DEVELOPMENT OF GENETIC MARKERS AND GENETIC ANALYSIS OF
BRUGIA MALAYI SUSCEPTIBILITY IN THE MOSQUITO AEDES AEGYPTI

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Eric Warner Chambers, B.A., M.S.

________________________________________
David W. Severson, Director

Graduate Program in Biological Sciences
Notre Dame, Indiana
January 2006
DEVELOPMENT OF GENETIC MARKERS AND GENETIC ANALYSIS OF 
BRUGIA MALAYI SUSCEPTIBILITY IN THE YELLOW FEVER 
MOSQUITO AEDES AEGYPTI

Abstract

by

Eric Warner Chambers

The mosquito, Aedes aegypti, is responsible for transmission of dengue and yellow fever and serves as an excellent model vector for transmission of malaria and lymphatic filariasis. Understanding the genetics of vector competence could allow for the development of vector-based control strategies that would alleviate mortality and morbidity worldwide.

In the first study we developed genetic markers based upon chromosomal regions conserved throughout the Culicidae. We designed 35 PCR primer pairs based upon orthologous exons in Aedes aegypti and Drosophila melanogaster or Anopheles gambiae. Twenty-three of the primers yielded a single PCR product in at least one Dipteran in addition to A. aegypti when screened with genomic DNA from seven Dipterans including five mosquito species. Eight of the primers amplified a single PCR
product in only *A. aegypti* while four primer pairs gave no PCR product in any species. The 23 successful CATS primer pairs gave broad genome coverage in *A. aegypti* and demonstrate an efficient strategy for developing comparative anchor marker loci for any species of Culicidae.

In the second study we isolated and identified microsatellite sequences from multiple genomic libraries for *A. aegypti*. We identified seven single-copy simple microsatellites from three plasmid libraries enriched for \((GA)_n\), \((AAT)_n\) and \((TAGA)_n\) motifs from *A. aegypti*. In addition, we identified 5 single-copy microsatellites from an *A. aegypti* cosmid library. Seven of these microsatellite markers were polymorphic and were genetically mapped in a segregating \(F_1\) intercross population. These markers greatly increase the number of microsatellite markers available for *A. aegypti* and provide additional tools for studying genetic variability of mosquito populations. Additionally, most *A. aegypti* microsatellites are closely associated with repetitive elements, accounting for our and other researchers limited success in developing an extensive panel of microsatellite marker loci.

In the third aim of this study we employed single nucleotide polymorphism (SNP) markers to analyze an advanced intercross line for quantitative trait loci associated with *Brugia malayi* susceptibility. Using single marker analysis of variance and linear regression we identified a single QTL at LF178 associated with susceptibility when evaluated as a quantitative phenotype and we identified one QTL for susceptibility when the phenotype was treated as binary trait.
For Renée and Benjamin
CONTENTS

FIGURES ...........................................................................................................................vi

TABLES............................................................................................................................vii

ACKNOWLEDGMENTS..........................................................................................................ix

CHAPTER 1 INTRODUCTION ...............................................................................................1
  1.1 Mosquitoes as disease vectors ............................................................................. 1
  1.2 History of mosquito borne diseases ................................................................ 2
    1.2.1 Filariasis ............................................................................................................. 3
    1.2.2 Malaria ............................................................................................................... 6
    1.2.3 Yellow fever ....................................................................................................... 7
    1.2.4 Dengue fever and dengue haemorrhagic fever ................................................... 9
  1.3 The Aedes aegypti mosquito .................................................................................. 10
    1.3.1 Aedes aegypti as a model organism ................................................................ 12
    1.3.2 Aedes aegypti genetics ..................................................................................... 12
  1.4 Heritable variation and disease within the Culicidae ............................................15
    1.4.1 Susceptibility to parasitic infections varies within and among .........................15
    1.4.2 Susceptibility to yellow fever infection varies within and among strains of A. aegypti ....................................................................................................................... 16
    1.4.3 Susceptibility to dengue virus infection varies within and among strains of A. aegypti ....................................................................................................................... 17
  1.5 Aedes aegypti and Filariasis .................................................................................. 18
  1.6 Quantitative genetics of Aedes aegypti susceptibility to pathogens ....................... 19
    1.6.1 Advanced intercross lines ................................................................................. 22
    1.6.2 The potential of genetic manipulation as a means of vector control .......... 24
  1.7 Comparative genomics .......................................................................................... 25
  1.8 Comparative genomics and vector competence in mosquitoes .......................... 26
  1.9 Specific aims ......................................................................................................... 28

CHAPTER 2 UTILITY OF COMPARATIVE ANCHOR-TAGGED SEQUENCES (CATS) AS PHYSICAL ANCHORS FOR COMPARATIVE GENOME ANALYSIS AMONG THE CULICIDAE ......................................................................................... 29
  2.1 Introduction ............................................................................................................ 29
  2.1 Materials and methods ......................................................................................... 32
    2.1.1 CAT candidate selection ................................................................................. 32
2.1.2 Primer design ................................................................. 32
2.1.3 DNA isolation .............................................................. 33
2.1.4 Primer optimization ...................................................... 33
2.1.5 Sequencing ................................................................. 34
2.2 Results ............................................................................. 34
2.3 Discussion ........................................................................ 45

CHAPTER 3 MICROSATELLITE ISOLATION AND LINKAGE GROUP IDENTIFICATION IN THE YELLOW FEVER MOSQUITO Aedes aegypti ....... 50
3.1 Introduction ........................................................................ 50
3.2 Materials and methods ...................................................... 53
  3.2.1 Microsatellite enriched library construction, sequence analysis, and primer design ........................... 53
  3.2.2 Existing library screening for microsatellites ................................................................. 54
  3.2.3 Cosmid subcloning, sequencing and primer design .......................................................... 55
  3.2.4 PCR amplification of microsatellites ........................................................................... 56
  3.2.5 Denaturing polyacrylamide gel electrophoresis .............................................................. 57
  3.2.6 Mosquito strains and crosses ....................................................................................... 57
  3.2.7 Restriction Length Fragment Polymorphism (RFLP) markers and Single Nucleotide Polymorphism (SNP) markers .......................................................... 57
  3.2.8 Linkage mapping ................................................................................................. 58
3.3 Results ................................................................................. 58
  3.3.1 Microsatellite enriched libraries ............................................................................... 58
  3.3.2 Microsatellite from cosmid subclones ........................................................................ 62
  3.3.3 Linkage analysis ................................................................................................. 62
3.4 Discussion ............................................................................. 67

CHAPTER 4 IDENTIFICATION OF QUANTITATIVE TRAIT LOCI ASSOCIATED WITH Filarial WORM SUSCEPTIBILITY IN AN Aedes aegypti ADVANCED INTERCROSS LINE .................................................................. 71
4.1 Introduction .......................................................................... 71
4.2 Materials and methods ...................................................... 74
  4.2.1 Advanced intercross lines .......................................................................................... 74
  4.2.2 Parasite exposure ................................................................................................... 74
  4.2.3 DNA isolation ........................................................................................................ 75
  4.2.4 Single Nucleotide Polymorphism marker development .............................................. 75
  4.2.5 DNA Sequencing .................................................................................................. 77
  4.2.6 Sequence analysis and identification of SNP specific restriction sites ..................... 78
  4.2.7 Linkage mapping .................................................................................................. 78
  4.2.8 Statistical analysis ............................................................................................... 79
4.3 Results .................................................................................. 80
  4.3.1 Infection of mosquitoes with B. malayi .................................................................... 80
4.3.2 Identification and development of genetic markers ........................................... 80
4.3.3 QTL determining susceptibility to B. malayi ............................................... 86
4.4 Discussion ........................................................................................................... 91

CHAPTER 5 SUMMARY ............................................................................................ 95

APPENDIX .................................................................................................................. 97

REFERENCES ........................................................................................................... 101
FIGURES

2.1 Representative agarose gels of single individual Dipterans after amplification with CATS PCR primer pairs

2.2 Relative distribution of CAT markers based upon RFLP loci on Aedes aegypti genetic map

2.3 Multiple sequence alignments of nucleotide sequences from CAT products in A. aegypti, D. melanogaster, A. subalbatus

2.4 Multiple sequence alignments of amino acid sequences from CAT products in A. aegypti, D. melanogaster, A. subalbatus

3.1 Linkage map for a. aegypti using microsatellite markers

4.1 Plot of average number of filarial worm in F8 A. aegypti with 0, 1, or 2 LIV parent alleles

4.2 Plot of filarial worm infection rate among F8 A. aegypti with 0, 1, or 2 LIV parent alleles
TABLES

2.1 CAT primers and PCR conditions ........................................................................................................37

2.2 CATS success rate in Diptera ............................................................................................................38

3.1 Analysis of enriched libraries with Repeatmasker and Sputnik ........................................60

3.2 Repeat Classifications associated with Aedes aegypti microsatellites ..........................61

3.3 Segregation of genetic markers in F1 intercross .................................................................63

3.4 Single copy microsatellites identified from enriched plasmids and selected cosmid subclones ..........................................................64

4.1 Primers used for direct cycle sequencing to identify single nucleotide polymorphism markers ............................................................................................................................76

4.2 SNP polymorphisms and INDELS ......................................................................................83

4.3 SNP marker assay information ..........................................................................................84

4.4 Segregation of Chromosome one SNP markers and Fis Values .....................................84

4.5 Single factor analysis of variance quantitative trait ....................................................86
4.6 Single factor analysis of variance binary trait

4.7 Regression statistics for SNP markers based on stepwise backward elimination analysis
ACKNOWLEDGMENTS

I wish to thank my dissertation advisor, Dr. David Severson, for his guidance, support, and patience throughout this project. I also wish to thank my graduate committee—Dr. Nora Besansky, Dr. Frank Collins, and Dr. Mike Ferdig—who provided me with constructive criticisms and insights that helped to improve my research. I am grateful also for my associates in the Severson lab: Dr. Akio Mori, Dr. Isabelle Morlais, Dr. Liza Jimenez, Dr. Mabel Berois, Dr. Consuelo Gomez, Dr. Chitra Chauhan (research faculty and post-docs); Yury Colton, Jennifer Meece, Justin Anderson, Jennifer Schneider, Sung-Jae Cha, Ryan Hemme, Irene Kasumba, and Dongyoung Shin (fellow graduate students). I also wish to thank Jim McGowan, a Notre Dame undergraduate student who helped me with the Aedes aegypti microsatellite project. I especially wish to acknowledge the invaluable support that Becky deBruyn and Diane Lovin have provided to me and to all the members of the lab. They keep our lab running smoothly and efficiently. I also wish to acknowledge that funding for this work was provided by the NIH grants AI33127 and AI34337. Finally, I also wish to thank my parents and family for supporting my desire pursue this work and above all I wish to thank my wife Renée for supporting me and encouraging me to complete this work.
CHAPTER 1

INTRODUCTION

1.1 Mosquitoes as Disease Vectors

Mosquitoes, members of the order Diptera and family Culicidae, are divided into three subfamilies: Culicinae, Anophelinae, and Toxorynchitinae (Knight and Stone 1977). A limited number of species within each of the former two subfamilies serve as obligate intermediate vectors for a variety of diseases such as malaria, dengue fever, yellow fever, and lymphatic filariasis. It is estimated that there were 515 million episodes of clinical *Plasmodium falciparum* malaria world wide in 2002 (Snow et al. 2005) causing an estimated 2.7 million deaths annually. Further, lymphatic filariasis is believed to afflict and disable upwards of 120 million people each year (WHO 2000) while dengue may strike 50 million individuals each year with an ever increasing number of accompanying cases of dengue haemorrhagic fever (DHF) (WHO 2002). There are an estimated 200,000 cases of yellow fever with 30,000 deaths annually (WHO 2001). In addition the residents of 33 African nations and 10 South American nations are believed to be at risk for this disease because of low immunization coverage (WHO 2001.)

Efforts to prevent and limit transmission of these diseases have proven difficult for a variety of reasons. Firstly, it has proven extremely difficult and costly to develop vaccines against many of these insect-borne pathogens, in particular anti-malarial
vaccines. Secondly, we have witnessed the independent emergence of several strains of *Plasmodium falciparum* parasites that are resistant to many of the most widely used antimalarial drugs. Thirdly, socio-economic conditions within the developing world have led to a loss of funding and subsequent collapse of once successful disease monitoring and control programs. Finally, the emergence of insecticide resistance in mosquito populations has led to the reduction in efficiency of a once effective means of controlling insect-borne disease transmission (Gubler 1988).

Despite these serious setbacks to disease control there still exists the potential to utilize vector control as a means to limit disease transmission. Years of genetic studies coupled with the recent sequencing of the human genome (Lander et al. 2001; Venter et al. 2001), as well as the genomes of *Plasmodium falciparum* (Gardner et al. 2002), *Anopheles gambiae* (Holt et al. 2002), and the forthcoming, *Aedes aegypti* (Severson et al. 2004) genome have provided us with a plethora of genomic tools which can be targeted towards limiting the effects of these newly reemerging but deadly diseases.

1.2 History of mosquito borne diseases

It has only been within the last 100 years that the connection between the spread of disease and insect vectors has become commonly accepted. Before the advent of modern biology it was much more difficult to establish a direct correlation between disease and insects. The very first hypotheses regarding malaria are well documented and they associated this disease with the foul odors caused by decay found within marshes, hence the derivation of the word malaria from the Latin *mal aria* or bad air (Busvine 1993). Nonetheless, as early as the 1st century A.D. the Roman soldier/farmer, Junious
Columella, wrote of the dangers of building structures too close to marshes because of the stinging/biting insects from which one would contract disease (Busvine 1993).

As the Renaissance period progressed, increased attention was focused on the origins of malaria. The Italian physician Giovanni Lancisci posited on the role that mosquitoes might play in the transmission of malaria (Busvine 1993). Throughout the 17th-18th centuries a number of scientists, clerics and physicians speculated that insects, primarily houseflies, might play a role in the spread of disease.

Observers in the 19th century began to focus more seriously on the possibility that insects might play a role in the transmission of disease. Louis Beauperthay and A.E.F King were two early figures that speculated that malaria was transmitted by mosquitoes (Busvine 1993). Unfortunately, neither had the requisite scientific reputation to convince the establishment of the time. In 1884, Sir Robert McCarrison, a British army surgeon serving in India, more famously known for his studies on nutrition, proposed the theory that malaria was transmitted by sandflies (Sinclair 1953).

As acceptance of the germ theory (as proposed by Robert Koch) became widespread in the mid-19th century, the foundation was now set for researchers to demonstrate that a variety of pathogens were transmitted by insects.

1.2.1 Filariasis

Lymphatic filariasis refers to infections caused by a variety of different nematode worms belonging to the superfamily Filarioidea (Heymann 2004). The three major forms of filariasis are Bancroftian, Brugian, and Timorean caused by *Wuchereria bancrofti*, *Brugia malayi*, and *Brugia timori* respectively. It is estimated that there are 128 million
cases of lymphatic filariasis worldwide with over 750 million additional individuals at risk, primarily those residing in Southeast Asia, Africa, the Western Pacific, Eastern Mediterranean and portions of Central and South America (Scott 2000).

Filariasis was the first disease in which it was clearly demonstrated that a mosquito vector could transmit a pathogen to humans. The parasitic worm was initially identified within the human host in 1863, when microfilariae were recovered from a female patient in Havana, Cuba (Busvine 1993). Subsequently, the immature and adult forms of the parasite were obtained from the blood of patients from India and Australia in the mid-to-late 1870’s. In 1878, Patrick Manson, the famous Scottish physician, identified mosquitoes as the vector for *W. bancrofti* (Scott 2000). He hypothesized that *Culex quinquefasciatus* could serve as a vector, and he subsequently allowed mosquitoes to feed on his gardener who was afflicted with filariasis. He dissected the blood fed mosquitoes and showed that the worms survived in the mosquito midgut. It is interesting to note that Manson believed that after developing within the mosquito vector the worms would return to the water and humans would in turn contract filariasis by drinking the infected water (Busvine 1993). It was at the beginning of the 20th century that Low (1900) demonstrated that the infective form of the worms moved from the thoracic muscles down the proboscis to the tip and then emerged to infect humans upon taking a blood meal.

The disease is manifested clinically in a complex manner due to the interplay of parasite, host immune response and fungal and bacterial infection (http://www.filariasis.org/index.pl?id=1769). There is chronic, acute, and asymptomatic presentation of this disease within at-risk populations. Chronic manifestations are
evidenced by hydrocoele, lymphoedema, and at its most extreme, elephantiasis. Additionally, chyluria, the discharge of intestinal lymph in the urine, is also a frequent acute clinical manifestation. This often leads to various nutritional deficiencies in the infected individual.

Filariasis can also manifest itself in an acute form most often reflected by (1) filarial fevers due to inflammation in the limbs or scrotum, (2) fevers due to lymph node infection caused by an immune-mediated response to dying adult worms, (3) tropical pulmonary eosinophilia, a hyper-immune response to filarial infection leading to high eosinophil levels, asthma and restrictive and/or obstructive lung disease, and (4) lymphangitis in expatriates exposed to the parasite for the first time.

Mosquitoes are the obligate intermediate hosts for filarial worms. Natural vectors of filarial worms are found within the genera *Aedes, Anopheles, Culex* and *Mansonial* (Bartholomay and Christensen 2002) and over 70 species and subspecies have been identified as vectors (Scott 2000). Filarial worms begin development within the mosquito when microfilariae (mf) are ingested by female mosquitoes while blood feeding. The mf travel to the midgut where within hours, they negotiate their way through the single cell layer of the midgut epithelium.

It is interesting to note that mf are encased within an acellular sheath which is primarily composed of an inner layer related to the egg shell and an outer layer derived from cells of the uterine lining during *in utero* growth (Scott 2000). It is this sheath encasing the mf that interfaces with both the definitive human host and the intermediate mosquito host. Exsheathment must occur for the mf to undergo further growth and
development and it appears that this can take place in the midgut (Chen and Shih 1988) or while the mf is in transit across the midgut (Christensen and Sutherland 1984). It is possible that a strategy of exsheathment that is completed in the hemocoel after transversing the midgut may be a way of deceiving the mosquito immune system and allowing the parasite to complete development. Once the mf has passed into the hemocoel it makes its way to the thoracic musculature and molts into the L2 stage, a squatter, sausage-shaped form of the parasite. This developmental stage takes place between days 6-10. The parasite undergoes a second molt to the L3 stage on days 11-13. Soon after, the L3 stage migrates from the flight muscles to the head of the mosquito. At this point in time the L3 parasites, which are 1.2-1.6 mm in length, are ready to leave the mosquito host and enter the human host. When the infected female takes a blood meal the L3 parasites are released from the proboscis onto the skin of the host where, presumably using elastases and collagenases (Petralanda et al. 1986) the parasite penetrates the host dermis at the puncture site and then makes it way to the host lymphatic system.

1.2.2 Malaria

A parasitic disease caused by protozoa belonging to the Apicomplexa, the French physician, Alphonse Laveran, was the first researcher to identify it as the causative agent of malaria when he observed exflagellation of the male gametocyte while viewing blood cells under the microscope (Busvine 1993). He shared the Nobel Prize in 1902 for his efforts with the British physician, Ronald Ross, who was the first researcher to determine that malaria was transmitted by “dapple-winged mosquitoes”, most likely *Anopheles*
when he identified malarial oocysts in the gut tissue of female mosquitoes (Vanderberg and Gwadz 1980).

There are four species of malaria parasites which can infect humans: *Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale,* and *Plasmodium malariae.* The former two species are the cause of most cases of malaria worldwide. Of these four species, *P. falciparum* is by far the most deadly causing and estimated 700,000-2.7 million deaths, primarily in sub-saharan Africa (Snow et al. 2005).

In its classical form malaria presents itself with a variety of symptoms beginning with fever, chills, profuse sweating, headache, muscle pain and then progressing to a more severe form reflected by cerebral malaria, kidney failure, respiratory distress, and coma leading to possible death (Heymann 2004).

There are approximately 430 species of Anopheline mosquitoes but only 30-40 of these species serve as vectors for human malaria. The great majority of cases of malaria occur in sub-Saharan Africa and are transmitted primarily by *A. gambiae s.s., A. funestus,* and *A. arabiensis.* Outside of Africa a variety of Anophelines serve as malaria vectors including, but not limited to, *A. albimanus, A. darlingi, A. maculatus, A. stephensi, A. barbirostis, A. punctulatus,* and *A. sinensis* (Kiszewski et al. 2004).

1.2.3 Yellow fever

Yellow fever is caused by a virus within the genus *Flavivirus* in the family Flaviviridae. In 1900 an American military physician, Walter Reed, assigned with the task of determining the cause of the yellow fever, tested the theory of Dr. Juan Carlos Finlay, a Cuban physician, who hypothesized that yellow fever was transmitted by the
mosquito *Culex fasciatus*, now known as *A. aegypti*. In 1901 Reed and his group demonstrated that mosquitoes were the intermediate hosts for yellow fever. Yellow fever is an acute disease of short duration but can vary in severity. Mild clinical symptoms are evidenced by fever, chills, headache, backache, muscle pain, nausea and vomiting (Heymann 2004). After about 5 days the disease often resolves itself but it can develop into a more dangerous phase in which jaundice, hemorrhagic symptoms and renal failure are seen. The case fatality rate among those who move into the jaundice phase is 20-50% (Heymann 2004). A vaccine against yellow fever, 17-D, was developed by researchers at the International Health Division at the Rockefeller Institute in 1937 and has been employed worldwide in endemic nations since that time.

Yellow fever can be transmitted in two different cycles. A sylvatic or jungle form transmitted by *Aedes* or *Haemagogus* mosquitoes to non-human primates can lead to the infection of humans who may work in these forests. Conversely an urban cycle of transmission can occur involving the domesticated mosquito, *Aedes aegypti*, and humans.

Sylvatic transmission is found primarily in the tropical regions of Africa and South America. Urban yellow fever was historically seen in many cities of the Americas. Although the 17-D yellow fever vaccine has been available for over 60 years, yellow fever is reemerging as a serious health problem for much of Africa, Central and South America. It is estimated that there are over 200,000 cases of yellow fever annually with an accompanying 30,000 deaths (WHO 2001).
1.2.4 Dengue fever and dengue haemorrhagic fever

Dengue is a viral disease of the genus *Flavivirus*, family Flaviviridae. It can be subdivided into 4 different serotypes: DEN-1, DEN-2, DEN-3, and DEN-4. Infection and subsequent recovery with one dengue serotype leads to partial immunity against only that particular serotype. It is believed that sequential infection to multiple serotypes can lead to dengue haemorrhagic fever (DHF).

There are two major transmission cycles. The first, an enzootic cycle involves transmission by a variety of canopy dwelling *Aedes* mosquitoes including but not limited to *A. aegypti*, *A. albopictus* and *A. polynesiensis*, and jungle dwelling primates (Gubler 1988). Sylvatic transmission of dengue within Africa involves *A. africanus*, *A. luteocephalus*, *A. opok*, *A. taylori*, and *A. furcifer* (Diallo et al. 2003), while within Malaysia *A. niveus* appears to play a major role in transmission of the sylvatic form of dengue (Rudnick 1965). Humans who enter these rainforest environments may often become infected and spark epidemics within rural villages and settlements.

The second transmission cycle, and of greater importance to humans, is an urban cycle in which the disease is maintained in an *A. aegypti*-human-*A. aegypti* cycle with *A. albopictus* acting as a secondary vector of dengue. This urban cycle of dengue transmission is believed to be both ecologically and evolutionarily separate from the sylvatic dengue cycles (Moncayo et al. 2004). Recent evidence seems to suggest that endemic and epidemic dengue virus transmission by these highly anthropophilic mosquitoes emerged when an ancestral, sylvatic form of dengue virus that was transmitted by non-human primates adapted to the more anthropophilic species of *Aedes* (Moncayo et al. 2004).
The virus has an extrinsic incubation period of 3-14 days (4-7 average) within the human host (Gubler 1998). Dengue fever is characterized by symptoms ranging from fever to frontal headache, retro-orbital pain, body aches, nausea and vomiting, joint pains, weakness, and rash. The severe pain often associated with dengue fever has also lead to it as being referred to as “break-bone fever”. DHF is characterized by the same initial symptoms as those seen in dengue fever infections followed by severe and continuous pain in abdomen, bleeding from the nose, mouth and gums or skin bruising, frequent vomiting with or without blood, stools like coal tar, excessive thirst (dry mouth) pale, cold skin restlessness, or sleepiness (WHO 2001).

It is estimated that there were 100 million case of dengue worldwide in 1998 as well as 500,000 cases of DHF and 25,000 deaths (Gubler, 1998).

1.3 The *Aedes aegypti* mosquito

*Aedes aegypti* or the yellow-fever mosquito is perhaps the most intensely studied and genetically characterized mosquito within the family Culicidae. It can be divided into two main forms. The first, *A. aegypti aegypti* is a light colored, domestic form found throughout the tropics and subtropics (Failloux et al. 2002). The second, *A. aegypti formosus*, is a darker form found primarily in tropical Africa, living in forests and breeding in tree-holes (Failloux et al. 2002). It is believed that the darker form migrated from tropical African forests to North Africa where it became domesticated and eventually differentiated into the lighter colored *A. aegypti aegypti*. Trade and commerce mostly likely led to the spread of this form to the Mediterranean basin and further diverse regions of the world.
*A. aegypti* is thought to have migrated from West Africa to the New World during the 15th-17th centuries as a result of the slave trade (Christophers 1960). The evolution of domestic traits in what was originally a sylvatic species allowed *A. aegypti* to survive in storage jars of potable water found in the holds of ships making the transatlantic journey from the Old World to the New World (Tabachnick 1991). Similarly, shipping and trade in the 18th-19th centuries most likely led to the establishment of this form throughout Asia (Smith 1956). Finally, troop movements during World War II in the mid-twentieth century led to the expansion of this species in the islands of the Pacific.

Upon arriving in the New World this species flourished in the tropical and temperate regions throughout the Americas. Soon after its arrival in the Americas, outbreaks of yellow fever occurred in many settlements. The first documented epidemic occurred in the Yucatan in 1648 (McNeil 1976), although it is believed to have been present on the island of Haiti as early as 1495 (Cloudsley-Thompson 1976). Yellow fever epidemics occurred frequently throughout the 17th-19th centuries along the Atlantic Coast, even as far North as New York and Philadelphia (Taylor 1951).

Because *A. aegypti* flourished and spread within the temperate and tropical zones of the Americas, an intensive eradication effort was attempted in the 1950’s and 1960’s, which led to its reduction and distribution but did not lead to complete elimination. Today, *A. aegypti* has reinfested the tropical/temperate regions of the Americas, and this reinfestation has been accompanied by the reemergence of dengue and dengue hemorrhagic fever (Gubler and 1995).
1.3.1 Aedes aegypti as a model organism

*A. aegypti* as noted by Craig and Hickey (1967) is an excellent model organism for genetic studies. It is easily cultured in the lab, reproduces in a robust manner and has a short reproductive cycle. Furthermore, the eggs can withstand long term storage (up to 12 months). It has a low chromosome number (three pairs) and exhibits a high degree of genetic variability. Finally, a great wealth of knowledge has been accumulated regarding the biology and physiology of this species. *A. aegypti* have proven invaluable in research efforts directed towards establishing the role of genetics in the susceptibility of mosquitoes to a wide variety of human pathogens.

1.3.2 Aedes aegypti genetics

As recently as 1953 there were no papers dealing with the formal genetics of *A. aegypti* (Kitzmiller 1953). In the latter part of that same decade Mattingly (1957,1958) speculated on the probability that a high degree of phenotypic variation was present among the disparate populations of this species. This genetic variability or “plasticity”, was first demonstrated by Craig and Hickey (1961). Subsequently increased efforts were undertaken to identify mutants which could be used to construct genetic maps.

By 1967, approximately 87 mutants had been isolated with the majority of these mutants affecting the color pattern or adult appendages (Craig and Hickey 1967). An initial linkage map was constructed which identified 28 loci residing on 3 chromosomes with a total map distance of 110 map units (Craig and Hickey 1967).

A revised version of this classical was generated more than a decade later consisting of 60 markers covering 156 map units (Munstermann and Craig 1979).
With the development of DNA-based genetic markers, genetic maps with dense marker coverage could be constructed. DNA-based markers allow for the detection of polymorphisms at the nucleotide level. They are excellent markers for genetics studies because they are neutral, often co-dominant and highly ubiquitous. A variety of these markers have been utilized and applied to *A. aegypti*.

Restriction fragment length polymorphisms (RFLPs) are molecular markers in which cloned DNA sequences are radioactively labeled and used as probes, which are hybridized to genomic DNA that has been restriction enzyme digested. The first DNA-based mosquito genetic map for *A. aegypti* was based upon restriction fragment length polymorphisms (RFLP) and consisted of 50 markers and covered 134 map units (Severson et al. 1993). Linkage was also established between several of these RFLP markers and morphological loci allowing for partial integration with the classical genetic map. The most recent map of *A. aegypti* consists of 146 markers across 205 cM (Severson et al. 2002).

Single strand conformation polymorphism markers (SSCP) are based upon the principle that single stranded nucleic acids that exhibit sequence divergence of even one nucleotide will have different secondary structures and thus differing motilities within non-denaturing polyacrylamide gels (Orita et al. 1989). Using this technique, a locus of interest from individuals of different genetic backgrounds can be PCR amplified and analyzed on the non-denaturing gel in order to identify any polymorphisms that exist between the populations.

Within *A. aegypti* linkage maps have been constructed using SSCP analysis of random amplified polymorphic DNA (RAPD) markers. RAPD markers utilize short
arbitrary 10-basepair primers to amplify regions of the genome of interest. RAPD markers can be generated quickly and inexpensively but often segregate as dominant alleles thus decreasing the amount of linkage information that can be obtained from them. Despite this drawback, Antolin et al. (1996) generated a linkage map based upon RAPD-SSCP markers consisting of 96 loci covering 168 cM. The SSCP approach of genetic analysis can also be applied to cDNA markers. Fulton et al. (2001) constructed an *A. aegypti* linkage map based upon SSCP analysis of cDNA markers consisting of 57 markers covering 134 cM.

Microsatellites are short stretches of DNA of 1-6 bases are repeated (Schlötterer 2000; Weber and May 1989). They often show high variability, co-dominant expression, and broad genome distribution (Tautz and Renz 1984; Tautz 1989; Weber and May 1989). The history and background of microsatellites within the Culicidae will be addressed in Chapter 3.

Single nucleotide polymorphisms (SNPs) are based upon point mutations that produce a single base-pair change at a particular chromosomal position. They have the potential to serve as excellent genetic markers because they are co-dominant in nature and appear to be highly ubiquitous; 90% of all genetic polymorphisms in the human genome are SNPs (Collins et al. 1998). They also appear to be abundant within the *A. aegypti* genome (Morlais and Severson 2003).
1.4 Heritable variation and disease within the Culicidae

1.4.1 Susceptibility to parasitic infections varies within and among mosquito species

It has long been held that different species of mosquitoes as well as different strains of the same species of mosquito exhibit varying degrees of susceptibility to *Plasmodium* infection. Over 70 years ago the susceptibility of *Culex pipiens molestus* to *Plasmodium cathemerium* was shown to potentially be an inherited trait when three generations of selection yielded a strain in which 64% of individuals from the third generation developed oocysts as opposed to only 28.1% of the individuals in the original parental strain (Huff 1929). Huff (1931) obtained similar results when he repeated these experiments using *Culex pipiens quinquefasciatus*. Furthermore, he demonstrated that susceptibility acted as a simple Mendelian recessive. Subsequent speculation focused on the possibility that intensity of avian *Plasmodium* infection in Culicine mosquitoes might be genetically controlled (Huff 1934, 1935).

Although an attractive hypothesis, a number of other researchers were unsuccessful in their efforts to increase rates of *Plasmodium* susceptibility through selection (Boyd and Russell 1943; Hovanitz 1947; Jeffery 1944; Kartman 1953). Ward (1963), while looking at *Plasmodium gallinaceum* susceptibility in *A. aegypti*, was able to separate mosquitoes from 11 different geographical locations into two groups based on mean oocyst counts. Although none of these strains of mosquitoes were completely refractory to infection, they did reflect average or high susceptibility. Ward was then able to select for a relatively refractory strain by conducting selection experiments over a
period of 26 generations. He was unable to select for a high susceptible strain using the same average susceptibility strain. He crossed his refractory strain with a susceptible parental strain and demonstrated that a single pair of genes was most likely responsible for this trait. Earlier, Trager (1942) had shown that susceptibility to *Plasmodium lophurae* was an inherited trait in *A. aegypti* when he selected for a strain that exhibited high susceptibility.

Kilama and Craig (1969) surveyed 19 strains of *A. aegypti* for susceptibility to *P. gallinaceum* and demonstrated that there was wide variation in the number of oocysts allowed to develop. They were also able to select for two refractory strains within one generation of selection and they demonstrated that refractoriness to *P. gallinaceum* in *A. aegypti* was controlled by a simple autosomal recessive gene, *pls*, located on chromosome 2.

Furthermore there are also instances where one population of mosquitoes may be susceptible to one species of parasite but refractory to another, closely related species of parasite. Susceptibility of *C. pipiens* to infection with *P. cathemerium*, *P. elongatum* and *P. relictum* was reflected in oocyst rates of 26.9%, 3.2% and 88.3% respectively (Huff 1930). *C. p. fatigans* was shown to exhibit varying rates of infection when exposed to the periodic and sub-periodic forms of *Wuchereria bancrofti* (Rosen 1955).

1.4.2 Susceptibility to yellow fever infection varies within and among strains of *A. aegypti*

Susceptibility to yellow fever virus in *A. aegypti* has also been shown to be controlled to some extent by genetic factors. Over 50 years ago Bruce-Chwatt (1950)
demonstrated that there were differences in infection rates between populations of *A. aegypti* to yellow fever. More recently, Wallis et al. (1985) artificially selected for strains of *A. aegypti* that exhibited susceptibility and refractoriness to oral infection with yellow fever, while Miller and Mitchell (1991), selected for two inbred lines of *A. aegypti* that also exhibited a resistant or susceptible phenotype. When they crossed the two lines, they produced progeny that were of intermediate susceptibility. They then backcrossed the F$_1$ progeny to each of the parents and demonstrated that susceptibility was controlled by one major locus that was modified by one or more minor genes. Furthermore when 28 populations of *A. aegypti*, representing a broad global distribution, were tested for their susceptibility to oral infection with yellow fever, isozyme analysis suggested that populations exhibiting genetic relatedness also exhibited similar infection rates (Tabachnick et al. 1985).

### 1.4.3 Susceptibility to dengue virus infection varies within and among strains of *A. aegypti*

Susceptibility to dengue virus has also been shown to vary within and among a variety of geographic strains of *A. aegypti*. Gubler et al. (1979) tested the susceptibility of 13 geographic strains of *A. aegypti* to oral infection with dengue virus and observed significant variation in susceptibility to the four dengue serotypes. Furthermore, strains of mosquitoes that were susceptible to one serotype tended to also exhibit susceptibility to the other serotypes. Tardieux et al. (1990) tested the susceptibility of 18 geographic strains of *A. aegypti* to dengue infection and also reported significant variation within and among strains. Twenty-four different geographic strains of *A. aegypti* in both Mexico
and the United States were tested for vector competence to the dengue 2 virus JAM 1409 and infections within these populations ranged from 24-83% (Bennett et al. 2002). Furthermore, they tested for both a midgut infection barrier (MIB) and a midgut escape barrier (MEB) within these populations. The MIB ranged from 14-59% within these mosquitoes while the MEB ranged from 4-53% (Bennett et al. 2002). Schneider et al. (2003 unpublished data) tested the susceptibility of 10 laboratory strains and one Trinidadian field strain of *A. aegypti* for susceptibility to the dengue 2 virus JAM1409 and observed infection levels between 8.3 and 52.5% between the various strains tested.

1.5 *Aedes aegypti* and Filariasis

Because it is relatively easy to rear *A. aegypti* in the laboratory, workers soon considered the possibility of utilizing this species to study the genetic basis for susceptibility to filarial worm infection in mosquitoes. Roubaud (1937) was perhaps the first worker to note that some strains of *A. aegypti* were more susceptible to infection with *Dirofilaria immitis* than other strains. Ramachandran et al. (1960) further explored this possibility when they showed that *A. aegypti* is a moderately susceptible host, when fed on cats infected with *B. malayi*. MacDonald (1962a) was successful in selecting for a strain of *A. aegypti* highly susceptible to semi-periodic *B. malayi*. These selection experiments were conducted using the Liverpool strain, originally from West Africa, which had been maintained in the Liverpool School of Tropical Medicine since 1936 and exhibited a mean susceptibility rate of 17-31%. Employing a single-pair sib mating strategy he selected for a strain with a mean susceptibility rate of 84.8% (MacDonald 1962a).
Susceptibility of *A. aegypti* to sub-periodic *B. malayi* was demonstrated to be controlled by a sex-linked recessive gene, which was designated \( f^m \) (MacDonald 1962b). This was achieved by crossing the highly susceptible Liverpool strain with the highly refractory Rangoon, Trinidad, and Malayan strains (MacDonald 1962b). MacDonald and Ramachandran (1965) followed up these studies with additional work which showed that the \( f^m \) gene also controls susceptibility to periodic *B. malayi*, and periodic and sub-periodic *Wuchereria bancrofti* but it does not control susceptibility to *Dirofilaria immitis*. They demonstrated that 80-90% of the filarial worm susceptible *Aedes* would allow development of subperiodic *B. malayi* or *B. pahangi* to the infective stage (MacDonald and Ramachandran 1965). Beckett and MacDonald (1971) later showed that up to 75% of *B. malayi* larvae die during development while only 25% of *B. pahangi* larvae die. There was no evidence of chitinization or melanization of parasites. They speculated that refractoriness in these parasites was linked to the environment of the thoracic musculature in which they resided. Nonetheless it is interesting to note that within these “susceptible” mosquitoes one could detect normal and abnormal parasites within a single muscles fiber (Beckett and MacDonald 1971).

Understanding the basis of filarial worm susceptibility in mosquitoes has also been aided by the development of a laboratory host, *Meriones unguiculatus*, which is susceptible to Brugian parasites (Ash and Riley 1970).

### 1.6 Quantitative genetics of *Aedes aegypti* susceptibility to pathogens

As with many phenotypic characters, parasite/pathogen susceptibility in mosquitoes often exhibits quantitative variation as opposed to the simple presence or
absence of susceptibility. Quantitative variation is often a sign that the trait of interest is under the control of multiple genes. With the aid of statistical procedures these traits, including susceptibility to parasites/pathogens, can be partitioned into discrete Mendelian components (Falconer 1989). Thus the observed phenotypic variance of a hypothetical trait, the number of oocysts on a mosquito midgut, for example, can be apportioned using the following equation.

$$V_P = V_A + V_D + V_I + V_E + V_{GE}$$

Where, $V_P =$ the phenotypic variance within a population while $V_A =$ the variance due to additive effects, $V_D =$ the variance due to dominance genetic effects, $V_I =$ the variance due to epistatic interactions between individual genetic effects, $V_E =$ the variance due to environmental effects and $V_{GE} =$ the variance due to the interaction between the genotype and the environment where $V_G = V_A + V_D + V_I$. By using an appropriately designed genetic cross and ensuring that all individuals are exposed to equivalent environmental effects the $V_P$ can be estimated based solely on the terms of $V_G$ (Severson 1994).

When the phenotypic variance is partitioned into its discrete components individual loci affecting the traits of interest can be identified. These loci are referred to as quantitative trait loci (QTL). Mosquito susceptibility to many pathogens is quantitative in nature, i.e. number of infective larvae, number of oocysts developing. Thoday (1961) first demonstrated that the map location of a QTL could be estimated using flanking markers. With the advent of DNA-based genetic markers more complete genetic maps with dense marker coverage have been developed within a wide variety of species. These dense genetic maps along with statistical methods based on regression analysis and maximum likelihood have allowed workers in many diverse fields to test for
the presence of QTL as well as estimate QTL parameters such as map location and gene effect (Jansen and Stram 1994; Knapp et al. 1990; Lander and Botstein 1989; Weller 1986; Zeng 1993, 1994)

More recently, Severson et al. (1994b) repeated the classical experiments of MacDonald. They constructed a linkage map based on restriction fragment length polymorphism (RFLP) markers and then employed QTL analysis to identify two QTL determining *B. malayi* susceptibility in *A. aegypti*. The first QTL, *fsb1*, demonstrated a major effect as it accounted for 22-43% of the phenotypic variance in the three mapping populations tested. The second QTL, *fsb2*, exhibited a minor effect as it accounted for 3-16% of the observed phenotypic variance. In a second QTL study, Beerntsen et al. (1995) identified a QTL that influenced filarial worm intensity. This QTL, *idb2*, which accounts for 6.8% of the phenotypic variance, plays a role in the ingestion of microfilaria as well as midgut penetration. This QTL was also found to be linked to the *fsb1* QTL associated with filarial worm susceptibility as well as the *pgs2* QTL associated with *P. gallinaceum* susceptibility (Severson et al., 1995) and it is associated with an isozyme locus putatively demonstrated to influence yellow fever virus susceptibility (Tabachnick et al. 1985).

Linkage map data based upon these same RFLP markers has also been utilized to identify two QTL controlling *P. gallinaceum* susceptibility in *A. aegypti*. One QTL, *pgs1*, found on chromosome 2, accounted for 49-65% of the phenotypic variance while a second QTL, *pgs2*, on chromosome 3, accounted for 10-14% of the phenotypic variance (Severson et al. 1995). Using a variety of single nucleotide polymorphism markers,
RFLP markers, and SSCP markers, Meece (2002) identified 5 additional QTL that played a role in susceptibility to *P. gallinaceum* when infection was treated as a binary trait.

Within the Culicines, quantitative trait loci in *A. aegypti* have also been identified that control vector competence to the dengue-2 virus (Bosio et al. 2000), that control dengue-2 virus dissemination (Bennett et al. 2005), and that affect dengue midgut infection barriers (Gomez-Machorro et al. 2004). Within *Ochlerotatus triseriatus*, a QTL conditioning transovarial transmission of La Crosse virus has also been identified (Graham et al. 2003). Furthermore, within interspecific hybrids of *Ochlerotatus hendersoni* and *O. triseriatus*, several QTL affecting oral transmission of LaCrosse virus have been identified (Anderson et al. 2005).

Within the Anophelines a number of QTL controlling various aspects of vector competence have been identified, including QTL associated with encapsulation of *Plasmodium cynomolgi B* (Zheng et al. 1997), *Plasmodium berghei* (Gorman et al. 1997), and *Plasmodium cynomolgi Ceylon* (Zheng et al. 2003).

### 1.6.1 Advanced Intercross Lines

Advanced intercross lines (AIL) are experimental populations that are initiated by crossing two inbred lines and then randomly crossing each subsequent generation (Darvasi and Soller 1995). The continued intercrossing of such a population should lead to a reduction in linkage disequilibrium and by extension, the rate of recombination between linked loci should approach 0.5 (Falconer 1989). An AIL population should allow the investigator to more accurately estimate the map location of quantitative trait loci (QTL) by decreasing the confidence interval associated with the QTL in question.
In theory, the proportion of recombinants in an AIL at the F\textsubscript{1} generation, \( r \), should equal \( r \), the proportion of recombinants in the F\textsubscript{2} generation, times \( t \), the number of generations in the AIL, divided by 2; as shown by the equation \( r = rt/2 \) (Darvasi and Soller 1995).

In a standard F\textsubscript{2} population, the confidence interval for the map location of a QTL is based upon the length of the chromosome, the QTL location in relation to the ends of the chromosome, the distance between genetic markers, the population size and the gene effect at the QTL (Darvasi et al. 1993). If these same conditions are met in an AIL, then \( C' \), the confidence interval of an AIL at the F\textsubscript{1} generation equals \( C \), the confidence interval of the F\textsubscript{2} generation divided by \( t/2 \), where \( t \) = the generation of the AIL.

\[
C' = C/t/2 \quad (\text{Darvasi and Soller 1995}).
\]

It is important to note that the theoretical reduction in QTL confidence interval generated by an AIL is only achieved if the aforementioned conditions are identical to those found in an F\textsubscript{2} population.

Advanced intercross lines have been applied to a number of mapping populations within a variety of species. Three QTL that influence survival following *Trypanosoma congolense* infection were fine mapped within two murine G\textsubscript{6} populations (Iraqi et al. 2000) and confidence intervals were reduced from 2.5-10 fold within these G\textsubscript{6} populations (Iraqi et al. 2000). An F\textsubscript{11} murine advanced intercross line was used to fine map QTL associated with pulmonary adenoma susceptibility (*Pas*) (Wang et al. 2003a). Development and application of AIL allowed the researchers to fine map one *Pas* QTL to \( \sim 1.0 \) cM or 1.3 Mb. Subsequently 27 candidate genes within the *Pas1* region were analyzed (Wang et al. 2003a). A murine F\textsubscript{11} AIL was used to resolve three QTL that regulate high-density lipoprotein cholesterol (HDL) concentrations (Wang et al. 2003b). The AIL approach also allowed for the identification of two separate QTLs on
chromosome 5 which a standard backcross had indicated was a single QTL (Wang et al. 2003b).

Advanced intercross lines have also been used in several studies involving *A. aegypti*. An F$_8$ advanced intercross line was used to identify 3 QTL on chromosome 2 that played a role in *P. gallinaceum* susceptibility on *A. aegypti* when susceptibility was treated as a binary trait (Meece 2002). Gomez-Machorro et al. (2004) used an F$_5$ AIL to identify QTL associated with a midgut infection barrier to dengue infection and identified two novel QTLs not previously identified from studies of F$_1$ intercrosses. Bennett et al. (2005) employed an F$_5$ intercross to identify multiple QTL associated with a midgut escape barrier following infection with dengue.

1.6.2 The potential of genetic manipulation as a means of vector control

Vector control by means of genetic manipulation of the insect vector is one possible means of disease control. In this scenario, mosquito populations would be genetically manipulated through the insertion of an effector molecule that would confer a refractory or resistant phenotype to the vector. A resistant phenotype is exhibited in vectors that kill the parasite or pathogen by an active immune response whereas a refractory phenotype is reflected in vectors that are incompatible with the parasite and do not allow it to develop (Christensen and Severson 1993). This transformed system could then be driven into populations of susceptible vectors and thus lead theoretically to decreased pathogen transmission (Alphey et al. 2002; Beaty 2000; Collins and Besansky 1994).
Germline transformation of a number of medically important mosquito vectors has already been accomplished (Sarkar et al. 1997; Lobo et al. 1999, 2002; Catteruccia et al. 2000; Grossman et al. 2001; Coates et al. 1998; Jasinskiene et al. 1998). Current efforts are underway to identify mechanisms that would allow for the effector gene to be driven in a non-Mendelian fashion through a susceptible mosquito population (Boete and Koella 2002; Gould and Schliekeman; Mori et al. 2004; O’Brochta et al. 2003; Rasgon and Gould 2005; Rasgon and Scott 2003; Sinkins 2004).

The identification of potential effector molecules or genes could be achieved through quantitative genetic studies that identify regions of the genome associated with susceptibility or refractoriness to a vector borne pathogen.

1.7 Comparative genomics

Comparative genomics involves the study of relationships between the genomes of different species. Homologous genes act as the landmarks upon which these relationships can be based. Genes that reside on the same chromosome exhibit synteny, while homologous genes that are syntenic demonstrate conserved synteny. If the homologous genes maintain their linear order then they exhibit conserved linkage (Erlich et al. 1997). The most informative relationship is that between orthologous genes of different species. As genome databases increase in size due to the increased number of organisms being sequenced comparative genomics will allow researchers to gain a greater understanding of chromosomal organization and evolution (Nadeau and Sankoff 1998).
1.8 Comparative genomics and vector competence in mosquitoes

Although the Culicidae contains over 3,500 species only *A. gambiae*, *A. aegypti* and *C. pipiens*, have been or are scheduled for complete genome analyses (Holt et al. 2002; Knudson et al. 2002). Employing a comparative genomics approach could allow for the comparison of genome information from sequenced species to other mosquito species that serve as the major vector for a particular pathogen.

Of particular interest are orthologous genes, which may play a role in determining vector competence, across diverse taxa. It is possible that a gene within one mosquito species that determines susceptibility to a given pathogen may have an ortholog in another species that plays an identical role. Conserved linkages between diverse mosquito taxa may increase understanding of the genetic basis of vector competence to a variety of mosquito borne pathogens.

Matthews and Munstermann (1994) analyzed isozyme linkage data within mosquito species representing three subfamilies and they identified sufficient linkage conservation to identify six syntenic groups. They speculated that these syntenic blocks represented whole chromosome arms and hypothesized that evolution within mosquitoes at the chromosomal level involved Robertsonian translocations and paracentric inversions.

Severson et al. (2004a) used an *in silico* approach to identify orthologs to genetically mapped and sequenced *A. aegypti* genes from the *D. melanogaster* and *A. gambiae* genomes databases. A comparison of chromosome positions of 73 orthologs between *D. melanogaster* and *A. aegypti* indicated that, although some ancestral chromosome elements had been retained, there had been large-scale reshuffling within
and between chromosome regions. The comparison of *A. gambiae* orthologs clearly indicated that *A. aegypti* chromosomal regions share broad homology to the five chromosomal arms of *A. gambiae*. Although there was extensive macrosynteny between the two mosquito species, and some evidence for microsynteny, gene-order or linkage conservation was limited primarily to tightly linked genes. This was unsurprising as the Anophelinae and Culicinae most likely diverged over 95 million years ago (Krzywinski et al. 2001) and additionally, recent evidence suggests that there is little conservation of gene order between *A. gambiae* and *A. funestus* even though these two species are believed to have diverged as recently as 5 million years ago (Sharakhov et al. 2002). Genome evolution within the Anophelinae appears to occur at a much higher rate than that reported for other eukaryotes (Ranz et al. 2001; Sharakhov et al. 2002).

Although there is little evidence for microsynteny or conservation of gene order for chromosomal segments longer than 50 kb between *A. aegypti* and *D. melanogaster* or *A. gambiae*, there is evidence for both synteny and linear order conservation within the Culicinae. Using cDNA based genetic markers for *A. aegypti* (Severson et al. 1993, 2002) comparative genetic maps have been constructed for *A. albopictus* (Severson et al. 1995), *Armigeres subalbatus* (Ferdig et al., 1998), *C. pipiens* (Mori et al. 1999), *Culex tritaeniorhynchus* (Mori et al. 2001), and *Ochlerotatus triseriatus* (Anderson et al. 2002). These studies demonstrated that inter- and intrachromosome rearrangements within the Culicinae do not disrupt gene order conservation within large chromosomal segments.
1.9 Specific aims

1. Develop a strategy for identifying Comparative Anchor-Tagged Sequences (CATS) which can be used for comparative mapping within the Culicidae. The development of comparative anchor tagged sequences should allow for the rapid identification of PCR-based genetic markers that provide physical anchor loci with broad application to genetically uncharacterized mosquito species, especially those mosquitoes within the Culicinae.

2. Isolate and develop simple sequence repeat (SSR) markers from three libraries enriched for (GA)$_n$, (AAT)$_n$ and (TAGA)$_n$ motifs from genomic DNA isolated from A. aegypti. Map these microsatellite markers as well as single-copy microsatellites previously identified in our laboratory on an A. aegypti F$_1$ intercross. The isolation and development of microsatellite markers within A. aegypti will provide workers with a new suite of genetic markers that can be applied to linkage analysis and population genetic studies.

3. Identify and develop chromosome one specific single nucleotide polymorphism (SNP) markers and apply these markers to an A. aegypti F$_8$ advanced intercross line for higher resolution mapping of chromosome one QTL affecting susceptibility to Brugia malayi. High resolution genetic mapping coupled with the ongoing A. aegypti sequencing project could allow for the identification of putative candidate genes contributing to B. malayi susceptibility.
CHAPTER 2

UTILITY OF COMPARATIVE ANCHOR-TAGGED SEQUENCES (CATS) AS PHYSICAL ANCHORS FOR COMPARATIVE GENOME ANALYSIS AMONG THE CULICIDAE

2.1 Introduction

The family Culicidae contains approximately 3500 species of mosquitoes (Knight and Stone 1977) and historically most genetic research efforts have been directed towards the two principal disease vector organisms, *Aedes aegypti* and *Anopheles gambiae*. The first detailed genetic map for *A. aegypti* consisted of 77 morphological, isozyme and insecticide resistance markers (Munstermann and Craig 1979). Early efforts to produce genetic maps of the *Anopheles* spp. were less successful due to a dearth of morphological and isozyme markers along with the difficulty of rearing these mosquitoes in the laboratory. In addition, within both the Culicinae and Anophelinae, these markers were not applicable for detailed genome coverage in a single mapping population and were often non-neutral in nature. Despite the difficulty of working with such markers, genetic linkage maps were constructed for *Aedes togoi, Ochlerotatus triseriatus, Aedes aegypti, Anopheles albimanus, Anopheles quadrimaculatus* and *Anopheles gambiae* (O’Brien 1990, 1993). Further, a limited comparison of isozyme loci suggested a high degree of synteny (Matthews and Munstermann 1994). Synteny refers to loci that remain physically linked in different species.
The advent of DNA-based genetic markers including Restriction Fragment Length Polymorphisms (RFLP), Random Amplified Polymorphic DNA (RAPD), Single Strand Conformation Polymorphisms (SSCP) and microsatellites has allowed for the construction of detailed linkage maps for a small number of mosquito species. The first such map, in *A. aegypti*, was generated using RFLP markers (Severson et al., 1993) and consisted of 50 DNA markers, based mostly on random cDNA’s, identifying 53 loci across the 3 linkage groups. The most recent map for this species consists of 146 markers across 205 cM (Severson et al. 2002). Two RAPD-SSCP linkage maps for *A. aegypti* were also constructed, the first consisting of 96 loci covering 168 cM (Antolin et al. 1996), while the second consisted of 94 SSCP markers based upon Single Nucleotide Polymorphisms (SNPs) and covered 134 cM (Fulton et al. 2001). Within the Anophelines, a DNA-marker-based genetic map of *An. gambiae* was constructed using 148 microsatellite markers covering 215 cM (Wang et al. 1999). These genetic maps have proven useful within *A. aegypti* and *A. gambiae* in efforts to identify genomic regions or genes associated with mosquito competence to transmit a variety of pathogens (Beernsten et al. 1995; Bosio et al. 2000; Gorman et al. 1997; Severson et al. 1994, 1995; Zheng et al. 1997).

Of interest is the construction of detailed genetic linkage maps in a greater number of mosquito species. Comparative RFLP-based linkage maps for the mosquitoes *Culex pipiens* (Mori et al. 1999), *Aedes albopictus* (Severson et al. 1995), *Ochlerotatus triseriatus* (formerly *Aedes*) (Anderson et al. 2001), *Culex tritaeniorhynchus* (Mori et al. 2001), and *Armigeres subalbatus* (Ferdig et al. 1998), were previously constructed using *A. aegypti* cDNA clones as probes to Southern blots. This strategy was successful
because many *A. aegypti* cDNA sequences are highly conserved across mosquito species (Severson et al. 1994). These maps allowed for the direct comparison of these species because the markers are based upon expressed genes (Type I markers) as opposed to microsatellites (Type II markers); that is, although microsatellites have proven useful useful in developing gene maps in many organisms, they are less useful when making comparisons between organisms as they are usually not conserved between species (Lyons et al. 1997).

Despite the success in constructing comparative gene maps for several mosquito species using RFLP markers, there has been no attempt to develop PCR-based DNA-markers that can be utilized across species within the Culicidae. The utility of orthologous gene comparisons for development as tools for PCR-based identification of conserved syntenies has been demonstrated for both birds and mammals (Lyons et al. 1997; Mazzarella et al. 1992; Smith et al. 2000). Markers developed in this manner have been referred to as Comparative Anchor-Tagged Sequences (CATS) and have been used to generate comparative genome maps in the human, dog and horse (Caetano et al. 1999; Chen et al. 1999; Lyons et al. 1999). The identification of such a set of markers within mosquitoes could prove to be extremely among those species which, although not genetically well characterized, nonetheless are of great importance medically.

In this study, we investigated the utility of the CATS marker strategy for developing comparative genome maps among mosquito species. We have evaluated 35 putative CATS primer sets across six mosquito species and two more distant species within the Diptera. Twenty-three of these CATS amplified a single PCR product in at least one dipteran species in addition to *A. aegypti*. This simple strategy in which
orthologous exons of two or more dipterans are identified, sequence alignments made, and PCR primers designed for highly conserved regions should facilitate the rapid identification of PCR-based genetic markers that provide physical anchor loci with broad application to genetically uncharacterized mosquito species and perhaps more distant dipteran species as well.

2.1 Materials and methods

2.1.1 CAT candidate selection

Identity searches were conducted on seventy-one previously mapped, partial cDNA clones or known genes from adult *A. aegypti* Liverpool females (Severson et al. 2002) by comparing nucleotide sequences against both the non-redundant as well as the non-human, non-mouse EST databases using the BLASTN computer program (Altschul et al., 1997). Thirty-five of the *A. aegypti* sequences showed significant nucleotide identities, based upon an expected value (E) of less than $e^{-15}$ as our critical value for defining sequences as orthologs with sequences from *D. melanogaster* and/or *A. gambiae*. Nucleotide sequences from the putative orthologs were aligned using the Multiple Alignment Program (MAP) on the Baylor College of Medicine (BCM) Search Launcher web site (Smith et al. 1996).

2.1.2 Primer design

Primers were designed using the PRIMER program (Whitehead Institute, Massachusetts Institute of Technology, Cambridge, Mass., U.S.A.). We first attempted to design primers within regions of the sequences which exhibited complete nucleotide
sequence identity between *Aedes aegypti*, *Drosophila melanogaster*, and *Anopheles gambiae*. In most instances, however, we designed primers to contain no more than 2-3 mismatches with complete sequence identity for the final 5-6 nucleotides at the 3’ end. When mismatches were unavoidable we biased nucleotide selection to the *A. aegypti* sequence. We designed primers to be 18-22 nucleotides in length with a G+C content of approximately 50%. When possible, we placed a GC clamp at the 3’ end. The expected PCR product size was between 100-400 bases. Whenever possible we avoided selecting primers with the 3’ nucleotide representing the third base position in a codon.

### 2.1.3 DNA isolation

DNA extractions from individuals representing *A. aegypti*, *O. triseriatus*, *A. togoi*, *A. gambiae*, *A. subalbatus* and *D. melanogaster* were performed as previously described (Severson 1997). Genomic DNA representing individuals from *C. pipiens* was kindly provided by Dr. Akio Mori and that from *Rhagoletis pomonella* was kindly provided by Dr. Jeff Feder (University of Notre Dame).

### 2.1.4 Primer optimization

PCR reactions were performed in 96-well plates utilizing a Hybaid Omnigene thermal cycler. Each 25ul PCR reaction mixture contained 1X Taq buffer (10mM KCl, 2mM Tris, pH 9.0, 0.02% TritonX) 1.5 or 3.0 mM MgCl₂, 0.4 mM each dATP, dCTP, dGTP, dTTP, 5-10 pmoles of each primer, 1 unit Taq DNA polymerase and 5 ng genomic DNA. Six primer sets were screened with eight dipteran species per plate. We screened each primer set at three anneal temperatures, 50°, 54° and 58° C. The standard PCR conditions were 94° C for 5 minutes, followed by 30 cycles of 94° C for 1 min, an
anneal step at one of the three previously mentioned temperatures for 1 min, 72° C for 2 min and a final extension at 72° C for 10 min. Primer sets with an expected PCR product size less than 200 bp were subjected to an alternate set of PCR conditions: 94° C for 10 min followed by 40 cycles at 94° C for 20 sec, anneal with one of the three temperatures for 20 sec, 72° C for 30 sec and a final extension for 10 min at 72° C. PCR products were size fractionated in 3% agarose gels and visualized with ethidium bromide under UV light.

2.1.5 Sequencing

Amplified PCR products were spin-column purified using the QIAquick PCR purification kit (Qiagen, Valencia, CA). The purified PCR products were then quantified visually on 3.0% agarose gels. PCR product templates were subjected to cycle sequencing using the ABI Prism Big Dye Terminator kit according to the suppliers recommendations (Applied Biosystems Inc., Foster City, CA) and an ABI Prism 310 Genetic Analyzer. Single-pass sequence information was obtained from each strand of each PCR product. Sequence data were then subjected to BLASTN and BLASTX searches of the GenBank non-redundant and EST databases using default settings (Altschulo 1997) as well as pairwise sequence alignment to the original A. aegypti clone sequence using the Blast2 program.

2.2 Results

From an initial screening of 71 A. aegypti cDNA’s we designed PCR primers for 35 putative comparative anchor tagged sequences (CATS). Of the thirty-six cDNA’s for which no primers were designed, 15 showed insufficient nucleotide identity with D.
based on our primer design criteria, to allow amplification of a minimum fragment size of at least 100 bp while the remaining twenty-one A. aegypti cDNA’s showed no nucleotide identity with either D. melanogaster or A. gambiae when screened against either the non-redundant or non-human non-mouse EST databases using the BLASTN program.

The 35 CATS primer sets were used to screen genomic DNA from six mosquito species including, A. aegypti, O. triseriatus, A. togoi, A. gambiae, C. pipiens, A. subalbatus, as well as two more distant species within the Dipterans: D. melanogaster, and R. pomonella. Twenty-three primer pairs (Table 2.1) yielded a single PCR product with at least one species in addition to A. aegypti, while 8 primer pairs produced a single PCR product only in A. aegypti. Four of the primer sets failed to produce a PCR product in all of the eight dipterans screened, including A. aegypti. Two typical CATS primer screening results are seen in Figure 2.1. The genetic locations in A. Aegypti (Severson et al., 2002) of all CATS for which we successfully obtained a PCR product in at least one non-A. aegypti species are shown in Figure 2.2.

We observed varying rates of success for amplification of individual CATS within each of the 8 species tested (Table 2.2). Amplification of a single PCR product varied among species relative to A. aegypti with a low of 13% seen in R. pomonella and a high of 60% seen in O. triseriatus. Thirteen of the twenty-three CATS primer pairs (56%) amplified with four or more of the species while the remaining 10 CATS primer pairs (44%) amplified between two to three different species.

We selected three CATS markers, LF178, LF158, and LF103 for sequence analysis and identity verification. They were selected as random representatives of each
of the three *A. aegypti* linkage groups. Also, these CATS amplified a single PCR product in at least 5 of the 8 species screened. We sequenced the PCR products from these CATS following amplification of genomic DNA from *A. aegypti, A. subalbatus* and/or *D. melanogaster*.
<table>
<thead>
<tr>
<th>CATS ID</th>
<th>Accession No.</th>
<th>PCR conditions</th>
<th>Predicted product Size (bp)</th>
<th>Primer sequences 5'-forward/reverse-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>LF90</td>
<td>T58320</td>
<td>50/1.5</td>
<td>110</td>
<td>GAAGAGGTCAGGTCTCGCT/ACAGATCCGTGACATGGAGC</td>
</tr>
<tr>
<td>LF92</td>
<td>BM005493</td>
<td>55/1.5</td>
<td>277</td>
<td>CCCAGATCAGCAAGCCTTG/TTACTTACCTCTTCTTCCG</td>
</tr>
<tr>
<td>LF96</td>
<td>BM005491</td>
<td>50/1.5</td>
<td>164</td>
<td>ATCAAGCTGCAGAAAGTC/GGGTTGTAGACGCGTCA</td>
</tr>
<tr>
<td>LF99</td>
<td>BM005477</td>
<td>54/1.5</td>
<td>356</td>
<td>AAGCGCAAGCAAGAAGAG/TGATCGACTTTCAGCT</td>
</tr>
<tr>
<td>LF103</td>
<td>BM005488</td>
<td>54/1.5</td>
<td>181</td>
<td>ATGCGCGATCTTCGGATC/AGTGGACGCAAATCTTCG</td>
</tr>
<tr>
<td>LF106</td>
<td>BM005490</td>
<td>54/1.5</td>
<td>175</td>
<td>CAAAAGCGCGCAAGAAAGA/GGGTGTACATTAGAG</td>
</tr>
<tr>
<td>LF108</td>
<td>T58322</td>
<td>50/1.5</td>
<td>265</td>
<td>AAGTGCGGCAAGCTGAGG/GCTTGTACCTCTTGAG</td>
</tr>
<tr>
<td>LF124</td>
<td>BM005518</td>
<td>54/1.5</td>
<td>103</td>
<td>GAGGAGAAGGCGAAAGC/TTGGCATGGCTGACAGC</td>
</tr>
<tr>
<td>LF129</td>
<td>BM005504</td>
<td>54/3.0</td>
<td>300</td>
<td>GCACCAGAAAGTGCGGAGATC/AGTACAACCGAGCGC</td>
</tr>
<tr>
<td>LF138</td>
<td>T58332</td>
<td>54/1.5</td>
<td>204</td>
<td>CGTGTTCAACTAGCACAT/CTCTTCAATGTAGATGAG</td>
</tr>
<tr>
<td>LF158</td>
<td>BM005485</td>
<td>50/3.0</td>
<td>239</td>
<td>GGAAGGTCTGCTGAGCACC/AGCACTTCTTGAGTTACC</td>
</tr>
<tr>
<td>LF168</td>
<td>R47184</td>
<td>54/1.5</td>
<td>113</td>
<td>GGAAGGCCTGCAAAGC/ACGAGTCTGCTGAGTACG</td>
</tr>
<tr>
<td>LF178</td>
<td>T58309</td>
<td>50/1.5</td>
<td>207</td>
<td>GCCGTGTTTGTCGAGATC/AGAGACGATACGGAGATG</td>
</tr>
<tr>
<td>LF179</td>
<td>BM005479</td>
<td>50/3.0</td>
<td>234</td>
<td>AACAAAGTGACCAACTGACT/ACACTTGACATTGCCC</td>
</tr>
<tr>
<td>LF188</td>
<td>BM005472</td>
<td>54/1.5</td>
<td>208</td>
<td>CTTTGAGACTTCGAGGCC/AGGAAAGCGGAAAGCC</td>
</tr>
<tr>
<td>LF218</td>
<td>BM005487</td>
<td>50/1.5</td>
<td>176</td>
<td>AAGGCGCGCATGCGAGTGGGGGTGCTTACG</td>
</tr>
<tr>
<td>LF231</td>
<td>BM005478</td>
<td>50/1.5</td>
<td>100</td>
<td>TTGCTGAGAGGTGTCAGAAG/TTGTCAGACTTCTCAGC</td>
</tr>
<tr>
<td>LF250</td>
<td>T58311</td>
<td>54/1.5</td>
<td>152</td>
<td>GTGGAGACTGGCAGG/TCAGGTTGAGTGTTACG</td>
</tr>
<tr>
<td>LF253</td>
<td>T58331</td>
<td>58/1.5</td>
<td>214</td>
<td>TGAAGACTTCTCAGAGC/CGAGTATACGGAAGTCT</td>
</tr>
<tr>
<td>LF272</td>
<td>BM005484</td>
<td>54/3.0</td>
<td>147</td>
<td>GTGCCGTTGATCCTCAGAA/GAGTGGTGTGGAGACGAGG</td>
</tr>
<tr>
<td>LF291</td>
<td>BM005482</td>
<td>54/1.5</td>
<td>250</td>
<td>GAACACCTGCTGCGGAGT/GTGACAGAGGAAACCAC</td>
</tr>
<tr>
<td>LF334</td>
<td>BM005506</td>
<td>54/1.5</td>
<td>200</td>
<td>GTCATGAGTACGAG/CTGAGGCTGACCTT</td>
</tr>
<tr>
<td>LF417</td>
<td>BM005499</td>
<td>54/1.5</td>
<td>228</td>
<td>GAAGACCAGAAGGTACGTT/GACCAGCGACCAGCTT</td>
</tr>
</tbody>
</table>
### TABLE 2.2

**CATS SUCCESS RATE IN DIPTERA**

<table>
<thead>
<tr>
<th>Genus species</th>
<th>Family</th>
<th>% Single product&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Aedes aegypti</em></td>
<td>Culicidae</td>
<td>100</td>
</tr>
<tr>
<td><em>Ochlerotatus triseriatus</em></td>
<td>Culicidae</td>
<td>65</td>
</tr>
<tr>
<td><em>Culex pipiens</em></td>
<td>Culicidae</td>
<td>61</td>
</tr>
<tr>
<td><em>Armigeres subalbatus</em></td>
<td>Culicidae</td>
<td>57</td>
</tr>
<tr>
<td><em>Anopheles gambiae</em></td>
<td>Culicidae</td>
<td>52</td>
</tr>
<tr>
<td><em>Drosophila melanogaster</em></td>
<td>Drosophilidae</td>
<td>39</td>
</tr>
<tr>
<td><em>Aedes togoi</em></td>
<td>Culicidae</td>
<td>30</td>
</tr>
<tr>
<td><em>Rhagoletis pomonella</em></td>
<td>Trypetidae</td>
<td>17</td>
</tr>
</tbody>
</table>

Note: "The percentage of the 23 primers listed in Table 2.1 that produced a single PCR product"
**FIGURE 2.1** Representative agarose gels of single individual Dipterans after amplification with CATS PCR primer pairs. Panel A shows amplification products with CATS LF178 primer pair. Lanes 1-3 show products from single *A. aegypti* individuals. Lanes 4-6 single *A. gambiae* individuals. Lanes 7-8 single *D. melanogaster* individuals. Lanes 9-11 single *C. pipiens* individuals. Lane 12 is a negative control lane. Lane 13 contains the 100 bp molecular weight marker. Panel B shows amplification products with CATS LF158 primer pair. Lane 1 contains the 100 bp molecular weight marker. Lanes 2-6 single *A. aegypti* individuals. Lanes 7-11 single *A. gambiae* individuals and lanes 12-16 single *D. melanogaster* individuals. Lane 17 is a negative control. The CAT LF158 PCR product from the *A. gambiae* and *D. melanogaster* individuals each contain an intron while the product from the *A. aegypti* individuals does not.
FIGURE 2.2 Relative distribution of CAT markers based on RFLP loci on Aedes aegypti genetic map. The numbers in ( ) indicate the species amplified by that marker: 1- Aedes aegypti; 2- Drosophila melanogaster; 3- Anopheles gambiae; 4- Aedes togoi; 5- Armigeres subalbatus; 6- Culex pipiens; 7- Ochlerotatus triseriatus; 8- Rhagoletis pomonella. Map distances are listed in Kosambi centiMorgans.
For LF178 we sequenced the PCR product obtained from one individual each from *A. aegypti*, *A. subalbatus*, and *D. melanogaster*. The CAT PCR products from *A. subalbatus* and *D. melanogaster* exhibited 92% and 83% nucleotide identity respectively when subjected to a pairwise alignment with the original *A. aegypti* LF178 cDNA sequence using the Blast2 program (Tatusova and Madden 1999). A representative nucleotide and amino acid sequence alignment for all CAT PCR products based upon LF178, LF158 and LF103 is shown in Figure 2.3.

For CAT PCR products designed from *A. aegypti* LF158, we sequenced one individual each from *A. aegypti* and *D. melanogaster*. The *D. melanogaster* sequence was interesting in that it was 394 nucleotides in length while the *A. aegypti* PCR product was only 241 nucleotides in length. BLAST analysis of the *D. melanogaster* PCR product showed 100% nucleic acid identity and 100% amino acid identity to the Drosophila Ribosomal Protein L12 gene, GenBank accession number AAF47152, which contains four introns. The *D. melanogaster* CAT sequence spanned the fourth intron. The first exon in the *D. melanogaster* CAT PCR product showed 82% nucleotide identity with the original *A. aegypti* LF158 clone (nucleotides 1-135) and the second exon showed 74% nucleotide identity with the same clone sequence (nucleotides 289-394). The CAT PCR product obtained from *An. gambiae* also seemed to contain an intron (Figure 1B).

Finally, for LF103 we sequenced PCR products obtained from one individual each of *A. aegypti* and *A. subalbatus*. The *A. subalbatus* CAT sequence showed 88% nucleotide identity with the original *A. aegypti* LF103 cDNA when the two sequences
**FIGURE 2.3 (a)** Nucleotide from *Aedes aegypti* (*Aa*), *Armigeres subalbatus* (*As*), *Drosophila melanogaster* (*Dm*) and *Aedes aegypti* LF178, 158 and 103 sequences. Black shading signifies complete nucleotide identity.
second exon
Aa158 136 TCAAGCACAACGGTAACATCACCTTCGATGAAGTTATCAGCATTGCCCGAGTGATGCGAC
Dm158 289 TCAAGCACAACGGTAACATCACCTTCGATGAAGTTATCAGCATTGCCCGAGTGATGCGAC
LF158 157 TCAAGCACAACGGTAACATCACCTTCGATGAAGTTATCAGCATTGCCCGAGTGATGCGAC

Aa158 196 CCCGCTCGATGCGCGAACTGTCCGGTACCTGCAAGGAAGTGCT
Dm158 349 CCCGCTCGATGCGCGAACTGTCCGGTACCTGCAAGGAAGTGCT
LF158 217 CCCGCTCGATGCGCGAACTGTCCGGTACCTGCAAGGAAGTGCT

As103 1 ATGCCGGAATCTTCGGGATCCG
Aa103 1 ATGCCGGAATCTTCGGGATCCGCAAGCTGTGCCTGAATATTTGCGTCGGTGAATCCGGTGAT
LF103 96 ATGCCGGAATCTTCGGGATCCGCAAGCTGTGCCTGAATATTTGCGTCGGTGAATCCGGTGAT

As103 61 AGGTGACCCCGTCGTAATATTGTTCGCGTGAATCCGGTGAT
Aa103 61 AGGTGACCCCGTCGTAATATTGTTCGCGTGAATCCGGTGAT
LF103 156 AGGTGACCCCGTCGTAATATTGTTCGCGTGAATCCGGTGAT

As103 121 AAGCGCTACATCTCGGCTGCAGTTGCGCTCAATCGAATCCGGTGAT
Aa103 121 AAGCGCTACATCTCGGCTGCAGTTGCGCTCAATCGAATCCGGTGAT
LF103 181 AAGCGCTACATCTCGGCTGCAGTTGCGCTCAATCGAATCCGGTGAT

FIGURE 2.3 (b) (continued)
**FIGURE 2.4** Deduced amino acid sequences from *Aedes aegypti* (*Aa*), *Armigeres subalbatus* (*As*), *Drosophila melanogaster* (*Dm*) and *Aedes aegypti* LF178, 158 and 103 sequences. Black shading signifies complete amino acid identity and similar amino acids are shaded in grey.
were subjected to a pairwise sequence alignment using the Blast2 computer program (Tatusova and Madden 1999).

2.3 Discussion

Mosquitoes, members of the insect order Diptera and family Culicidae, are divided into three subfamilies: Culicinae, Anophelinae, and Toxorychnitinae (Knight and Stone 1977). A limited number of species within each of the former two subfamilies serve as obligate intermediate vectors for a variety of diseases such as malaria, dengue fever, yellow fever, and lymphatic filariasis. Malaria alone accounts for 2.7 million deaths annually, lymphatic filariasis afflicts and disables upwards of 120 million people each year (WHO 1997), and dengue strikes 50 million individuals each year with an ever increasing number of accompanying cases of dengue haemorrhagic fever (DHF) (WHO 1997).

The development of CATS markers with utility across a wide range of mosquito species could prove to be very useful for identifying species-specific as well as common genes involved in vector competence for malaria, filariasis and arboviruses. That is, a comparative genomics approach facilitated by such markers would allow us to rapidly determine if the same genomic regions play a role in vector competence. Quantitative Trait Loci (QTL) have been identified that are associated with vector competence of *Aedes aegypti* to *Brugia malayi*, a filarioid nematode (Severson et al. 1994b, 1999) to *Plasmodium gallinaceum* (Severson et al. 1995) and to the Dengue-2 virus (Bosio et al. 2000). In *Anopheles gambiae*, QTL have been associated with encapsulation of
Positional cloning efforts to isolate and characterize the individual genes from these regions could then be better directed toward mosquito species that are better characterized genetically and more amenable to laboratory manipulations. Many of the mosquito species responsible for disease transmission to humans are extremely difficult to colonize in the laboratory or have never been successfully colonized.

Here we have developed a PCR-based strategy for producing genetic markers that can be used as a comparative core set for rapid genetic mapping of multiple species within the Culicidae. We developed these markers by utilizing a comparative genomics approach based upon comparative anchor tagged sequences. Utilizing available sequence data from *A. aegypti* EST’s and genes we were able to develop 23 CATS primer pairs which generated a single-copy PCR product in *A. aegypti* and at least one of seven additional mosquito or more distantly related dipteran species. These markers exhibit broad genome coverage within *A. aegypti* and thus broad potential coverage within other dipterans and clearly demonstrate the utility of a CATS-based strategy for marker development within the diptera.

Failure of the CATS primers to amplify a PCR product from genomic DNA was likely due to sufficient nucleotide divergence from the *A. aegypti* sequence leading to the selection of primers within those regions of nucleotide divergence or to the presence of an intron. Interference by introns could include intron/exon boundaries residing within the selected primer sequence or due to instances in which an intron was too large to allow for amplification.
Our CATS markers can be easily applied to genetic studies in previously uncharacterized mosquito species. For example, the CAT PCR products can rapidly be screened for polymorphism and subsequent mapping using Single Strand Conformation Polymorphism (SSCP) screening with subsequent silver staining (Hiss et al. 1994; Fulton et al. 2001). Single nucleotide polymorphisms (SNP’s) can be identified within CATS and genotyping performed through the employment of various techniques including Melting Curve analysis of Single Nucleotide Polymorphism’s (McSNP) (Akey et al. 2001).

The development and subsequent mapping of CATS within the Diptera should also be useful in identifying syntenic relationships among species. Comparisons of linkage relationships of biochemical and morphological markers suggested that linkage group conservation existed among the higher flies, *Ceratitis capitata*, *D. melanogaster*, and *Musca domestica*, but the relationship of mosquito linkage groups to these species was not clear (Weller and Foster 1993). Recent reports within the *Drosophila* indicates that while syntenic regions are common there has been extensive rearrangement of gene orders such that the estimated length of conserved chromosomal fragments between any two species of *Drosophila* is expected to be no greater than ca. 20-600 kb (Ranz et al. 1999, 2001). A comparative genomic analysis of *D. melanogaster* and *A. gambiae* uncovered microsyntenic blocks of 2-3 genes between these two species that were estimated to correspond to ca. 50-80 kb genomic segments (Bolshakov et al. 2002). A further analysis of the full *D. melanogaster* and *A. gambiae* genomes revealed the presence of 948 microsyntenic blocks, with a microsyntenic block being defined as having at least two orthologous groups within the region (Zdobnov et al. 2002). The total
fraction of *Anopheles* orthologs assigned to these microsyntenic blocks is 34% and constitutes a significant level of local conservation (Zdobnov et al. 2002). Previously, six microsyntenic clusters composed of 17 orthologues between *A. gambiae* and *D. melanogaster* were identified within the *Pen1* region of *An. gambiae* thus demonstrating that gene arrangements in their last common ancestor persist in these two species to some degree (Thomasova et al. 2002). When 157 *A. funestus* cDNAs were physically mapped to polytene chromosomes of this species as well as mapped *in silico* to *A. gambiae*, perfect preservation of synteny was observed but substantial shuffling of gene order was also seen (Sharakhov et al. 2002). In contrast to this, comparisons within the *Culicinae* show conservation of gene order (Anderson et al. 2001).

The development of high-density CAT maps could provide an excellent means for eventually delineating the boundaries of syntenic regions (Lyons et al. 1997) and identifying regions of additional microsynteny which may exist between dipterans in general and particularly among the great majority of mosquito species who are members of the sub-family *Culicinae*.

Additionally, CATS could prove valuable for phylogenetic studies. Using CATS, sequence comparisons can be made between a wide variety of mosquito species, which would contribute toward a greater understanding of systematic and taxonomic relationships between mosquitoes. It is believed that the lower diptera (mosquitoes, midges, gnats) diverged from the higher flies (*Drosophila* and house flies) about 250 million years ago (Ross 1951). Traditional evolutionary systematics in mosquitoes was based on morphological characters (Ross 1951) that placed the Anophelinae at the basal position, with Toxoryhnchitinae intermediate and Culicinae the most recently derived
(Ross 1951). Recent molecular data has supported this assertion as evidenced by Besansky and Fahey’s (1997) utilization of the white gene to evaluate taxonomic relationships among the Culicidae. It is interesting to note that our success in amplification with CATS primer pairs was highly correlated with the reported evolutionary relationship among the various species tested. Those species that had diverged more recently from A. aegypti, e.g. O. triseriatus and C. pipiens exhibited greater success in amplification as CATS, while those which diverged earlier, e.g. A. gambiae, D. melanogaster and R. pomonella, were less successful in amplifying with the CATS primer pairs.

Given that the total genome sequence is available for D. melanogaster (Myers et al. 2000), and just recently for A. gambiae (Holt et al. 2002), and an A. aegypti genome project has recently been initiated (Severson et al. 2004a), the application of comparative genomics to other Culicidae as well as to the Diptera in general is likely to grow rapidly (Kaufman et al. 2002). Our demonstration of CATS utility for inter-specific comparisons should facilitate that process.
CHAPTER 3

MICROSATELLITE ISOLATION AND LINKAGE GROUP IDENTIFICATION IN THE YELLOW FEVER MOSQUITO AEDES AEGYPTI

3.1 Introduction

Microsatellites or simple sequence repeats (SSRs) are short stretches of DNA sequence in which motifs of 1-6 bases are tandemly repeated (Schlötterer 2000; Weber and May 1989). They have become the genetic marker of choice for many eukaryotic species because of their high variability, co-dominant expression, and broad genome distribution (Tautz and Renz 1984; Tautz 1989; Webber and May 1989). In addition they are easily developed into PCR-based molecular markers that are particularly useful for small organisms with limited DNA.

Microsatellites have been used as genetic markers for a number of arthropod vectors for human disease. One hundred-fifty polymorphic microsatellite loci have been characterized in the African malaria vector, Anopheles gambiae s.l. (Zheng et al. 1993, 1996). They have been employed to evaluate population genetics and structure, as well as gene flow within and among A. gambiae s.l. (Kamau et al. 1998, 1999; Lanzaro et al. 1998; Lehman et al. 1996, 1997), A. gambiae s.s. (Carnahan et al. 2002; Tripet et al. 2001, 2005; Zheng et al. 1996) and Anopheles arabiensis (Donnelly and Towson 2000; Nyanjom et al. 2003). Microsatellites have also been mapped to the A. gambiae s.s
genome (Wang et al. 1999; Zheng et al. 1993, 1996) and have proven useful in QTL studies of susceptibility to *Plasmodium* (Gorman et al. 1997; Zheng et al. 1997) and of permethrin resistance (Ranson et al. 2004). Microsatellites have also been characterized in *Anopheles darlingi* (Conn et al. 2001) and *Anopheles funestus* (Baginets et al. 2003; Sharakhov et al. 2004, 2001; Sinkins et al. 2000), and have been used to investigate population genetics, structure and gene flow within *Anopheles maculatus* (Rongnoparut et al. 1996, 1999) and *Anopheles albimanus* (Molina-Cruz et al. 2004).

Microsatellites have also been characterized within the family Glossinidae (Luna et al. 2001; Solano et al. 1998, 1999) that serve as vectors for African trypanosomiasis. Microsatellites within the Glossinidae have been used for population genetic studies and evaluation of genetic structure and gene flow in *Glossina morsitans s.l.* (Krafsur and Endsley 2002), *Glossina pallidipes* (Krafur 2002), *Glossina morsitans centralis* (Krafsur et al. 2001), and *Glossina palpalis gambiensis* (Solano et al. 1999).

Additionally, within other arthropods microsatellite marker loci have been identified and employed to study population genetics and ecology, including investigations of gene flow, dispersal, migration, relatedness and parentage within ants (Gertsch et al. 1995; Tsutsui et al. 2003), honeybees (Estoup et al. 1995; Franck et al. 2001), parasitic wasps (Butcher et al. 2000), yellow-jacket wasps (Thoren et al. 1995), butterflies and moths (Anthony et al. 2001; Bogdanowicz et al. 1997; Keyghobadi et al. 2002), and aphids (Sloane et al. 2001).

In contrast to most other arthropods, microsatellites appear to be underrepresented within some members of the mosquito subfamily Culicinae. A limited number have been identified in *Culex pipiens* and *Culex pipiens quinquefasciatus* (Fonseca et al. 1998;
Keyghobadi et al. 2004; Smith et al. 2005). Within Aedes aegypti, the primary vector for the yellow fever and dengue viruses, six unique microsatellites have been identified using sequence databases (Barbazan et al. 1999; Huber et al. 2001; Ravel et al. 2001). Huber et al. (1999) isolated four microsatellites from an enriched partial library of 120 recombinant clones and successfully designed PCR primers for two of the microsatellites. They later screened an enriched-(CAA)$_n$ library and successfully designed primers for two additional microsatellites (Huber et al. 2001). Fagerberger et al. (2001) utilized an enrichment procedure to generate a number of microsatellite-containing libraries. They isolated 52 putative microsatellite containing clones and successfully designed primers for 7 putative microsatellites from three different enriched libraries. Six of these putative microsatellites were monomorphic in the F$_1$ intercross families they tested. Only one locus, TAG66, was polymorphic in the F$_1$ intercross families tested (Fagerberger et al. 2001). Still, these limited numbers of microsatellites have been employed for several studies of A. aegypti population structure (Huber et al. 2004, 2002a, 2002b; Ravel et al. 2001, 2002).

The development of additional microsatellite markers for A. aegypti would greatly benefit both genetic and populations studies, and thereby assist efforts to investigate the genetic basis for dengue vector competence and to assess population variation subsequent to developing genetic control strategies. Here, we report on our efforts to isolate and develop microsatellite markers from three libraries enriched for (GA)$_n$, (AAT)$_n$ and (TAGA)$_n$ motifs from genomic DNA isolated from A. aegypti, as well as by the direct screening of selected clones from three different A. aegypti derived libraries. We identified 12 new microsatellites that segregate as single locus, co-dominant markers, and
determined the genetic map positions for six of these markers. We also describe the close
association of microsatellites in *A. aegypti* with known repetitive elements, which likely
accounts for some of the limited success in developing an extensive panel of
microsatellite marker loci from the present study, as well as previous attempts by other
investigators (Fagerberger et al. 2001).

3.2 Materials and methods

3.2.1 *Microsatellite enriched library construction, sequence analysis, and primer
design*

Genomic DNA was extracted from ca. 700 pupae from our *A. aegypti* Trinidad
strain. Pupae were gently homogenized using a glass homogenizer in equal parts DEB
(0.5% SDS, 0.2 M NaCl, 25 mM EDTA, 10 mM Tris, pH 8.0) and phenol followed by
phenol/chloroform extraction, including incubation with RNaseA (Severson 1997).

Genomic DNA was enriched for three microsatellite motifs (GA, AAT, and
TAGA) using a proprietary technique by Genetic Identification Services (Chatsworth,
CA). The enriched genomic fragments were ligated into the *Hind*III site of the pUC19
plasmid vector and transformed into *E. coli* strain DH5α cells via electroporation. White
recombinant colonies were arrayed manually into 96-well microplates and stored at -
80°C. Sequencing was performed using an ABI 3700 DNA sequencer (PE Applied
Biosystems, Crescent City, CA) at the Purdue University Genomics Center.

Sequence data were analyzed using RepeatMasker (Smith et al. 2004) and Sputnik
(Abajian 1994). RepeatMasker screens DNA sequences against a library of known
repetitive elements and returns a masked query sequence ready for database searches as
well as a table annotating the masked regions. Sputnik utilizes a recursive algorithm to search for two to five base repeated patterns of nucleotides. Identity searches were conducted on all clones by BLASTN analysis against the non-redundant GenBank database (Altschul et al. 1997).

Primers for putative unique sequences flanking microsatellites were designed using PRIMER3 (Rozen and Skaletsky 2000). Primers were designed to be 18-27 nt in length with a G+C content of 50%, and optimum anneal temperature of 60°C, with a range of 57-63°C. When possible, we included a GC clamp and selected primers that would yield a 150-300 bp product. A sequence similarity and homology search was conducted on all putative PCR primers using the fasta3 computer program (Pearson 1990). This program was employed as it allows the user to set the KTUP (word size) to 1, allowing for a more sensitive search that facilitates avoidance of primer sequences that may have homology to known repetitive elements.

3.2.2 Existing library screening for microsatellites

Three *A. aegypti* derived libraries were screened with three dinucleotide motifs; GA, CA, GC, and five trinucleotide repeats; ATA, AGT, ATC, CAA and ATG that included: (i) 85 cDNA and genomic DNA plasmid clones previously mapped as RFLP markers (Severson et al. 2002), (ii) 207 cosmids from the *A. aegypti* ATC-10 cell line that had been previously mapped using Fluorescent In Situ Hybridization (FISH) (Brown et al. 1995 and unpublished data) and (iii) 36 cosmids prepared from the *A. aegypti* Liverpool strain known to contain mapped RFLP marker sequences.
Clones were manually arrayed in 96-well plates, transferred to nylon membranes (NEN™ Life Science Products, Boston, MA) using a 96-well plate replicator and grown overnight at 37°C on LB plates selected with ampicillin. Clones were fixed to the nylon membranes by placing them sequentially onto Whatman paper drenched with 10% SDS for 3 minutes, 0.5 M NaOH for 5 minutes, and 1.5 M NaCl/0.5 M Tris pH 8.0 for 5 minutes. Membranes were allowed to dry for 1 hour at room temperature and UV crosslinked at 1200 x 100 µJ/cm². Membranes were then hybridized overnight at the probe specific temperature in a hybridization oven (Hybaid, Milford, MA) in a pre-hybridization solution consisting of 0.1 M Sodium Phosphate buffer pH 7.8, 20 mM Sodium pyrophosphate, 5 mM EDTA pH 8.0, 0.1% SDS, 10% (w/v) Sodium dextran sulfate, 1.0 mM 0-phenoanthroline, 500 µg/ml Heparine sulfate, 50 µg/ml Yeast RNA.

Di- and trinucleotide oligonucleotides were end-labeled using [γ-³²P] ATP and bacteriophage T4 polynucleotide kinase, and purified by ethanol precipitation as described by Sambrook et al. (1989). Following addition of the labeled probe, membranes were hybridized overnight in a hybridization oven. Membranes were then washed for 15 min each in 0.1% SDS/2X SSC and 0.1% SDS/0.2X SSC, and exposed to X-ray film for 4-8 hours.

3.2.3 Cosmid subcloning, sequencing and primer design

Purified cosmid DNA from clones selected with oligonucleotide probes was isolated using the alkaline lysis miniprep protocol (Birnboim and Doly 1979). The purified DNA was then digested with Sau3AI according to supplier’s recommendations (Promega, Madison, WI). One half of each digest was size fractionated in 0.9% agarose
gels at 1.8-2.0 V/cm. Gels were UV nicked in a crosslinker at 800 x 100 µJ/cm², Southern blotted, and hybridized with the appropriate end-labeled oligonucleotide as described above. The remaining clone digest was then size fractionated in 0.9% LMT agarose and fragments positive for microsatellites were identified by comparison with positive fragments identified by autoradiography. Microsatellite containing bands were excised and subcloned into the pGEM®-3Z (Promega) cloning vector. Subclones were plated onto LB ampicillin plates. DNA from the recombinant clones was transferred to nylon membranes (NEN™ Life Science Products), and hybridized with the appropriate end-labeled oligonucleotide.

Sequencing of positive subclones was performed using an ABI 310 DNA sequencer (PE Applied Biosystems). Microsatellites were identified by visual inspection and subjected to BLASTX analysis (Altschul et al. 1997) against the GenBank database to identify sequences containing known A. aegypti repetitive elements. Primers for unique sequences flanking the microsatellites were designed using Primer3 (Rozen and Skaletsky 2000).

3.2.4 PCR amplification of microsatellites

Putative single-copy microsatellite loci were amplified from individual mosquitoes among four strains of A. aegypti (Liverpool, Moyo-R, Red, and Trinidad). PCR reactions were performed in 25 ul volumes containing 1X Taq buffer (10 mM KCl, 2 mM Tris, pH 9.0, 0.02% TritonX), 1.5 or 3.0 mM MgCl₂, 0.4 mM each dATP, dCTP, dGTP, dTTP, 5 pmoles of each primer, 1 U Taq DNA polymerase and 5 ng genomic DNA. PCR reactions for microsatellite loci A10, H08, B07, F06, and G11 were
performed under the following conditions: 94°C for five minutes, followed by 30 cycles of 94°C for 1 min, 60°C anneal for 1 min, 72°C extension for 2 min, and a final 72°C extension for 10 min. PCR reactions for B19, B6, M205 and M313 were performed under the following conditions: 94°C for 5 minutes, followed by 39 cycles at 94°C for 20 sec, 55°C for 20 sec, and 72°C for 30 sec, and a final 72°C extension for 10 min. PCR products were size fractionated in 2% agarose gels and visualized with ethidium bromide under UV light.

3.2.5 Denaturing polyacrylamide gel electrophoresis

Polymorphisms were resolved on 31.0 cm x 38.5 cm x 0.44 mm denaturing gels (4.0% acrylamide (19:1 polyacrylamide:bis) 1X TBE, 38% urea), preheated to 50°C. Three microliters of PCR product was mixed with 5 ul of denaturing loading mix (10 mM NaOH, 95% formamide, 0.05% bromophenol blue, and 0.05% xylene cyanol). The gels were run at 2200 V for 45 minutes. The gels were fixed for 20 minutes in 10% acetic acid and the DNA was then visualized by silver staining using the genePrint STR system (Promega).

3.2.6 Mosquito strains and crosses

A segregating F1 intercross mapping population, SDAM2, produced by pairwise matings between *A. aegypti* LIV*SD* strain females and MOYO strain males was previously described (Severson et al. 1999).
3.2.7 Restriction Length Fragment Polymorphism (RFLP) markers and Single Nucleotide Polymorphism (SNP) markers

RFLP and SNP anchor loci were selected from the existing linkage map for integrating SSR loci. Five RFLP markers were mapped on linkage group one while seven RFLP markers were selected to give broad coverage over linkage group 2. These markers represented either random *A. aegypti* cDNA clones, random RAPD clones or known *A. aegypti* genes (Severson et al. 1999; Severson et al. 2002). Three SNP markers, *ApolipoII*, *Cyp9*, and LF108 (Severson et al. 2002), were selected as anchor loci for integrating SSR into linkage group 3.

3.2.8 Linkage mapping

Multi-point linkage analysis was performed using the MAPMAKER/EXP 3.0 computer package (Lander et al. 1987) with the linkage LOD score set at 3.0 with the Kosambi (1944) mapping function. The linkage map was drawn using the DrawMap computer program (van Ooijen 1994).

3.3 Results

3.3.1 Microsatellite enriched libraries

We obtained sequence data for 253 putative microsatellite-containing GA, AAT, or TAGA-enriched clones from three microsatellite enriched plasmid libraries prepared from the Trinidad strain of the yellow fever mosquito *A. aegypti*. These data were screened against a database of known *A. aegypti* interspersed repeats and low complexity DNA sequences using RepeatMasker. This analysis indicated that 110 of the 253 clones
(43.5%) contained known repetitive elements (Table 1), and because 182 repetitive elements were identified (Table 2), many clones contained multiple elements. The 182 repetitive elements consisted of seven different repeat class family types represented by 15 different repeat types. The most common repeat type was the *Feilai* element, a member of the short interspersed repetitive element (SINE) class (Tu 1999). Sputnik analyses identified 127 clones (50.2%) that contained a microsatellite. However, for 59 of these clones (46.5%) the microsatellites were within or immediately adjacent to known repetitive elements or low-complexity DNA regions (Table 1). Of the remaining 68 clones, several were duplications of the same sequence resulting in a final total of 30 putative unique sequence clones containing a microsatellite.

We were able to design PCR-primers for unique sequences flanking the microsatellite for 12 of the 30 putative unique sequences. The remaining sequences were unsuitable for primer design because the microsatellites were too close to the clone end or the base composition of the flanking regions was unsuitable for quality primer design. The 12 putative single-copy microsatellite loci were tested for amplification with genomic DNA from individual mosquitoes representing four strains of *A. aegypti*. Eight of the primer pairs amplified a single PCR product when visualized on agarose gel, while the remaining four primer pairs amplified multiple PCR products indicating that they likely recognized repetitive sequences. SDS-PAGE gel analysis indicated that seven of these eight microsatellites were single locus markers. Five were polymorphic within and among the strains tested while one, G11, was single locus and polymorphic within the Red strain and between the Red and Liverpool strain, but was monomorphic between Liverpool and MoyoR, and the other marker, F06, was also single locus, but
monomorphic within the strains tested. The seven single locus microsatellites consisted of one perfect dinucleotide repeat (B07), two imperfect dinucleotide repeats (A10, B6), three perfect trinucleotide repeats (B19, G11, H08), and one perfect tetranucleotide repeat (F06) (Table 3). One of the trinucleotide repeats (B19) also contained an imperfect dinucleotide repeat.

TABLE 3.1
ANALYSIS OF ENRICHED LIBRARIES WITH REPEATMASKER\textsuperscript{a} and SPUTNIK\textsuperscript{b}

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Total Clones Sequenced</th>
<th>No. Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>RepeatMasker</td>
<td>Clones positive for low complexity and interspersed repeat sequences</td>
<td>110</td>
</tr>
<tr>
<td>Sputnik</td>
<td>Clones positive for microsatellite sequences</td>
<td>127</td>
</tr>
<tr>
<td>Sputnik and RepeatMasker</td>
<td>Clones positive for microsatellite sequences and negative for interspersed repeats</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Clones positive for microsatellite sequences and interspersed repeats</td>
<td>59</td>
</tr>
</tbody>
</table>

Note: \textsuperscript{a}Screens DNA sequences for interspersed repeats and low complexity DNA sequences (Abajian 1994).
\textsuperscript{b}Searches DNA sequences for microsatellites of between 2-5 nucleotides in length using a recursive algorithm (Smith et al. 2004).
TABLE 3.2

REPEAT CLASSIFICATIONS ASSOCIATED WITH *Aedes aegypti*

MICROSATELLITES (AFTER SEVERSON ET AL. 2001)

<table>
<thead>
<tr>
<th>Repeat Class/Family</th>
<th>Element</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTR retrotransposons</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>Non-LTR retrotransposons</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LINES</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Zebedee</em></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td><em>Jam1</em></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td><em>Mosquil</em></td>
<td>22</td>
</tr>
<tr>
<td></td>
<td><em>JuanA</em></td>
<td>14</td>
</tr>
<tr>
<td></td>
<td><em>Lion</em></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td><em>heF-A-like</em></td>
<td>1</td>
</tr>
<tr>
<td>SINE</td>
<td></td>
<td>86</td>
</tr>
<tr>
<td></td>
<td><em>Feilai</em></td>
<td></td>
</tr>
<tr>
<td>DNA/LTR</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td><em>Schelli-like</em></td>
<td>1</td>
</tr>
<tr>
<td>DNA/MITE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Dufu</em></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td><em>Homey</em></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td><em>Pony</em></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td><em>Wukong</em></td>
<td>13</td>
</tr>
<tr>
<td></td>
<td><em>Youzi</em></td>
<td>7</td>
</tr>
<tr>
<td>DNA/MSITE</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td><em>Microuli</em></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td><em>P2405</em></td>
<td>182</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.3.2 Microsatellite from cosmid subclones

We isolated a total of 9 microsatellites by screening and subcloning a selected set of cosmid clones. SDS-PAGE gel analysis indicated that M201, M205, and M313 were single locus and polymorphic between the four strains screened. However, the subclone M313 sequence was identical to the TAG66 locus reported by Fagerberger et al. (2001). M106 and M307 were single locus but monomorphic. We were unable to identify primer sequences suitable for PCR amplification from subclones M203 and M102, and the microsatellites in subclones M209 and M310 were located too close to one end of the subclone insert for primer design. All five of these single locus microsatellites were trinucleotide repeats (Table 3). Three were imperfect repeats (M205, M307, M313), while two were perfect repeats (M106, M201).

3.3.3 Linkage analysis

We performed linkage analyses with 21 *A. aegypti* genetic markers. Twelve of these were RFLP markers, 2 were SNP markers while five were *A. aegypti* microsatellite markers isolated from the enriched plasmid libraries (A10, B07, B6, B19, H08), and two were microsatellite markers obtained by screening an *A. aegypti* cosmid library (M205, M313). Significant deviations from the expected 1:2:1 ratio were observed with 19% of the loci examined in this F$_1$ intercross (Table 3.3). Six of the microsatellite markers mapped to either chromosome 2 or 3 on the *A. aegypti* linkage map (Fig. 1), while one microsatellite, B6, while segregating in our mapping population, did not show linkage to any of the three linkage groups. Two of the microsatellites, F06 and G11 were not
polymorphic in our mapping population. We have submitted sequence data for all
twelve single-copy microsatellites isolated in this study to GenBank (Table 3.4).

**TABLE 3.3**

SEGREGATION OF RFLP, MICROSATellite AND SNP MARKERS IN THE F1 INTERCROSS USED TO CONSTRUCT AN *Aedes aegypti* LINKAGE MAP

<table>
<thead>
<tr>
<th>Marker&lt;sup&gt;a&lt;/sup&gt;</th>
<th>No. of individuals&lt;sup&gt;b&lt;/sup&gt;</th>
<th>χ²&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L</td>
<td>H</td>
</tr>
<tr>
<td><strong>Chromosome 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LF230</td>
<td>20</td>
<td>49</td>
</tr>
<tr>
<td>LF198</td>
<td>17</td>
<td>54</td>
</tr>
<tr>
<td>B8L260</td>
<td>16</td>
<td>52</td>
</tr>
<tr>
<td>LF178</td>
<td>16</td>
<td>52</td>
</tr>
<tr>
<td><em>LAP</em></td>
<td>15</td>
<td>52</td>
</tr>
<tr>
<td><strong>Chromosome 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F17M590</td>
<td>21</td>
<td>58</td>
</tr>
<tr>
<td><em>Aricl</em></td>
<td>22</td>
<td>56</td>
</tr>
<tr>
<td>B8M980</td>
<td>22</td>
<td>55</td>
</tr>
<tr>
<td>B07</td>
<td>11</td>
<td>44</td>
</tr>
<tr>
<td>H08</td>
<td>13</td>
<td>38</td>
</tr>
<tr>
<td>LF282</td>
<td>25</td>
<td>52</td>
</tr>
<tr>
<td>A13975</td>
<td>25</td>
<td>53</td>
</tr>
<tr>
<td>M205</td>
<td>8</td>
<td>29</td>
</tr>
<tr>
<td>LF98</td>
<td>19</td>
<td>56</td>
</tr>
<tr>
<td>LF338</td>
<td>20</td>
<td>55</td>
</tr>
<tr>
<td><strong>Chromosome 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A10</td>
<td>18</td>
<td>39</td>
</tr>
<tr>
<td>B19</td>
<td>22</td>
<td>36</td>
</tr>
<tr>
<td><em>Apolip</em></td>
<td>17</td>
<td>39</td>
</tr>
<tr>
<td>LF108</td>
<td>11</td>
<td>31</td>
</tr>
<tr>
<td>M313</td>
<td>15</td>
<td>32</td>
</tr>
<tr>
<td><em>Cyp9</em></td>
<td>20</td>
<td>38</td>
</tr>
</tbody>
</table>

Note: <sup>a</sup>Marker loci are arranged in chromosomal order  
<sup>b</sup>(L) LiverpoolSD strain; (H) heterozygote; (M) Moyo-R strain  
<sup>c</sup>(*) P < 0.05; (**) P < 0.01 (loci tested for expected 1:2:1 ratio)
### TABLE 3.4

SINGLE-COPY MICROSATELLITE SEQUENCES IDENTIFIED FROM ENRICHED PLASMID LIBRARIES AND SELECTED COSMID SUBCLONES

<table>
<thead>
<tr>
<th>Library Source</th>
<th>Sequence ID</th>
<th>Repeat Motif</th>
<th>Linkage Group</th>
<th>Primers 5’-3’</th>
<th>Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>B&lt;sup&gt;a&lt;/sup&gt;</td>
<td>A10</td>
<td>CT&lt;sub&gt;10&lt;/sub&gt;(TT)CT</td>
<td>III</td>
<td>ATCCCGAAAAACAAATCGTGA ATCGAACATCGCTTCCAACCT</td>
<td>DU169901</td>
</tr>
<tr>
<td>B&lt;sup&gt;a&lt;/sup&gt;</td>
<td>B07</td>
<td>GA&lt;sub&gt;15&lt;/sub&gt;</td>
<td>II</td>
<td>CAAAACCAACATGCTCAGC GAAGATCGGCTCCTTACGTT</td>
<td>DU169902</td>
</tr>
<tr>
<td>B&lt;sup&gt;a&lt;/sup&gt;</td>
<td>B6</td>
<td>CT&lt;sub&gt;17&lt;/sub&gt;(T)CT&lt;sub&gt;2&lt;/sub&gt;</td>
<td>unk</td>
<td>GATAAAGGCTCAAGGTAACG CATTTGGCATAGGAACAGC</td>
<td>DU169904</td>
</tr>
<tr>
<td>B&lt;sup&gt;a&lt;/sup&gt;</td>
<td>B19</td>
<td>CAT&lt;sub&gt;7&lt;/sub&gt;</td>
<td>III</td>
<td>ATGGGCGTGAGAACATTTTG GAGGAGTGGAGCATAGGAGTG</td>
<td>DU169905</td>
</tr>
<tr>
<td>B&lt;sup&gt;a&lt;/sup&gt;</td>
<td>H08</td>
<td>TCG&lt;sub&gt;7&lt;/sub&gt;</td>
<td>II</td>
<td>AAAAACCACGATACCAGGAAG ACGCGATACACACTGAAAATG</td>
<td>DU169903</td>
</tr>
<tr>
<td>C&lt;sup&gt;b&lt;/sup&gt;</td>
<td>G11</td>
<td>TTA&lt;sub&gt;16&lt;/sub&gt;</td>
<td>unk</td>
<td>TGTCTCATGGATGGCCTTATT GTCAAGAATTITGGGGACCA</td>
<td>DU169906</td>
</tr>
<tr>
<td>D&lt;sup&gt;c&lt;/sup&gt;</td>
<td>F06</td>
<td>TAGA&lt;sub&gt;8&lt;/sub&gt;</td>
<td>unk</td>
<td>GCCAAAAACCAACAAACACGG AATCGACCCGACCAAATAAACA</td>
<td>DU169907</td>
</tr>
<tr>
<td>113.9G&lt;sub&gt;7d&lt;/sub&gt;</td>
<td>M205</td>
<td>GAA(GTA)GAA&lt;sub&gt;4&lt;/sub&gt;</td>
<td>II</td>
<td>CTATTAGTCCCAAGTCTCCGG TGTGATAGGGATAAGAAGCTGC</td>
<td>DU169908</td>
</tr>
<tr>
<td>126.8E&lt;sub&gt;8e&lt;/sub&gt;</td>
<td>M313</td>
<td>ATG&lt;sub&gt;3&lt;/sub&gt;(ATA)ATG</td>
<td>III</td>
<td>CACCTCGTGACATAAACACACC AGTACCAAGCCACGTCACA</td>
<td>DU169909</td>
</tr>
<tr>
<td>102.1D&lt;sub&gt;5f&lt;/sub&gt;</td>
<td>M201</td>
<td>ATA&lt;sub&gt;36&lt;/sub&gt;</td>
<td>II</td>
<td>GGAGCATTCATAGGAATTGTCA GAGATGAACACAGTCATAAGG</td>
<td>DU635091</td>
</tr>
<tr>
<td>Library Source</td>
<td>Sequence ID</td>
<td>Repeat Motif</td>
<td>Linkage Group</td>
<td>Primers 5’-3’</td>
<td>Accession No.</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td>--------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
</tr>
<tr>
<td>125.4D7^g</td>
<td>M106</td>
<td>ATG₅</td>
<td>I</td>
<td>TGTACTCAGTTCATCGGGGA TTTTCATGCGAGAGATGACG CAGCTCCACCCCATCAGATTT GTTTTCCCAGTCACGACGTT</td>
<td>DU635090</td>
</tr>
<tr>
<td>6.3F1^h</td>
<td>M307</td>
<td>CAT₃(CGT)CAT(CGT)CAT(CGT)</td>
<td>III</td>
<td>CAGCTCCACCCCATCAGATTT GTTTTCCCAGTCACGACGTT</td>
<td>DU635092</td>
</tr>
</tbody>
</table>

Note: ^a isolated from plasmid libraries enriched for (GA)ₙ motifs
^b isolated from plasmid libraries enriched for (AAT)ₙ motifs
^c isolated from plasmid libraries enriched for (TAGA)ₙ motifs
^d isolated from *A. aegypti* Liverpool cosmid clone containing previously mapped RAPD marker A13L975
^e isolated from *A. aegypti* Liverpool cosmid clone containing previously mapped *def* marker
^f isolated from *A. aegypti* Liverpool cosmid clone containing previously mapped *vcp* marker
^g isolated from *A. aegypti* Liverpool cosmid clone containing previously mapped LF198 marker
^h isolated from *A. aegypti* ATC-10 cell line cosmid clone physically mapped to chromosome 3 as a fluorescent *in situ* hybridization (FISH) probe.
FIGURE 3.1 Linkage map for *Aedes aegypti*. The map positions of microsatellites isolated from either an enriched library or an *Aedes aegypti* cosmid library are indicated by asterisk. Map distances are given in centiMorgans.
3.4 Discussion

Despite their demonstrated utility among a wide variety of arthropods, microsatellites are not highly abundant nor easily isolated within all arthropod species. Within four different species of phlebotomine sandflies, *Lutzomyia whitmani*, *Lutzomyia longipalpis*, *Phlebotomus papatasi*, and *Phlebotomus langeroni*, microsatellites containing the (TG)$_n$ motif varied greatly in abundance, with the average frequency of occurrence ranging from 1 per 100 kb in *P. langeroni* to 1 per 1000 kb in *P. papatasi* and *L. longipalpis* (Day and Ready 1999). In contrast, the (GT)$_n$ motif repeats occur about once every 26 kb in *A. gambiae* (Knudson et al. 1996), once every 68 kb in *A. maculatus* (Rongnoparut et al. 1996), once every 88-224 kb in *Drosophila melanogaster* (Schug et al. 1998) and once every 10,000 kb in *A. aegypti* (Knudson et al. 1996). The frequency of repeats exhibiting the (TC)$_n$ motif were estimated to occur at a rate of 1 per 25 kb in *P. langeroni* and 1 per 250 in other sandfly species (Day and Ready 1999). Finally, (AAT)$_n$ repeats were found to occur at a frequency of 1 per 12.5 kb in *P. papatasi* and at a frequency of 1 per 125 kb in other sandfly species (Day and Ready 1999). Two species of mites, *Amblyseius fallacies* and *Tetranychus urticae*, when tested for 10 repeat motifs, yielded considerably fewer microsatellites than the typical vertebrate species, and trinucleotide repeats were more abundant than dinucleotide repeats (Navajas et al. 1998). It is clear that the frequency and types of microsatellites can occur at widely varying rates within many arthropod species.

Microsatellites appear to be underrepresented and, therefore, of limited utility in *A. aegypti*. In this study, we screened genomic libraries enriched for three repeat motifs (GA)$_n$, (AAT)$_n$, (TAGA)$_n$ and selected cosmids clones for useful, single locus
microsatellite sequences. We have identified and characterized 12 single locus microsatellite loci within the *A. aegypti* genome and, therefore, they should prove useful for genetic mapping studies as well as tools for studying population genetics. With this study, we have increased the number of documented single-copy microsatellite markers in *A. aegypti* from nine to twenty-one. Fagerberger et al. (2001) employed an enrichment protocol in order to identify microsatellites within *A. aegypti* and failed to identify length variation at any single microsatellite locus, while Huber et al. (1999), using a similar enrichment protocol designed to identify (CAA)$_n$ microsatellites isolated four imperfect repeats. Microsatellites appear to be underrepresented within other culicine mosquitoes as both di- and trinucleotide repeats are underrepresented within *Culex pipiens* (Keyghobadi et al. 2004; Smith et al. 2005) and *Culex pipiens quinquefasciatus* (Fonseca et al. 1998).

Our results also indicate that microsatellites are not only underrepresented in the *A. aegypti* genome, but most of those that are present are associated with repetitive sequences. Over 46% (59/127) of the microsatellites identified by Sputnik were found directly within or closely linked with known *A. aegypti* repetitive elements. Further, only 30 of the remaining 68 putative single-copy microsatellite loci were unique. Following PCR amplification and screening, only seven microsatellites from our original 127 were demonstrated to be single-copy and polymorphic, while an additional five amplified as multiple-copy loci, even though RepeatMasker did not detect repetitive elements in sequences flanking the microsatellites. This was not unexpected as our RepeatMasker database only contains known repetitive elements and many novel repetitive elements remain to be identified.
There were a number of retroelements and, to a lesser degree, DNA elements that were highly associated with microsatellites (Table 2). It was not surprising that the SINE element, Feilai, was frequently associated with microsatellites. These elements alone represented 47% of all repetitive element types associated with microsatellites in our study. Feilai elements are found throughout the A. aegypti genome, and account for up to 2% of the genome (Tu 1999). The 500-bp regions flanking these elements have been shown to be exceptionally AT-rich (Tu 999), and this may explain why the microsatellite libraries enriched for AAT and TAGA sequences yielded only two single-copy repeat loci while those enriched for GA yielded five. Other common repeat types associated with microsatellites included the Mosqi (Tu and Hill 1999) and Juan (Mouches et al. 1992), members of the non-long terminal-repeat (nonLTR) or long interspersed repetitive element (LINE) family, and Zebedee (Warren et al. 1997), a member of the long terminal repeat (LTR) family. Among the DNA elements most frequently associated with microsatellites were Wukong (Tu 1997) and Pony (Tu 2000), both miniature inverted-repeat transposable elements (MITEs), which together accounted for almost 11% of the repetitive elements associated with the microsatellites isolated in our study. Finally, several microsatellites were associated with Microuli, a miniatiure subterminal inverted-repeat transposable element (MSITEs) (Tu and Orphanidis 2001).

It is likely not surprising that microsatellites within some Culicines, and within A. aegypti in particular, would be associated frequently with a variety of dispersed repetitive elements. Within humans, a large proportion of microsatellites are associated with Alu and other SINEs (Arcot et al. 1995; Nadir et al. 1996). Within plants, a large proportion of microsatellites from Hordeum vulgare have flanking sequences exhibiting homology.
to a number of retrotransposons and dispersed repetitive elements (Ramsay et al. 1999), while within *Oryza sativa*, the *micron* element which resembles a MITE is intimately associated with microsatellites (Akagi et al. 2001). Finally, within the Diptera a novel retrotransposon, *mini-me*, is associated with microsatellite genesis and dispersal. It provides a model for how microsatellites can be associated with both the 5’ and internal regions of retrotransposons (Wilder and Hollacher 2001).

Despite the limited number, the nine previously identified microsatellites have proven quite useful in evaluating the population genetics of *A. aegypti* in a number of populations including, Ho Chi Minh City, Vietnam (Huber et al. 2002a, 2002b), Phnom Penh, Cambodia (Huber et al. 2004), Cote d’ Ivoire, Africa (Ravel et al. 2002), and Guaymas, Mexico (Ravel et al. 2001). The twelve new microsatellite markers that we have identified in this study should provide additional tools for understanding population structure and gene flow within *A. aegypti* and thereby facilitate efforts to understand disease transmission within these populations. Finally, the ongoing *A. aegypti* genome project (Severson et al. 2004a) will undoubtedly reveal additional single copy microsatellites for development as useful genetic markers, although our results indicate the number of useful microsatellite loci is likely to remain low.
CHAPTER 4

IDENTIFICATION OF QUANTITATIVE TRAIT LOCI ASSOCIATED WITH
FILARIAL WORM SUSCEPTIBILITY IN AN AEDES AEGYPTI ADVANCED
INTERCROSS LINE

4.1 Introduction

Lymphatic filariasis, caused by the nematode worms Wuchereria bancrofti, Brugia malayi, and Brugia timori infects over 120 million people worldwide in 83 different countries (Michael et al. 1996). Mosquitoes are the obligate intermediate hosts for filarial worms, and natural vectors representing over 70 species and subspecies are found within the genera Aedes, Anopheles, Culex and Mansonia (Bartholomay and Christensen 2002; Scott 2000). A program currently exists for the elimination of lymphatic filariasis, The Global Alliance to eliminate Lymphatic Filariasis. This program is a collaborative effort involving over 80 national ministries of health, as well as a variety of non-governmental organizations (NGOs) and entities from the private sector. The goal of the Alliance is the elimination of lymphatic filariasis through an annual series of mass drug administrations (MDA). Although drug treatment is highly effective it is possible that drug treatment in the absence of vector control may not lead to the elimination of transmission. Elimination campaigns based upon a single strategy are often less than successful because of short windows of opportunity, decreasing compliance, and development of drug resistance (Burkot et al. 2002)
The development of vector-based control measures for filariasis is dependent upon an understanding of how the parasite develops within, and is transmitted by, the intermediate mosquito host. Mosquito vectors of disease can be classified as susceptible, resistant or refractory in terms of their relationship with a potential pathogen (Christensen and Severson 1993). Susceptible mosquitoes permit the full development of the parasite or pathogen whereas resistant mosquitoes mount an active immune response against the pathogen. Refractory mosquitoes are incompatible with the pathogen such that the pathogen dies before it completes development. It has been demonstrated that mosquitoes exhibit variation in vector competence for filarial worm infection. In a classic series of genetic studies, MacDonald (1962a, 1962b) and other co-workers (MacDonald and Ramachandran 1965; MacDonald and Shepard 1965) demonstrated that susceptibility of *A. aegypti* to infection with the parasite *B. malayi* was under the primary control of a sex-linked recessive gene, $f^m$. McDonald and Ramachandran (1965) later demonstrated that this gene also controlled susceptibility of *A. aegypti* to several other filarial worms. It was later demonstrated that the $f^m$ gene was not entirely responsible for susceptibility in *A. aegypti* and this suggested that other modifying genes must also play a role (MacDonald, 1963).

The development of DNA-based genetic markers has proven to be an excellent tool to partition complex genetic traits, such as filarial worm susceptibility in *A. aegypti*, into discrete Mendelian components. A linkage map based on restriction fragment polymorphism (RFLP), single nucleotide polymorphism (SNP), and single strand conformation polymorphism (SSCP) markers has been developed for *A. aegypti* (Severson et al. 2002). These markers have been used to identify quantitative trait loci
that affect filarial worm susceptibility in *A. aegypti* (Severson et al. 1994), as well as quantitative trait loci influencing filarial worm intensity (Beerntsen et al. 1995).

Severson et al. (1994b), identified two QTL affecting *B. malayi* susceptibility in *A. aegypti*. The first QTL, *fsb1*, located on chromosome 1 had a major effect on phenotypic variation and accounted for 22-43% of that seen in the three populations tested. The second QTL, *fsb2*, found on chromosome 2, accounted for 3-16% of the phenotypic variation. The previously described *fm* locus appears to map within the *fsb1* QTL region.

Severson et al. (1999) used bulk segregant analysis for targeted marker development within the *fsb1* QTL. They identified 5 RAPD clones that exhibited linkage to *fsb1*, including one clone, B8L260 which mapped within the flanking markers delimiting this QTL. This resulted in a narrowing of *fsb1* from 10.4 cM to 2.9 cM.

Although the *fsb1* QTL interval has been reduced, the physical size of this region is still an obstacle in map-based cloning efforts to identify the gene or genes affecting filarial worm susceptibility in *A. aegypti*. Because polytene chromosome physical maps are difficult to construct in Culicines and because recombination levels in *A. aegypti* appear to be low for its estimated gene size (Severson et al. 1999), advanced intercross lines (AIL) are one means to increase the mapping resolution of QTLs (Darvasi and Soller 1995). AILs can provide more accurate estimates of QTL map location than standard mapping populations because multi-generation intercrosses increase the probability of recombination between any two loci and thus serve to statistically stretch the genome and increase mapping resolution (Darvasi and Soller 1995).

Here we describe the nature, development, and use of single nucleotide polymorphism (SNP) markers to re-examine susceptibility to *B. malayi* in an *A. aegypti*
advanced intercross line. We identify a QTL on chromosome 1 associated with *B. malayi* susceptibility when this phenotype is treated as either a quantitative or binary trait using single marker analysis of variance and linear regression.

4.2 Materials and Methods

4.2.1 Advanced intercross lines

An advanced intercross line was developed by crossing a single Liverpool (LIV) female, susceptible to *B. malayi*, with a single Moyo-R male, refractory to *B. malayi*. The Liverpool strain was originally colonized at the Liverpool school of Tropical Medicine and Hygiene in 1937 and a strain susceptible to *B. malayi* was selected for in 1963 (MacDonald 1962a). The Moyo-R strain was selected from the MOYO (Moyo-In-Dry) strain by Thathy et al. (1994). The MOYO strain was originally collected from the Shauri Moyo Village, Mombasa Kenya in 1974. We allowed the progeny of each subsequent generation to randomly and sequentially intercross for eight generations, insuring that the size of the breeding population of each generation was > 500. Mosquitoes were reared in environmental chambers at 26°C and 84% relative humidity with a 16-hr light/8-hr dark cycle and a one-hour crepuscular period at the beginning and end of each light cycle. Adult mosquitoes were maintained on a 5% sucrose solution.

4.2.2 Parasite exposure

*F*₈ female mosquitoes were allowed to engorge on *B. malayi* infected jirds (*Meriones unguiculatus*) that were anesthetized with a mixture of ketamine and Rompun (7:3) administered i.m. at a dosage of 0.1 ml/100 g of body weight. Microfilaremia of
gerbils used in this study ranged from 82-198 microfilariae per 20 ul of blood. Female mosquitoes with engorged abdomens were considered to have blood-fed and were separated from unfed females. Blood-fed female mosquitoes were dissected 14 days after blood feeding to determine whether they were permissive for *B. malayi* larval development. We counted the number of filarial worms that successfully developed to infective third-stage larvae (L3) in each individual.

**4.2.3 DNA isolation**

Following dissections to determine the filarial worm susceptibility phenotype, DNA extractions were performed on each mosquito carcass by homogenization in 200 ul of DNA extraction buffer using a Tissue Tearer (Biospec, Bartlesville, OK), followed by a standard phenol/chloroform extraction and ethanol precipitation as previously described (Severson 1997).

**4.2.4 Single Nucleotide Polymorphism marker development and PCR amplification of SNP markers**

Primers for putative single nucleotide polymorphism markers (SNPs) were designed based upon *A. aegypti* RFLP markers, known genes, and BAC end sequences. Accession numbers for putative markers are found in Table 4.1. PCR primers were designed using PRIMER3 (Rozen and Skaletsky 2000). We sought to design primers that were 18-27 nt in length with a G+C content of 50%, and an optimum anneal temperature of 60°C ± 3°C. When possible we included a GC clamp and selected primers that would yield a 300-800 bp product. A sequence similarity and homology search was conducted
<table>
<thead>
<tr>
<th>Sequence origin</th>
<th>Acc no.</th>
<th>Forward /Reverse Primer sequence 5’-3’</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ttsf</em></td>
<td>AY064541</td>
<td>TAGTGGCAGTGCTTCTGCTGAGGCCACGTACATGTCTTCCCTCCGGAAACTCTCCCGTTGTA</td>
<td>232</td>
</tr>
<tr>
<td><em>APN</em></td>
<td>AF390100</td>
<td>CAAGCAACAAAACCTGGTTATATTATCGGGGCCACGCTGGCCTGAGATTAGGTGGCGTAACTGCGAC</td>
<td>802</td>
</tr>
<tr>
<td>LF198</td>
<td>T58319</td>
<td>CTACACCGCTGGGCTAGATTAGGTGGCGTAACCTGCGAC</td>
<td>233</td>
</tr>
<tr>
<td>LF178</td>
<td>T58309</td>
<td>ATAAGCCTACCCGAAGTCGACCCAGAGCCTCGATAACCT</td>
<td>716</td>
</tr>
<tr>
<td>B8L260</td>
<td>BH214532</td>
<td>CACACGGTATACACATGAAAACACTGGTGCACAATGTAACCT</td>
<td>109</td>
</tr>
<tr>
<td>ND41C1t7</td>
<td>BH817084</td>
<td>TCGAACCTTTTCTTGGCTGGTTTGTCTTTCTTGCCCTCC</td>
<td>433</td>
</tr>
<tr>
<td>ND41C1sp6</td>
<td>BH817083</td>
<td>CGAAAAACGGGACAAGCTTGATTCTACGCAGTCTGTTTCTCGCATTCTACGCAGTCTGCTGAC</td>
<td>413</td>
</tr>
<tr>
<td>27TV</td>
<td>AACKD27TV</td>
<td>TTAGAACGGGAATTATGCCAACCGGGATCCAGGAAATGAAAAT</td>
<td>484</td>
</tr>
<tr>
<td>27TP</td>
<td>AACKD27TP</td>
<td>CATCTGAGCGAGGATCAACAAGAACACGTGACGCTTACC</td>
<td>650</td>
</tr>
<tr>
<td>81TV</td>
<td>AACER81TV</td>
<td>GCTGCACCTTTATGGCTGTTACGAAGAACGTTGGATGTTACC</td>
<td>601</td>
</tr>
<tr>
<td>80TV</td>
<td>AACEN80TV</td>
<td>TTGGCACTCTGTTGAGTATTTACACCATCCAGCAAGACG</td>
<td>688</td>
</tr>
<tr>
<td>Actin*</td>
<td>U20287</td>
<td>CTCTATCTACCTCCAGGCTATGACGCTGACAAGTATCACAA</td>
<td>295</td>
</tr>
</tbody>
</table>

Note: *Primer sequence from Gomez-Machorro et al. 2004.
on all putative PCR primers using the FASTA3 computer program (Pearson 1990). This program was employed as it allows the user to set the KTUP (word size) to 1, allowing for a more sensitive search that facilitates avoidance of primer sequences that may have homology to known repetitive elements.

PCR amplifications of chromosome 1 marker loci were performed on P₀ parents used to construct the F₈ (AIL). Reactions were carried out in 25 ul PCR reaction mixtures containing 1X Taq buffer (10mM KCl, 2mM Tris, pH 9.0, 0.02% TritonX) 3.0 mM MgCl₂, 0.4 mM each dNTP (dATP, dCTP, dGTP, dTTP), 10 pmoles of each primer, 1 unit Taq DNA polymerase and 6 ng genomic DNA. Reactions were carried out in Eppendorf Mastercycler thermal cyclers (Hamburg, Germany) under the following PCR parameters: DNA denaturation 94°C for 10 min, 30 cycles of 94°C for 1 min, 60°C for 1 min, 72°C for 2 min, and a final extension of 72°C for 10 min. PCR products were size fractionated in 2% agarose gels and visualized with ethidium bromide under UV light.

4.2.5 DNA sequencing

Amplified PCR products were spin-column purified using the QIAquick PCR purification kit (Qiagen, Valencia, CA). The purified PCR products were then quantified visually on 2.0% agarose gels. PCR product templates were subjected to direct cycle sequencing using the ABI Prism Big Dye Terminator kit v. 3.1 (Applied Biosystems Inc., Foster City, CA) and the ABI Prism 377 Genetic Analyzer. Single-pass sequence information was obtained from each strand of each PCR product. Sequence data were then subjected to BLASTN and BLASTX searches of the GenBank non-redundant and EST databases using default settings (Altschul et al 1999).
4.2.6 Sequence analysis and identification of SNP specific restriction sites

Sequence data generated from amplified PCR products were aligned using the default settings of ClustalW 1.8 multiple sequence alignment program found on the Baylor College of Medicine (BCM) Search Launcher web site (http://searchlauncher.bcm.tmc.edu/) (Smith et al. 1996) and strain specific SNPs were identified. Restriction sites that were SNP specific were identified by sequence analysis using NEBcutter version 2.0 (http://tools.neb.com/NEBcutter2/index.php) (Vincze et al. 2003). SNP-specific primers were designed for the APN and Tsf loci in order to amplify DNA fragments that could be analyzed by melting-curve differences. New primers were also designed for some loci in order to ensure that restriction digests were strain-specific. Primers used for SNP analysis are found in Table 4.2 and were designed using the same conditions as outlined in section 4.2.4. Restriction digests were performed in 25 ul reaction volumes consisting of 10 ul of PCR product, 4 units of restriction enzyme, 2.5 ul of 10X reaction buffer, 0.25 ul BSA, and 11.85 ul ddH₂O. SNP analysis was performed following the McSNP protocol (Akey et al. 2001). McSNP products were scored using a gel-free system (Hybaid DASH system, Thermo Hybaid, Franklin MA). When this was not possible the restriction digested PCR products were genotyped by size fractionation on 3% agarose gels and visualized with ethidium bromide under UV light.

4.2.7 Linkage mapping

Multi-point linkage analysis was performed using the MAPMAKER/EXP 3.0 computer package (Lander et al. 1987) with the linkage LOD score set at 3.0 with the Kosambi (1944) mapping function.
4.2.8 Statistical analysis

To determine which SNP markers were linked to loci contributing to *B. malayi* susceptibility in the F₈ population that data were first analyzed by single factor analysis of variance. The effect of each marker was examined individually. Each marker is considered a genetic treatment and the three genetics states (maternal homozygotes, heterozygotes, and paternal homozygotes) are “levels” within the treatment. Although these procedures cannot accurately define QTL location, they are an excellent means for identifying significant associations between marker genotypes and phenotypes (Armstrong et al. 1992). The data was then analyzed using stepwise backward elimination multiple regression. We evaluated markers for both linear and quadratic effects. This selected the best set of markers that explained the greatest degree of phenotypic variance. We analyzed phenotypic data as both a quantitative and binary trait. We included analysis as a binary trait because binary trait loci (BTL) are believed to represent one or two loci exhibiting a major effect on phenotype and penetrance is restricted to values of 0 or 1. That is, analysis as a binary trait assumes the phenotype reflects an underlying continuous trait that is evident only when a threshold is exceeded. Previous research seemed to indicate that *B. malayi* susceptibility in *A. aegypti* may meet these criteria (Severson et al. 1994b; Beernsten et al. 1995). We performed ANOVA and stepwise regression using the STATA 8.0 computer package (STATA Corp. 2003).
4.3 Results

4.3.1 Infection of mosquitoes with B. malayi

We measured the response of 504 female F<sub>8</sub> AIL progeny obtained from a single cross between a LIV<sup>SB</sup> female and a MOYO-R male to infection with B. malayi. We allowed 5-7 day-old F<sub>8</sub> females to blood feed on B. malayi–infected gerbils. Following a 14-day incubation period we dissected the mosquitoes and counted the number of infective L3 larvae present in the head and thorax of the female mosquitoes. We observed L3 larvae in 13.1% (66/504) of all female mosquitoes. Female mosquitoes which were infected had an average of 3.91 ± 5.32 L3 larvae per mosquito (with a range of 1-39 L3).

4.3.2 Identification and development of genetic markers

We screened chromosome one loci for informative SNP markers that would provide broad coverage over the entire chromosome length while simultaneously placing special emphasis on identifying markers that would map within the previously identified fsb1 QTL in A. aegypti (Severson et al. 1994b). The Tsf, FerH, APN, and LF198 markers were selected because these loci were outside of the fsb1 QTL region and served as anchor loci. The LF178 and actin loci were selected because they had previously been mapped within the fsb1 QTL interval. We screened additional loci in order to identify SNP markers within the fsb1 QTL. PCR primers for candidate marker loci are found in Table 4.1. PCR primers were based upon known chromosome one A. aegypti genes and RFLP markers. In addition, two short physical contigs representing the LF178 and B8L260 genetic markers were constructed using BAC clones from the NDL BAC library.
(Jimenez and deBruyn unpublished data). The NDL BAC library was prepared from the
*A. aegypti* Liverpool strain and has an average insert size of 122 Kb giving an estimated
7.87–fold genome coverage (Jimenez et al. 2004). These contigs were constructed as part
of a targeted effort to develop new markers within the *fsb1* QTL. BAC clones that
overlapped the two contigs were not identified. We designed PCR primers based upon
BAC end sequence of overlapping clones found within each of the contigs. PCR primers
were initially based upon BAC end sequence data generated in our laboratory at the
University of Notre Dame. In cases where BAC end sequence data generated markers
that were not informative, we performed BLASTN analysis of our NDL sequence data
against the full NDL BAC library sequence data set at The Institute for Genomic
Research (TIGR) ([http://www.tigr.org/msc/aedes/aedes.shtml](http://www.tigr.org/msc/aedes/aedes.shtml)) in order to identify
additional sequence which could be screened for informative SNPs.

The candidate markers were amplified in the P₀ parents of our F₈ AIL and
subjected to direct cycle sequencing in both the forward and reverse direction. We
identified putative SNPs and INDELS following multiple sequence alignment analysis of
the P₀ parents (Table 4.2). Not surprisingly, transitions were more common than
transversions by a ratio of approximately 2:1.

Loci that exhibited strain-specific SNP, were then subjected to analysis with
NEBcutter 2.0 (Vincze et al. 2002) to identify SNPs that were restriction enzyme
specific. We identified eight markers that were informative and gave broad coverage
over chromosome one. PCR primers and restriction enzymes used for SNP analysis are
found in Table 4.3.
We examined genotypic ratios at each marker locus for all individuals and tested for Hardy-Weinberg ratios for all eight loci using Genepop 3.4. We estimated Wrights inbreeding coefficient (F_{IS}) (Weir and Cockerham 1984) at each locus and performed a χ² goodness of fit test with 1 d.f. The F_{IS} value allows us to determine if loci are in Hardy-Weinberg proportions. If F_{IS} ~ 0 then genotypes are in Hardy-Weinberg proportions, if F_{IS} < 0 then there is an excess of heterozygotes and if F_{IS} > 0 then there is an excess of homozygotes.

Four of the eight markers on chromosome one diverged from Hardy-Weinberg proportions (Table 4.4). All four exhibited F < 0 denoting an excess of heterozygotes. Three of these markers, LF198, Actin, and LF178 were significant at the P < 0.001 level while APN was significant at the P < 0.05 level. Two other markers, 80TV and 81TV also exhibited a slight excess of heterozygotes but neither marker reached the P < 0.05 level of significance. These results were not surprising in that the loci that exhibited the greatest deviation from Hardy-Weinberg proportions were those most tightly linked to the sex-locus. Thus the excess of heterozygotes and maternal homozygotes was not unexpected.
**TABLE 4.2**

SNP POLYMORPHISMS AND INDELS IDENTIFIED FROM *Aedes aegypti*

CHROMOSOME ONE cDNAs, KNOWN GENES, BAC END SEQUENCES, AND WHOLE GENOME SEQUENCE

<table>
<thead>
<tr>
<th>Marker ID</th>
<th>Product</th>
<th>Size</th>
<th>#SNPs</th>
<th>Indels</th>
<th># Transitions</th>
<th># Transversions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tsf</td>
<td>232</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>APN</td>
<td>722</td>
<td>22</td>
<td>0</td>
<td>14</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>LF198</td>
<td>270</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>LF178</td>
<td>716</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>B8L260</td>
<td>109</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ND41C1T7</td>
<td>433</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ND41C1SP6</td>
<td>413</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>AACKD27TV</td>
<td>484</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>AACKD27TP</td>
<td>650</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>AACER81TV</td>
<td>601</td>
<td>33</td>
<td>6</td>
<td>20</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>AACEN80TV</td>
<td>688</td>
<td>8</td>
<td>0</td>
<td>7</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Actin</td>
<td>295</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Marker</td>
<td>SNP restriction site</td>
<td>Restriction enzyme</td>
<td>Strain restricted</td>
<td>Forward/Reverse Primer sequence 5’-3’</td>
<td>Fragment sizes (bp) uncut, cut</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----------------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>----------------------------------------</td>
<td>---------------------------------</td>
<td></td>
</tr>
<tr>
<td>Tsf</td>
<td>C’CATGG</td>
<td>NcoI</td>
<td>LIV</td>
<td>AGTGCTTCTTGCTGATCGGGTC CGGAACGCACAATTTGAAC</td>
<td>72, 39, 33</td>
<td></td>
</tr>
<tr>
<td>LF198</td>
<td>GTT’AAC</td>
<td>HpaI</td>
<td>LIV</td>
<td>CTACACCGCTGGCGTAGATT AGGTGGCGTAACTTCGAC</td>
<td>270, 210, 60</td>
<td></td>
</tr>
<tr>
<td>Actin</td>
<td>23 bp indel</td>
<td>NA</td>
<td>NA</td>
<td>CTCTATCTACCTTCCAGGCTAT GACGTGACAAGTATCACAA</td>
<td>295/272</td>
<td></td>
</tr>
<tr>
<td>LF178</td>
<td>GT’AC</td>
<td>Rsal</td>
<td>MoyoR</td>
<td>CAACGGTTCTCCTGG TGACCGTTTGTATTGCTAGC GCTGCACCTATTGGTGCGTT ACGAAGAACGGATGGCTAC</td>
<td>131, 88, 43</td>
<td></td>
</tr>
<tr>
<td>81TV</td>
<td>GG’CC</td>
<td>HaeIII</td>
<td>MoyoR</td>
<td>GCTGCACCTATTGGTGCGTT ACGAAGAACGGATGGCTAC AC</td>
<td>601, 425, 176</td>
<td></td>
</tr>
<tr>
<td>80TV</td>
<td>ATGCA’T</td>
<td>NsiI</td>
<td>MoyoR</td>
<td>AATGACGGCACTGCAAATG CCATATTATGCACCCGACA</td>
<td>209, 157, 52</td>
<td></td>
</tr>
<tr>
<td>FerH</td>
<td>GG’CC</td>
<td>HaeIII</td>
<td>MoyoR</td>
<td>AAACCGCCCTACAGAAGGAG GTCATTGCTGGCATTTCG</td>
<td>80, 57, 23</td>
<td></td>
</tr>
<tr>
<td>APN</td>
<td>C’CGG</td>
<td>HpaII</td>
<td>LIV</td>
<td>GACATCGATCCAAACCTGAAG GTCATGCGATCCAATA</td>
<td>93, 57, 36</td>
<td></td>
</tr>
</tbody>
</table>
## TABLE 4.4

SEgregation of *Aedes aegypti* Chromosome 1 Markers and F Values

<table>
<thead>
<tr>
<th>Marker</th>
<th>S ( ^a )</th>
<th>H</th>
<th>R</th>
<th>F(_{is} )</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Tsf</em></td>
<td>123</td>
<td>252</td>
<td>113</td>
<td>-0.0322</td>
</tr>
<tr>
<td>LF198</td>
<td>126</td>
<td>294</td>
<td>79</td>
<td>-0.1879(^b)</td>
</tr>
<tr>
<td><em>Actin</em></td>
<td>93</td>
<td>260</td>
<td>60</td>
<td>-0.2660(^b)</td>
</tr>
<tr>
<td>LF178</td>
<td>107</td>
<td>289</td>
<td>97</td>
<td>-0.1719(^b)</td>
</tr>
<tr>
<td>80TV</td>
<td>146</td>
<td>222</td>
<td>57</td>
<td>-0.0915</td>
</tr>
<tr>
<td>81TV</td>
<td>160</td>
<td>238</td>
<td>63</td>
<td>-0.0793</td>
</tr>
<tr>
<td><em>FerH</em></td>
<td>133</td>
<td>237</td>
<td>121</td>
<td>0.0351</td>
</tr>
<tr>
<td><em>APN</em></td>
<td>135</td>
<td>269</td>
<td>89</td>
<td>-0.0999(^c)</td>
</tr>
</tbody>
</table>

Note: \(^a\) Genotypic designations: S = filarial worm susceptible parental type; H = heterozygote; R = filarial worm refractory parental type.

\(^b\) P < 0.001. Deviations from expected Hardy-Weinberg ratios were assessed by \( \chi^2 \) analysis.

\(^c\) P < 0.05. Deviations from expected Hardy-Weinberg ratios were assessed by \( \chi^2 \) analysis.
4.3.3 QTL determining susceptibility to B. malayi

Two sets of phenotypic data were analyzed: (1) number of L3 as a quantitative trait; (2) presence of L3 as a binary trait. The data were analyzed using single factor analysis of variance (Table 4.5 and Table 4.6) and multiple regression using stepwise backward elimination (Table 4.7).

<table>
<thead>
<tr>
<th>Marker</th>
<th>F-score</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tsf</td>
<td>0.05</td>
<td>0.9475</td>
</tr>
<tr>
<td>LF198</td>
<td>1.35</td>
<td>0.2596</td>
</tr>
<tr>
<td>Actin</td>
<td>0.44</td>
<td>0.6456</td>
</tr>
<tr>
<td>LF178</td>
<td>12.07</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>80TV</td>
<td>0.65</td>
<td>0.5232</td>
</tr>
<tr>
<td>81TV</td>
<td>0.34</td>
<td>0.7109</td>
</tr>
<tr>
<td>FerH</td>
<td>0.56</td>
<td>0.5691</td>
</tr>
<tr>
<td>APN</td>
<td>2.52</td>
<td>0.0816</td>
</tr>
</tbody>
</table>

TABLE 4.5
RESULTS OF SINGLE FACTOR ANALYSIS OF VARIANCE ON SNP DATA FOR 504 INDIVIDUALS FROM THE LIV X MOYO-R F8 POPULATION WHEN Filarial worm susceptibility is considered a quantitative trait.
### TABLE 4.6

RESULTS OF SINGLE FACTOR ANALYSIS OF VARIANCE ON SNP DATA

FOR 504 INDIVIDUALS FROM THE LIV X MOYO-R F₈ POPULATION

WHEN FILARIAL WORM SUSCEPTIBILITY IS CONSIDERED A BINARY TRAIT

<table>
<thead>
<tr>
<th>Marker</th>
<th>F-score</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tsf</td>
<td>0.95</td>
<td>0.3891</td>
</tr>
<tr>
<td>LF198</td>
<td>1.37</td>
<td>0.2560</td>
</tr>
<tr>
<td>Actin</td>
<td>0.20</td>
<td>0.8182</td>
</tr>
<tr>
<td>LF178</td>
<td>23.52</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>80TV</td>
<td>4.26</td>
<td>0.0148</td>
</tr>
<tr>
<td>81TV</td>
<td>0.88</td>
<td>0.4152</td>
</tr>
<tr>
<td>FerH</td>
<td>0.21</td>
<td>0.8103</td>
</tr>
<tr>
<td>APN</td>
<td>0.35</td>
<td>0.7046</td>
</tr>
</tbody>
</table>
TABLE 4.7
REGRESSION STATISTICS FOR SNP MARKERS DETERMINED BY STEP-WISE BACKWARD ELIMINATION ANALYSIS FOR PHENOTYPIC DATA TREATED AS A QUANTITATIVE TRAIT AND AS A BINARY TRAIT IN AN *Aedes aegypti* F\textsubscript{8} POPULATION.

<table>
<thead>
<tr>
<th>SNP marker</th>
<th>Effect</th>
<th>Partial regression coefficient</th>
<th>Standard error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Filarial worm susceptibility as a quantitative trait (multiple $R^2 = 0.06$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LF178</td>
<td>Quadratic</td>
<td>0.456</td>
<td>0.099</td>
<td>4.58*</td>
</tr>
<tr>
<td>B. Filarial worm susceptibility as a binary trait (multiple $R^2 = 0.14$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LF198</td>
<td>Linear</td>
<td>0.059</td>
<td>0.030</td>
<td>1.98**</td>
</tr>
<tr>
<td>LF178</td>
<td>Quadratic</td>
<td>0.0.084</td>
<td>0.012</td>
<td>6.53*</td>
</tr>
<tr>
<td>80TV</td>
<td>Quadratic</td>
<td>0.022</td>
<td>0.012</td>
<td>1.81**</td>
</tr>
</tbody>
</table>

Note: *The multiple $R^2$ statistic is an estimate of the proportion of the phenotypic variance jointly explained by the select marker set.*  
*P < 0.001; **P < 0.10

Single factor analysis of variance identified one SNP marker, LF178, that exhibited a significant association with filarial worm susceptibility ($P < 0.0001$) when the susceptibility phenotype was treated as a quantitative trait (Table 4.5).

When the susceptibility phenotype was treated as a binary trait single factor analysis of variance identified two SNP markers, LF178 and 80TV that exhibited a significant association with filarial worm susceptibility.
The SNP marker LF178 explained 6% of the observed phenotypic variance when susceptibility was treated as a quantitative trait (Table 4.7). LF178 was the only marker selected by regression analysis and was shown to have a significant quadratic effect.

A set of three SNP markers (LF178, LF198, 80TV) were selected by stepwise regression with backward elimination. Together, these markers explained 14% of the observed phenotypic variance when susceptibility to *B. malayi* was treated as a binary trait. Only LF178 was seen as significant at the $P < 0.05$ level.

We evaluated infection rates among $F_8$ mosquitoes when infection with L3 worms was treated as both a quantitative and binary trait. Figure 4.1 shows the average number of infective L3 per mosquito as a function of the number of alleles inherited from the LIV parent at the LF178 locus. The average number of infective L3 at the LF178 locus was $1.46 \pm 4.41$ when both alleles are inherited from the LIV parent, $0.33 \pm 1.34$ when one allele was inherited from each of the parents and $0.01 \pm 0.10$ when both alleles are inherited from the Moyo-R parent.

Figure 4.2 shows the percent of mosquitoes infected with the L3 stage of *B. malayi* as a function of the number of alleles inherited from the LIV parent at the LF178 locus. Infection rates for mosquitoes at the LF178 locus were 31% when both alleles were inherited from the LIV parent, 10.4% when one allele was inherited from each parent and 1% when both alleles were inherited from the Moyo-R parent.
**FIGURE 4.1** Plot of average number of filarial worms per mosquito as a function of the number of LIV alleles present

**FIGURE 4.2** Plot of filarial worm infection rate among $F_8 A. aegypti$ with 0, 1, or 2 LIV alleles
4.4 Discussion

The purpose of this study was to re-examine the previously defined fsb1 QTL (Severson et al. 1994b) and to identify any other potential QTL found on chromosome one as part of an effort to perform fine scale mapping and minimize the genomic regions that need to be screened for candidate loci. We successfully re-confirmed a QTL influencing B. malayi susceptibility in A. aegypti associated with the genetic marker LF178, that was also found to be associated with the fsb1 QTL in previous studies (Severson et al. 1994b).

We genotyped 8 SNP genetic markers in our F₈ AIL, and then performed tentative linkage examinations between markers using Mapmaker 3.0. We note that while these calculations are inappropriate for the F₈, they can provide information on the level of recombination and linear order among markers. Linkage associations were only observed between Tsf and LF198 and between 80TV and 81TV. This was not surprising as advanced intercross lines will greatly increase the recombinational distance between linked markers causing them to approach recombinational values of 0.5 in an asymptotic fashion (Darvasi and Soller 1995). This relationship between increased generation time and increased recombination can be represented by the equation \( r_t = rt/2 \), where \( r_t \), the proportion of recombinants in the F₄ generation should equal \( r \), the proportion of recombinants in the F₂ generation times \( t \), the number of generations in the AIL, divided by 2 (Darvasi and Soller 1995). Furthermore, Meece (2002), demonstrated that recombination between linked markers in an A. aegypti AIL can even be higher than theoretical predictions. Several markers in her advanced intercross line exhibited recombinational values many times that of the standard F₂, including markers less than 1
cM apart in a standard F\textsubscript{1} intercross. Widely varying rates of recombination could be due to the presence of recombinational hot spots. These hot spots are regions of the chromosome where homologous recombination occurs at a higher than average frequency (Lichten and Goldman 1995).

In many studies the lack of calculated linkage between markers in an AIL population is overcome by analyzing the phenotypic and genotypic data of the AIL while using the mapping data from a standard F\textsubscript{1} intercross. Here we were unable to do this as we did not know the map positions of either 80TV or 81TV, as they have never been mapped in an F\textsubscript{1} intercross. Neither marker demonstrated linkage to LF178 or \textit{Actin}. Markers 80TV and 81TV nonetheless represented key markers in our study in that they were isolated from different BAC clones of the NDL BAC library that made up a short physical contig containing the LF178 marker. Based on the estimated size of the average BAC clone in this library (122 Kb) we would expect these markers to be less than 1 Mb from LF178. Again, as noted above, recombination rates between even tightly linked loci in AIL populations can be extremely high and thus it was not entirely surprising that neither 80TV or 81TV showed linkage with LF178.

Because of the limited marker numbers and uncertain recombination rates, we performed QTL analysis using single marker analysis of variance and linear regression as opposed to more commonly used methods such as interval mapping (IM) or composite interval mapping (CIM). This approach has a number of advantages in that it is computationally simple, does not require a genetic map and it can be extended to include multiple regression models to account for multiple markers if need be. The disadvantages are that it may not provide accurate information regarding exact QTL
location, and it loses power when there are insufficient markers available. Here, these disadvantages were negligible, because we did have information regarding the map location of the majority of our markers and we felt that the increased recombinational power of an AIL coupled with what we felt was sufficient marker coverage on chromosome one would provide us adequate power needed for QTL detection.

We were unable to identify an informative SNP marker at the B8L260 locus, a locus previously shown to map within the \textit{fsb1} QTL, nor were we able to develop informative markers from any of the BAC clones found within the B8L260 physical contig. We were, however, successful in developing an informative marker based upon a 20 bp INDEL at the \textit{Actin} locus that is found within the original \textit{fsb1} QTL interval. It is interesting to note that this marker was not associated with the filarial worm susceptibility phenotype whereas the LF178 marker, which is tightly linked to \textit{Actin} is significantly associated with this phenotype. This seems to indicate that the \textit{fsb1} QTL loci, believed to be, perhaps, a single gene modified by additional loci on chromosome 2 (Beernsten et al. 1995; MacDonald 1963; Severson et al. 1994b) is tightly linked with LF178. With the forthcoming release of the \textit{A. aegypti} sequence assembly we should be in an excellent position to screen the contigs and supercontigs near LF178 for candidate genes that may play a role in \textit{B. malayi} susceptibility. These data should facilitate the development of additional genetic markers that will allow us to reduce the genome interval containing the \textit{fsb1} QTL and reduce the number of candidate genes to be evaluated.

Knowledge and understanding of the genetic basis for filarial worm susceptibility in \textit{A. aegypti} should prove extremely useful when applied to other important natural
vectors that are less tractable for laboratory studies. As the practicality for controlling lymphatic filariasis disease transmission using mass drug treatment administrations of anti-filarial drugs remains uncertain, an understanding of mosquito-filarial worm interactions may prove critical in developing future integrated control programs for this disease.
CHAPTER 5

SUMMARY

There were three primary goals of this dissertation: (1) Identify and develop PCR-based genetic markers that could be used for comparative mapping within the Culicidae, (2) Isolate and develop microsatellite markers for *A. aegypti* from enriched libraries and perform linkage mapping in an F$_1$ intercross, and (3) identify PCR-based genetic markers from chromosome one of *A. aegypti* and employ these markers to identify quantitative trait loci for *B. malayi* susceptibility in an advanced intercross line (AIL) of *A. aegypti*.

In Chapter 2, we developed 23 comparative anchor-tagged loci (CATS) based upon conserved chromosomal regions within the Culicidae. We developed PCR primers based upon conserved regions of orthologous exons from *A. aegypti*, *A. gambiae* and *D. melanogaster*. We screened these primers in 8 different dipterans and were successful in amplifying single-copy PCR products. These 23 CATS primer pairs gave broad genome coverage in *A. aegypti* and demonstrated a potential method for developing comparative anchor marker loci for a variety of species within the Culicidae, in particular genetically uncharacterized mosquito species.

In Chapter 3 we isolated and developed microsatellite markers from three libraries enriched for (GA)$_n$, (AAT)$_n$ and (TAGA)$_n$ motifs from genomic DNA isolated from *A. aegypti*. We identified seven novel microsatellites from these enriched libraries. Meece (2002), had previously identified 5 microsatellite markers through the direct screening of
selected clones from three different A. aegypti derived libraries. Together our lab has identified 12 new microsatellites that segregate as single locus, co-dominant markers, and determined the genetic map positions for six of these markers. In addition we described the close association of microsatellites in *A. aegypti* with known repetitive elements, providing insight into the limited success workers have had in developing extensive panels of microsatellite marker loci (Fagerberger et al 2001).

Finally, in Chapter 4, we used an advanced intercross line to identify QTL for *B. malayi* susceptibility. We developed PCR-based single nucleotide polymorphism (SNP) markers based on existing chromosome one markers as well as SNP markers based upon BAC clones believed to be associated with physical contigs based upon marker loci flanking a known QTL for *B. malayi* susceptibility (Severson et al. 1994b). We demonstrated that genetic recombination was substantially increased within our AIL. We identified QTL for filarial worm susceptibility based on single marker analysis of variance and linear regression. We identified a QTL significantly associated with LF178 when susceptibility was treated as both a quantitative trait and binary. This marker, LF178 was found to have a significant quadratic effect.
APPENDIX A

CHAPTER 3 MAPPING DATA FOR LIV x MOYO-R F₁ INTERCROSS

CHROMOSOME ONE BEST ORDER OF MARKERS

Best 20 orders:
1:  6 2 10 1 5  Like:  0.00
2:  6 2 10 5 1  Like:  0.00
3:  6 2 5 10 1  Like: -0.13
4:  6 2 5 10 1  Like: -0.13
5:  6 2 10 5 1  Like: -0.78
6:  6 2 1 5 10  Like: -0.78
7:  2 6 10 1 5  Like: -1.41
8:  2 6 1 10 5  Like: -1.41
9:  2 6 5 10 1  Like: -1.54
10: 2 6 5 10 1  Like: -1.54
11: 2 6 10 5 1  Like: -2.20
12: 2 6 1 5 10  Like: -2.20
13: 6 5 10 1 2  Like: -21.24
14: 6 5 1 10 2  Like: -21.24
15: 6 10 1 5 2  Like: -21.26
16: 6 1 10 5 2  Like: -21.26
17: 6 10 5 1 2  Like: -21.90
18: 6 1 5 10 2  Like: -21.90
19: 5 6 2 10 1  Like: -31.14
20: 5 6 2 1 10  Like: -31.14

Order 1 is set

=======================================================================
Map:                        Apriori
Markers          Distance   Prob  Candidate Errors
6  LF230         6.0 cM     -
2  LF198        23.1 cM   1.0%  [#17 H-B-H 1.20]  [#50 B-H-B 1.15]
10 B81260        0.0 cM   1.0%  -
1  LF178        1.5 cM   1.0%  -
5  LAP          -------   -

30.7 cM 5 markers  log-likelihood= -105.2
CHROMOSOME TWO BEST ORDER OF MARKERS

26> try 19

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td>-----</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td>------</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>------</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td>------</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td>------</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td>------</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>------</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td>------</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>------</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td>------</td>
</tr>
<tr>
<td>INF</td>
<td></td>
<td></td>
<td></td>
<td>------</td>
</tr>
</tbody>
</table>

BEST -187.42

<table>
<thead>
<tr>
<th>Markers</th>
<th>Distance</th>
<th>Prob</th>
<th>Candidate Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>F17m590</td>
<td>1.8 cM</td>
<td>1.0%</td>
<td>[#50 A-H-A 1.87]</td>
</tr>
<tr>
<td>ARC1</td>
<td>16.3 cM</td>
<td>1.0%</td>
<td>[#88 A-B-H 1.56] [#98 H-B-H 1.07]</td>
</tr>
<tr>
<td>B8m980</td>
<td>11.3 cM</td>
<td>1.0%</td>
<td>[#3 A-B-A 4.23]  [#2 H-B-- 2.01]</td>
</tr>
<tr>
<td>B07</td>
<td>4.0 cM</td>
<td>1.0%</td>
<td>[#23 A-B-A 5.00] [#22 H-B-A 2.62]</td>
</tr>
</tbody>
</table>
| H08    | 4.9 cM   | 1.0%  | [
| LF282  | 1.5 cM   | 1.0%  | - |
| A131975| 0.0 cM   | 1.0%  | - |
| M205   | 2.4 cM   | 1.0%  | [#49 H-A-H 4.33] [#10 B-H-B 4.32] |
| LF98   | 7.3 cM   | 1.0%  | [#88 A-H-A 2.11] [#94 --B-H 2.11] |
| LF338  |          |       |                  |

49.4 cM 10 markers log-likelihood= -187.42
CHROMOSOME THREE BEST ORDER OF MARKERS

Best 20 orders:
1: 17 20 22 16 18 15 Like: 0.00
2: 20 17 22 16 18 15 Like: -1.03
3: 17 20 15 18 16 22 Like: -1.50
4: 20 17 15 18 16 22 Like: -2.12
5: 17 20 16 22 18 15 Like: -2.30
6: 17 20 15 18 22 16 Like: -2.69
7: 15 18 16 20 17 22 Like: -2.97
8: 17 20 16 18 15 22 Like: -3.03
9: 17 20 22 15 18 16 Like: -3.20
10: 17 20 22 16 15 18 Like: -3.22
11: 20 17 15 18 22 16 Like: -3.26
12: 20 17 16 22 18 15 Like: -3.42
13: 15 18 16 17 20 22 Like: -3.62
14: 16 18 15 20 17 22 Like: -3.65
15: 16 18 15 17 20 22 Like: -3.81
16: 20 17 22 15 18 16 Like: -3.97
17: 20 17 16 18 15 22 Like: -4.19
18: 20 17 22 16 15 18 Like: -4.23
19: 17 20 16 22 15 18 Like: -4.53
20: 17 20 22 18 15 16 Like: -4.54

Order 1 is set

Map: Apriori
Markers | Distance | Prob | Candidate Errors
--|--|--|---|
15 A10  | 13.6 cM  |   |   |
18 B19  | 12.4 cM  | 1.0% | #98 B-A-H 1.66 | #89 H-B-H 1.11 |
16 Apolipo | 10.7 cM | 1.0% | #33 H-A-H 1.21 | #81 H-A-H 1.21 |
22 LF108 | 14.9 cM  | 1.0% | #23 H-B-H 1.13 | #97 B-H-B 1.10 |
20 m313 | 5.8 cM   | 1.0% | #39 H-B-H 1.40 | #36 H-B-H 1.40 |
17 Cyp9  | -------- |   |   |
57.3 cM | 6 markers | log-likelihood= -111.86

=======================================================================
Map:                        Apriori
Markers          Distance   Prob  Candidate Errors
15  A10          13.6 cM     |   |
18  B19          12.4 cM     | 1.0% | #98 B-A-H 1.66 | #89 H-B-H 1.11 |
16  Apolipo      10.7 cM     | 1.0% | #33 H-A-H 1.21 | #81 H-A-H 1.21 |
22  LF108        14.9 cM     | 1.0% | #23 H-B-H 1.13 | #97 B-H-B 1.10 |
20  m313         5.8 cM      | 1.0% | #39 H-B-H 1.40 | #36 H-B-H 1.40 |
17  Cyp9         --------    |   |
57.3 cM         6 markers    | log-likelihood= -111.86

=======================================================================

99
REFERENCES


World Health Organization (WHO), Dengue and dengue haemorrhagic fever fact sheet No. 117 (Revised April 2002)


