP) as a liver cell expressed receptor element for dengue virus serotype 2.* Archives of virology, 2004. **149**(5): p. 915-27.

53. Chen, Y.C., S.Y. Wang, and C.C. King, *Bacterial lipopolysaccharide inhibits dengue virus infection of primary human monocytes/macrophages by blockade of virus entry via a CD14-dependent mechanism.* Journal of virology, 1999. **73**(4): p. 2650-7.
54. Thepparit, C. and D.R. Smith, *Serotype-specific entry of dengue virus into liver cells: identification of the 37-kilodalton/67-kilodalton high-affinity laminin receptor as a dengue virus serotype 1 receptor.* Journal of virology, 2004. **78**(22): p. 12647-56.
55. Navarro-Sanchez, E., et al., *Dendritic-cell-specific ICAM3-grabbing non-integrin is essential for the productive infection of human dendritic cells by mosquito-cell-derived dengue viruses.* EMBO reports, 2003. **4**(7): p. 723-8.
56. Tassaneetrithep, B., et al., *DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells.* The Journal of experimental medicine, 2003. **197**(7): p. 823-9.
57. Davis, C.W., et al., *The location of asparagine-linked glycans on West Nile virions controls their interactions with CD209 (dendritic cell-specific ICAM-3 grabbing nonintegrin).* The Journal of biological chemistry, 2006. **281**(48): p. 37183-94.
58. Davis, C.W., et al., *West Nile virus discriminates between DC-SIGN and DC-SIGNR for cellular attachment and infection.* Journal of virology, 2006. **80**(3): p. 1290-301.
59. Mondotte, J.A., et al., *Essential role of dengue virus envelope protein N glycosylation at asparagine-67 during viral propagation.* Journal of virology, 2007. **81**(13): p. 7136-48.
60. Dejnirattisai, W., et al., *Lectin switching during dengue virus infection.* The Journal of infectious diseases, 2011. **203**(12): p. 1775-83.
61. Chen, S.T., et al., *CLEC5A is critical for dengue-virus-induced lethal disease.* Nature, 2008. **453**(7195): p. 672-6.
62. Chen, S.T., et al., *CLEC5A regulates Japanese encephalitis virus-induced neuroinflammation and lethality.* PLoS pathogens, 2012. **8**(4): p. e1002655.
63. Wu, M.F., et al., *CLEC5A is critical for dengue virus-induced inflammasome activation in human macrophages.* Blood, 2013. **121**(1): p. 95-106.
64. Lee, J.W., J.J. Chu, and M.L. Ng, *Quantifying the specific binding between West Nile virus envelope domain III protein and the cellular receptor alphaVbeta3 integrin.* The Journal of biological chemistry, 2006. **281**(3): p. 1352-60.
65. Chu, J.J. and M.L. Ng, *Interaction of West Nile virus with alpha v beta 3 integrin mediates virus entry into cells.* The Journal of biological chemistry, 2004. **279**(52): p. 54533-41.
66. Chu, J.J. and M.L. Ng, *Characterization of a 105-kDa plasma membrane associated glycoprotein that is involved in West Nile virus binding and infection.* Virology, 2003. **312**(2): p. 458-69.
67. Li, Y., et al., *Human apolipoprotein A-I is associated with dengue virus and enhances virus infection through SR-BI.* PloS one, 2013. **8**(7): p. e70390.
68. Che, P., H. Tang, and Q. Li, *The interaction between claudin-1 and dengue viral prM/M protein for its entry.* Virology, 2013. **446**(1-2): p. 303-13.
69. Gao, F., et al., *Novel binding between pre-membrane protein and claudin-1 is required for efficient dengue virus entry.* Biochemical and biophysical research communications, 2010. **391**(1): p. 952-7.
70. Hershkovitz, O., et al., *NKp44 receptor mediates interaction of the envelope glycoproteins from the West Nile and dengue viruses with NK cells.* Journal of immunology, 2009. **183**(4): p. 2610-21.

71. Olson, K.E., et al., *Development of a Sindbis virus expression system that efficiently expresses green fluorescent protein in midguts of Aedes aegypti following per os infection.* Insect molecular biology, 2000. **9**(1): p. 57-65.
72. Pierro, D.J., et al., *Development of an orally infectious Sindbis virus transducing system that efficiently disseminates and expresses green fluorescent protein in Aedes aegypti.* Insect molecular biology, 2003. **12**(2): p. 107-16.
73. Romoser, W.S., et al., *Evidence for arbovirus dissemination conduits from the mosquito (Diptera: Culicidae) midgut.* Journal of medical entomology, 2004. **41**(3): p. 467-75.
74. Scholle, F., et al., *trans-Packaged West Nile virus-like particles: infectious properties in vitro and in infected mosquito vectors.* Journal of virology, 2004. **78**(21): p. 11605-14.
75. Salazar, M.I., et al., *Dengue virus type 2: replication and tropisms in orally infected Aedes aegypti mosquitoes.* BMC microbiology, 2007. **7**: p. 9.
76. Smith, D.R., et al., *Venezuelan equine encephalitis virus in the mosquito vector Aedes taeniorhynchus: infection initiated by a small number of susceptible epithelial cells and a population bottleneck.* Virology, 2008. **372**(1): p. 176-86.
77. Cheng, G., et al., *A C-type lectin collaborates with a CD45 phosphatase homolog to facilitate West Nile virus infection of mosquitoes.* Cell, 2010. **142**(5): p. 714-25.
78. Liu, Y., et al., *Transmission-blocking antibodies against mosquito C-type lectins for dengue prevention.* PLoS pathogens, 2014. **10**(2): p. e1003931.
79. Sakoonwatanyoo, P., V. Boonsanay, and D.R. Smith, *Growth and production of the dengue virus in C6/36 cells and identification of a laminin-binding protein as a candidate serotype 3 and 4 receptor protein.* Intervirology, 2006. **49**(3): p. 161-72.
80. Salas-Benito, J., et al., *Evidence that the 45-kD glycoprotein, part of a putative dengue virus receptor complex in the mosquito cell line C6/36, is a heat-shock related protein.* The American journal of tropical medicine and hygiene, 2007. **77**(2): p. 283-90.
81. Kuadkitkan, A., et al., *Identification and characterization of prohibitin as a receptor protein mediating DENV-2 entry into insect cells.* Virology, 2010. **406**(1): p. 149-61.
82. Chee, H.Y. and S. AbuBakar, *Identification of a 48kDa tubulin or tubulin-like C6/36 mosquito cells protein that binds dengue virus 2 using mass spectrometry.* Biochemical and biophysical research communications, 2004. **320**(1): p. 11-7.
83. Wichit, S., et al., *Dengue virus type 2 recognizes the carbohydrate moiety of neutral glycosphingolipids in mammalian and mosquito cells.* Microbiology and immunology, 2011. **55**(2): p. 135-40.
84. Reyes-del Valle, J. and R.M. del Angel, *Isolation of putative dengue virus receptor molecules by affinity chromatography using a recombinant E protein ligand.* Journal of virological methods, 2004. **116**(1): p. 95-102.
85. Yazi Mendoza, M., et al., *A putative receptor for dengue virus in mosquito tissues: localization of a 45-kDa glycoprotein.* The American journal of tropical medicine and hygiene, 2002. **67**(1): p. 76-84.
86. Salas-Benito, J.S. and R.M. del Angel, *Identification of two surface proteins from C6/36 cells that bind dengue type 4 virus.* Journal of virology, 1997. **71**(10): p. 7246-52.
87. Munoz, M.L., et al., *Putative dengue virus receptors from mosquito cells.* FEMS microbiology letters, 1998. **168**(2): p. 251-8.
88. Mercado-Curiel, R.F., et al., *The four serotypes of dengue recognize the same putative receptors in Aedes aegypti midgut and Ae. albopictus cells.* BMC microbiology, 2006. **6**: p. 85.

89. Mercado-Curiel, R.F., W.C.t. Black, and L. Munoz Mde, *A dengue receptor as possible genetic marker of vector competence in Aedes aegypti.* BMC Microbiol, 2008. **8**: p. 118.
90. Cao-Lormeau, V.M., *Dengue viruses binding proteins from Aedes aegypti and Aedes polynesiensis salivary glands.* Virology journal, 2009. **6**: p. 35.
91. Gollins, S.W. and J.S. Porterfield, *Flavivirus infection enhancement in macrophages: an electron microscopic study of viral cellular entry.* The Journal of general virology, 1985. **66 (Pt 9)**: p. 1969-82.
92. van der Schaar, H.M., et al., *Dissecting the cell entry pathway of dengue virus by single-particle tracking in living cells.* PLoS pathogens, 2008. **4**(12): p. e1000244.
93. Krishnan, M.N., et al., *Rab 5 is required for the cellular entry of dengue and West Nile viruses.* Journal of virology, 2007. **81**(9): p. 4881-5.
94. Ang, F., et al., *Small interference RNA profiling reveals the essential role of human membrane trafficking genes in mediating the infectious entry of dengue virus.* Virology journal, 2010. **7**: p. 24.
95. Alhoot, M.A., S.M. Wang, and S.D. Sekaran, *Inhibition of dengue virus entry and multiplication into monocytes using RNA interference.* PLoS neglected tropical diseases, 2011. **5**(11): p. e1410.
96. Modis, Y., et al., *Structure of the dengue virus envelope protein after membrane fusion.* Nature, 2004. **427**(6972): p. 313-9.
97. Gollins, S.W. and J.S. Porterfield, *The uncoating and infectivity of the flavivirus West Nile on interaction with cells: effects of pH and ammonium chloride.* The Journal of general virology, 1986. **67 (Pt 9)**: p. 1941-50.
98. Gollins, S.W. and J.S. Porterfield, *pH-dependent fusion between the flavivirus West Nile and liposomal model membranes.* The Journal of general virology, 1986. **67 (Pt 1)**: p. 157-66.
99. Allison, S.L., et al., *Oligomeric rearrangement of tick-borne encephalitis virus envelope proteins induced by an acidic pH.* Journal of virology, 1995. **69**(2): p. 695-700.
100. Zaitseva, E., et al., *Dengue virus ensures its fusion in late endosomes using compartment-specific lipids.* PLoS pathogens, 2010. **6**(10): p. e1001131.
101. Stiasny, K., et al., *Structural requirements for low-pH-induced rearrangements in the envelope glycoprotein of tick-borne encephalitis virus.* Journal of virology, 1996. **70**(11): p. 8142-7.
102. Gillespie, L.K., et al., *The endoplasmic reticulum provides the membrane platform for biogenesis of the flavivirus replication complex.* Journal of virology, 2010. **84**(20): p. 10438-47.
103. Welsch, S., et al., *Composition and three-dimensional architecture of the dengue virus replication and assembly sites.* Cell host & microbe, 2009. **5**(4): p. 365-75.
104. Mackenzie, J., *Wrapping things up about virus RNA replication.* Traffic, 2005. **6**(11): p. 967-77.
105. Chu, P.W. and E.G. Westaway, *Molecular and ultrastructural analysis of heavy membrane fractions associated with the replication of Kunjin virus RNA.* Archives of virology, 1992. **125**(1-4): p. 177-91.
106. Junjhon, J., et al., *Ultrastructural characterization and three-dimensional architecture of replication sites in dengue virus-infected mosquito cells.* Journal of virology, 2014. **88**(9): p. 4687-97.

107. Heaton, N.S., et al., *Dengue virus nonstructural protein 3 redistributes fatty acid synthase to sites of viral replication and increases cellular fatty acid synthesis.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(40): p. 17345-50.
108. Miller, S., et al., *The non-structural protein 4A of dengue virus is an integral membrane protein inducing membrane alterations in a 2K-regulated manner.* The Journal of biological chemistry, 2007. **282**(12): p. 8873-82.
109. Plevka, P., et al., *Maturation of flaviviruses starts from one or more icosahedrally independent nucleation centres.* EMBO reports, 2011. **12**(6): p. 602-6.
110. Yu, I.M., et al., *Structure of the immature dengue virus at low pH primes proteolytic maturation.* Science, 2008. **319**(5871): p. 1834-7.
111. Elshuber, S., et al., *Cleavage of protein prM is necessary for infection of BHK-21 cells by tick-borne encephalitis virus.* The Journal of general virology, 2003. **84**(Pt 1): p. 183-91.
112. Stadler, K., et al., *Proteolytic activation of tick-borne encephalitis virus by furin.* Journal of virology, 1997. **71**(11): p. 8475-81.
113. Smit, J.M., et al., *Flavivirus cell entry and membrane fusion.* Viruses, 2011. **3**(2): p. 160-71.
114. Junjhon, J., et al., *Influence of pr-M cleavage on the heterogeneity of extracellular dengue virus particles.* Journal of virology, 2010. **84**(16): p. 8353-8.
115. Rico-Hesse, R., *Molecular evolution and distribution of dengue viruses type 1 and 2 in nature.* Virology, 1990. **174**(2): p. 479-93.
116. Vasilakis, N., et al., *Fever from the forest: prospects for the continued emergence of sylvatic dengue virus and its impact on public health.* Nature reviews. Microbiology, 2011. **9**(7): p. 532-41.
117. Vasilakis, N., et al., *Sylvatic dengue viruses share the pathogenic potential of urban/endemic dengue viruses.* Journal of virology, 2010. **84**(7): p. 3726-7; author reply 3727-8.
118. Franco, L., et al., *First report of sylvatic DENV-2-associated dengue hemorrhagic fever in West Africa.* PLoS neglected tropical diseases, 2011. **5**(8): p. e1251.
119. Vasilakis, N., et al., *Potential of ancestral sylvatic dengue-2 viruses to re-emerge.* Virology, 2007. **358**(2): p. 402-12.
120. Diallo, D., et al., *Landscape ecology of sylvatic chikungunya virus and mosquito vectors in southeastern Senegal.* PLoS neglected tropical diseases, 2012. **6**(6): p. e1649.
121. Diallo, M., et al., *Vectors of Chikungunya virus in Senegal: current data and transmission cycles.* The American journal of tropical medicine and hygiene, 1999. **60**(2): p. 281-6.
122. Hanley, K.A., et al., *Fever versus fever: the role of host and vector susceptibility and interspecific competition in shaping the current and future distributions of the sylvatic cycles of dengue virus and yellow fever virus.* Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 2013. **19**: p. 292-311.
123. Mutebi, J.P., et al., *Phylogenetic and evolutionary relationships among yellow fever virus isolates in Africa.* Journal of virology, 2001. **75**(15): p. 6999-7008.
124. Mutebi, J.P. and A.D. Barrett, *The epidemiology of yellow fever in Africa.* Microbes and infection / Institut Pasteur, 2002. **4**(14): p. 1459-68.
125. Vainio, J., Cutts, F. , *Yellow Fever*, WHO/EPI/GEN, Editor 1998.

126. Reiter, P., et al., *First recorded outbreak of yellow fever in Kenya, 1992-1993. II. Entomologic investigations.* The American journal of tropical medicine and hygiene, 1998. **59**(4): p. 650-6.
127. Haddow, A.J., K.C. Smithburn, and et al., *Implication of the mosquito Aedes (Stegomyia) africanus Theobald in the forest cycle of yellow fever in Uganda.* Annals of Tropical Medicine and Parasitology, 1948. **42**(2): p. 218-23.
128. Germain, M., et al., *Yellow fever in the Gambia, 1978--1979: entomological aspects and epidemiological correlations.* The American journal of tropical medicine and hygiene, 1980. **29**(5): p. 929-40.
129. Diallo, D., et al., *Patterns of a sylvatic yellow Fever virus amplification in southeastern senegal, 2010.* The American journal of tropical medicine and hygiene, 2014. **90**(6): p. 1003-13.
130. Germain, M., Sureau, P., Herve, J.P., Fabre, J., Mouchet, J., Robin, Y., Geoffroy, B., *Isolements du virus de la fievre juane a partir d' Aedes du groupe A. africanus (Theobald) en Republique Centralafricaine. Importance des savanes humides et semi-humides en tant que zone d' emergence du virus amaril.* Notebooks ORSTOM.Série Medical Entomology and Parasitology, 1976. **14**: p. 125-139.
131. Diallo, M., et al., *Potential role of sylvatic and domestic African mosquito species in dengue emergence.* The American journal of tropical medicine and hygiene, 2005. **73**(2): p. 445-9.
132. Morlais, I. and D.W. Severson, *Intraspecific DNA variation in nuclear genes of the mosquito Aedes aegypti.* Insect molecular biology, 2003. **12**(6): p. 631-9.
133. Wallis, G.P., W.J. Tabachnick, and J.R. Powell, *Macrogeographic genetic variation in a human commensal: Aedes aegypti, the yellow fever mosquito.* Genetical research, 1983. **41**(3): p. 241-58.
134. Powell, J.R., W.J. Tabachnick, and J. Arnold, *Genetics and the origin of a vector population: Aedes aegypti, a case study.* Science, 1980. **208**(4450): p. 1385-7.
135. Tabachnick, W.J. and J.R. Powell, *A world-wide survey of genetic variation in the yellow fever mosquito, Aedes aegypti.* Genetical research, 1979. **34**(3): p. 215-29.
136. Trpis, M., Hausermann, W., *Demonstration of differential domesticity of Aedes aegypti (L.) (Diptera, Culicidae) in Africa by mark-release-recapture.* Bulletin of Entomological Research, 1975. **65**(2): p. 199-208.
137. McClelland, G.A.H., *A worldwide survery of variation in scale pattern of the abdominal tergum of Aedes aegypti (L.) (Diptera: Culicidae).* The Transactions of the Royal Entomological Society of London, 1974. **126**: p. 239-259.
138. Mattingly, P.F., *Genetical aspects of the Aedes aegypti problem, I. Taxonomy and bionomics.* Annals of Tropical Medicine and Parasitology, 1958. **51**: p. 392-408.
139. Mattingly, P.F., *Genetical aspects of the Aedes aegypti problem. II. Disease relationships, genetics and control.* Annals of Tropical Medicine and Parasitology, 1958. **52**: p. 5-17.
140. Van Someren E.C.C, H.R.B., Furlong M., *Observations on the behaviour of some mosquitoes of the Kenya coast.* Bull. ent. Res., 1958. **49**: p. 643-660.
141. Van Someren E.C.C., T.C., Furlong M, *The mosquitoes of the Kenya coast; records of occurence, behaviour and habitat.* Bull. ent. Res., 1955. **46**: p. 463-493.

142. Paupy, C., et al., *Gene flow between domestic and sylvan populations of Aedes aegypti (Diptera: Culicidae) in North Cameroon.* Journal of medical entomology, 2008. **45**(3): p. 391-400.
143. Paupy, C., et al., *Comparative role of Aedes albopictus and Aedes aegypti in the emergence of Dengue and Chikungunya in central Africa.* Vector borne and zoonotic diseases, 2010. **10**(3): p. 259-66.
144. Huber, K., et al., *Aedes aegypti in Senegal: genetic diversity and genetic structure of domestic and sylvatic populations.* Am J Trop Med Hyg, 2008. **79**(2): p. 218-29.
145. Sylla, M., et al., *Gene flow, subspecies composition, and dengue virus-2 susceptibility among Aedes aegypti collections in Senegal.* PLoS Negl Trop Dis, 2009. **3**(4): p. e408.
146. Brown, J.E., et al., *Worldwide patterns of genetic differentiation imply multiple 'domestications' of Aedes aegypti, a major vector of human diseases.* Proceedings. Biological sciences / The Royal Society, 2011. **278**(1717): p. 2446-54.
147. Paupy, C., et al., *Morphological and genetic variability within Aedes aegypti in Niakhar, Senegal.* Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 2010. **10**(4): p. 473-80.
148. Bhalla, S.C. and G.B. Craig, Jr., *Linkage analysis of chromosome I of Aedes aegypti.* Canadian journal of genetics and cytology. Journal canadien de genetique et de cytologie, 1970. **12**(3): p. 425-35.
149. Munstermann, L.E., Craig, G.B., *Genetics of Aedes aegypti: Updating the linkage map.* Journal of Heridity, 1979. **70**: p. 291-296.
150. Severson, D.W., et al., *Linkage map for Aedes aegypti using restriction fragment length polymorphisms.* The Journal of heredity, 1993. **84**(4): p. 241-7.
151. Fulton, R.E., et al., *SSCP analysis of cDNA markers provides a dense linkage map of the Aedes aegypti genome.* Genetics, 2001. **158**(2): p. 715-26.
152. Antolin, M.F., et al., *Intensive linkage mapping in a wasp (Bracon hebetor) and a mosquito (Aedes aegypti) with single-strand conformation polymorphism analysis of random amplified polymorphic DNA markers.* Genetics, 1996. **143**(4): p. 1727-38.
153. Severson, D.W., et al., *Linkage map organization of expressed sequence tags and sequence tagged sites in the mosquito, Aedes aegypti.* Insect molecular biology, 2002. **11**(4): p. 371-8.
154. Zhong, D., et al., *Amplified fragment length polymorphism mapping of quantitative trait loci for malaria parasite susceptibility in the yellow fever mosquito Aedes aegypti.* Genetics, 2006. **173**(3): p. 1337-45.
155. Chambers, E.W., et al., *Microsatellite isolation and linkage group identification in the yellow fever mosquito Aedes aegypti.* The Journal of heredity, 2007. **98**(3): p. 202-10.
156. Severson, D.W., et al., *Restriction fragment length polymorphism mapping of quantitative trait loci for malaria parasite susceptibility in the mosquito Aedes aegypti.* Genetics, 1995. **139**(4): p. 1711-7.
157. Severson, D.W., et al., *Chromosomal mapping of two loci affecting filarial worm susceptibility in Aedes aegypti.* Insect molecular biology, 1994. **3**(2): p. 67-72.
158. Bennett, K.E., et al., *Quantitative trait loci that control dengue-2 virus dissemination in the mosquito Aedes aegypti.* Genetics, 2005. **170**(1): p. 185-94.
159. Gomez-Machorro, C., et al., *Quantitative trait loci affecting dengue midgut infection barriers in an advanced intercross line of Aedes aegypti.* Insect Mol Biol, 2004. **13**(6): p. 637-48.

160. Bosio, C.F., et al., *Quantitative trait loci that control vector competence for dengue-2 virus in the mosquito Aedes aegypti.* Genetics, 2000. **156**(2): p. 687-98.
161. Fansiri, T., et al., *Genetic mapping of specific interactions between Aedes aegypti mosquitoes and dengue viruses.* PLoS genetics, 2013. **9**(8): p. e1003621.
162. Brown, S.E., et al., *Integration of the Aedes aegypti mosquito genetic linkage and physical maps.* Genetics, 2001. **157**(3): p. 1299-305.
163. Timoshevskiy, V.A., et al., *An integrated linkage, chromosome, and genome map for the yellow fever mosquito Aedes aegypti.* PLoS Negl Trop Dis. **7**(2): p. e2052.
164. Juneja, P., et al., *Assembly of the genome of the disease vector Aedes aegypti onto a genetic linkage map allows mapping of genes affecting disease transmission.* PLoS neglected tropical diseases, 2014. **8**(1): p. e2652.
165. Nene, V., et al., *Genome sequence of Aedes aegypti, a major arbovirus vector.* Science, 2007. **316**(5832): p. 1718-23.
166. Severson, D.W. and S.K. Behura, *Mosquito genomics: progress and challenges.* Annual review of entomology, 2012. **57**: p. 143-66.
167. Tabachnick, W.J., *Geographic and temporal patterns of genetic variation of Aedes aegypti in New Orleans.* The American journal of tropical medicine and hygiene, 1982. **31**(4): p. 849-53.
168. Tabachnick, W.J., Munstermann, L.E., Powell, J.R., *Genetic distinctness of sympatric forms of Aedes-Aegypti in East-Africa.* Evolution, 1979. **33**: p. 287-295.
169. Tabachnick, W.J., et al., *Oral infection of Aedes aegypti with yellow fever virus: geographic variation and genetic considerations.* Am J Trop Med Hyg, 1985. **34**(6): p. 1219-24.
170. Wallis, G.P., et al., *Selection for susceptibility and refractoriness of Aedes aegypti to oral infection with yellow fever virus.* Am J Trop Med Hyg, 1985. **34**(6): p. 1225-31.
171. Wallis, G.P. and W.J. Tabachnick, *Genetic analysis of rock hole and domestic Aedes aegypti on the Caribbean island of Anguilla.* Journal of the American Mosquito Control Association, 1990. **6**(4): p. 625-30.
172. Wallis, G.P., W.J. Tabachnick, and J.R. Powell, *Genetic heterogeneity among Caribbean populations of Aedes aegypti.* The American journal of tropical medicine and hygiene, 1984. **33**(3): p. 492-8.
173. Lorenz, L., et al., *The effect of colonization upon aedes aegypti susceptibility to oral infection with yellow fever virus.* The American journal of tropical medicine and hygiene, 1984. **33**(4): p. 690-4.
174. Gorrochotegui-Escalante, N., et al., *Breeding structure of Aedes aegypti populations in Mexico varies by region.* The American journal of tropical medicine and hygiene, 2002. **66**(2): p. 213-22.
175. Gorrochotegui-Escalante, N., et al., *Genetic isolation by distance among Aedes aegypti populations along the northeastern coast of Mexico.* The American journal of tropical medicine and hygiene, 2000. **62**(2): p. 200-9.
176. Urdaneta-Marquez, L., et al., *Genetic relationships among Aedes aegypti collections in Venezuela as determined by mitochondrial DNA variation and nuclear single nucleotide polymorphisms.* The American journal of tropical medicine and hygiene, 2008. **78**(3): p. 479-91.
177. Moore, M., et al., *Dual African origins of global Aedes aegypti s.l. populations revealed by mitochondrial DNA.* PLoS neglected tropical diseases, 2013. **7**(4): p. e2175.

178. Brown, J.E., et al., *Human impacts have shaped historical and recent evolution in Aedes aegypti, the dengue and yellow fever mosquito.* Evolution; international journal of organic evolution, 2014. **68**(2): p. 514-25.
179. Tabachnick, W.J. and J.R. Powell, *Genetic structure of the East African domestic populations of Aedes aegypti.* Nature, 1978. **272**(5653): p. 535-7.
180. Coyne, J., Orr, HA, *Speciation*2004, Sunderland, MA: Sinauer Associates.
181. Mayr, E., *Systematics and the Origin of Species from the Viewpoint of a Zoologist*1942: Columbia University Press.
182. Kirkpatrick, M. and V. Ravigne, *Speciation by natural and sexual selection: models and experiments.* The American naturalist, 2002. **159 Suppl 3**: p. S22-35.
183. Jiggins, C.D., et al., *Sex-linked hybrid sterility in a butterfly.* Evolution; international journal of organic evolution, 2001. **55**(8): p. 1631-8.
184. Wright, S., *Character change, speciation, and the higher taxa.* Evolution, 1982. **36**: p. 427-433.
185. Darwin, C., *On the Origin of Species by Means of Natural Selection or the Preservation of Favored Races in the Struggle for Life*1859, London: J. Murray.
186. Via, S., *Natural selection in action during speciation.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106 Suppl 1**: p. 9939-46.
187. Pinho, C., Hey, J., *Divergence with gene flow: models and data.* Annual Review of Ecology, Evolution, and Systematics, 2010. **41**: p. 215-230.
188. Wu, C.-I., *The genic view of the process of speciation.* Jornal of Evolutionary Biology, 2001. **14**: p. 851-865.
189. Favia, G., et al., *Molecular identification of sympatric chromosomal forms of Anopheles gambiae and further evidence of their reproductive isolation.* Insect molecular biology, 1997. **6**(4): p. 377-83.
190. Fanello, C., et al., *The pyrethroid knock-down resistance gene in the Anopheles gambiae complex in Mali and further indication of incipient speciation within An. gambiae s.s.* Insect molecular biology, 2003. **12**(3): p. 241-5.
191. della Torre, A., Z. Tu, and V. Petrarca, *On the distribution and genetic differentiation of Anopheles gambiae s.s. molecular forms.* Insect biochemistry and molecular biology, 2005. **35**(7): p. 755-69.
192. Simard, F., et al., *Ecological niche partitioning between Anopheles gambiae molecular forms in Cameroon: the ecological side of speciation.* BMC ecology, 2009. **9**: p. 17.
193. Tripet, F., et al., *DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of Anopheles gambiae.* Molecular ecology, 2001. **10**(7): p. 1725-32.
194. Coetzee, M., Hunt, R.H., Wilkerson, R., Della Torre, A., Coulibaly, M.B., Besansky, N.J. , *Anopheles coluzzii and Anopheles amharicus, new members of the Anopheles gambiae complex.* Zootaxa, 2013. **3619**.
195. Diabate, A., et al., *Spatial swarm segregation and reproductive isolation between the molecular forms of Anopheles gambiae.* Proceedings. Biological sciences / The Royal Society, 2009. **276**(1676): p. 4215-22.
196. Costantini, C., et al., *Living at the edge: biogeographic patterns of habitat segregation conform to speciation by niche expansion in Anopheles gambiae.* BMC ecology, 2009. **9**: p. 16.

197. Gimonneau, G., et al., *Larval habitat segregation between the molecular forms of the mosquito Anopheles gambiae in a rice field area of Burkina Faso, West Africa.* Medical and veterinary entomology, 2012. **26**(1): p. 9-17.
198. Marsden, C.D., et al., *Asymmetric introgression between the M and S forms of the malaria vector, Anopheles gambiae, maintains divergence despite extensive hybridization.* Molecular ecology, 2011. **20**(23): p. 4983-94.
199. Stapleford, K.A., et al., *Emergence and transmission of arbovirus evolutionary intermediates with epidemic potential.* Cell host & microbe, 2014. **15**(6): p. 706-16.
200. Gubler, D.J. and L. Rosen, *Variation among geographic strains of Aedes albopictus in susceptibility to infection with dengue viruses.* The American journal of tropical medicine and hygiene, 1976. **25**(2): p. 318-25.
201. Tesh, R.B., D.J. Gubler, and L. Rosen, *Variation among goegraphic strains of Aedes albopictus in susceptibility to infection with chikungunya virus.* The American journal of tropical medicine and hygiene, 1976. **25**(2): p. 326-35.
202. Grimstad, P.R., et al., *Aedes triseriatus and LA crosse virus: geographic variation in vector susceptibility and ability to transmit.* The American journal of tropical medicine and hygiene, 1977. **26**(5 Pt 1): p. 990-6.
203. Takahashi, M., *Variation in susceptibility among colony strains of Culex tritaeniorhynchus to Japanese encephalitis virus infection.* Japanese journal of medical science & biology, 1980. **33**(6): p. 321-9.
204. Mitchell, C.J., D.J. Gubler, and T.P. Monath, *Variation in infectivity of Saint Louis encephalitis viral strains for Culex pipiens quinquefasciatus (Diptera: Culicidae).* Journal of medical entomology, 1983. **20**(5): p. 526-33.
205. Kay, B.H., I.D. Fanning, and J.G. Carley, *The vector competence of Australian Culex annulirostris with Murray Valley encephalitis and Kunjin viruses.* The Australian journal of experimental biology and medical science, 1984. **62 (Pt 5)**: p. 641-50.
206. Boromisa, R.D., K.S. Rai, and P.R. Grimstad, *Variation in the vector competence of geographic strains of Aedes albopictus for dengue 1 virus.* Journal of the American Mosquito Control Association, 1987. **3**(3): p. 378-86.
207. Tardieux, I., et al., *Variation among strains of Aedes aegypti in susceptibility to oral infection with dengue virus type 2.* The American journal of tropical medicine and hygiene, 1990. **43**(3): p. 308-13.
208. Vazeille-Falcoz, M., et al., *Variation in oral susceptibility to dengue type 2 virus of populations of Aedes aegypti from the islands of Tahiti and Moorea, French Polynesia.* The American journal of tropical medicine and hygiene, 1999. **60**(2): p. 292-9.
209. Bennett, K.E., et al., *Variation in vector competence for dengue 2 virus among 24 collections of Aedes aegypti from Mexico and the United States.* Am J Trop Med Hyg, 2002. **67**(1): p. 85-92.
210. Johnson, B.W., et al., *Vector competence of Brazilian Aedes aegypti and Ae. albopictus for a Brazilian yellow fever virus isolate.* Transactions of the Royal Society of Tropical Medicine and Hygiene, 2002. **96**(6): p. 611-3.
211. Paupy, C., et al., *Variation over space and time of Aedes aegypti in Phnom Penh (Cambodia): genetic structure and oral susceptibility to a dengue virus.* Genetical research, 2003. **82**(3): p. 171-82.

212. Paupy, C., et al., *Population structure of Aedes albopictus from La Reunion Island (Indian Ocean) with respect to susceptibility to a dengue virus.* Heredity, 2001. **87**(Pt 3): p. 273-83.
213. Huber, K., et al., *Aedes aegypti in south Vietnam: ecology, genetic structure, vectorial competence and resistance to insecticides.* The Southeast Asian journal of tropical medicine and public health, 2003. **34**(1): p. 81-6.
214. Knox, T.B., et al., *Enhanced vector competence of Aedes aegypti (Diptera: Culicidae) from the Torres Strait compared with mainland Australia for dengue 2 and 4 viruses.* Journal of medical entomology, 2003. **40**(6): p. 950-6.
215. Vaidyanathan, R. and T.W. Scott, *Geographic variation in vector competence for West Nile virus in the Culex pipiens (Diptera: Culicidae) complex in California.* Vector borne and zoonotic diseases, 2007. **7**(2): p. 193-8.
216. Diallo, M., et al., *Vector competence of Aedes aegypti populations from Senegal for sylvatic and epidemic dengue 2 virus isolated in West Africa.* Transactions of the Royal Society of Tropical Medicine and Hygiene, 2008. **102**(5): p. 493-8.
217. van den Hurk, A.F., et al., *Vector competence of Australian mosquitoes for yellow fever virus.* The American journal of tropical medicine and hygiene, 2011. **85**(3): p. 446-51.
218. Black, W.C.t., et al., *Flavivirus susceptibility in Aedes aegypti.* Archives of medical research, 2002. **33**(4): p. 379-88.
219. Mousson, L., et al., *Genetic structure of Aedes aegypti populations in Chiang Mai (Thailand) and relation with dengue transmission.* Tropical medicine & international health : TM & IH, 2002. **7**(10): p. 865-72.
220. Huber, K., et al., *Genetic differentiation of the dengue vector, Aedes aegypti (Ho Chi Minh City, Vietnam) using microsatellite markers.* Molecular ecology, 2002. **11**(9): p. 1629-35.
221. Bracco, J.E., et al., *Genetic variability of Aedes aegypti in the Americas using a mitochondrial gene: evidence of multiple introductions.* Memorias do Instituto Oswaldo Cruz, 2007. **102**(5): p. 573-80.
222. Urdaneta-Marquez, L. and A.B. Failloux, *Population genetic structure of Aedes aegypti, the principal vector of dengue viruses.* Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 2011. **11**(2): p. 253-61.
223. Bosio, C.F., et al., *Genetic structure of Aedes aegypti populations in Thailand using mitochondrial DNA.* The American journal of tropical medicine and hygiene, 2005. **72**(4): p. 434-42.
224. Kambhampati, S., W.C.t. Black, and K.S. Rai, *Geographic origin of the US and Brazilian Aedes albopictus inferred from allozyme analysis.* Heredity, 1991. **67 (Pt 1)**: p. 85-93.
225. Urbanelli, S., et al., *Population structure of Aedes albopictus (Skuse): the mosquito which is colonizing Mediterranean countries.* Heredity, 2000. **84 (Pt 3)**: p. 331-7.
226. Usmani-Brown, S., L. Cohnstaedt, and L.E. Munstermann, *Population Genetics of Aedes albopictus (Diptera: Culicidae) Invading Populations, Using Mitochondrial nicotinamide Adenine Dinucleotide Dehydrogenase Subunit 5 Sequences.* Annals of the Entomological Society of America, 2009. **102**(1): p. 144-150.
227. Kamgang, B., et al., *Genetic structure of the tiger mosquito, Aedes albopictus, in Cameroon (Central Africa).* PloS one, 2011. **6**(5): p. e20257.

228. Hardy, J.L., et al., *Selection of a strain of Culex tarsalis highly resistant to infection following ingestion of western equine encephalomyelitis virus.* The American journal of tropical medicine and hygiene, 1978. **27**(2 Pt 1): p. 313-21.
229. Hardy, J.L., et al., *Temporal variations in the susceptibility of a semi-isolated population of Culex tarsalis to peroral infection with western equine encephalomyelitis and St. Louis encephalitis viruses.* The American journal of tropical medicine and hygiene, 1990. **42**(5): p. 500-11.
230. Hardy, J.L., W.C. Reeves, and R.D. Sjogren, *Variations in the susceptiblility of field and laboratory populations of Culex tarsalis to experimental infection with western equine encephalomyelitis virus.* American journal of epidemiology, 1976. **103**(5): p. 498-505.
231. Kramer, L.D., J.L. Hardy, and S.B. Presser, *Effect of temperature of extrinsic incubation on the vector competence of Culex tarsalis for western equine encephalomyelitis virus.* The American journal of tropical medicine and hygiene, 1983. **32**(5): p. 1130-9.
232. Kramer, L.D., et al., *Dissemination barriers for western equine encephalomyelitis virus in Culex tarsalis infected after ingestion of low viral doses.* The American journal of tropical medicine and hygiene, 1981. **30**(1): p. 190-7.
233. Meyer, R.P., J.L. Hardy, and S.B. Presser, *Comparative vector competence of Culex tarsalis and Culex quinquefasciatus from the coachella, imperial, and San Joaquin Valleys of California for St. Louis encephalitis virus.* The American journal of tropical medicine and hygiene, 1983. **32**(2): p. 305-11.
234. Meyer, R.P., et al., *Comparative arboviral susceptibility of female Culex tarsalis (Diptera: Culicidae) collected in CO2-baited traps and reared from field-collected pupae.* Journal of medical entomology, 1983. **20**(1): p. 56-61.
235. Reisen, W.K., J.L. Hardy, and S.B. Presser, *Effects of water quality on the vector competence of Culex tarsalis (Diptera: Culicidae) for western equine encephalomyelitis (Togaviridae) and St. Louis encephalitis (Flaviviridae) viruses.* Journal of medical entomology, 1997. **34**(6): p. 631-43.
236. Reisen, W.K., et al., *Seasonal variation in the vector competence of Culex tarsalis (Diptera:Culicidae) from the Coachella Valley of California for western equine encephalomyelitis and St. Louis encephalitis viruses.* Journal of medical entomology, 1996. **33**(3): p. 433-7.
237. Reisen, W.K., et al., *Effect of temperature on the transmission of western equine encephalomyelitis and St. Louis encephalitis viruses by Culex tarsalis (Diptera: Culicidae).* Journal of medical entomology, 1993. **30**(1): p. 151-60.
238. Cui, F., et al., *Genetic differentiation of Culex pipiens (Diptera: Culicidae) in China.* Bulletin of entomological research, 2007. **97**(3): p. 291-7.
239. Huang, S., G. Molaei, and T.G. Andreadis, *Genetic insights into the population structure of Culex pipiens (Diptera: Culicidae) in the Northeastern United States by using microsatellite analysis.* The American journal of tropical medicine and hygiene, 2008. **79**(4): p. 518-27.
240. Edillo, F., et al., *Effects of latitude and longitude on the population structure of Culex pipiens s.l., vectors of West Nile virus in North America.* The American journal of tropical medicine and hygiene, 2009. **81**(5): p. 842-8.
241. Fonseca, D.M., et al., *Population genetics of the mosquito Culex pipiens pallens reveals sex-linked asymmetric introgression by Culex quinquefasciatus.* Infection, genetics and

evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 2009. **9**(6): p. 1197-203.
242. Sharma, A.K., et al., *Genetic variability in geographical populations of Culex quinquefasciatus Say (Diptera: Culicidae) from India based on random amplified polymorphic DNA analysis.* Acta tropica, 2009. **112**(1): p. 71-6.
243. Weitzel, T., et al., *Genetic differentiation of populations within the Culex pipiens complex and phylogeny of related species.* Journal of the American Mosquito Control Association, 2009. **25**(1): p. 6-17.
244. Beerntsen, B.T., A.A. James, and B.M. Christensen, *Genetics of mosquito vector competence.* Microbiology and molecular biology reviews : MMBR, 2000. **64**(1): p. 115-37.
245. Miller, B.R. and C.J. Mitchell, *Genetic selection of a flavivirus-refractory strain of the yellow fever mosquito Aedes aegypti.* Am J Trop Med Hyg, 1991. **45**(4): p. 399-407.
246. Tardieux, I., et al., *Analysis of inheritance of oral susceptibility of Aedes aegypti (Diptera: Culicidae) to dengue-2 virus using isofemale lines.* Journal of medical entomology, 1991. **28**(4): p. 518-21.
247. Mourya, D.T., et al., *Inheritance of oral susceptibility of Aedes aegypti to Chikungunya virus.* The American journal of tropical medicine and hygiene, 1994. **51**(3): p. 295-300.
248. Bennett, K.E., B.J. Beaty, and W.C.t. Black, *Selection of D2S3, an Aedes aegypti (Diptera: Culicidae) strain with high oral susceptibility to Dengue 2 virus and D2MEB, a strain with a midgut barrier to Dengue 2 escape.* J Med Entomol, 2005. **42**(2): p. 110-9.
249. Bosio, C.F., B.J. Beaty, and W.C.t. Black, *Quantitative genetics of vector competence for dengue-2 virus in Aedes aegypti.* Am J Trop Med Hyg, 1998. **59**(6): p. 965-70.
250. Black, W.C., IV; Tabachnick, W.J, *Population genetics of disease vectors*, in *Biology of Disease Vectors*, W.C. Marqaurdt, Editor 2005, Elsevier Academic Press: Amsterdam, NL. p. 465-504.
251. Tchankouo-Nguetcheu, S., et al., *Differential protein modulation in midguts of Aedes aegypti infected with chikungunya and dengue 2 viruses.* PloS one, 2010. **5**(10).
252. Behura, S.K., et al., *Global cross-talk of genes of the mosquito Aedes aegypti in response to dengue virus infection.* PLoS neglected tropical diseases, 2011. **5**(11): p. e1385.
253. Colpitts, T.M., et al., *Alterations in the Aedes aegypti transcriptome during infection with West Nile, dengue and yellow fever viruses.* PLoS Pathog. **7**(9): p. e1002189.
254. Bonizzoni, M., et al., *Strain Variation in the Transcriptome of the Dengue Fever Vector, Aedes aegypti.* G3, 2012. **2**(1): p. 103-14.
255. Lambrechts, L., *Quantitative genetics of Aedes aegypti vector competence for dengue viruses: towards a new paradigm?* Trends Parasitol. **27**(3): p. 111-4.
256. Lambrechts, L., et al., *Genetic specificity and potential for local adaptation between dengue viruses and mosquito vectors.* BMC evolutionary biology, 2009. **9**: p. 160.
257. Gubler, D.J., et al., *Variation in susceptibility to oral infection with dengue viruses among geographic strains of Aedes aegypti.* The American journal of tropical medicine and hygiene, 1979. **28**(6): p. 1045-52.
258. Miller, B.R. and C.J. Mitchell, *Passage of yellow fever virus: its effect on infection and transmission rates in Aedes aegypti.* Am J Trop Med Hyg, 1986. **35**(6): p. 1302-9.
259. Armstrong, P.M. and R. Rico-Hesse, *Differential susceptibility of Aedes aegypti to infection by the American and Southeast Asian genotypes of dengue type 2 virus.* Vector borne and zoonotic diseases, 2001. **1**(2): p. 159-68.

260. Armstrong, P.M. and R. Rico-Hesse, *Efficiency of dengue serotype 2 virus strains to infect and disseminate in Aedes aegypti.* The American journal of tropical medicine and hygiene, 2003. **68**(5): p. 539-44.
261. Mitchell, C.J., B.R. Miller, and D.J. Gubler, *Vector competence of Aedes albopictus from Houston, Texas, for dengue serotypes 1 to 4, yellow fever and Ross River viruses.* Journal of the American Mosquito Control Association, 1987. **3**(3): p. 460-5.
262. Turell, M.J., G.V. Ludwig, and J.R. Beaman, *Transmission of Venezuelan equine encephalomyelitis virus by Aedes sollicitans and Aedes taeniorhynchus (Diptera: Culicidae).* Journal of medical entomology, 1992. **29**(1): p. 62-5.
263. Mecham, J.O. and R.A. Nunamaker, *Complex interactions between vectors and pathogens: Culicoides variipennis sonorensis (Diptera: Ceratopogonidae) infection rates with bluetongue viruses.* Journal of medical entomology, 1994. **31**(6): p. 903-7.
264. Brault, A.C., et al., *Venezuelan equine encephalitis emergence: enhanced vector infection from a single amino acid substitution in the envelope glycoprotein.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(31): p. 11344-9.
265. Vanlandingham, D.L., et al., *Short report: comparison of oral infectious dose of West Nile virus isolates representing three distinct genotypes in Culex quinquefasciatus.* The American journal of tropical medicine and hygiene, 2008. **79**(6): p. 951-4.
266. McElroy, K.L., et al., *Role of the yellow fever virus structural protein genes in viral dissemination from the Aedes aegypti mosquito midgut.* J Gen Virol, 2006. **87**(Pt 10): p. 2993-3001.
267. Moudy, R.M., et al., *A newly emergent genotype of West Nile virus is transmitted earlier and more efficiently by Culex mosquitoes.* The American journal of tropical medicine and hygiene, 2007. **77**(2): p. 365-70.
268. Rico-Hesse, R., *Dengue virus virulence and transmission determinants.* Current topics in microbiology and immunology, 2010. **338**: p. 45-55.
269. Deardorff, E.R. and S.C. Weaver, *Vector competence of Culex (Melanoconion) taeniopus for equine-virulent subtype IE strains of Venezuelan equine encephalitis virus.* The American journal of tropical medicine and hygiene, 2010. **82**(6): p. 1047-52.
270. Christofferson, R.C. and C.N. Mores, *Estimating the magnitude and direction of altered arbovirus transmission due to viral phenotype.* PloS one, 2011. **6**(1): p. e16298.
271. Andrade, C.C., et al., *North American West Nile virus genotype isolates demonstrate differential replicative capacities in response to temperature.* The Journal of general virology, 2011. **92**(Pt 11): p. 2523-33.
272. Mahmood, F., et al., *Effects of time after infection, mosquito genotype, and infectious viral dose on the dynamics of Culex tarsalis vector competence for western equine encephalomyelitis virus.* Journal of the American Mosquito Control Association, 2006. **22**(2): p. 272-81.
273. Styer, L.M., K.A. Bernard, and L.D. Kramer, *Enhanced early West Nile virus infection in young chickens infected by mosquito bite: effect of viral dose.* The American journal of tropical medicine and hygiene, 2006. **75**(2): p. 337-45.
274. Richards, S.L., et al., *Impact of West Nile virus dose and incubation period on vector competence of Culex nigripalpus (Diptera: Culicidae).* Vector borne and zoonotic diseases, 2011. **11**(11): p. 1487-91.

275. Pesko, K., et al., *Effects of infectious virus dose and bloodmeal delivery method on susceptibility of Aedes aegypti and Aedes albopictus to chikungunya virus.* Journal of medical entomology, 2009. **46**(2): p. 395-9.
276. Anderson, S.L., et al., *Effects of West Nile virus dose and extrinsic incubation temperature on temporal progression of vector competence in Culex pipiens quinquefasciatus.* Journal of the American Mosquito Control Association, 2010. **26**(1): p. 103-7.
277. Holmes, E., *The Evolution and Emergence of RNA Viruses*2009, New York: Oxford University Press.
278. Weaver, S.C., et al., *Alphaviruses: population genetics and determinants of emergence.* Antiviral research, 2012. **94**(3): p. 242-57.
279. Tsetsarkin, K.A., et al., *A single mutation in chikungunya virus affects vector specificity and epidemic potential.* PLoS pathogens, 2007. **3**(12): p. e201.
280. Tsetsarkin, K.A. and S.C. Weaver, *Sequential adaptive mutations enhance efficient vector switching by Chikungunya virus and its epidemic emergence.* PLoS pathogens, 2011. **7**(12): p. e1002412.
281. Tsetsarkin, K.A., et al., *Chikungunya virus: evolution and genetic determinants of emergence.* Current opinion in virology, 2011. **1**(4): p. 310-7.
282. Vazeille, M., et al., *Two Chikungunya isolates from the outbreak of La Reunion (Indian Ocean) exhibit different patterns of infection in the mosquito, Aedes albopictus.* PloS one, 2007. **2**(11): p. e1168.
283. McMullen, A.R., et al., *Evolution of new genotype of West Nile virus in North America.* Emerging infectious diseases, 2011. **17**(5): p. 785-93.
284. Ebel, G.D., et al., *Genetic and phenotypic variation of West Nile virus in New York, 2000-2003.* The American journal of tropical medicine and hygiene, 2004. **71**(4): p. 493-500.
285. Anderson, J.F., et al., *Horizontal and vertical transmission of West Nile virus genotype NY99 by Culex salinarius and genotypes NY99 and WN02 by Culex tarsalis.* The American journal of tropical medicine and hygiene, 2012. **86**(1): p. 134-9.
286. Dohm, D.J., M.L. O'Guinn, and M.J. Turell, *Effect of environmental temperature on the ability of Culex pipiens (Diptera: Culicidae) to transmit West Nile virus.* Journal of medical entomology, 2002. **39**(1): p. 221-5.
287. Hurlburt, H.S., *The effect of environmental temperature upon the transmission of St. Louis encephalitis virus by Culex pipiens quinquefasciatus.* Jornal of Medical Entomology 1973. **10**: p. 1-12.
288. Watts, D.M., et al., *Effect of temperature on the vector efficiency of Aedes aegypti for dengue 2 virus.* The American journal of tropical medicine and hygiene, 1987. **36**(1): p. 143-52.
289. Turell, M.J., *Effect of environmental temperature on the vector competence of Aedes taeniorhynchus for Rift Valley fever and Venezuelan equine encephalitis viruses.* The American journal of tropical medicine and hygiene, 1993. **49**(6): p. 672-6.
290. Turell, M.J. and J.O. Lundstrom, *Effect of environmental temperature on the vector competence of Aedes aegypti and Ae. taeniorhynchus for Ockelbo virus.* The American journal of tropical medicine and hygiene, 1990. **43**(5): p. 543-50.

291. Cornel, A.J., P.G. Jupp, and N.K. Blackburn, *Environmental temperature on the vector competence of Culex univittatus (Diptera: Culicidae) for West Nile virus.* Journal of medical entomology, 1993. **30**(2): p. 449-56.
292. Thu, H.M., K.M. Aye, and S. Thein, *The effect of temperature and humidity on dengue virus propagation in Aedes aegypti mosquitos.* The Southeast Asian journal of tropical medicine and public health, 1998. **29**(2): p. 280-4.
293. Brubaker, J.F. and M.J. Turell, *Effect of environmental temperature on the susceptibility of Culex pipiens (Diptera: Culicidae) to Rift Valley fever virus.* Journal of medical entomology, 1998. **35**(6): p. 918-21.
294. Kilpatrick, A.M., et al., *Spatial and temporal variation in vector competence of Culex pipiens and Cx. restuans mosquitoes for West Nile virus.* The American journal of tropical medicine and hygiene, 2010. **83**(3): p. 607-13.
295. Kramer, L.D., J.L. Hardy, and S.B. Presser, *Characterization of modulation of western equine encephalomyelitis virus by Culex tarsalis (Diptera: Culicidae) maintained at 32 degrees C following parenteral infection.* Journal of medical entomology, 1998. **35**(3): p. 289-95.
296. Lambrechts, L., et al., *Impact of daily temperature fluctuations on dengue virus transmission by Aedes aegypti.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(18): p. 7460-5.
297. Kay, B.H. and C.D. Jennings, *Enhancement or modulation of the vector competence of Ochlerotatus vigilax (Diptera: Culicidae) for ross river virus by temperature.* Journal of medical entomology, 2002. **39**(1): p. 99-105.
298. Mourya, D.T., P. Yadav, and A.C. Mishra, *Effect of temperature stress on immature stages and susceptibility of Aedes aegypti mosquitoes to chikungunya virus.* The American journal of tropical medicine and hygiene, 2004. **70**(4): p. 346-50.
299. Reisen, W.K., Y. Fang, and V.M. Martinez, *Effects of temperature on the transmission of west nile virus by Culex tarsalis (Diptera: Culicidae).* Journal of medical entomology, 2006. **43**(2): p. 309-17.
300. Rohani, A., et al., *The effect of extrinsic incubation temperature on development of dengue serotype 2 and 4 viruses in Aedes aegypti (L.).* The Southeast Asian journal of tropical medicine and public health, 2009. **40**(5): p. 942-50.
301. Alto, B.W., et al., *Larval competition alters susceptibility of adult Aedes mosquitoes to dengue infection.* Proceedings. Biological sciences / The Royal Society, 2008. **275**(1633): p. 463-71.
302. Grimstad, P.R. and L.D. Haramis, *Aedes triseriatus (Diptera: Culicidae) and La Crosse virus. III. Enhanced oral transmission by nutrition-deprived mosquitoes.* Journal of medical entomology, 1984. **21**(3): p. 249-56.
303. Kay, B.H., I.D. Fanning, and P. Mottram, *Rearing temperature influences flavivirus vector competence of mosquitoes.* Medical and veterinary entomology, 1989. **3**(4): p. 415-22.
304. Westbrook, C.J., et al., *Larval environmental temperature and the susceptibility of Aedes albopictus Skuse (Diptera: Culicidae) to Chikungunya virus.* Vector borne and zoonotic diseases, 2010. **10**(3): p. 241-7.
305. Muturi, E.J. and B.W. Alto, *Larval environmental temperature and insecticide exposure alter Aedes aegypti competence for arboviruses.* Vector borne and zoonotic diseases, 2011. **11**(8): p. 1157-63.

306. Muturi, E.J., et al., *Can pesticides and larval competition alter susceptibility of Aedes mosquitoes (Diptera: Culicidae) to arbovirus infection?* Journal of medical entomology, 2011. **48**(2): p. 429-36.
307. Pan, X., et al., *Wolbachia induces reactive oxygen species (ROS)-dependent activation of the Toll pathway to control dengue virus in the mosquito Aedes aegypti.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(1): p. E23-31.
308. Rances, E., et al., *The relative importance of innate immune priming in Wolbachia-mediated dengue interference.* PLoS pathogens, 2012. **8**(2): p. e1002548.
309. Zug, R. and P. Hammerstein, *Still a host of hosts for Wolbachia: analysis of recent data suggests that 40% of terrestrial arthropod species are infected.* PloS one, 2012. **7**(6): p. e38544.
310. Bian, G., et al., *The endosymbiotic bacterium Wolbachia induces resistance to dengue virus in Aedes aegypti.* PLoS pathogens, 2010. **6**(4): p. e1000833.
311. Moreira, L.A., et al., *A Wolbachia symbiont in Aedes aegypti limits infection with dengue, Chikungunya, and Plasmodium.* Cell, 2009. **139**(7): p. 1268-78.
312. Walker, T., et al., *The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations.* Nature, 2011. **476**(7361): p. 450-3.
313. Breeuwer, J.A. and J.H. Werren, *Microorganisms associated with chromosome destruction and reproductive isolation between two insect species.* Nature, 1990. **346**(6284): p. 558-60.
314. Landmann, F., et al., *Wolbachia-mediated cytoplasmic incompatibility is associated with impaired histone deposition in the male pronucleus.* PLoS pathogens, 2009. **5**(3): p. e1000343.
315. Frentiu, F.D., et al., *Limited dengue virus replication in field-collected Aedes aegypti mosquitoes infected with Wolbachia.* PLoS neglected tropical diseases, 2014. **8**(2): p. e2688.
316. Tabachnick, W.J., et al., *Oral infection of Aedes aegypti with yellow fever virus: geographic variation and genetic considerations.* The American journal of tropical medicine and hygiene, 1985. **34**(6): p. 1219-24.
317. Miller, B.R., et al., *Epidemic yellow fever caused by an incompetent mosquito vector.* Tropical medicine and parasitology : official organ of Deutsche Tropenmedizinische Gesellschaft and of Deutsche Gesellschaft fur Technische Zusammenarbeit, 1989. **40**(4): p. 396-9.
318. Blair, C.D., *Mosquito RNAi is the major innate immune pathway controlling arbovirus infection and transmission.* Future microbiology, 2011. **6**(3): p. 265-77.
319. Rodrigues, J., et al., *Hemocyte differentiation mediates innate immune memory in Anopheles gambiae mosquitoes.* Science, 2010. **329**(5997): p. 1353-5.
320. Kurtz, J. and S.A. Armitage, *Alternative adaptive immunity in invertebrates.* Trends in immunology, 2006. **27**(11): p. 493-6.
321. Dong, Y., et al., *Anopheles NF-kappaB-regulated splicing factors direct pathogen-specific repertoires of the hypervariable pattern recognition receptor AgDscam.* Cell host & microbe, 2012. **12**(4): p. 521-30.
322. Lindbo, J.A., et al., *Induction of a Highly Specific Antiviral State in Transgenic Plants: Implications for Regulation of Gene Expression and Virus Resistance.* The Plant cell, 1993. **5**(12): p. 1749-1759.

323. Ratcliff, F., B.D. Harrison, and D.C. Baulcombe, *A similarity between viral defense and gene silencing in plants.* Science, 1997. **276**(5318): p. 1558-60.
324. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**(6669): p. 806-11.
325. Kennerdell, J.R. and R.W. Carthew, *Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway.* Cell, 1998. **95**(7): p. 1017-26.
326. Weber, F., et al., *Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses.* Journal of virology, 2006. **80**(10): p. 5059-64.
327. Liu, Q., et al., *R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway.* Science, 2003. **301**(5641): p. 1921-5.
328. Okamura, K., et al., *Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways.* Genes & development, 2004. **18**(14): p. 1655-66.
329. Rand, T.A., et al., *Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(40): p. 14385-9.
330. Pham, J.W., et al., *A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila.* Cell, 2004. **117**(1): p. 83-94.
331. Waterhouse, R.M., et al., *Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes.* Science, 2007. **316**(5832): p. 1738-43.
332. Sanders, H.R., et al., *Sindbis virus induces transport processes and alters expression of innate immunity pathway genes in the midgut of the disease vector, Aedes aegypti.* Insect biochemistry and molecular biology, 2005. **35**(11): p. 1293-307.
333. Sim, S. and G. Dimopoulos, *Dengue virus inhibits immune responses in Aedes aegypti cells.* PloS one, 2010. **5**(5): p. e10678.
334. Xi, Z., J.L. Ramirez, and G. Dimopoulos, *The Aedes aegypti toll pathway controls dengue virus infection.* PLoS pathogens, 2008. **4**(7): p. e1000098.
335. Souza-Neto, J.A., S. Sim, and G. Dimopoulos, *An evolutionary conserved function of the JAK-STAT pathway in anti-dengue defense.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(42): p. 17841-6.
336. Hess, A.M., et al., *Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense.* BMC microbiology, 2011. **11**: p. 45.
337. Scott, J.C., et al., *Comparison of dengue virus type 2-specific small RNAs from RNA interference-competent and -incompetent mosquito cells.* PLoS neglected tropical diseases, 2010. **4**(10): p. e848.
338. Brackney, D.E., et al., *West Nile virus genetic diversity is maintained during transmission by Culex pipiens quinquefasciatus mosquitoes.* PloS one, 2011. **6**(9): p. e24466.
339. Brackney, D.E., et al., *C6/36 Aedes albopictus cells have a dysfunctional antiviral RNA interference response.* PLoS neglected tropical diseases, 2010. **4**(10): p. e856.
340. Galiana-Arnoux, D., et al., *Essential function in vivo for Dicer-2 in host defense against RNA viruses in drosophila.* Nature immunology, 2006. **7**(6): p. 590-7.
341. van Rij, R.P., et al., *The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster.* Genes & development, 2006. **20**(21): p. 2985-95.

342. Wang, X.H., et al., *RNA interference directs innate immunity against viruses in adult Drosophila.* Science, 2006. **312**(5772): p. 452-4.
343. Zambon, R.A., V.N. Vakharia, and L.P. Wu, *RNAi is an antiviral immune response against a dsRNA virus in Drosophila melanogaster.* Cellular microbiology, 2006. **8**(5): p. 880-9.
344. Kemp, C. and J.L. Imler, *Antiviral immunity in drosophila.* Current opinion in immunology, 2009. **21**(1): p. 3-9.
345. Olson, K.E., et al., *Genetically engineered resistance to dengue-2 virus transmission in mosquitoes.* Science, 1996. **272**(5263): p. 884-6.
346. Campbell, C.L., et al., *Aedes aegypti uses RNA interference in defense against Sindbis virus infection.* BMC microbiology, 2008. **8**: p. 47.
347. Sanchez-Vargas, I., et al., *Dengue virus type 2 infections of Aedes aegypti are modulated by the mosquito's RNA interference pathway.* PLoS pathogens, 2009. **5**(2): p. e1000299.
348. Gaines, P.J., et al., *Pathogen-derived resistance to dengue type 2 virus in mosquito cells by expression of the premembrane coding region of the viral genome.* Journal of virology, 1996. **70**(4): p. 2132-7.
349. Adelman, Z.N., et al., *Sindbis virus-induced silencing of dengue viruses in mosquitoes.* Insect molecular biology, 2001. **10**(3): p. 265-73.
350. Keene, K.M., et al., *RNA interference acts as a natural antiviral response to O'nyong-nyong virus (Alphavirus; Togaviridae) infection of Anopheles gambiae.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(49): p. 17240-5.
351. Sanchez-Vargas, I., et al., *RNA interference, arthropod-borne viruses, and mosquitoes.* Virus research, 2004. **102**(1): p. 65-74.
352. Cirimotich, C.M., et al., *Suppression of RNA interference increases alphavirus replication and virus-associated mortality in Aedes aegypti mosquitoes.* BMC microbiology, 2009. **9**: p. 49.
353. Franz, A.W., et al., *Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified Aedes aegypti.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(11): p. 4198-203.
354. Lee, Y.S., et al., *Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways.* Cell, 2004. **117**(1): p. 69-81.
355. Carthew, R.W. and E.J. Sontheimer, *Origins and Mechanisms of miRNAs and siRNAs.* Cell, 2009. **136**(4): p. 642-55.
356. Carmell, M.A. and G.J. Hannon, *RNase III enzymes and the initiation of gene silencing.* Nature structural & molecular biology, 2004. **11**(3): p. 214-8.
357. Cenik, E.S., et al., *Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease.* Molecular cell, 2011. **42**(2): p. 172-84.
358. Welker, N.C., et al., *Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode.* Molecular cell, 2011. **41**(5): p. 589-99.
359. Wu, B., et al., *Structural basis for dsRNA recognition, filament formation, and antiviral signal activation by MDA5.* Cell, 2013. **152**(1-2): p. 276-89.
360. Jiang, F., et al., *Structural basis of RNA recognition and activation by innate immune receptor RIG-I.* Nature, 2011. **479**(7373): p. 423-7.
361. Kowalinski, E., et al., *Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA.* Cell, 2011. **147**(2): p. 423-35.

362. Luo, D., et al., *Structural insights into RNA recognition by RIG-I.* Cell, 2011. **147**(2): p. 409-22.
363. Luo, D., et al., *Visualizing the determinants of viral RNA recognition by innate immune sensor RIG-I.* Structure, 2012. **20**(11): p. 1983-8.
364. Deddouche, S., et al., *The DExD/H-box helicase Dicer-2 mediates the induction of antiviral activity in drosophila.* Nature immunology, 2008. **9**(12): p. 1425-32.
365. Paradkar, P.N., et al., *Secreted Vago restricts West Nile virus infection in Culex mosquito cells by activating the Jak-STAT pathway.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(46): p. 18915-20.
366. Paradkar, P.N., et al., *Dicer-2-dependent activation of Culex Vago occurs via the TRAF-Rel2 signaling pathway.* PLoS neglected tropical diseases, 2014. **8**(4): p. e2823.
367. Lauring, A.S. and R. Andino, *Quasispecies theory and the behavior of RNA viruses.* PLoS pathogens, 2010. **6**(7): p. e1001005.
368. Ciota, A.T., et al., *Quantification of intrahost bottlenecks of West Nile virus in Culex pipiens mosquitoes using an artificial mutant swarm.* Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 2012. **12**(3): p. 557-64.
369. Brackney, D.E., J.E. Beane, and G.D. Ebel, *RNAi targeting of West Nile virus in mosquito midguts promotes virus diversification.* PLoS pathogens, 2009. **5**(7): p. e1000502.
370. Fitzpatrick, K.A., et al., *Population variation of West Nile virus confers a host-specific fitness benefit in mosquitoes.* Virology, 2010. **404**(1): p. 89-95.
371. Bernhardt, S.A., et al., *Rapid intraspecific evolution of miRNA and siRNA genes in the mosquito Aedes aegypti.* PloS one, 2012. **7**(9): p. e44198.
372. Li, F. and S.W. Ding, *Virus counterdefense: diverse strategies for evading the RNA-silencing immunity.* Annual review of microbiology, 2006. **60**: p. 503-31.
373. Obbard, D.J., et al., *Natural selection drives extremely rapid evolution in antiviral RNAi genes.* Current biology : CB, 2006. **16**(6): p. 580-5.
374. Lambrechts, L. and T.W. Scott, *Mode of transmission and the evolution of arbovirus virulence in mosquito vectors.* Proceedings. Biological sciences / The Royal Society, 2009. **276**(1660): p. 1369-78.
375. Chow, V.T., et al., *Monitoring of dengue viruses in field-caught Aedes aegypti and Aedes albopictus mosquitoes by a type-specific polymerase chain reaction and cycle sequencing.* The American journal of tropical medicine and hygiene, 1998. **58**(5): p. 578-86.
376. Chung, Y.K. and F.Y. Pang, *Dengue virus infection rate in field populations of female Aedes aegypti and Aedes albopictus in Singapore.* Tropical medicine & international health : TM & IH, 2002. **7**(4): p. 322-30.
377. Urdaneta, L., et al., *Detection of dengue viruses in field-caught Aedes aegypti (Diptera: Culicidae) in Maracay, Aragua state, Venezuela by type-specific polymerase chain reaction.* Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 2005. **5**(2): p. 177-84.
378. Ding, S.W. and O. Voinnet, *Antiviral immunity directed by small RNAs.* Cell, 2007. **130**(3): p. 413-26.

379. Attarzadeh-Yazdi, G., et al., *Cell-to-cell spread of the RNA interference response suppresses Semliki Forest virus (SFV) infection of mosquito cell cultures and cannot be antagonized by SFV.* Journal of virology, 2009. **83**(11): p. 5735-48.
380. Li, H.W. and S.W. Ding, *Antiviral silencing in animals.* FEBS letters, 2005. **579**(26): p. 5965-73.
381. Blakqori, G., et al., *La Crosse bunyavirus nonstructural protein NSs serves to suppress the type I interferon system of mammalian hosts.* Journal of virology, 2007. **81**(10): p. 4991-9.
382. Kakumani, P.K., et al., *Role of RNA interference (RNAi) in dengue virus replication and identification of NS4B as an RNAi suppressor.* Journal of virology, 2013. **87**(16): p. 8870-83.
383. Pijlman, G.P., et al., *A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity.* Cell host & microbe, 2008. **4**(6): p. 579-91.
384. Schnettler, E., et al., *Noncoding flavivirus RNA displays RNA interference suppressor activity in insect and Mammalian cells.* Journal of virology, 2012. **86**(24): p. 13486-500.
385. Moon, S.L., et al., *A noncoding RNA produced by arthropod-borne flaviviruses inhibits the cellular exoribonuclease XRN1 and alters host mRNA stability.* RNA, 2012. **18**(11): p. 2029-40.
386. Nurbaev, S.D. and E.V. Balanovskaia, *[Interpopulation diversity of the gene pool: beta distribution of Wright's F(ST) statistics]*. Genetika, 1998. **34**(7): p. 1004-8.
387. Webb, H.E., et al., *Aerobic Fitness Affects Cortisol Responses to Concurrent Challenges.* Medicine and Science in Sports and Exercise, 2013. **45**(2): p. 379-386.
388. Sylla, M., Ndiaye, M., Black, W.C., *Aedes species in treeholes and fruit husks between dry and wet seasons in southeastern Senegal.* Journal of Vector Ecology, 2013. **38**(2): p. 237-244.
389. Timoshevskiy, V.A., et al., *Genomic composition and evolution of Aedes aegypti chromosomes revealed by the analysis of physically mapped supercontigs.* BMC biology, 2014. **12**: p. 27.
390. Wu, T.D. and S. Nacu, *Fast and SNP-tolerant detection of complex variants and splicing in short reads.* Bioinformatics, 2010. **26**(7): p. 873-81.
391. Gorrochotegui-Escalante, N., et al., *Association mapping of segregating sites in the early trypsin gene and susceptibility to dengue-2 virus in the mosquito Aedes aegypti.* Insect biochemistry and molecular biology, 2005. **35**(7): p. 771-88.
392. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
393. Fumagalli, M., et al., *Quantifying population genetic differentiation from next-generation sequencing data.* Genetics, 2013. **195**(3): p. 979-92.
394. Team, R.C. *R: A language and environment for statistical computing*. 2013; Available from: http://www.R-project.org/.
395. Via, S., *Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2012. **367**(1587): p. 451-60.
396. Kimura, M., *Evolutionary rate at the molecular level.* Nature, 1968. **217**(5129): p. 624-6.
397. Ohta, T., *Slightly deleterious mutant substitutions in evolution.* Nature, 1973. **246**(5428): p. 96-8.

398. King, J.L. and T.H. Jukes, *Non-Darwinian evolution.* Science, 1969. **164**(3881): p. 788-98.
399. Biswas, S. and J.M. Akey, *Genomic insights into positive selection.* Trends in genetics : TIG, 2006. **22**(8): p. 437-46.
400. Olafsdottir, G.A., M.G. Ritchie, and S.S. Snorrason, *Positive assortative mating between recently described sympatric morphs of Icelandic sticklebacks.* Biology letters, 2006. **2**(2): p. 250-2.
401. Linn, C.E., Jr., et al., *Postzygotic isolating factor in sympatric speciation in Rhagoletis flies: reduced response of hybrids to parental host-fruit odors.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(51): p. 17753-8.
402. Marin, I. and B.S. Baker, *The evolutionary dynamics of sex determination.* Science, 1998. **281**(5385): p. 1990-4.
403. Whiting, P.W., *Selective fertilization and sex-determination in Hymenoptera.* Science, 1933. **78**: p. 537-538.
404. Hediger, M., et al., *Sex determination in Drosophila melanogaster and Musca domestica converges at the level of the terminal regulator doublesex.* Development genes and evolution, 2004. **214**(1): p. 29-42.
405. Gempe, T., et al., *Sex determination in honeybees: two separate mechanisms induce and maintain the female pathway.* PLoS biology, 2009. **7**(10): p. e1000222.
406. Hasselmann, M., et al., *Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees.* Nature, 2008. **454**(7203): p. 519-22.
407. Lagos, D., et al., *The transformer gene in Bactrocera oleae: the genetic switch that determines its sex fate.* Insect molecular biology, 2007. **16**(2): p. 221-30.
408. Pane, A., et al., *The transformer gene in Ceratitis capitata provides a genetic basis for selecting and remembering the sexual fate.* Development, 2002. **129**(15): p. 3715-25.
409. Hediger, M., et al., *Molecular characterization of the key switch F provides a basis for understanding the rapid divergence of the sex-determining pathway in the housefly.* Genetics, 2010. **184**(1): p. 155-70.
410. Salvemini, M., et al., *Ceratitis capitata transformer-2 gene is required to establish and maintain the autoregulation of Cctra, the master gene for female sex determination.* The International journal of developmental biology, 2009. **53**(1): p. 109-20.
411. Concha, C. and M.J. Scott, *Sexual development in Lucilia cuprina (Diptera, Calliphoridae) is controlled by the transformer gene.* Genetics, 2009. **182**(3): p. 785-98.
412. Verhulst, E.C., L.W. Beukeboom, and L. van de Zande, *Maternal control of haplodiploid sex determination in the wasp Nasonia.* Science, 2010. **328**(5978): p. 620-3.
413. Traut, W., et al., *Phylogeny of the sex-determining gene Sex-lethal in insects.* Genome / National Research Council Canada = Genome / Conseil national de recherches Canada, 2006. **49**(3): p. 254-62.
414. Hasselmann, M. and M. Beye, *Signatures of selection among sex-determining alleles of the honey bee.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(14): p. 4888-93.
415. Hasselmann, M., et al., *Evidence for convergent nucleotide evolution and high allelic turnover rates at the complementary sex determiner gene of Western and Asian honeybees.* Molecular biology and evolution, 2008. **25**(4): p. 696-708.

416. Boggs, R.T., et al., *Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the transformer gene.* Cell, 1987. **50**(5): p. 739-47.
417. Nagoshi, R.N., et al., *The control of alternative splicing at genes regulating sexual differentiation in D. melanogaster.* Cell, 1988. **53**(2): p. 229-36.
418. Burtis, K.C. and B.S. Baker, *Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides.* Cell, 1989. **56**(6): p. 997-1010.
419. McClelland, G.A., *Sex-linkage at two loci affecting eye pigment in the mosquito Aedes aegypti (diptera: culicidae).* Canadian journal of genetics and cytology. Journal canadien de genetique et de cytologie, 1966. **8**(2): p. 192-8.
420. McDonald, P.T. and K.S. Rai, *Correlation of linkage groups with chromosomes in the mosquito, Aedes aegypti.* Genetics, 1970. **66**(3): p. 475-85.
421. Craig, G.B., Jr. and W.A. Hickey, *Current status of the formal genetics of Aedes aegypti.* Bulletin of the World Health Organization, 1967. **36**(4): p. 559-62.
422. Hall, A.B., et al., *Insights into the preservation of the homomorphic sex-determining chromosome of Aedes aegypti from the discovery of a male-biased gene tightly linked to the M-locus.* Genome biology and evolution, 2014. **6**(1): p. 179-91.
423. Salvemini, M., et al., *The orthologue of the fruitfly sex behaviour gene fruitless in the mosquito Aedes aegypti: evolution of genomic organisation and alternative splicing.* PloS one, 2013. **8**(2): p. e48554.
424. Salvemini, M., et al., *Genomic organization and splicing evolution of the doublesex gene, a Drosophila regulator of sexual differentiation, in the dengue and yellow fever mosquito Aedes aegypti.* BMC evolutionary biology, 2011. **11**: p. 41.
425. Sarno, F., et al., *The gene transformer-2 of Anastrepha fruit flies (Diptera, Tephritidae) and its evolution in insects.* BMC evolutionary biology, 2010. **10**: p. 140.
426. Koboldt, D.C., et al., *VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing.* Genome research, 2012. **22**(3): p. 568-76.
427. Grabherr, M.G., et al., *Full-length transcriptome assembly from RNA-Seq data without a reference genome.* Nature biotechnology, 2011. **29**(7): p. 644-52.
428. A, S.V.a.O., *DNAcopy: DNA copy number data analysis*, 2014.
429. Olshen, A.B., et al., *Circular binary segmentation for the analysis of array-based DNA copy number data.* Biostatistics, 2004. **5**(4): p. 557-72.
430. Black WC, D.N., *RAPD-PCR and SSCP analysis for insect population genetic studies*, in *The Molecular Biology of Insect Disease Vectors: a Methods Manual*, B.C. Crampton J, Louis C, Editor 1997, Chapman & Hall: New York. p. 361–373.
431. Gempe, T. and M. Beye, *Function and evolution of sex determination mechanisms, genes and pathways in insects.* BioEssays : news and reviews in molecular, cellular and developmental biology, 2011. **33**(1): p. 52-60.
432. Franco, M.G., P.G. Rubini, and M. Vecchi, *Sex-determinants and their distribution in various populations of Musca domestica L. of Western Europe.* Genetical research, 1982. **40**(3): p. 279-93.
433. Dubendorfer, A., et al., *Musca domestica, a window on the evolution of sex-determining mechanisms in insects.* The International journal of developmental biology, 2002. **46**(1): p. 75-9.

434. Inoue, H., Fukumori, Y., Hiroyoshi, T., *Mapping of autosomal male-determining factors of the housefly, Musca domestica L., by means of sex-reversal.* Japanese Journal of Genetics, 1983. **58**: p. 451-461.
435. McDonald, I., Evenson, P., Nickel, CA., Johnson, OA., *Housefly genetics: isolation of a female determining factor on chromosome 4.* Annals of the Entomological Society of America, 1978. **71**(692): p. 694.
436. Craig JGB, H.A., *Genetics of Insect Vectors of Disease*, ed. P.R. Wright WJ1967, New York: Elsevier.
437. McClelland, G.A.H., *Sex-linkage in Aedes aegypti*. Transactions of the Royal Society of Tropical Medicine and Hygiene, 1962. **56**.
438. Conrad, B. and S.E. Antonarakis, *Gene duplication: a drive for phenotypic diversity and cause of human disease.* Annual review of genomics and human genetics, 2007. **8**: p. 17-35.
439. Organization, T.W.H. *Yellow fever*. 2014; Available from: [http://www.who.int/mediacentre/factsheets/fs100/en/.
440. Moncayo, A.C., et al., *Dengue emergence and adaptation to peridomestic mosquitoes.* Emerging infectious diseases, 2004. **10**(10): p. 1790-6.
441. Zeller, H.G., et al., *Dengue-2 virus isolation from humans during an epizootic in southeastern Senegal in November, 1990.* Research in virology, 1992. **143**(2): p. 101-2.
442. Saluzzo, J.F., et al., *[Dengue 2 in eastern Senegal: serologic survey in simian and human populations. 1974-85]*. Bulletin de la Societe de pathologie exotique et de ses filiales, 1986. **79**(3): p. 313-22.
443. Cornet, M.S., J.F.; Hervy, J.P.; Digoutte, J.P.; Germain, M.; Chauvancy, M.F.; Eyraud, M.; and L.H. Ferrara, G.; Legros, F., *Dengue 2 au Senegal oriental: Une poussee epizootique en milieu selvatique; isolements du virus a partir des moustiques et d'un singe et considerations epidemiologiques.* Cah. ORSTOM. ser Ent. Med. et Parasitol, 1984. **22**: p. 313–323.
444. Diallo, M., et al., *Amplification of the sylvatic cycle of dengue virus type 2, Senegal, 1999-2000: entomologic findings and epidemiologic considerations.* Emerging infectious diseases, 2003. **9**(3): p. 362-7.
445. Weaver, S.C. and N. Vasilakis, *Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease.* Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 2009. **9**(4): p. 523-40.
446. Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D., *WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility.* Stat. Comput, 2000. **10**: p. 325-337.
447. McCarthy, M.A., *Bayesian Methods for Ecology*2012, New York: Cambridge University Press.
448. Bolling, B.G., et al., *Transmission dynamics of an insect-specific flavivirus in a naturally infected Culex pipiens laboratory colony and effects of co-infection on vector competence for West Nile virus.* Virology, 2012. **427**(2): p. 90-7.
449. Lambrechts, L., et al., *Specificity of resistance to dengue virus isolates is associated with genotypes of the mosquito antiviral gene Dicer-2.* Proc Biol Sci. **280**(1751): p. 20122437.
450. Doerge, R.W. and G.A. Churchill, *Permutation tests for multiple loci affecting a quantitative character.* Genetics, 1996. **142**(1): p. 285-94.

451. Librado, P. and J. Rozas, *DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.* Bioinformatics, 2009. **25**(11): p. 1451-2.
452. Nei, M., *Molecular Evolutionary Genetics.*1987, New Yor: Columbia Univ. Press.
453. WEIR, B.S.a.C.C.C., *Estimating F-statistics for the analysis of population structure.* Evolution, 1984. **38**: p. 1358-1370.
454. Dickson, L.B., Sanchez-Vargas, I., Sylla, M., Fleming, K., Black, W.C., *Vector competence in West African Aedes aegypti is flavivirus species and genotype dependent.* PLoS Negl Trop Dis, In press.