UNIVERSITY OF CALIFORNIA
RIVERSIDE

Evolutionary and Population Genetics of Mosquito Disease Vectors

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

Doctor of Philosophy

in

Genetics, Genomics, and Bioinformatics

by

Eric Andrew Smith

June 2017

Dissertation Committee:
- Dr. Bradley White, Chairperson
- Dr. Omar Akbari
- Dr. Adler Dillman
- Dr. Jason Stajich

The Dissertation of Eric Andrew Smith is approved:

_____

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGMENTS

I have many people to thank for helping me along the way. First, I'd like to thank my advisor, Dr. Bradley White, for the opportunity to work in his lab and for advice and direction along the way. He has allowed me to grow tremendously as a researcher and independent thinker throughout my time at UCR. I would also like to thank the other members of my dissertation committee, Drs. Omar Akbari, Adler Dillman, and Jason Stajich for research and career advice along the way. They have been instrumental in my progress to this point. My appreciation also goes out to former members of Dr. White's lab, Dr. David Turrissini and Stephanie Gamez. Without David's help and advice, I wouldn't be nearly the bioinformatician I am today. Stephanie put in many, many hours preparing sequencing libraries for the projects in chapters 1 and 2. These projects never would have come to fruition without her hard work and dedication. Thank you to my undergraduate mentor, Dr. Dalma Martinovic-Weigelt for bringing me on in her lab and providing me with the necessary foundation to be successful in graduate school. Thank you also to Dr. Nancy Hartung, who was an amazing genetics and evolution professor that pushed me to pursue my interests in those fields.

I would like to thank many friends who made my time in graduate school and Riverside enjoyable. Tim Ngo was a tremendous source of support during some very difficult times. I've been an honorary member of the entomology graduate student cohort that came in to UCR at the same time as me for the last few years. Many good times were

had with all of you. Special thanks go to my partner, Amelia Lindsey, who has provided an endless supply of love and support, and who has kept me on track when I've started meandering. Amelia has also provided valuable and honest scientific feedback. Thanks go to my family, as well. My parents have been nothing but supportive of every decision I've made, regardless of what they thought of it. Without their support, I would not have found my way to where I am today, nor grown into the person I am today. Their recognition of the value of education, both institutional and otherwise, has been critical in my development as a human being. My sister, through all of her ups and downs, has been a great source of support during difficult times for me. My brother has provided good times whenever I travel back to Minnesota to see family and makes life much more enjoyable. There are many more people with whom I have connected over the years that also deserve thanks here.

Thanks go out to my collaborators. Chapter 1 collaborators include Drs. Roberto Barrera and Ryan Hemme of the Centers for Disease Control and Prevention, and my advisor, Dr. Bradley White. Chapter 2 was a collaboration between myself, Dr. David Turrissini – currently at Acestry.com, Inc. – and Dr. Bradley White. Dr. White was also a collaborator on chapter 3.

The work in this dissertation and my time at UCR was supported by the National Institutes of Health; the Centers for Disease Control and Prevention; the University of

California, Riverside; and through a Chancellor's Distinguished Fellowship. Without the financial support provided by these institutions, none of this work would have been possible.

ABSTRACT OF THE DISSERTATION

Evolutionary and Population Genetics of Mosquito Disease Vectors

by

Eric Andrew Smith

Doctor of Philosophy, Genetics, Genomics, and Bioinformatics
University of California, Riverside, June 2017
Dr. Bradley White, Chairperson

Mosquitoes are widely regarded as the deadliest animals on the planet because of the many diseases they vector. Malaria alone killed 429,000 people in 2015. To date, the most effective method at reducing the burden imposed by mosquito-borne illnesses is vector control. Traditional vector control strategies have largely relied on the use of chemical insecticides to reduce vector populations and ease disease burden. The most effective current methods continue to use insecticides, but they are being incorporated in programs known as integrated vector management that aim to incorporate multiple vector control strategies and closely monitor success and failures. Recent advances in genetic technology are now making it possible to modify and control vector populations via the introduction of genetic elements that will spread through the target population. While there has been great success in using insecticides and the use of genetic elements shows great promise, there is still much we do not yet know about vector evolution and population genomics. The following research projects fill in some of the gaps in our knowledge. I have modeled effective population size changes *Aedes aegypti* in response

to the deployment of novel traps in Puerto Rico, generated fine-scale recombination rate maps for *Anopheles gambiae,* and reconstructed a genomic region of high divergence between *An. gambiae* and *An. coluzzii* to gain insight into the evolution of genes important in the modulation of the mosquito's immune response to malaria infection. My results indicate that despite success in *An. gambiae*, approximate Bayesian computation may not be an appropriate method for estimating effective population size in *Ae. aegypti*; I show that recombination rate in *An. gambiae* varies dependent on sex and the presence of chromosomal inversions; and I describe a previously unannotated gene in *An. gambiae* that may play an important role in mosquito immunity, as well as show that it arose as a chimera of two neighboring genes, much like another gene in this region. Taken together, these projects fill important holes in our knowledge of mosquito vector evolution and population genetics and will help inform future vector control strategies.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

INTRODUCTION

The devastating diseases that mosquitoes vector have led to them being widely considered as the deadliest animal on the planet. The most deadly of these diseases, malaria, was responsible for 212 million cases and 429,000 deaths in 2015 (WHO, 2016). Dengue, deadliest of the many diseases vectored by the mosquitoes in the genus *Aedes*, accounted for an additional 390 million cases and 20,000-25,000 deaths (Bhatt et al. 2013) In addition to malaria and dengue, mosquitoes vector pathogens such as dengue virus, Chikungunya virus, Zika virus, yellow fever virus, West Nile virus, and a number of encephalitis viruses, among others. In total, the World Health Organization estimates that at least half of the world's population is at risk for contracting a mosquito-borne disease every year (WHO, 2017b). Despite progress in the development of vaccines to prevent these diseases, the most effective approach to curtail mosquito-borne illnesses is vector control (WHO, 2017a). Major advances in vector control strategies have occurred over the last few decades; however, there is still a long way to go to fully eradicate mosquito-borne diseases.

Historical vector control

In what has widely been considered the most successful vector control program in history, Fred L. Soper and his team fully eradicated the malaria-carrying mosquito *Anopheles arabiensis* within a decade of its introduction in Brazil in the 1930s (Killeen 2003). This was done primarily by using the larvicide Paris green and several adulticides

to first confine the range of the mosquito, and then eradicate it from that range (Griffing et al. 2015; Soper and Wilson 1943).

The use of chemicals for mosquito control has been standard practice since the early nineteenth century and has remained a critical vector control strategy since (Breman 2001). In the second half of the nineteenth century and the first half of the twentieth century, numerous compounds were discovered to have an insecticidal effect, including, but not limited to: mercuric chloride, Paris green, phenols and cresols, naphthalene, calcium arsenate, and nicotine sulfate (Raghavendra and Subbarao 2002). Starting in the late 1930s, synthetic compounds – such as dichlorodiphenyltrichloroethane (DDT), organophosphates, carbamates, and pyrethroids – began to replace natural insecticides (Aktar et al. 2009; Lerner and Lerner 2002). Historically and currently, insecticides have been applied in four ways: indoor residual spraying (IRS), fogging, insecticide-treated bed nets (ITNs), and long-lasting insecticidal nets (LLINs).

IRS is the process of spraying indoor surfaces (generally walls and roof) with an environmentally persistent insecticide that mosquitoes will then come in contact with when they rest indoors. The use of IRS in the control of *Anopheles* mosquitoes that transmit malaria has been credited with eradication of the disease from several locales (Wakabi 2007; Mabaso et al. 2004; Shiff 2002; Trigg and Kondrachine 1998).

Fogging – also referred to as space spraying – consists of rapid heating of a chemical insecticide to form a very fine mist that is then sprayed in areas where adult mosquitoes are likely to be found. This technique is less commonly used and is often reserved only for use during emergency situations, such as epidemics. However, fogging is a popular vector control strategy in Korea (Lee et al. 2009).

The use of insecticide treated bed nets has been on the rise over the last few decades. As the name implies, this control method uses bed nets that have been treated with insecticides to kill mosquitoes as they try to feed on sleeping people. The first generation of bed nets used pyrethroid-based insecticides. Reports indicate that they have been highly effective at reducing populations of several species of *Anopheline* mosquitoes (Batra et al. 2005; Sampath et al. 1998; Jana-Kara et al. 1995; Sharma and Yadav 1995). However, the use of ITNs has suffered from the fact that the insecticides often wear off quickly and are not re-applied (Lines 1996). In an effort to combat the relatively short lifespan of ITNs, LLINs have recently been introduced. The insecticidal properties of these nets generally last at least 3 years, about the same amount of time that the net itself remains useful (Pulkki-Brännström et al. 2012). Like their early predecessors, LLINs have been shown to reduce the incidence of mosquitoes and disease in a number of localities (Sharma et al. 2009; Sreehari et al. 2009; Stich et al. 1994; Beach et al. 1993).

In conjunction with adult control programs, control of mosquito larvae has largely been accomplished with the use of chemical larvicides, but also involves reduction of suitable larval habitats by removing or covering sources of standing water (reviewed in Fillinger and Lindsay 2011). In addition to chemical control strategies, the use of ovitraps has gained favor in recent decades and is a common component of control programs targeting *Aedes* mosquitoes (cf. Long et al. 2015; Barrera et al. 2014; 2013a; 2013b; Norzahira et al. 2011; Reiter and Colon 1991; Cheng et al. 1982).

Current and future vector control interventions

Current vector control methods continue to employ the strategies mentioned above, but have also developed into programs that incorporate multiple control measures in a coordinated effort known as integrated vector management (IVM), (WHO, 2017c). These methods aim utilize multiple methods of vector control (i.e. adulticiding and larval source reduction), in collaboration with local communities and health departments to provide effective vector control with minimal environmental impact. There are often multiple lines of research occurring in conjunction with control measures to inform decision-making.

Multiple methods are currently in development/testing that will add to the vector control repertoire, often relying on genetic manipulation of vector populations. One of the more prominent of these technologies is sterile insect technique (SIT), whereby

genetically modified, sterile males are released into a population that will then mate with females but produce inviable offspring, resulting a population crash (Catteruccia et al. 2005). Oxitec has shown modest successes releasing sterile *Aedes aegypti* males for population control (Gorman et al. 2016; Carvalho et al. 2015). A more favored approach for population suppression and replacement includes the introduction of a symbiotic bacterium, *Wolbachia*, that affects mating compatibility between infected and uninfected individuals, eventually driving itself through the population (Turelli and Hoffmann 1995). This bacterium also has the ability to block pathogen infection and replication in the host mosquito, rendering it incapable of vectoring disease (Bian et al. 2010; Moreira et al. 2009; McMeniman et al. 2009; Kambris et al. 2009). Releases of *Wolbachia*-infected mosquitoes have been carried out successfully in a few different locations and early results indicate successful population replacement (Schmidt et al. 2017; Turelli and Barton, 2017; Eliminate Dengue, 2016; 2015).

Further genetic approaches are currently being developed that rely on gene drive systems to drive a pathogen-resistance gene through a population of mosquitoes. One of these methods fuses a resistance gene to an active transposable element to replicate and insert the transgene into the genome of the host mosquito (O'Brochta et al. 2003). Another method involves the use of a homing endonuclease to introduce resistance genes in the genome of the mosquito and drive them through a population (Windbichler et al. 2011). Another drive element, known as “maternal-effect dominant embryonic

arrest" (*Medea*), causes offspring that do not inherit an element-bearing chromosome to be inviable (Beeman et al. 1992). This system has been shown to drive population replacement in *Drosophila* (Chen et al. 2007), and theory indicates that it could drive population suppression as well (Akbari et al. 2014). One more system – "release of insects carrying a dominant lethal" (RIDL) – involves engineering mosquitoes to express a repressible sex-specific lethal gene (Thomas et al. 2000). This construct can be repressed in laboratory rearing settings, but upon release into wild populations, the repressor is removed and the dominant gene is then expressed, causing lethality in female offspring.

Problems facing vector control

While the previously mentioned vector control strategies have seen great success in recent decades, there are several issues that must be addressed to maintain their effectiveness and ensure that we continue to reduce the burden imposed by mosquitoes.

Evolution of insecticide resistance

Perhaps the most pressing of issues facing the vector control community is the evolution of resistance to commonly used insecticides. The evolution of knock-down resistance (*kdr*) to pyrethroids, due to a mutation in the gene *para* (or VSG), has been described (Ffrench-Constant 2013; Jones et al. 2012). The mutations responsible for

resistance have arisen multiple times in *Anopheles gambiae*, the primary vector of malaria in sub-Saharan Africa, and other insect species (Martinez Torres et al. 1998; Williamson et al. 1996). Further resistance to pyrethroids has been reported in *Anopheles funestus* mosquitoes carrying a mutant copy of another gene, *GSTe2*, which also provides resistance to DDT (Riveron et al. 2014). Alleles of *GSTe4* also provide some pyrethroid resistance and have been shown to have arisen independently in multiple *Anopheles* species (Wilding et al. 2015). Mutations at a locus known as resistance to dieldrin (*Rdl*) have been shown arise quite rapidly in *Anopheles* populations, within 18 months in one case (Elliot and Ramakrishna 1956). It has since been shown that the alleles for resistance arose independently in *An. gambiae* and *Anopheles arabiensis* (Du et al. 2005). In addition to physiological resistance, changes in mosquito biting behavior and host preference in response to insecticide use have been documented, as well (reviewed in Gatton et al. 2013).

Cryptic population structure

Sub-population structure within mosquito populations also poses a problem for vector control programs. A cryptic population of *Anopheles gambiae* that is highly susceptible to the malaria parasite and is primarily exophilic (ecologically independent of humans) was reported in west Africa in 2011 (Riehle et al. 2011). This subgroup is not affected by indoor intervention strategies, such as IRS or LLINs, because of its exophilic behavior and its high susceptibility to the malaria parasite makes it a potentially dangerous vector.

Ecological niche partitioning has been described in *An. gambiae*, *An. coluzzii*, and *An. arabiensis* previously, another potential way for dangerous vectors to remain hidden from control interventions (Simard et al. 2009). Likewise, *Aedes aegypti* populations in Indonesia show geographic population structuring, suggesting that vector control strategies that work for one population may not work for other populations (Rašić et al. 2015). In fact, *Ae. aegypti* has been shown to have significant differences in genetic structure at distances as small as 500m (Hlaing et al. 2010). These findings indicate that complete eradication of vectors will need to rely on programs tailored to each locale in which they are deployed, instead of one blanket strategy for a whole region. In contrast to the small local structuring mentioned above, it has also been demonstrated that *Ae. aegypti* demonstrates a broad homogeneity in some regions of the world, further complicating the picture of population structure in this important disease vector (Crawford et al. 2017). It is also possible that population structure not only complicates vector control, but is a product of it, as has been shown in *Anopheline* species (Kamdem et al. 2017).

Cryptic habitats

In order for vector control strategies to be effective, they must be employed and targeted towards areas where mosquitoes are present. This becomes especially difficult in areas where habitats are not easily identifiable or may go unnoticed. *Ae. aegypti* larvae have been found in roof gutters (Gustave et al. 2012), stormwater drains (Arana-

Guardia et al. 2014), and septic tanks (Barrera et al. 2008), among others (cf. Pilger et al. 2011; Russell et al. 2009; González et al. 2007; Kay et al. 2000). *Anopheles* spp. have been found in polluted waters that were previously thought to be uninhabitable (Kamdem et al. 2017). These cryptic habitats can act as reservoirs for vectors and negate the effects of even the most well-intentioned control programs. Recognition of all of the potential breeding sites and habitats is critical for vector control programs to be effective. This is far from an exhaustive list of the issues facing current vector control programs, but is meant to provide an idea of some of the more pressing issues.

The dissertation

The chapters of this dissertation aim to address holes in our current understanding of vector evolution and population genetics, and will provide valuable insight to help inform future vector control strategies.

Studies have shown marked decreases in the effective population sizes of several populations of *An. gambiae* in west Africa in response to the implementation of vector control programs (Hodges et al. 2013; Athrey et al. 2012). In Puerto Rico, *Ae. aegypti* census populations have decreased dramatically in response to the use of new traps for control (Barrera et al. 2014; 2013a); however, it is not yet known whether the decrease in census population size also corresponds to a decrease in effective population size. The effective population size, in short, is the number of breeding individuals that

contribute to the next generation. Knowing the effective size of a population is critical to knowing how much genetic variation exists within that population and how quickly it might respond to new control measures (Athrey et al. 2012). The first chapter of this dissertation aims to fill this gap in our current knowledge of how the effective population size of *Ae. aegypti* in Puerto Rico responds to control measures that drastically affect census population size.

The second chapter provides recombination rate maps for all chromosomes of *An. gambiae* lab crosses. Knowing the recombination rate – and how it varies across the genome and between strains – is important for two reasons. The first is that recombination influences the efficacy of selection, which in turn affects the likelihood and speed with which an introduced genetic element will move through a population (Connallon and Knowles 2007; Haddrill et al. 2007; Roze and Barton 2006; Presgraves 2005; Barton and Charlesworth 1998; Begun and Aquadro 1992; Felsenstein 1974). The second reason is that quantitative information on recombination rate will greatly improve estimates of gene flow between populations (Bossart and Pashley Prowell 1998; Hudson et al. 1992), which can help predict if and how transgenes or resistance alleles will spread between populations.

The third chapter looks at a region of the genome that has been identified as highly divergent between sister species of *Anopheles* – *An. gambiae* sensu stricto and *An.*

coluzzii (White et al. 2011; Lawniczak et al. 2010). This region contains the gene *TEP1*, a gene important in the mosquito's immune response to the malaria parasite (Blandin et al. 2009; 2004). It has been shown that *TEP1* first arose as a chimera of two other TEP genes – *TEP5* and *TEP6* (Obbard et al. 2008). However, this region contains at least five other TEP genes, the function and evolution of which is unknown. I aim to elucidate factors that may have contributed to the evolution of this region and determine whether the chimerization of genes is a common feature of evolution in this region that may have been underappreciated.

Mosquito control has come a long way since its inception, and tremendous strides have been made in reducing the burden of diseases vectored by them. However, despite recent progress, there is still a long way to go to fully eradicate these diseases. Current control programs suffer from the evolution of insecticide resistance (both physiological and behavioral), the presence of cryptic or previously unrecognized habitats, and a lack of critical information on population evolution and structure. By addressing some of the gaps in this knowledgebase, I hope to improve our understanding of vector populations and improve vector control strategies.

REFERENCES

Akbari OS, Chen C-H, Marshall JM, Huang H, Antoshechkin I, Hay BA. 2014. Novel Synthetic MedeaSelfish Genetic Elements Drive Population Replacement in Drosophila; a Theoretical Exploration of Medea-Dependent Population Suppression. *ACS Synth Biol* **3**: 915–928.

Aktar MW, Sengupta D, Chowdhury A. 2009. Impact of pesticides use in agriculture: their benefits and hazards. *Interdiscip Toxicol* **2**: 1–12.

Arana-Guardia R, Baak-Baak CM, Loroño-Pino MA, Machain-Williams C, Beaty BJ, Eisen L, García-Rejón JE. 2014. Stormwater drains and catch basins as sources for production of Aedes aegypti and Culex quinquefasciatus. *Acta Tropica* **134**: 33–42.

Athrey G, Hodges TK, Reddy MR, Overgaard HJ, Matias A, Ridl FC, Kleinschmidt I, Caccone A, Slotman MA. 2012. The Effective Population Size of Malaria Mosquitoes: Large Impact of Vector Control ed. D.A. Petrov. *PLoS Genet* **8**: e1003097.

Barrera R, Amador M, Acevedo V, Hemme RR, Félix G. 2014. Sustained, area-wide control of Aedes aegypti using CDC autocidal gravid ovitraps. *Am J Trop Med Hyg* **91**: 1269–1276.

Barrera R, Amador M, Acevedo V, Caban B, Félix G, Mackay AJ. 2013a. Use of the CDC Autocidal Gravid Ovitrap to Control and Prevent Outbreaks of Aedes aegypti (Diptera: Culicidae). *J Med Entomol* **51**: 145–154.

Barrera R, Mackay AJ, Amador M. 2013b. An Improved Trap to Capture Adult Container-Inhabiting Mosquitoes. *http://dxdoiorg/102987/13-63431* **29**: 358–368.

Barrera R, Amador M, Diaz A, Smith J, Munoz-Jordan JL, Rosario Y. 2008. Unusual productivity of Aedes aegypti in septic tanks and its implications for dengue control. *Med Vet Entomol* **22**: 62–69.

Barton NH, Charlesworth B. 1998. Why sex and recombination? *Science* **281**: 1986–1990.

Batra CP, Raghavendra K, Adak T. 2005. Evaluation of bifenthrin treated mosquito nets against anopheline & culicine mosquitoes. *Indian Journal of …*.

Beach RF, Ruebush TK, Sexton JD, Bright PL, Hightower AW, Breman JG, Mount DL, Oloo AJ. 1993. Effectiveness of permethrin-impregnated bed nets and curtains for malaria control in a holoendemic area of western Kenya. *Am J Trop Med Hyg* **49**: 290–300.

Beeman RW, Friesen KS, Denell RE. 1992. Maternal-effect selfish genes in flour beetles. *Science* **256**: 89–92.

Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. *Nature* **356**: 519–520.

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, et al. 2013. The global distribution and burden of dengue. *Nature* **496**: 504–507.

Bian G, Xu Y, Lu P, Xie Y, Xi Z. 2010. The endosymbiotic bacterium Wolbachia induces resistance to dengue virus in Aedes aegypti. ed. D.S. Schneider. *PLoS Pathog* **6**: e1000833.

Blandin S, Shiao SH, Moita LF, Janse CJ, Waters AP, Kafatos FC, Levashina EA. 2004. Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector Anopheles gambiae. *Cell* **116**: 661–670.

Blandin SA, Wang-Sattler R, Lamacchia M, Gagneur J, Lycett G, Ning Y, Levashina EA, Steinmetz LM. 2009. Dissecting the genetic basis of resistance to malaria parasites in Anopheles gambiae. *Science* **326**: 147–150.

Bossart JL, Pashley Prowell D. 1998. Genetic estimates of population structure and gene flow: Limitations, lessons and new directions. *Trends Ecol Evol (Amst)* **13**: 202–206.

Breman JG. 2001. The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *Am J Trop Med Hyg* **64**: 1–11.

Carvalho DO, McKemey AR, Garziera L, Lacroix R, Donnelly CA, Alphey L, Malavasi A, Capurro ML. 2015. Suppression of a Field Population of Aedes aegypti in Brazil by Sustained Release of Transgenic Male Mosquitoes. ed. K.E. Olson. *PLoS Negl Trop Dis* **9**: e0003864.

Catteruccia F, Brown AE, Petris E, Scali C. 2005. Development of a toolkit for manipulating malaria vectors. *… Modified Mosquitoes for …*.

Chen C-H, Huang H, Ward CM, Su JT, Schaeffer LV, Guo M, Hay BA. 2007. A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila. *Science* **316**: 597–600.

Cheng ML, Ho BC, Bartnett RE, Goodwin N. 1982. Role of a modified ovitrap in the control of Aedes aegypti in Houston, Texas, USA. *Bull World Health Organ* **60**: 291–296.

Connallon T, Knowles LL. 2007. Recombination rate and protein evolution in yeast. *BMC Evol Biol* **7**: 235.

Crawford JE, Alves JM, Palmer WJ, Day JP, Sylla M, Ramasamy R, Surendran SN, Black WC, Pain A, Jiggins FM. 2017. Population genomics reveals that an anthropophilic population of Aedes aegypti mosquitoes in West Africa recently gave rise to American and Asian populations of this major disease vector. *BMC Biology 2017 15:1* **15**: 16.

Du W, Awolola TS, Howell P, Koekemoer LL, Brooke BD, Benedict MQ, Coetzee M, Zheng L. 2005. Independent mutations in the Rdl locus confer dieldrin resistance to Anopheles gambiae and An. arabiensis. *Insect Molecular Biology* **14**: 179–183.

Eliminate Dengue. 2016. Townsville Field Trial Update. (http://www.eliminatedengue.com/library/newsletter/document/townsville/20160426_tsv_trial_update.pdf). Accessed 22 May 2017.

Eliminate Dengue. 2015. Cairns Field Trial Update. (http://www.eliminatedengue.com/library/publication/document/cairns/ed_cairns_field_trial_update_dec.pdf). Accessed 22 May 2017.

Elliot R, Ramakrishna V. 1956. Insecticide resistance in Anopheles gambiae Giles. *Nature* **177**: 532–533.

Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* **78**: 737–756.

Ffrench-Constant RH. 2013. The molecular genetics of insecticide resistance. *Genetics* **194**: 807–815.

Fillinger U, Lindsay SW. 2011. Larval source management for malaria control in Africa: myths and reality. *Malar J* **10**: 353.

Gatton ML, Chitnis N, Churcher T, Donnelly MJ, Ghani AC, Godfray HCJ, Gould F, Hastings I, Marshall J, Ranson H, et al. 2013. The importance of mosquito behavioural adaptations to malaria control in Africa. *Evolution* **67**: 1218–1230.

González R, Gamboa R, Perafán O, Suárez MF. 2007. *Experience of an entomological analysis of the breeding sites of Aedes aegypti and Culex quinquefasciatus in Cali, Colombia*. Rev Colomb Entomol.

Gorman K, Young J, Pineda L, Márquez R, Sosa N, Bernal D, Torres R, Soto Y, Lacroix R, Naish N, et al. 2016. Short-term suppression of Aedes aegypti using genetic control does not facilitate Aedes albopictus. *Pest Management Science* **72**: 618–628.

Griffing SM, Tauil PL, Udhayakumar V, Silva-Flannery L. 2015. A historical perspective on malaria control in Brazil. *Memórias do Instituto Oswaldo Cruz* **110**: 701–718.

Gustave J, Fouque F, Cassadou S, Leon L, Anicet G, Ramdini C, Sonor F. 2012. Increasing Role of Roof Gutters as Aedes aegypti (Diptera: Culicidae) Breeding Sites in Guadeloupe (French West Indies) and Consequences on Dengue Transmission and Vector Control. *Journal of Tropical Medicine* **2012**: 1–3.

Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the Drosophila genome that lack crossing over. *Genome Biol* **8**: R18.

Hlaing T, Tun Lin W, Somboon P, Socheat D, Setha T, Min S, Thaung S, Anyaele O, De Silva B, Chang MS, et al. 2010. Spatial genetic structure of Aedes aegypti mosquitoes in mainland Southeast Asia. *Evolutionary Applications* **3**: 319–339.

Hodges TK, Athrey G, Deitz KC, Overgaard HJ, Matias A, Caccone A, Slotman MA. 2013. Large fluctuations in the effective population size of the malaria mosquito Anopheles gambiae s.s. during vector control cycle. *Evolutionary Applications* **6**: 1171–1183.

Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.

Jana-Kara BR, Jihullah WA, Shahi B, Dev V, Curtis CF, Sharma VP. 1995. Deltamethrin impregnated bednets against Anopheles minimus transmitted malaria in Assam, India. *J Trop Med Hyg* **98**: 73–83.

Jones CM, Liyanapathirana M, Agossa FR, Weetman D, Ranson H, Donnelly MJ, Wilding CS. 2012. Footprints of positive selection associated with a mutation (N1575Y) in the voltage-gated sodium channel of Anopheles gambiae. *Proc Natl Acad Sci U S A* **109**: 6614–6619.

Kambris Z, Cook PE, Phuc HK, Sinkins SP. 2009. Immune Activation by Life-Shortening Wolbachia and Reduced Filarial Competence in Mosquitoes. *Science* **326**: 134–136.

Kamdem C, Fouet C, Gamez S, White BJ. 2017. Pollutants and Insecticides Drive Local Adaptation in African Malaria Mosquitoes. *Mol Biol Evol* **34**: 1261–1275.

Kay BH, Ryan PA, Russell BM, Holt JS, Lyons SA, Foley PN. 2000. The importance of subterranean mosquito habitat to arbovirus vector control strategies in north Queensland, Australia. *J Med Entomol* **37**: 846–853.

Killeen GF. 2003. Following in Soper's footsteps: northeast Brazil 63 years after eradication of Anopheles gambiae. *Lancet Infect Dis* **3**: 663–666.

Lawniczak MKN, Emrich SJ, Holloway AK, Regier AP, Olson M, White B, Redmond S, Fulton L, Appelbaum E, Godfrey J, et al. 2010. Widespread divergence between incipient Anopheles gambiae species revealed by whole genome sequences. *Science* **330**: 512–514.

Lee HI, Seo BY, Shin E-H, Burkett DA, Lee J-K, Shin YH. 2009. Efficiency evaluation of Nozawa-style black light trap for control of anopheline mosquitoes. *Korean J Parasitol* **47**: 159–165.

Lerner KL, Lerner BW. 2002. *World of anatomy and physiology*. Detroit : Thomson/Gale.

Lines J. 1996. Mosquito nets and insecticides for net treatment: a discussion of existing and potential distribution systems in Africa. *Tropical Medicine & International Health* **1**: 616–632.

Long SA, Jacups SP, Ritchie SA. 2015. Lethal ovitrap deployment for Aedes aegypti control: potential implications for non-target organisms. *Journal of Vector Ecology* **40**: 139–145.

Mabaso MLH, Sharp B, Lengeler C. 2004. Historical review of malarial control in southern African with emphasis on the use of indoor residual house-spraying. *Trop Med Int Health* **9**: 846–856.

Martinez Torres D, Chandre F, Williamson MS, Darriet F, Bergé JB, Devonshire AL, Guillet P, Pasteur N, Pauron D. 1998. Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector Anopheles gambiae s.s. *Insect Molecular Biology* **7**: 179–184.

McMeniman CJ, Lane RV, Cass BN, Fong AWC, Sidhu M, Wang Y-F, O'Neill SL. 2009. Stable Introduction of a Life-Shortening Wolbachia Infection into the Mosquito Aedes aegypti. *Science* **323**: 141–144.

Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT, Hedges LM, Rocha BC, Hall-Mendelin S, Day A, Riegler M, et al. 2009. A Wolbachia Symbiont in Aedes aegypti Limits Infection with Dengue, Chikungunya, and Plasmodium. *Cell* **139**: 1268–1278.

Norzahira R, Hidayatulfathi O, Wong HM, Cheryl A. 2011. Ovitrap surveillance of the dengue vectors, Aedes (Stegomyia) aegypti (L.) and Aedes (Stegomyia) albopictus Skuse in selected areas in Bentong, Pahang, …. *Trop Biomed*.

O'Brochta DA, Sethuraman N, Wilson R, Hice RH, Pinkerton AC, Levesque CS, Bideshi DK, Jasinskiene N, Coates CJ, James AA, et al. 2003. Gene vector and transposable element behavior in mosquitoes. *Journal of Experimental Biology* **206**: 3823–3834.

Obbard DJ, Callister DM, Jiggins FM, Soares DC, Yan G, Little TJ. 2008. The evolution of TEP1, an exceptionally polymorphic immunity gene in Anopheles gambiae. *BMC Evol Biol* **8**: 274.

Pilger D, Lenhart A, Manrique Saide P, Siqueira JB, da Rocha WT, Kroeger A. 2011. Is routine dengue vector surveillance in central Brazil able to accurately monitor the Aedes aegypti population? Results from a pupal productivity survey. *Tropical Medicine & International Health* **16**: 1143–1150.

Presgraves DC. 2005. Recombination enhances protein adaptation in Drosophila melanogaster. *Curr Biol* **15**: 1651–1656.

Pulkki-Brännström A-M, Wolff C, Brännström N, Skordis-Worrall J. 2012. Cost and cost effectiveness of long-lasting insecticide-treated bed nets - a model-based analysis. *Cost Eff Resour Alloc* **10**: 5.

Raghavendra K, Subbarao SK. 2002. *Chemical insecticides in malaria vector control in India*. ICMR bull.

Rašić G, Endersby-Harshman N, Tantowijoyo W, Goundar A, White V, Yang Q, Filipović I, Johnson P, Hoffmann AA, Arguni E. 2015. Aedes aegypti has spatially structured and seasonally stable populations in Yogyakarta, Indonesia. *Parasit Vectors* **8**: 610.

Reiter P, Colon M. 1991. ENHANCEMENT OF THE CDC OVITRAP WITH HAY INFUSIONS FOR DAILY MONITORING OF AEDES AEGYP?/POPULATIONS.

Riehle MM, Guelbeogo WM, Gneme A, Eiglmeier K, Holm I, Bischoff E, Garnier T, Snyder GM, Li X, Markianos K, et al. 2011. A Cryptic Subgroup of Anopheles gambiae Is Highly Susceptible to Human Malaria Parasites. *Science* **331**: 596–598.

Riveron JM, Yunta C, Ibrahim SS, Djouaka R, Irving H, Menze BD, Ismail HM, Hemingway J, Ranson H, Albert A, et al. 2014. A single mutation in the GSTe2 gene allows tracking of metabolically based insecticide resistance in a major malaria vector. *Genome Biol* **15**: R27.

Roze D, Barton NH. 2006. The Hill-Robertson effect and the evolution of recombination. *Genetics* **173**: 1793–1811.

Russell BM, McBride WJH, Mullner H, Kay BH. 2009. Epidemiological Significance of Subterranean Aedes aegypti (Diptera: Culicidae) Breeding Sites to Dengue Virus Infection in Charters Towers, 1993. *http://dxdoiorg/101603/0022-2585-391143*.

Sampath TR, Yadav RS, Sharma VP, Adak T. 1998. Evaluation of lambdacyhalothrin-impregnated bednets in a malaria endemic area of India. Part 2. Impact on malaria vectors. *J Am Mosq Control Assoc* **14**: 437–443.

Schmidt, T. L., Barton, N. H., Rašić, G., Turley, A. P., Montgomery, B. L., Iturbe-Ormaetxe, I., et al. (2017). Local introduction and heterogeneous spatial spread of dengue-suppressing Wolbachia through an urban population of Aedes aegypti. *Plos Biology*, *15*(5), e2001894.

Sharma SK, Upadhyay AK, Haque MA, Tyagi PK, Mohanty SS, Raghavendra K, Dash AP. 2009. Field evaluation of Olyset nets: a long-lasting insecticidal net against malaria vectors Anopheles culicifacies and Anopheles fluviatilis in a hyperendemic tribal area of Orissa, India. *J Med Entomol* **46**: 342–350.

Sharma VP, Yadav RS. 1995. *Cyfluthrin impregnated mosquito nets to control malaria in the mining settlement of Orissa, India*. Public Health.

Shiff C. 2002. Integrated approach to malaria control. *Clin Microbiol Rev* **15**: 278–293.

Simard F, Ayala D, Kamdem GC, Pombi M, Etouna J, Ose K, Fotsing J-M, Fontenille D, Besansky NJ, Costantini C. 2009. Ecological niche partitioning between Anopheles gambiae molecular forms in Cameroon: the ecological side of speciation. *BMC Ecol* **9**: 17.

Soper FL, Wilson DB. 1943. Anopheles gambiae in Brazil, 1930 to 1940.

Sreehari U, Raghavendra K, Rizvi MMA, Dash AP. 2009. Wash resistance and efficacy of three long-lasting insecticidal nets assessed from bioassays on Anopheles culicifacies and Anopheles stephensi. *Tropical Medicine & International Health* **14**: 597–602.

Stich AH, Maxwell CA, Haji AA, Haji DM, Machano AY, Mussa JK, Matteelli A, Haji H, Curtis CF. 1994. Insecticide-impregnated bed nets reduce malaria transmission in rural Zanzibar. *Trans R Soc Trop Med Hyg* **88**: 150–154.

Thomas DD, Donnelly CA, Wood RJ, Alphey LS. 2000. Insect population control using a dominant, repressible, lethal genetic system. *Science* **287**: 2474–2476.

Trigg PI, Kondrachine AV. 1998. Commentary: malaria control in the 1990s. *Bull World Health Organ* **76**: 11–16.

Turelli, M., & Barton, N. H. (2017). Deploying dengue-suppressing Wolbachia : Robust models predict slow but effective spatial spread in Aedes aegypti. *Theoretical Population Biology*, *115*, 45–60.

Turelli M, Hoffmann AA. 1995. Cytoplasmic incompatibility in Drosophila simulans: dynamics and parameter estimates from natural populations. *Genetics* **140**: 1319–1338.

Wakabi W. 2007. Africa counts greater successes against malaria. *Lancet*, December 8.

White BJ, Lawniczak MKN, Cheng C, Coulibaly MB, Wilson MD, Sagnon N, Costantini C, Simard F, Christophides GK, Besansky NJ, et al. 2011. Adaptive divergence between incipient species of Anopheles gambiae increases resistance to Plasmodium. *Proc Natl Acad Sci U S A* **108**: 244–249.

WHO. 2017a. Dengue and severe dengue fact sheet. (http://www.who.int/mediacentre/factsheets/fs117/en/). Accessed 12 May 2017.

WHO. 2017b. Mosquito-borne diseases. (http://www.who.int/neglected_diseases/vector_ecology/mosquito-borne-diseases/en/). Accessed 17 May 2017.

WHO. 2017c. Neglected Tropical Diseases: Integrated Vector Management. (http://www.who.int/neglected_diseases/vector_ecology/ivm_concept/en/). Accessed 22 May 2017.

WHO. 2016. World Malaria Report.

Wilding CS, Weetman D, Rippon EJ, Steen K, Mawejje HD, Barsukov I, Donnelly MJ. 2015. Parallel evolution or purifying selection, not introgression, explains similarity in the pyrethroid detoxification linked GSTE4 of Anopheles gambiae and An. arabiensis. *Mol Genet Genomics* **290**: 201–215.

Williamson MS, Martinez Torres D, Hick CA. 1996. Identification of mutations in the houseflypara-type sodium channel gene associated with knockdown resistance (kdr) to pyrethroid insecticides. *Molecular and General …*.

Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, Ulge UY, Hovde BT, Baker D, Monnat RJ, Burt A, et al. 2011. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* **473**: 212–215.

CHAPTER 1

Effective Population Size Fluctuation in *Aedes aegypti* in Response to Vector Control Strategies in Puerto Rico

ABSTRACT

Aedes aegypti is a major vector of several arboviruses of significant medical importance around the world. As a result, it has been the target of many vector control strategies with varying success rates. Most strategies involve one of two foci: eliminating the adult stage of mosquitoes or eliminating the juvenile stages of the mosquito. Eliminating the adult stage of mosquitoes is limited by the evolution of insecticide resistance and the fact that mosquitoes may not be killed before they are able to reproduce and spread disease. Eliminating juvenile stages of mosquitoes suffers from the evolution of larvicide resistance and the presence of cryptic larval habitats that are not able to be targeted. In Puerto Rico, a new control measure has shown promise. autocidal gravid ovitraps have shown a reduction in census population size of ~80% in areas where the traps have been deployed.

As with all vector control strategies, census population size decreases are favorable; however, the ultimate goal is a reduction in the effective population size, broadly the number of individuals in the population that actually reproduce and contribute to the gene pool. Effective population size is a major factor in determining genetic variation within the population, which ultimately predicts how well a population can adapt to a new challenge. Studies in *Anopheles gambiae* have shown that Approximate Bayesian Computation (ABC) is an effective method for predicting effective population size changes in response to vector control. Here, I show that ABC gives mixed results, at

best, in showing concomitant effective population size decreases in response to the vector control measures being deployed. I posit reasons why *Ae aegypti* may not respond in the same way as *An. gambiae* to vector control and why ABC may not be an effective tool for estimating effective population size in *Ae. aegypti*. Additionally, I show that spatially distinct populations of *Ae. aegypti* in this region are not genetically distinct, possibly indicating high levels of migration between geographic populations or a recent population bottleneck affecting all populations.

INTRODUCTION

Arboviruses vectored by the mosquito *Aedes aegypti* – including dengue virus, chikungunya virus, yellow fever virus, and Zika virus – continue to impose a tremendous burden on human populations in tropical and sub-tropical climates. Dengue virus alone causes an estimated 390 million cases every year, with 20,000-25,000 of those resulting in death (Bhatt et al. 2013). While vaccines are currently under development, vector control remains the most effective method of reducing disease transmission (WHO, 2017). Vector control strategies commonly rely on controlling either immature or adult mosquitoes. Immature control strategies involve removing oviposition sites, such as containers that may hold standing water after rainfall, and larviciding; adult control strategies most commonly rely on pesticide sprays, though some locations are starting to implement indoor residual spraying programs (Vazquez-Prokopec et al. 2010). As a result of evolution of resistance to pesticides and the fact that adulticidal strategies may not kill mosquitoes before they transmit disease, insecticide spraying is often not a viable long-term vector control strategy (Gubler et al. 2014). Likewise, larviciding programs can be plagued by the evolution of resistance (Grisales et al. 2013; Marcombe et al. 2011; Ranson et al. 2010) and the presence of cryptic larval habitats (Arana-Guardia et al. 2014; Gustave et al. 2012; Pilger et al. 2011). For these reasons, Barrera et al. 2013 implemented a new mosquito trap – the autocidcal gravid ovitrap (AGO) – and tested its efficacy in Puerto Rico. Over the course of the study, there was an ~80% reduction in the number of mosquitoes caught per trap, per week relative to pre-trap

placement levels (Barrera et al. 2014; 2013). While reductions in the gross number of mosquitoes are important for vector control, the ultimate determinant of the long-term efficacy of an intervention strategy is a reduction in effective population size (N_e).

The effective population size of a population is the ideal population size at which genetic drift, the accumulation of inbreeding, and the loss of variation reflect what is seen in the actual population (Nei and Takahata 1993; Hill 1979). N_e is an important factor in the determination of levels of migration, drift, and selection in a population, which ultimately determine the amount of genetic variation in a population (Athrey et al. 2012). High levels of genetic variation in a population will allow the population to easily adapt to new challenges, while low levels can make a population susceptible to crashes in the face of a new challenge to which it is not well adapted. Studies have shown that Approximate Bayesian Computation (ABC, Beaumont et al. 2002) can be used to understand evolutionary histories of populations by estimating parameters of interest, such as effective population size changes (Beaumont 2010; Estoup et al. 2010; Guillemaud et al. 2009; Palero et al. 2009). In *Anopheles gambiae*, ABC has been used with microsatellite markers to demonstrate that vector control strategies that have reduced census population size also cause a concomitant reduction in effective population size (Hodges et al. 2013; Athrey et al. 2012). Using single nucleotide polymorphism (SNP) data and ABC, I set out to determine whether the 80% reduction in census population size seen in Puerto Rico in response to AGO deployment also caused

a reduction in effective population size. Additionally, I sought to determine whether genetic population structure in *Ae. aegypti* in the region reflects geographic population structure and if that could be used to detect mosquito migration from village to village.

MATERIALS AND METHODS

Trapping and collection of mosquitoes

Mosquito trapping and collections were carried out as part of Barrera et al. 2014 and Barrera et al. 2013. Briefly, two villages in southern Puerto Rico were chosen as intervention and surveillance sites (La Margarita and Villodas, respectively) to test the efficacy of AGO traps in vector control. Both sites have reliable sanitation services (i.e. piped water, trash collection services, and sewerage). Prior to installation of three to four traps per home in La Margarita, trappings were carried out in both locations to assess the baseline mosquito populations. After baseline trappings were completed, source reduction, larviciding, and egg removal were carried out in both areas and AGO traps were installed in La Margarita. Mosquito density in each location was assessed weekly from December 2011 to October 2012 by counting the number of mosquitoes collected in sentinel AGO (SAGO) and BG-Sentinel traps placed uniformly throughout each village. AGO traps were deployed in Villodas in February 2013 as a continuation of the original phase of the trial. AGO traps were thus in deployment in La Margarita from December 2011 to February 2014 and in Villodas from February 2013 to February 2014. Thirty mosquitoes from each location collected during weeks 4, 7, 30, 33, 50, 53, 157,

and 160 of the study period (beginning October 17, 2011) were preserved for genotyping and population genetic analysis. To assess whether migration between La Margarita and surrounding areas was occurring, an additional 20 mosquitoes from La Margarita and four neighboring urban areas – separated from La Margarita by 200m of vegetation – were collected and preserved for genotyping in week 158 of the study period.

ddRADseq library construction

Double-digest restriction-associated DNA sequencing (ddRADseq) libraries were constructed, largely following (Turissini et al. 2014). DNA was first extracted from adult mosquitoes using the ZR-96 Quick-gDNA kit (Zymo Research; Irvine, CA). Extracted DNA was digested for three hours at 37^{o}C with *MluC1* and *NlaIII* (New England Biolabs [NEB]; Ipswich, MA). The digested DNA was then purified using homemade 'Ampure' magnetic beads. A barcoded adapter (1 of 48) was then ligated to the sticky end produced by *NlaIII*, and a universal adapter was ligated to the sticky end produced by *MluC1* using T4 DNA ligase (NEB). Reactions from uniquely barcoded individuals were then pooled together, concentrated, and cleaned using two rounds of magnetic bead purification. These sub-libraries were size selected on a Blue Pippin (Sage Science; Beverly, MA) on a 1.5% gel cassette with the 'tight' setting and a 400 bp fragment target. Size-selected DNA from each sub-library was then PCR amplified using one of 12 indexed primers. Each PCR reaction consisted of 1x NEB Q5 PCR buffer, 10mM each DNTP, 20pmol of the

universal and indexed primer, 0.25 U of NEB Q5 DNA polymerase, and 4 µl of template. The reactions were carried out as follows: 98^{o}C for 1 minute; 12 cycles of 98^{o}C for 8 seconds, 68^{o}C for 20 seconds, and 72^{o}C for 20 seconds; and a final extension at 72^{o}C for 2 minutes. Following the PCR reaction, there are 576 (12 indexed primers x 48 barcodes) uniquely barcoded/indexed individuals that can be multiplexed for sequencing. This was done by first checking the quality and quantity of each of the twelve sub-libraries on a BioAnalyzer 2100 (Agilent Technologies; Santa Clara, CA), after which equal molar amounts of each sub-library were pooled to create a final sequencing library. This library was then sequenced on three lanes of the Illumina HiSeq 2500 (Illumina; San Diego, CA) at the UCR genomics core using 100 bp paired-end reads.

Read mapping and SNP calling

After library construction and replicate sequencing runs on three lanes of the Illumina HiSeq 2500, samples were demultiplexed and reads were filtered and trimmed for quality using custom scripts. Reads were then mapped to the *Aedes aegypti* reference genome using the Burrows-Wheeler aligner (BWA, Li and Durbin 2009) BWA-MEM algorithm with default settings. SNPs were called on a per population basis using the Genome Analysis ToolKit (GATK, DePristo et al. 2011) HaplotypeCaller algorithm with default settings.

Estimation of N_e

The SNP data were used to estimate N_e at each sampling time point in Villodas and LaMargarita using approximate Bayesian computation, implemented in the program DIYABC (Cornuet et al. 2014). This method draws model parameters from a uniform prior distribution, simulates a dataset using a given model and the parameters drawn from the prior distribution, calculates a set of summary statistics, and calculates how close the summary statistics of the simulated dataset are to the given dataset to determine whether the simulation is acceptable or not. A posterior distribution for each of the parameters is then drawn, based on which parameter values resulted in acceptable simulations under different models. It can then be determined which model best describes the given data and summary statistics (i.e. N_e) can be calculated from the posterior distributions. For each population, I ran 100,000 simulations for each of 6 different models. The models were: 1) fluctuating population size throughout the sampling period, 2) contracting population size throughout the sampling period, 3) expanding population size throughout the sampling period, and 4-6) population bottlenecks at varying times throughout the sampling period. Because of limitations in computing power, I ran multiple sets of simulations using a random sample of 10,000 sites for each set and compared results across sets of simulations. Additionally, because of computing limitations and because sampling was carried out in multiple sets of samples taken 3 weeks apart (e.g. samples were taken at weeks 4 and 7, weeks 30 and 33, weeks 50 and 53, and weeks 157 and 160), I 1) ran simulations using samples taken

first in each pair of time points (e.g. samples taken at weeks 4, 30, 50, and 157), 2) ran simulations taken second in each pair of time points (weeks 7, 33, 53, and 160), 3) averaged N_e for each pair of sampling events after simulations, and 4) merged each pair of sampling events prior to simulations (so each pair of sampling events was treated as a single event). To determine whether there was an effect of missing data, I ran additional simulations in which I filtered the data to include only those sites for which I had data for 75% or more of individuals and again merged pairs of sampling events.

Analysis of the genetic structure of populations

To determine whether individual mosquitoes are migrating from neighboring villages into LaMargarita, additional sampling was carried out in LaMargarita and 4 other nearby villages (Pueblo, San Juan, Villa Cofresi, and Villa Esperanza). Sequencing, read mapping, and SNP calling were performed as described above. I assessed whether each of these villages represented a unique genetic population using STRUCTURE (Pritchard et al. 2000), a program that will bin individuals into a user defined number of populations (*K*) based on genetic differences between individuals, as well as report how much of each individual's genome belongs in each population. I first subsampled our set of SNPs so that only sites with data for at least 75% of individuals were included. STRUCTURE was run for *K*=1-10 on each of 10 different random samples of 20,000 SNPs from this filtered set. After an initial burn-in of 100,000 replicates I ran 100,000 analysis replicates for

each *K* and then determined the best fitting *K* value for each random sample of SNPs by calculating the relative probability, lnPr(*X*|*K*), of each *K*.

RESULTS

Read mapping and SNP calling

A total of 568,989,132 reads across three lanes of sequencing passed filter and were successfully assigned to individuals (~87%), based on barcode and index sequences, allowing for up to one mismatch per barcode and index. This results in an average of 989,546 (median: 525,969) reads assigned per individual (Figure 1.1a). The Villodas population had a total of 5,187,366 polymorphic sites with individuals having data for, on average, 526,706 sites (median: 352,342) (Figure 1.1b). The samples obtained from LaMargarita had 5,518,965 total polymorphic sites and an average of 619,909 (median: 414,758) sites with data per individual. The control populations had a total of 1,778,915 polymorphic sites and an average of 313,659 (median: 248,845) sites with data per individual.

Estimation of N_e

Villodas

For all simulations run, the model allowing for fluctuating N_e throughout the sampling period was the best fit (Figure 1.2). For simulations run with each sampling time point as a separate event, the minimum N_e in Villodas occurred at week 4 and was 1,617

individuals, while the maximum N_e occurred at week 50 and was 95,160 individuals. When the sampling time points were averaged after simulations, the minimum came at weeks 4/7, with 29,619 individuals, and the maximum was 69,465 at weeks 50/53. The simulations in which the paired sampling time points were merged prior to simulation, the minimum N_e was 24,900 at weeks 4/7 and the maximum was 83,000 at weeks 30/33.

LaMargarita

As in Villodas, the model allowing for fluctuating N_e throughout the sampling period was the best fit for simulations run with each sampling time point as a separate event; however, in the simulations in which paired sampling time points were merged prior to the simulations, a model in which the population declines from weeks 4/7 to weeks 50/53 and rebounds by week 157/160 was supported (Figure 1.3). In LaMargarita, the minimum N_e when all sampling time points were treated as separate events was 26,190 at week 4 and the maximum occurred at week 33 and was 94,910. When pairs of sampling time points were averaged after simulations, the minimum N_e was 31,725 at weeks 4/7 and the maximum was 86,830 at weeks 50/53. In the merged sampling simulations, N_e hit a minimum at weeks 50/53 with an estimate of 39,480 mosquitoes, while the maximum occurred at weeks 157/160 with 80,450 mosquitoes.

Analysis of the genetic structure of populations

For all runs, the most appropriate *k* estimated by STRUCTURE was 2 (relative lnPr(*X*|*K*) for *k* = 2 was >0.99 for all replicates, Figure 1.4a). All individuals in all five villages are grouped into one genetic cluster, with a small number of individuals from each village showing a very small fraction (maximum 3.8%) of their genome belonging to a secondary cluster; individuals without a secondary fraction of their genome account for 52% of all samples. This suggests a high degree of migration and admixture between these five villages. When *k* is increased to 5, which would be expected to group individuals by their sampling site if there were genetic population structure between the villages, I again see that all individuals in all five villages have >95% of their genome belonging to the first genetic cluster, with a few individuals showing a small fraction (maximum 3.9%) of their genome belonging to secondary clusters (Figure 1.4b). In total, 45% of individuals in this scenario show no trace of a secondary genetic cluster, while 31% of individuals show traces of multiple secondary clusters. To test whether I could pick up genetic population structure of more geographically separated populations, I ran STRUCTURE on the samples collected from La Margarita and Villodas – which are separated by 20 Km – at week 160. Using 10 random samples of 10,000 SNPs, I ran STRUCTURE for *k* = 1-10 and for all replicates the most likely *k* was 1 (relative lnPr(*X*|*K*) for *k* = 1 was >0.999 for all replicates), meaning that these two villages comprise 1 genetic population.

DISCUSSION

Estimating N_e

While the use of ABC to track N_e changes in *An. gambiae* proved to be useful and showed concomitant N_e changes in response to vector control interventions (Hodges et al. 2013; Athrey et al. 2012), it did not demonstrate the same results in this study on *Ae. aegypti*. This could be for several reasons, the first being that effective population size in these villages may not respond to vector control in the same way that it does in *An. gambiae*. The genetic homogeneity of these populations – as demonstrated by our STRUCTURE results – may mean that N_e is subject to fluctuations caused by stochastic processes, rather than interventions for vector control. Because mosquitoes can breed in very small amounts of water, I also looked at rainfall data to determine whether it may be having an effect on effective population size that could be confounding the results, but saw no correlation between rainfall and the fluctuations seen in effective population size. Additionally, the samples used in this study came from points that were separated in time by as much as two years, making it difficult for us to track N_e fluctuation over the time scales that trappings showed decreases in census population size. This study also includes only 30 individuals per time point at each sampling site, whereas the studies conducted in *An. gambiae* that showed much more clear results included an average of ~70 mosquitoes per sampling event. It has been shown that a sufficient number of samples is required to properly estimate effective population size, and our sampling strategy may not have met this minimum requirement (Wang 2001).

The *Anopheles* studies used microsatellite data, resulting in many fewer markers per individual and potentially many more alleles per site, which could make calculating N_e more straightforward than using thousands of SNP markers – which can only have a finite number of alleles – per individual, as I have done here. *Anopheles* also has a much smaller effective population size than what I have estimated here for *Ae. aegypti*, frequently numbering in the few hundred individuals, whereas our lowest estimate of N_e in this study is a few thousand individuals (Athrey et al. 2012). Genome size and content differences may also explain the differences in ability to detect N_e changes. *Ae. aegypti* has a genome size of 1.3 Gb and nearly half of it (47%) is composed of transposable element sequence (Nene et al. 2007). *An. gambiae*, on the other hand, has a genome size of 273 Mb and only ~20% is transposable element sequence (Fernández-Medina et al. 2011; Holt et al. 2002). The relatively high degree of transposable element sequence in *Ae. aegypti* may make it difficult to accurately estimate N_e. Further complicating matters is the fact that our ddRADseq protocol was optimized for *An. gambiae* and the restriction enzymes used may not be as suitable for sequencing *Ae. aegypti.*

Analyzing the genetic structure of populations

While previous studies in other locales have demonstrated that *Ae. aegypti* has spatially structured populations over similar distances as our sampling sites (Rašić et al. 2015; Olanratmanee et al. 2013; Hlaing et al. 2010; Herrera et al. 2006), I was not able to

detect a correlation between genetic and spatial population structure in our samples. Even at a distance of 20 Km, there was no genetic structure between samples taken from La Margarita and Villodas. While this may again have to do with our sequencing protocol not necessarily being optimized for *Aedes*, our results corroborate recent findings indicating that *Ae. aegypti* populations in the Americas likely arose out of a very small founding event consisting of mosquitoes from Senegal a few hundred years ago, and as such, spatial population differentiation is difficult, if not impossible, to detect (Crawford et al. 2017). High levels of migration between geographic populations may also account for the lack of genetic structure between them, which may explain the STRUCTURE results for La Margarita and the nearby villages, though this is unlikely at the distance between La Margarita and Villodas (Honório et al. 2003).

Conclusions

AGO traps have been shown to be effective at reducing census population size in *Ae. aegypti* in Puerto Rico (Barrera et al. 2014; 2013). However, unlike results for *An. gambiae* (Hodges et al. 2013; Athrey et al. 2012), I was unable to detect a concomitant reduction in N_e in response to this vector control strategy. There are a number of mechanistic reasons why this may be the case – outlined above – however, it may also be the case that AGO trappings do not effectively reduce N_e, meaning that these populations might still maintain a high degree of adaptability to new vector control strategies and future control efforts will need to take this into account. Furthermore,

this study corroborates recent findings that *Ae. aegypti* in the Americas show little geographic differentiation at the genomic level, making it difficult to accurately predict levels of migration between spatially close populations, further complicating vector control. Going forward, an effective method for monitoring *Ae. aegypti* effective population size needs to be developed so that vector control strategies can be accurately monitored. A good first step in this pursuit would be to perform a similar analysis as I have done here, but to sample more individuals per site, per time point. Additionally, it may be possible to determine whether microsatellite sequence data have been captured in this study and use that data to re-evaluate the scenarios and simulations carried out here. Further elucidating population structure in this important region is also critical, as is determining rates of migration from population to population throughout Puerto Rico. It is clear from this study that there is still a long way to go to fully understand the population genetics and vector biology of this important vector species and it will continue to be difficult to implement effective long-term vector control strategies without further information on how *Ae. aegypti* responds to those efforts.

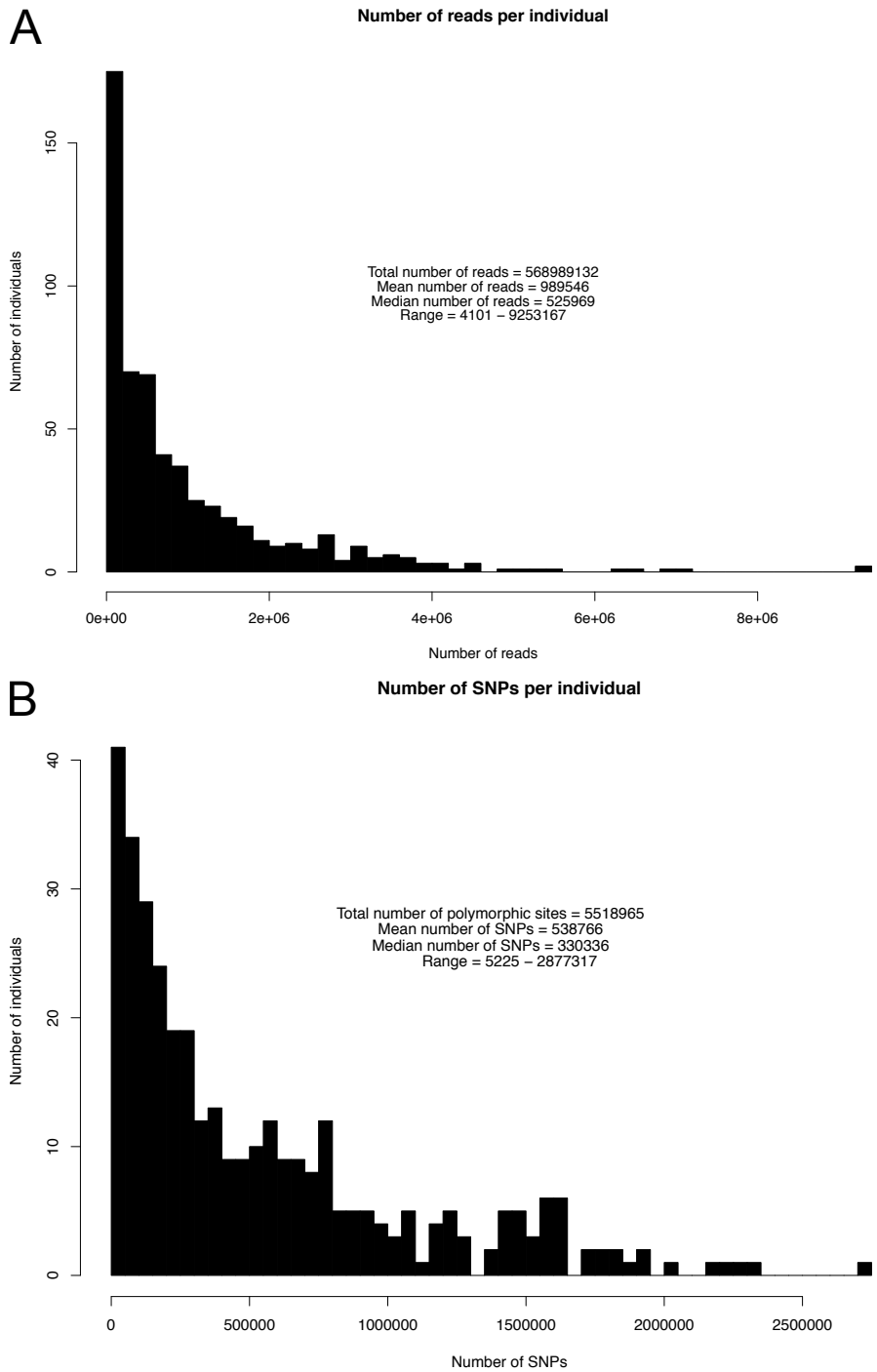

Figure 1.1 Reads and SNPs assigned per individual The number of reads assigned (A) and SNPs called (B) per individual over all sampling time points and locations.

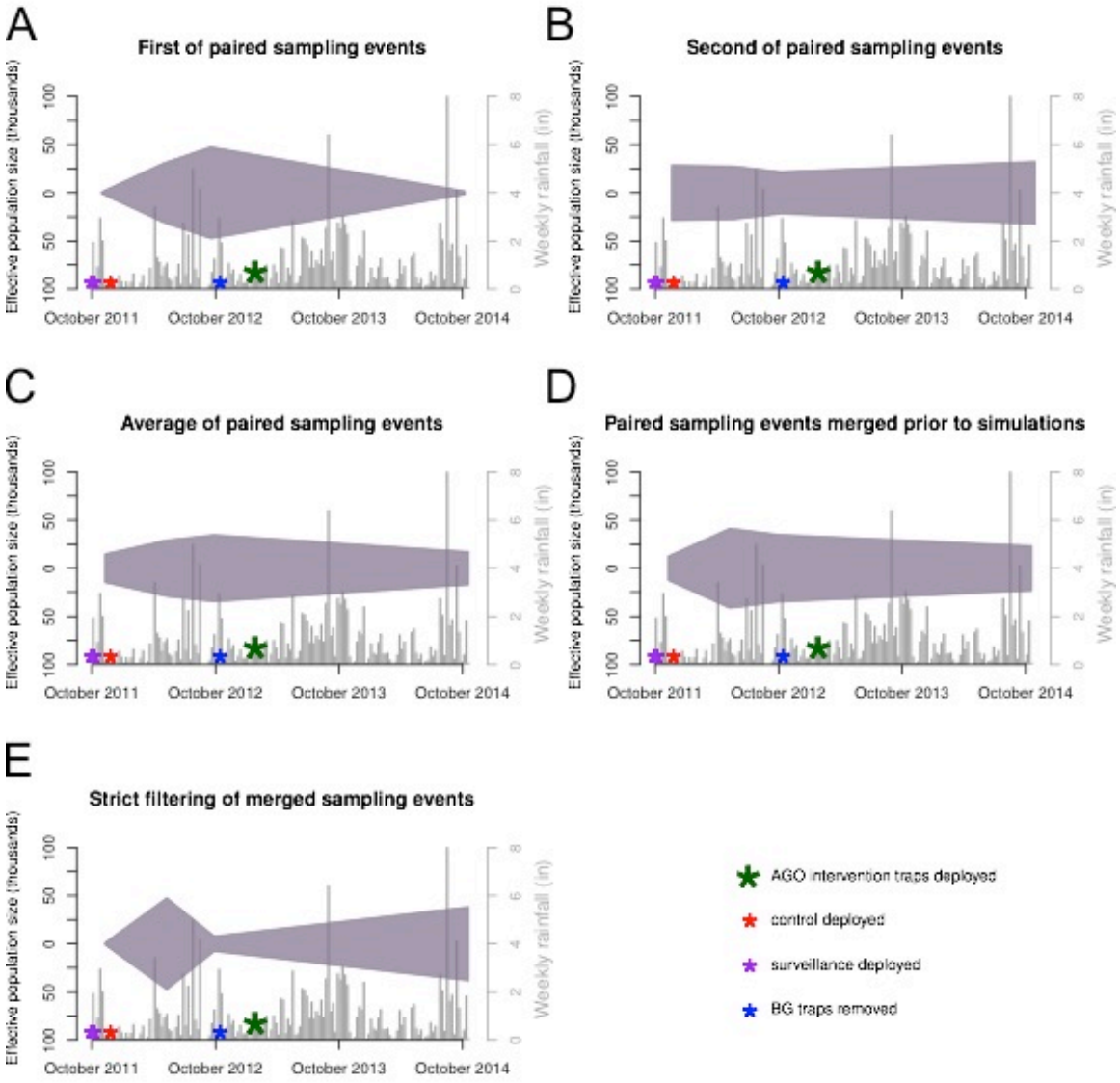
A
First of paired sampling events
B
Second of paired sampling events
C
Average of paired sampling events
D
Paired sampling events merged prior to simulations
E
Strict filtering of merged sampling events
Effective population size (thousands)
Weekly rainfall (in)
October 2011
October 2012
October 2013
October 2014
AGO intervention traps deployed
control deployed
surveillance deployed
BG traps removed

Figure 1.2 Effective population size in Villodas as estimated by ABC Effective population size estimates in Villodas as estimated by various runs of ABC. A) first sampling events, referring to samples taken at weeks 4, 30, 50, and 157; B) second sampling events, referring to samples taken at weeks 7, 33, 53, and 160; C) average of paired sampling events, referring to the averages of weeks 4 and 7, 30 and 33, 50 and 53, and 157 and 160 from the first two sets of simulations; D) paired sampling events merged prior to simulations, referring to simulations run in which the pairs of sampling events mentioned above were merged into single sampling events prior to running simulations; and e) strict filtering of merged sampling events, referring to simulations run in which the individuals with less than 1 SNP per 10,000 bases were filtered out, after which sites with data for less than 75% of the remaining individuals were filtered out and pairs of sampling events were merged as above. “surveillance deployed” refers to the placement of BG-Sentinel and Sentinel AGO traps; “control deployed” refers to the application of larvicides and pesticides and the removal of larval habitats; “AGO intervention deployed” refers to the placement of AGO traps; and “BG traps removed” refers to the removal of BG-Sentinel traps. The gray bar plot in the background is weekly rainfall data for the area, collected by the National Oceanographic and Atmospheric Administration (NOAA).

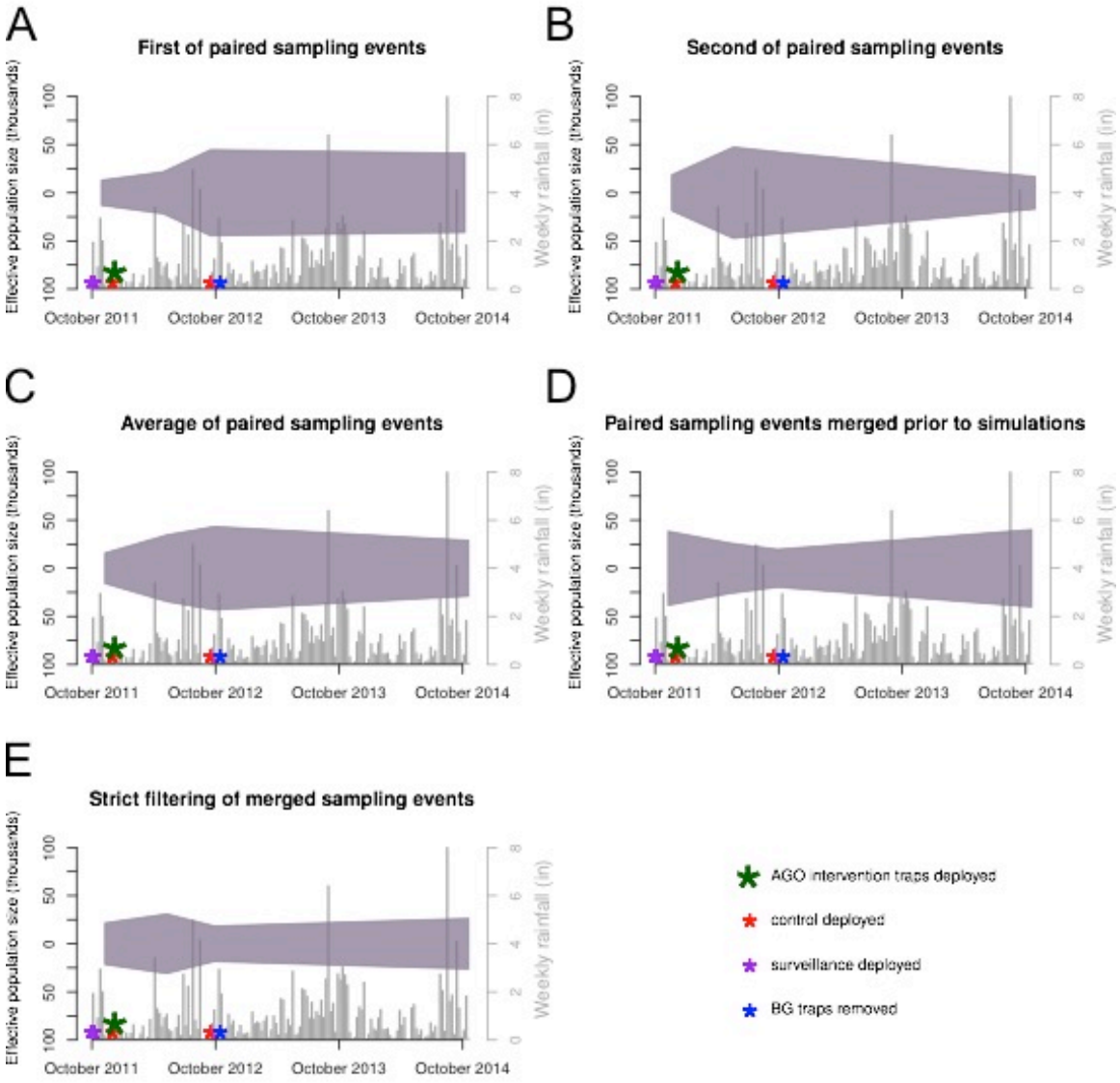
A
First of paired sampling events
B
Second of paired sampling events
C
Average of paired sampling events
D
Paired sampling events merged prior to simulations
E
Strict filtering of merged sampling events
Effective population size (thousands)
Weekly rainfall (in)
October 2011
October 2012
October 2013
October 2014
AGO intervention traps deployed
control deployed
surveillance deployed
BG traps removed

Figure 1.3 Effective population size in La Margarita as estimated by ABC Effective population size estimates in La Margarita as estimated by various runs of ABC. A) first sampling events, referring to samples taken at weeks 4, 30, 50, and 157; B) second sampling events, referring to samples taken at weeks 7, 33, 53, and 160; C) average of paired sampling events, referring to the averages of weeks 4 and 7, 30 and 33, 50 and 53, and 157 and 160 from the first two sets of simulations; D) paired sampling events merged prior to simulations, referring to simulations run in which the pairs of sampling events mentioned above were merged into single sampling events prior to running simulations; and e) strict filtering of merged sampling events, referring to simulations run in which the individuals with less than 1 SNP per 10,000 bases were filtered out, after which sites with data for less than 75% of the remaining individuals were filtered out and pairs of sampling events were merged as above. “surveillance deployed” refers to the placement of BG-Sentinel and Sentinel AGO traps; “control deployed” refers to the application of larvicides and pesticides and the removal of larval habitats; “AGO intervention deployed” refers to the placement of AGO traps; and “BG traps removed” refers to the removal of BG-Sentinel traps. The gray bar plot in the background is weekly rainfall data for the area, collected by the National Oceanographic and Atmospheric Administration (NOAA).

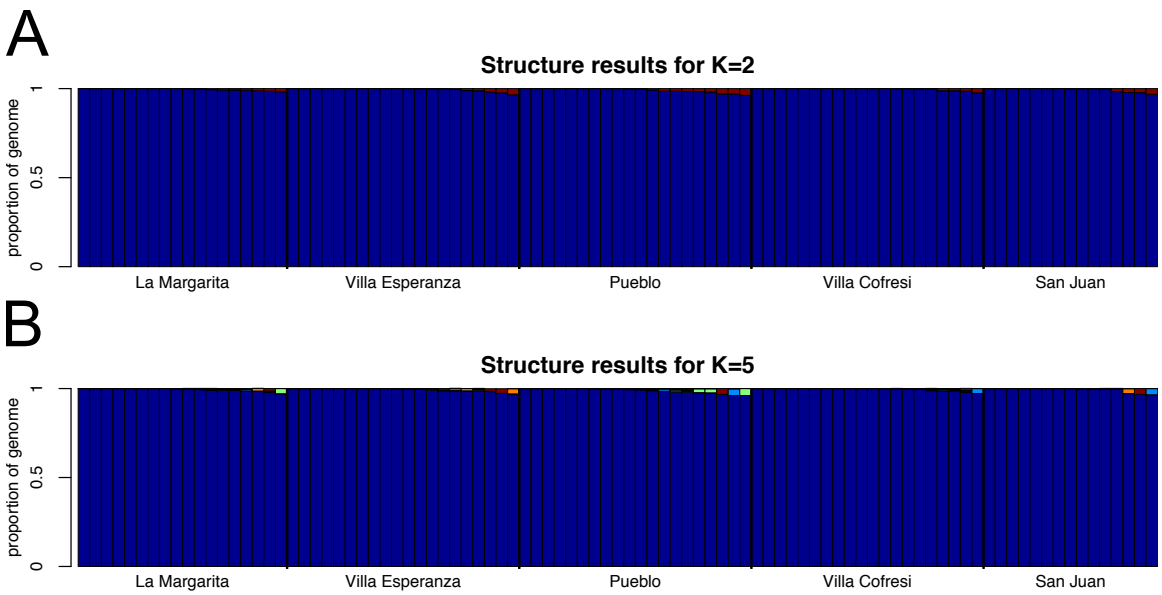

Figure 1.4 STRUCTURE results for two selected *k* values STRUCTURE results for *k*=2 (A) and *k*=5. Plots are representative runs from 10 replicates of random sampling of 20,000 SNPs run from *k*=1-10. Each bar represents one individual and different colors represent the portion of the genome that belongs to a given genetic cluster.

REFERENCES

Arana-Guardia R, Baak-Baak CM, Loroño-Pino MA, Machain-Williams C, Beaty BJ, Eisen L, García-Rejón JE. 2014. Stormwater drains and catch basins as sources for production of Aedes aegypti and Culex quinquefasciatus. *Acta Tropica* **134**: 33–42.

Athrey G, Hodges TK, Reddy MR, Overgaard HJ, Matias A, Ridl FC, Kleinschmidt I, Caccone A, Slotman MA. 2012. The Effective Population Size of Malaria Mosquitoes: Large Impact of Vector Control ed. D.A. Petrov. *PLoS Genet* **8**: e1003097.

Barrera R, Amador M, Acevedo V, Caban B, Félix G, Mackay AJ. 2013. Use of the CDC Autocidal Gravid Ovitrap to Control and Prevent Outbreaks of Aedes aegypti (Diptera: Culicidae). *J Med Entomol* **51**: 145–154.

Barrera R, Amador M, Acevedo V, Hemme RR, Félix G. 2014. Sustained, area-wide control of Aedes aegypti using CDC autocidal gravid ovitraps. *Am J Trop Med Hyg* **91**: 1269–1276.

Beaumont MA. 2010. Approximate Bayesian Computation in Evolution and Ecology. *http://dxdoiorg/101146/annurev-ecolsys-102209-144621* **41**: 379–406.

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian Computation in Population Genetics. *Genetics* **162**: 2025–2035.

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, et al. 2013. The global distribution and burden of dengue. *Nature* **496**: 504–507.

Cornuet J-M, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin J-M, Estoup A. 2014. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* **30**: 1187–1189.

Crawford JE, Alves JM, Palmer WJ, Day JP, Sylla M, Ramasamy R, Surendran SN, Black WC, Pain A, Jiggins FM. 2017. Population genomics reveals that an anthropophilic population of Aedes aegypti mosquitoes in West Africa recently gave rise to American and Asian populations of this major disease vector. *BMC Biology 2017 15:1* **15**: 16.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**: 491–498.

Estoup A, Baird SJE, Ray N, Currat M, Cornuet J-M, Santos F, Beaumont MA, Excoffier L. 2010. Combining genetic, historical and geographical data to reconstruct the dynamics of bioinvasions: application to the cane toad Bufo marinus. *Molecular Ecology Resources* **10**: 886–901.

Fernández-Medina RD, Struchiner CJ, Ribeiro JMC. 2011. Novel transposable elements from Anopheles gambiae. *BMC Genomics* **12**: 260.

Grisales N, Poupardin R, Gomez S, Fonseca-Gonzalez I, Ranson H, Lenhart A. 2013. Temephos Resistance in Aedes aegypti in Colombia Compromises Dengue Vector Control ed. P.F. Pimenta. *PLoS Negl Trop Dis* **7**: e2438.

Gubler DJ, Ooi EE, Vasudevan S, Farrar J. 2014. Dengue and Dengue Hemorrhagic Fever, 2nd Edition. CABI.

Guillemaud T, Beaumont MA, Ciosi M, Cornuet JM, Estoup A. 2009. Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity* **104**: 88–99.

Gustave J, Fouque F, Cassadou S, Leon L, Anicet G, Ramdini C, Sonor F. 2012. Increasing Role of Roof Gutters as Aedes aegypti (Diptera: Culicidae) Breeding Sites in Guadeloupe (French West Indies) and Consequences on Dengue Transmission and Vector Control. *Journal of Tropical Medicine* **2012**: 1–3.

Herrera F, Urdaneta L, Rivero J, Zoghbi N, Ruiz J, Carrasquel G, Martínez JA, Pernalete M, Villegas P, Montoya A, et al. 2006. Population genetic structure of the dengue mosquito Aedes aegypti in Venezuela. *Memórias do Instituto Oswaldo Cruz* **101**: 625–633.

Hill WG. 1979. A note on effective population size with overlapping generations. *Genetics* **92**: 317–322.

Hlaing T, Tun Lin W, Somboon P, Socheat D, Setha T, Min S, Thaung S, Anyaele O, De Silva B, Chang MS, et al. 2010. Spatial genetic structure of Aedes aegypti mosquitoes in mainland Southeast Asia. *Evolutionary Applications* **3**: 319–339.

Hodges TK, Athrey G, Deitz KC, Overgaard HJ, Matias A, Caccone A, Slotman MA. 2013. Large fluctuations in the effective population size of the malaria mosquito Anopheles gambiae s.s. during vector control cycle. *Evolutionary Applications* **6**: 1171–1183.

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JMC, Wides R, et al. 2002. The genome sequence of the malaria mosquito Anopheles gambiae. *Science* **298**: 129–149.

Honório NA, Silva WDC, Leite PJ, Gonçalves JM, Lounibos LP, Lourenço-de-Oliveira R. 2003. Dispersal of Aedes aegypti and Aedes albopictus (Diptera: Culicidae) in an urban endemic dengue area in the State of Rio de Janeiro, Brazil. *Memórias do Instituto Oswaldo Cruz* **98**: 191–198.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Marcombe S, Darriet F, Agnew P, Etienne M, Yp-Tcha M-M, Yébakima A, Corbel V. 2011. Field Efficacy of New Larvicide Products for Control of Multi-Resistant Aedes aegypti Populations in Martinique (French West Indies). *Am J Trop Med Hyg* **84**: 118–126.

Nei M, Takahata N. 1993. Effective population size, genetic diversity, and coalescence time in subdivided populations. *J Mol Evol*.

Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, Loftus B, Xi Z, Megy K, Grabherr M, et al. 2007. Genome Sequence of Aedes aegypti, a Major Arbovirus Vector. *Science* **316**: 1718–1723.

Olanratmanee P, Kittayapong P, Chansang C, Hoffmann AA, Weeks AR, Endersby NM. 2013. Population Genetic Structure of Aedes (Stegomyia) aegypti (L.) at a Micro-Spatial Scale in Thailand: Implications for a Dengue Suppression Strategy ed. K. Hirayama. *PLoS Negl Trop Dis* **7**: e1913.

Palero F, Lopes J, Abelló P, Macpherson E, Pascual M, Beaumont MA. 2009. Rapid radiation in spiny lobsters (Palinurus spp) as revealed by classic and ABC methods using mtDNA and microsatellite data. *BMC Evol Biol* **9**: 263.

Pilger D, Lenhart A, Manrique Saide P, Siqueira JB, da Rocha WT, Kroeger A. 2011. Is routine dengue vector surveillance in central Brazil able to accurately monitor the Aedes aegypti population? Results from a pupal productivity survey. *Tropical Medicine & International Health* **16**: 1143–1150.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Ranson H, Burhani J, Lumjuan N, Black WC IV. 2010. Insecticide resistance in dengue vectors. *TropIKAnet* **1**: 0–0.

Rašić G, Endersby-Harshman N, Tantowijoyo W, Goundar A, White V, Yang Q, Filipović I, Johnson P, Hoffmann AA, Arguni E. 2015. Aedes aegypti has spatially structured and seasonally stable populations in Yogyakarta, Indonesia. *Parasit Vectors* **8**: 610.

Turissini DA, Gamez S, White BJ. 2014. Genome-Wide Patterns of Polymorphism in an Inbred Line of the African Malaria Mosquito Anopheles gambiae. *Genome Biol Evol* **6**: 3094–3104.

Vazquez-Prokopec GM, Kitron U, Montgomery B, Horne P, Ritchie SA. 2010. Quantifying the Spatial Dimension of Dengue Virus Epidemic Spread within a Tropical Urban Environment ed. D.J. Gubler. *PLoS Negl Trop Dis* **4**: e920.

Wang J. 2001. A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genetics Research* **78**: 243–257.

WHO. 2017. Dengue and severe dengue fact sheet. (http://www.who.int/mediacentre/factsheets/fs117/en/). Accessed 12 May 2017.

CHAPTER 2

Recombination Rate Mapping in the African Malaria Mosquito, *Anopheles gambiae*

ABSTRACT

Despite progress in reducing malaria transmission with insecticide-based vector control, the disease continues to claim nearly 500,000 lives per year, most of which are African children. A significant impediment to control in Africa is the presence *of Anopheles gambiae* – a uniquely efficient mosquito vector endemic to the continent. While great advancements have been made in the molecular biology and genetics of *An. gambiae* over the past decade, inadequate data on recombination rate in this species prevents novel and traditional vector control strategies from being deployed with maximum effectiveness. Recombination is a fundamental biological process with profound evolutionary implications. In mosquitoes and other sexual eukaryotes, recombination between homologous chromosomes is required for both the proper formation of haploid gametes from diploid germ cells and the production of new combinations of alleles. However, the rate at which recombination occurs varies with genomic position, sex, and the presence of chromosomal inversions. Such variation in recombination rate influences a myriad of evolutionary processes including the efficacy of natural selection, levels of standing diversity, and the elimination of deleterious mutations. Using a backcrossing strategy, I have developed high-resolution recombination rate maps for both male and female *An. gambiae* constructed using data from thousands of mosquitoes, each sequenced at thousands of loci. These maps indicate that there is a significant reduction in recombination in chromosomal inversions and that there are significant differences in recombination rate between male and female mosquitoes. The

maps developed in this study give us insight into some of the basic evolutionary processes that make *An. gambiae* such an efficient and adaptable vector. These maps can be used to aid in the development of novel vector control strategies, as well as to help predict of the efficacy of these vector control strategies.

INTRODUCTION

Malaria continues to be the most devastating insect-borne disease in the world, with 429,000 deaths and 212 million cases reported in 2015 (WHO, 2016). Mosquitoes in the genus *Anopheles* are the exclusive vectors of human malaria; however, not all *Anopheline* mosquitoes are equally competent vectors. Of the ~500 *Anopheles* species, only about two dozen are major vectors of malaria (Collins & Paskewitz, 1995). Within this small number of vectors, one species stands out as particularly dangerous: *Anopheles gambiae*. *An. gambiae*'s extraordinarily high vectorial capacity can be attributed to two important factors. First, it has been able to quickly adapt to, and thrive in, the different climactic conditions present in sub-Saharan Africa, allowing it to track human populations as they have expanded out of central Africa and across the continent (Ayala & Coluzzi, 2005). Second, *An. gambiae* has evolved to be in close contact with humans at every stage of its life cycle: the vast majority of its blood meals come from humans; while digesting a blood meal, it rests indoors for protection; and oviposition sites are most often water sources created by anthropogenic modification of the environment (Coluzzi, 1999; Simard et al., 2009). The presence of this vector in sub-Saharan African is one of the primary reasons that *Plasmodium falciparum* malaria, the deadliest form of the disease, continues to occur at such an exceptionally high rate in the region.

Sustainable control of *An. gambiae* with insecticides and other traditional techniques has proven difficult; therefore, vector biologists are looking to a number of new control strategies (e.g. transposons (James, 2005), homing endonucleases (Windbichler et al., 2011), Medea (Akbari et al., 2014), RIDL (Thomas et al., 2000)) that rely on introducing genetic elements into the mosquito genome. All of these approaches require that the transgene spread through the target population, with the aim of reducing population size, shortening the mosquito's lifespan, or making mosquitoes resistant to infection with *Plasmodium*. Understanding and modeling how these transgenes will spread through the population is a critical first step in determining whether a control strategy will be viable and effective. However, an important variable in such modeling remains poorly understood in *An. gambiae* – local and global recombination rate. Understanding recombination rates is critical for two reasons. First, recombination strongly influences the efficacy of selection, which in turn influences the likelihood and speed with which an introduced genetic element will spread through the target population (Connallon & Knowles, 2007; Haddrill et al., 2007; Roze & Barton, 2006; Presgraves 2005; Barton & Charlesworth, 1998; Begun & Aquadro, 1992). Additionally, many gene drive mechanisms require multiple genetic elements be inherited simultaneously, and recombination has the potential to break-down the linkage between the components necessary for proper inheritance. Therefore, sites with low or no recombination are often selected for genetic engineering, as the likelihood of breakdown via recombination is less likely in these regions; for this reason, proper identification of

recombination rate across the genome is hugely important. Second, quantitative information on recombination rate will greatly improve estimates of gene flow between populations (Bossart & Pashley Prowell, 1998; Hudson et al., 1992), which will be critical in predicting if and how transgenes will spread through the highly structured populations of *An. gambiae*. High-resolution recombination rate maps will also help direct researchers to areas of the genome where local recombination rates are best suited to their particular transgene approach.

In addition to the utility in vector control outlined above, a high-resolution recombination rate map will improve population genetics in *An. gambiae*. Recombination rate influences levels of standing variation (Nachman et al. 1998; Begun & Aquadro, 1992) , the "signature" left by natural selection (McGaugh et al., 2012; Nielsen, 2005), and the degree of linkage disequilibrium (Hartl, 1980), among myriad other genomic patterns. As a result, current population genetic studies in *An. gambiae* are severely limited and many assume no recombination (Weetman et al., 2012; White et al., 2011; Crawford & Lazzaro, 2010; Obbard et al., 2009). This has some important effects when it comes to vector control. First, precisely estimating gene flow rates between populations is not possible (Bossart and Pashley Prowell 1998; Hudson et al., 1992); this makes it very difficult to predict whether alleles for insecticide resistance, for example, or heritable genetic modifications will spread into other populations. Second, the absence of recombination rate data makes it impossible to predict the precise

genomic location of historical selection, the strength of that selection, and when it acted (Barrett & Schluter, 2008; Nielsen, 2005; Przeworski et al., 2005; Kim & Stephan, 2002; Sabeti et al., 2002). The ability to accurately predict these factors will give insight into the evolution of human blood preference, vector competence, and many other important traits (Garret-Jones & Shidrawi, 1969). Finally, estimating effective population size in the absence of recombination rate data is difficult. Without the ability to estimate effective population size, evaluating the efficacy of control methods aimed at reducing population size must rely on labor intensive methods (e.g. mark-recapture) or secondary indicators (e.g. number of malaria cases) that are prone to influence from factors other than mosquito population density.

I have generated the above-mentioned high-resolution recombination rate map for *An. gambiae* using a backcrossing scheme with lab-reared mosquito colonies and a genotyping-by-sequencing approach. My recombination maps include thousands of meioses and have allowed us to closely look at several factors that influence recombination rate. First, I have created separate maps for males and females, allowing us to compare local and global recombination between the sexes, which has previously been done only in humans (Kong et al., 2010; Coop et al., 2008). Differences in recombination rate between the sexes is widespread in both plants and animals, but remains a genetic enigma (Lenormand, 2003). By identifying regions of conservation and divergence in recombination rate between the sexes, we can begin to understand how

and why recombination rate differs between the sexes. Second, my crosses have used mosquito colonies that are fixed for one of two inversion karyotypes on chromosome 2L. It has long been known that inversions suppress recombination because single crossovers in heterokaryotypic regions result in unbalanced products of meiosis (Ashburner & Novitski, 1976). My crosses have allowed me to quantify the magnitude and spatial distribution of inversion-mediated recombination suppression, as well as explore the hypothesis that recombination suppression in inversions leads to increased recombination rates in collinear portions of the genome (Joyce & Mckim, 2010; Schultz & Redfield, 1951). The maps and analyses generated here will prove useful for vector biologists and evolutionary biologists alike.

MATERIALS AND METHODS

Mosquito rearing and crossing

I used three colonies of *An. gambiae* for crosses: the Mali-NIH strain, which is fixed for the 2La inversion, was used as one parent in all initial crosses; the second parent was from one of two SUA-strain colonies – one fixed for the 2La chromosomal arrangement and one fixed for the 2L+ chromosomal arrangement. Mosquitoes were reared in accordance with standard protocols (White et al., 2013). In brief, mosquitoes were maintained in insectaries under controlled conditions of 27°C, 65% relative humidity, and a 12h:12h light:dark cycle with 1 h dawn and dusk transitions. Larvae were reared in freshwater (dH2O) at a density of 200 larvae/L of water. Each larval tray was fed ~100

mg per day of a 4:1 mixture of finely ground fish pellets to baker's yeast. F1 crosses were performed with 200-400 virgin mosquitoes from each colony. Backcrosses were then performed with similar numbers of mosquitoes and female offspring from these backcrosses were used for sequencing and all downstream analyses.

ddRADseq library construction

In an effort to reliably genotype thousands of mosquitoes at a reproducible set of markers throughout the genome, I employed double-digest restriction-associated DNA sequencing (ddRADseq). Libraries were constructed largely following (Turissini, Gamez, & White, 2014). DNA was first extracted from adult mosquitoes using the ZR-96 Quick-gDNA kit (Zymo Research; Irvine, CA). Extracted DNA was digested for three hours at 37^{o}C with *MluC1* and *NlaIII* (New England Biolabs [NEB]; Ipswich, MA). The digested DNA was then purified using homemade 'Ampure' magnetic beads. A barcoded adapter (1 of 48) was then ligated to the sticky end produced by *NlaIII*, and a universal adapter was ligated to the sticky end produced by *MluC1* using T4 DNA ligase (NEB). Reactions from uniquely barcoded individuals were then pooled together, concentrated, and cleaned using two rounds of magnetic bead purification. These sub-libraries were size selected on a Blue Pippin (Sage Science; Beverly, MA) on a 1.5% gel cassette with the 'tight' setting and a 400 bp fragment target. Size-selected DNA from each sub-library was then PCR amplified using one of 12 indexed primers. Each PCR reaction consisted of 1x NEB Q5 PCR buffer, 10mM each DNTP, 20pmol of the universal and indexed primer, 0.25 U of

NEB Q5 DNA polymerase, and 4 µl of template. The reactions were carried out as follows: 98^{o}C for 1 minute; 12 cycles of 98^{o}C for 8 seconds, 68^{o}C for 20 seconds, and 72^{o}C for 20 seconds; and a final extension at 72^{o}C for 2 minutes. Following the PCR reaction, there are 576 (12 indexed primers x 48 barcodes) uniquely barcoded/indexed individuals that can be multiplexed for sequencing. This was done by first checking the quality and quantity of each of the twelve sub-libraries on a BioAnalyzer 2100 (Agilent Technologies; Santa Clara, CA), after which equal molar amounts of each sub-library were pooled to create a final sequencing library. Final libraries were then sequenced on the Illumina HiSeq 2500 (Illumina; San Diego, CA) at the UCR genomics core using 100 bp single-end reads.

Genotyping and ancestry reconstruction

I was able to successfully assign >98% of reads to individuals using barcode and adapter sequences and allowing up to one mismatch (all barcodes/adapters are separated by 3 mutational steps). Reads were then mapped to the *An. gambiae* PEST reference genome (AgamP4.2) using the Burrows-Wheeler aligner (BWA, Li & Durbin, 2009, v1.2.3), BWA-MEM algorithm with n=8 and otherwise default parameters. Following read mapping, BAM files for individuals from each cross were merged and genotypes were called using the Genome Analysis ToolKit (GATK, DePristo et al., 2011), with default parameters. I then extracted genotypes for individuals from the parental lines and identified single-nucleotide polymorphisms (SNPs) that had an F_{ST} of 1 and data for at least 16 individuals

from each parental line to use as markers for ancestry reconstruction. To correct for mapping error in repetitive regions, sites with coverage greater than 5 standard deviations from the mean coverage were filtered out on a per individual basis. Any SNPs matching any of the following criteria were filtered out prior to ancestry reconstruction: 1) frequency of less than 20% in backcross individuals, 2) the backcross parent allele was missing in an individual, 3) the non-backcross parental allele called as homozygous in >20% of individuals, and 4) x-linked sites where <90% of individuals were called as homozygous in crosses where F1 males were backcrossed to the female parental line of the original cross; and, in individuals with lower than expected heterozygous genotype calls, x-linked sites where <30% or >70% of individuals were called as heterozygous in crosses where F1 males were backcrossed to the male paternal line of the original cross – which would be expected to be completely heterozygous on the X chromosome.

As a result of variation in coverage from SNP to SNP and individual to individual, GATK genotype calls are not infallible. Therefore, I wrote a Hidden-Markov model (HMM) to impute ancestry and call recombination breakpoints in each individual (Andolfatto et al., 2011; Elliott, Aggoun, & Moore, 1995). A single SNP with the highest coverage from each RAD fragment was identified and used in the analysis in order to reduce noise. As a result of the design of my crosses, the HMM only had two states: homozygous for the backcross parent or heterozygous. The transition probability for the HMM was:

$$a = 2\times 10^{-8}(x_i - x_{i-1})e^{-2\times 10^{-8}(x_i - x_{i-1})}$$

where $x_i - x_{i-1}$ is the distance between neighboring SNPs. The emission probability was:

$$e = \sum_{i=0}^{n} n! p^i (1-p)^{n-i} \sum_{j=\max(0,k-i)}^{\min(k,n-i)} \frac{p_{ab}^{2j-k+i}(1-p_{ab})^{n+k-i-2j}}{j!(n-i-j)!(j-k+i)!(k-j)!}$$

where k is the read depth of the backcross parental allele, n is the total coverage, p is the expected allele frequency for the state (p = 1 for homozygous sites, p = 0.5 for heterozygous sites), and p_{ab} is the sequencing error rate between the two alleles (set to 0.01).

I filtered out individuals that had any chromosome with zero markers and individuals with less than 200 total markers. I called crossover breakpoints at the midpoint between two makers in which the probability of a given ancestry state dropped below 50%. After application of the HMM, multiple 'peaks' where ancestry quickly transitioned from one ancestry state to the other and back again over the space of a few SNPs were present. The expectation is that crossover interference should repress multiple crossovers over such a small distance (reviewed in (Hillers, 2004). As most 'peaks' were present in multiple individuals, I concluded that they were likely the result of mapping error, although some of them may have represented true gene conversion events. I identified SNPs in all such 'peaks' and filtered them out prior to ancestry map generation and recombination rate analysis. Recombination maps were then made by binning crossover

breakpoints into 1 Mb intervals and fitting the derivative of a spline to the cumulative map using a smoothing parameter of 1.

Statistical analysis

To determine significant differences in recombination rate between crosses, sexes, and chromosomes, I randomly sampled 100 individuals from each cross and recalculated recombination rate for each chromosome using those 100 individuals. This was repeated 10,000 times to generate a distribution of recombination rates for each cross/sex/chromosome combination. An ANOVA was then run, with recombination rate as the response variable and cross, sex, and chromosome – and the interactions between them – as fixed effects. I then computed Tukey's Honest Significant Difference for each pairwise comparison of cross/sex/chromosome. All statistical analyses were carried out in R (Team, 2016), v3.3.2).

RESULTS

Read mapping and SNP calling

I sequenced female offspring from a total of 4 backcrosses: Mali-NIH x SUA2La F1 males backcrossed to Mali-NIH (hereafter referred to as 2La/2La males), Mali-NIH x SUA2La F1 females backcrossed to Mali-NIH (2La/2La females), Mali-NIH x SUA2L+ F1 males backcrossed to Mali-NIH (2La/2L+ males), and Mali-NIH x SUA2L+ F1 males backcrossed to Mali-NIH (2La/2L+ females). I mapped an average of 254,788 reads per individual

across the four crosses (Figure 2.1, 2La/2La males: mean 241,488 reads and 1536 individuals; 2La/2La females: mean 196,845 reads and 576 individuals; 2La/2L+ males: mean 217,083 reads and 2208 individuals; and 2La/2L+ females: mean 349,967 reads and 1440 individuals). There was an average of 10,062 SNPs called per individual, with an average of 2,362 SNPs per individual passing filter (Figure 2.1, 2La/2La males: 8,449 SNPs per individual, 2,721 passing filter; 2La/2La females: 4,899 SNPs per individual, 1,496 passing filter; 2La/2L+ males: 10,964 SNPs per individual, 2,181 passing filter; 2La/2L+ females: 12,485 SNPs per individual, 2,494 passing filter).

Ancestry reconstruction and recombination rate mapping

In total, I constructed ancestry maps for 4,264 individuals (2La/2La males: 1,213; 2La/2La females: 315; 2La/2L+ males: 1,377; 2La/2L+ females: 1,359). Transitions in ancestry were used to create recombination rate maps and calculate map distance on a per chromosome, per cross basis. Statistical analyses showed that all pairwise comparisons of cross/sex/chromosome combinations were significantly different, except for the comparison of chromosome 3 in 2La/2L+ females to chromosome 3 in 2La/2La females and the comparison of chromosome 3 in 2La/2L+ males to chromosome 2 in 2La/2L+ females.

Chromosome 2

The average map distance for chromosome 2 across all four crosses was 75.5 centimorgans (cM), but varied substantially depending on the sex and inversion karyotype of each individual, with females having a higher recombination rate than males and individuals that were heterokaryotypic for the 2La inversion showing a lower average recombination rate across the chromosome than sex-matched individuals that were homokaryotypic for the inversion (Table 2.2, Figure 2.1). In the 2La/2La crosses, females had a 53.5% higher recombination rate than males at 0.962 centimorgans per megabase (cM/Mb) compared to 0.627 cM/Mb for males. In the 2La/2L+ crosses, recombination rate in females was 0.851 cM/Mb, 71% higher than the male recombination rate of 0.498 cM/Mb. The chromosome-wide decrease in recombination rate in the 2La/2L+ crosses was much more pronounced in males than in females (a 2.75-fold reduction in males vs. a 1.13-fold reduction in females).

Chromosome 3

As expected, chromosome 3 – lacking fixed chromosomal inversions in my crosses – showed a more uniform recombination rate between crosses, though rate differences between sexes in the same cross were still substantial (Table 2.2, Figure 2.2). The average map distance across all four crosses for chromosome 3 was 88.8 cM. In both sets of crosses, recombination rate in females was considerably higher than that of males; 2La/2La females had a recombination rate of 1.158 cM/Mb, 61% higher than the

males' recombination rate of 0.717 cM/Mb, while 2La/2L+ females had a recombination rate of 1.157 cM/Mb, which was 36% higher than the 0.852 cM/Mb recombination rate of males in the same cross. Interestingly, females had nearly identical recombination rates on chromosome 3, regardless of inversion karyotype on chromosome 2. Meanwhile, 2L inversion heterokaryotypic males showed a ~19% higher recombination rate than 2L inversion homokaryotypic males.

X Chromosome

X-Y recombination has been noted in *An. gambiae* (Hall et al., 2016); however, it occurs only periodically and my power to detect it without also increasing the number of false-positive recombination events annotated is limited. Additionally, I only sequenced females so do not have information on Y-linked alleles that could be used for detecting recombination between the X and Y chromosomes. Therefore, my comparison of X chromosome recombination is restricted to only females of each cross. The average map distance for the X chromosome is 47.4 cM. In 2La/2La females, the recombination rate was 1.913 cM/Mb, while in 2La/2L+ females, the recombination rate was 1.952 cM/Mb (Table 2.2, Figure 2.3).

DISCUSSION

Local and global patterns of recombination

In general, all four crosses showed a positive correlation between recombination rate and distance from the centromere, particularly in the absence of the confounding effects of chromosomal inversions, a pattern that is more pronounced in females. This correlation is especially pronounced on the X chromosome, which also has the highest rates of local recombination in the genome, with some regions showing rates >6 cM/Mb. In the absence of chromosomal inversions (i.e. on chromosome 3), there is remarkable similarity in the pattern of recombination between crosses of the same sex (i.e. 2La/2La males are remarkably similar to 2La/2L+ males, but less similar to either female cross), suggesting that in the absence of the effects of chromosomal architecture, sex may be a more important determinant of recombination patterns and rate. The presence of chromosomal inversions on one chromosome seems to have no influence on the patterns of recombination of other chromosomes, as evidenced by the similarity in chromosomal recombination patterns on chromosome 3 between crosses of the same sex. I do see a marked decrease in recombination rate in regions of inversion in inversion heterokaryotypes, though recombination suppression in these regions is not complete.

Sex differences in recombination patterns and rates

On both chromosome 2 and chromosome 3, females had significantly higher global recombination rates than males of the same cross type. However, the degree of difference between the sexes varied depending on chromosome and inversion karyotype. In 2La/2La crosses, the discrepancy between males and females was more pronounced on chromosome 3 than on chromosome 2 (61% and 53% greater recombination rate in females per chromosome, respectively); whereas in 2La/2L+ crosses, the discrepancy between male and female recombination rates was much higher on chromosome 2 than on chromosome 3 (71% and 36% greater recombination rate in females per chromosome, respectively). The reversal of these discrepancies may be the result of increased selection pressure on males relative to females to completely abolish recombination within the inversion. When recombination occurs within an inversion in males, 50% of sperm are affected by the aneuploidy that results. However, because females only produce one egg with three polar bodies during meiosis, it may be possible that females are able to shunt aneuploid cells to the polar body and use one of the two euploid cells as the ova. This selection pressure may “spillover” to the rest of chromosome 2, resulting in greater suppression of recombination chromosome-wide in heterokaryotypic males relative to heterokaryotypic females. The reduction in male recombination seen on chromosome 3 can not be explained by the same hypothesis; however, sex-specific reduced or abolished recombination occurs widely in nature and is nearly always in the heterogametic sex, a phenomenon known as the Haldane-Huxely

rule (Haldane, 1922; Huxley, 1928). There are a number of competing hypotheses to explain this effect (reviewed in Lenormand, 2003) and my data unfortunately do not lend support or contradict any of them. It seems that, at least for now, sexual dimorphism in recombination rates will remain an enigma.

The effect of inversions on local and global recombination

My results indicate that the presence of the 2La inversion can have a large effect on recombination chromosome-wide. Indeed, chromosome-wide recombination in inversion homokaryotypic males was 26% higher than in inversion heterokaryotypic males; in females, this number was 13%. Recombination within the inversion was not completely abolished in heterokaryotypic individuals, however, and in males recombination within the inversion of 2La/2L+ individuals was roughly 1/3 (32%) of what was seen in 2La/2La individuals, while in females 2La/2L+ individuals had a recombination rate in the inversion that was roughly 1/4 (24%) of what was seen in 2La/2La individuals. In most instances, the recombination events that I observe within the 2La inversion occur near the end of the inversion, where crossing over may not result in aneuploidy. In instances where I see recombination occurring in the "heart" of the inversion, it is most likely the result of gene conversion, as these tracks of switched ancestry are usually quite short and include only a few SNPs. Unfortunately, my markers are not dense enough to definitively say whether I am observing gene conversion or double crossover events within the inversion; however, the distances over which I see

recombination within the inversion seem to rule out double crossovers because crossover interference (reviewed in (Berchowitz & Copenhaver, 2010) would be expected to prevent double crossovers over the distances I observe.

It has been hypothesized that inversions on one chromosome will increase recombination rate on other chromosomes, as the result of increased time at the pachytene checkpoint (Joyce & Mckim, 2010; Schultz & Redfield, 1951; Washington & Bridges, 1919). My data lend support to this hypothesis in males only. There was no significant difference in recombination rate on chromosome 3 in females, regardless of inversion karyotype; however, 2La/2L+ males had a 19% higher recombination rate on chromosome 3 than 2La/2La males. 2La/2L+ females did have a significantly higher recombination rate on the X chromosome than 2La/2La females; however, the difference was only 2% and may be the result of random effects. The increased recombination seen on chromosome 3 in 2La/2L+ males may also be a compensatory effect for the lack of recombination on chromosome 2. If recombination increases offspring fitness, it stands to reason that a decrease in recombination in one region of the genome may exert selection pressure for the increase of recombination rate elsewhere in the genome; in this case, the increase in recombination rate on chromosome 3 in 2La/2L+ males may not just be the mechanistic result of the increase in time for chromosomes to line up during the pachytene checkpoint, and may actually be selectively advantageous.

Recombination on the X chromosome

In general, the pattern of recombination seen on the X chromosome is the same, regardless of cross. I did observe a higher recombination rate on the X in 2La/2L+ females relative to 2La/2La females, but again, this effect was modest. The recombination rate at the telomere of the X chromosome was the highest of any place in the genome, at >6 cM/Mb in some places. This huge recombination rate may be because X chromosome recombination can only occur in one sex (although X-Y recombination has been observed in *An. gambiae,* as mentioned above), so increased recombination on the X in females may compensate for that.

Conclusions

It remains to be seen whether the patterns of recombination seen in this study hold true for natural populations, or even other lab crosses. The data generated here can be used to quantify local differences between male and female recombination rates, especially whether male and female hotspots and cold spots are correlated or if large differences between the sexes exist. Additionally, it may be possible to correlate hot- and cold-spots with genes and gain a clearer picture of how male and female mosquitoes may respond differently to different vector control strategies. It is possible that differing selection pressures on males on females may manifest in disparities between areas of increased or reduced recombination in the two sexes. Given that there is recombination between

the *An. gambiae* X- and Y-chromosomes, it will also be important to measure the rate of this recombination and how that may affect the efficacy of vector control strategies. Further dissecting the basis of recombination in inversions is also an avenue of further exploration. It is not yet known how the size of a given inversion affects recombination within it. It also may be possible to engineer genetic drive elements to induce an inversion and thus reduce the likelihood that recombination will breakdown the drive element. This is one study in one species of mosquito; it is thus important to study the phenomenon of recombination in other mosquito species, as well. *Aedes aegypti*, for instance, has completely collinear (homomorphic) sex chromosomes (Nene et al., 2007), which will certainly have an effect on recombination in the sex chromosomes, but may also have an effect on genome-wide recombination, as well. Given the vector status of *Ae. aegypti*, elucidating patterns of recombination in this species and how the collinearity of its sex chromosomes influences these patterns is of utmost importance.

I have generated the first comprehensive, fine-scale recombination maps for *An. gambiae*. These tools are essential for vector control and population genetics alike. With accurate measures of recombination, it is now possible to accurately model the spread of important alleles through and between populations (i.e. engineered genetic elements for vector control, insecticide resistance alleles, etc.). These maps will also be useful for determining what signatures of selection look like in *An. gambiae*, measuring linkage disequilibrium in natural populations, and measuring levels of standing variation. The

maps can also be used to improve estimates of where in the genome historical selection has acted, the strength of that selection, and when it occurred. These are all important to provide insight into the evolution of vector competence, human blood preference, and additional phenotypes of medical importance in this species.

TABLES AND FIGURES

Table 2.1 Sequencing and SNP calling statistics

	Individuals Sequenced	Reads/individual	Raw SNPs/individual	SNPs/individual Passing Filter	Ancestry Maps Created
2La/2La Males	1,536	241,488	8,449	2,721	1,213
2La/2La Females	576	196,845	4,899	1,496	315
2La/2L+ Males	2,208	217,083	10,964	2,181	1,377
2La/2L+ Females	1,440	349,967	12,485	2,494	1,359

Table 2.1 Various sequencing and SNP calling statistics for all crosses. Note that the large discrepancy between the number of individuals sequenced and the number of ancestry maps created for the 2La/2L+ Males was the result of a larger number of individuals needing to be filtered out because of a contaminated library preparation in one of my sequencing runs.

Table 2.2 Recombination rate and map distance for each chromosome and cross

Chromosome

	2		**3**		**X**	
	Map Distance (cM)	Recombination Rate (cM/Mb)	Map Distance (cM)	Recombination Rate (cM/Mb)	Map Distance (cM)	Recombination Rate (cM/Mb)
2La/2La Males	69.5	0.627	68.3	0.717	N/A	N/A
2La/2La Females	106.7	0.962	110.2	1.158	46.7	1.913
2La/2L+ Males	55.2	0.498	81.0	0.852	N/A	N/A
2La/2L+ Females	94.3	0.851	110.1	1.157	47.6	1.952

Table 2.2 Recombination rate and map distance for each cross/sex/chromosome combination. All differences are significant except for 2La/2L+ females, chromosome 2 compared to 2La/2L+ males, chromosome 3; and 2La/2L+ females, chromosome 3 compared to 2La/2La females, chromosome 3; as determined by Tukey's test of Honest Significant Differences.

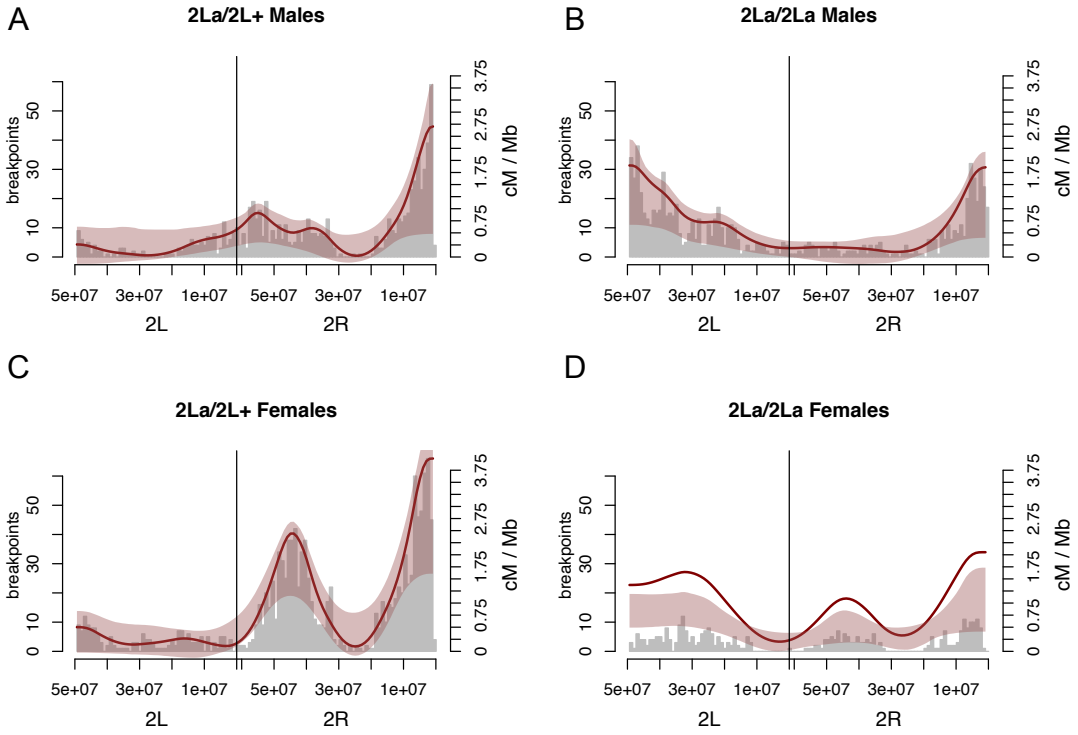

Figure 2.1 Recombination rate maps for chromosome 2 for all four crosses. A) 2La/2L+ males, B) 2La/2La males, C) 2La/2L+ females, and D) 2La/2La females. Red lines indicate recombination rate calculated in 1 Mb windows across the chromosome and red shading indicates the bounds between which recombination rate fell for 95% of subsampled datasets. Gray bars are histograms showing number of recombination breakpoints in 1 Mb bins across the chromosome. Vertical lines indicate the centromere, and numbers on the x-axis correspond to reference genome coordinates. Note that we sequenced many fewer individuals in the 2La/2La female cross, but we chose to use consistent axes for all four plots, explaining why the histogram bars for that cross are so much smaller than the other crosses.

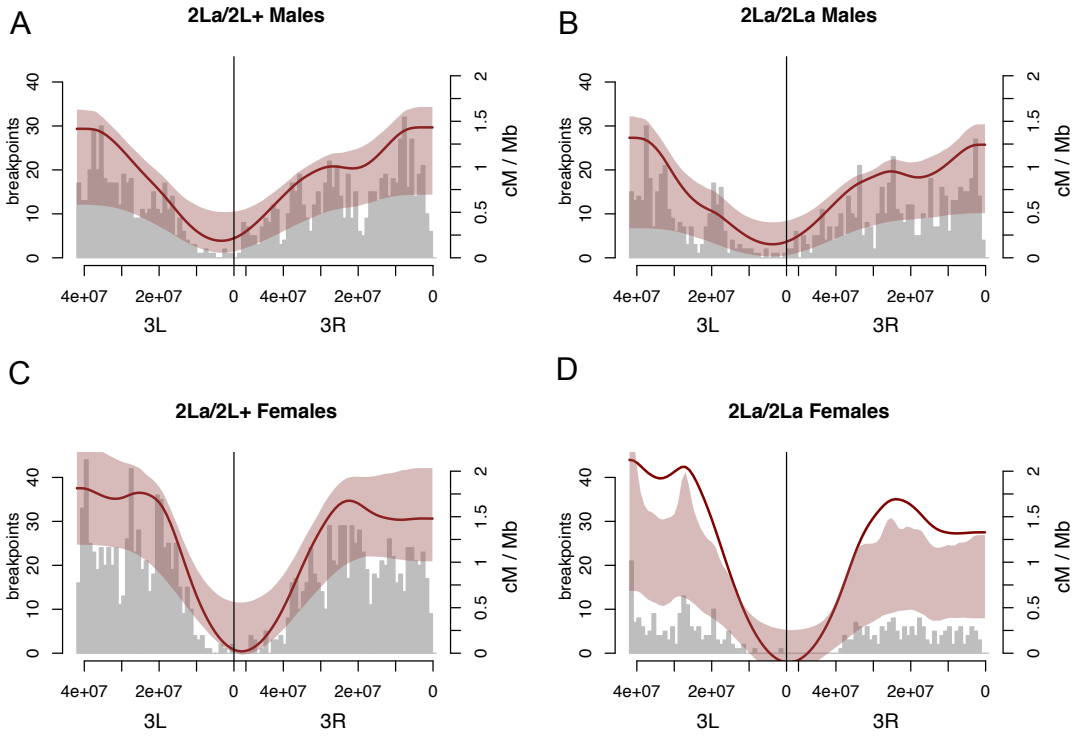

Figure 2.2 Recombination rate maps for chromosome 3 for all four crosses. A) 2La/2L+ males, B) 2La/2La males, C) 2La/2L+ females, and D) 2La/2La females. Red lines indicate recombination rate calculated in 1 Mb windows across the chromosome and red shading indicates the bounds between which recombination rate fell for 95% of subsampled datasets. Gray bars are histograms showing number of recombination breakpoints in 1 Mb bins across the chromosome. Vertical lines indicate the centromere, and numbers on the x-axis correspond to reference genome coordinates. Note that we sequenced many fewer individuals in the 2La/2La female cross, but we chose to use consistent axes for all four plots, explaining why the histogram bars for that cross are so much smaller than the other crosses.

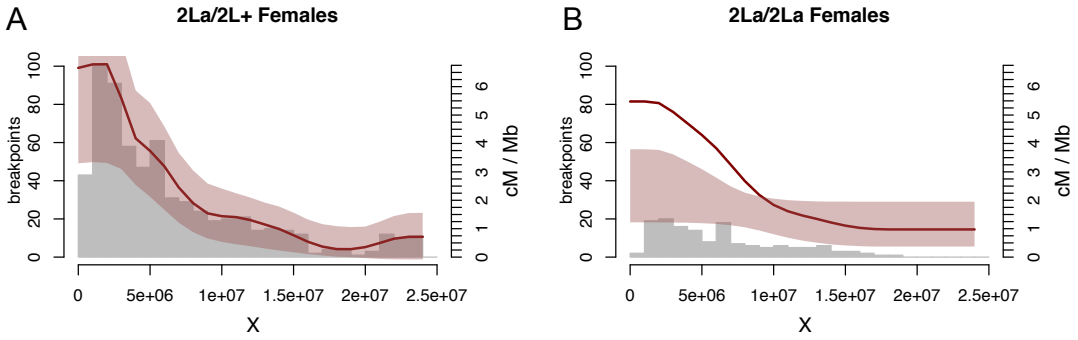

Figure 2.3 Recombination rate maps for the X chromosome for both female crosses. A) 2La/2L+ females and B) 2La/2La females. Red lines indicate recombination rate calculated in 1 Mb windows across the chromosome and red shading indicates the bounds between which recombination rate fell for 95% of subsampled datasets. Gray bars are histograms showing number of recombination breakpoints in 1 Mb bins across the chromosome. Because of how the reference genome is ordinated, the centromere at the right end of the x-axis and the telomere is at the left end of the x-axis. Note that we sequenced many fewer individuals in the 2La/2La female cross, but we chose to use consistent axes for both plots, explaining why the histogram bars for that cross are so much smaller than the other crosses.

REFERENCES

Akbari OS, Chen C-H, Marshall JM, Huang H, Antoshechkin I, Hay BA. 2014. Novel Synthetic Medea Selfish Genetic Elements Drive Population Replacement in Drosophila; a Theoretical Exploration of Medea-Dependent Population Suppression. *ACS Synth Biol* **3**: 915–928.

Andolfatto P, Davison D, Erezyilmaz D, Hu TT, Mast J, Sunayama-Morita T, Stern DL. 2011. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res* **21**: 610–617.

Ashburner M, Novitski E. 1976. *The Genetics and biology of Drosophila*. London; New York: Academic Press.

Ayala FJ, Coluzzi M. 2005. Chromosome speciation: Humans, Drosophila, and mosquitoes. *Proc Natl Acad Sci U S A* **102**: 6535–6542.

Barrett RDH, Schluter D. 2008. Adaptation from standing genetic variation. *Trends Ecol Evol (Amst)* **23**: 38–44.

Barton NH, Charlesworth B. 1998. Why sex and recombination? *Science* **281**: 1986–1990.

Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. *Nature* **356**: 519–520.

Berchowitz LE, Copenhaver GP. 2010. Genetic Interference: Don't Stand So Close to Me. *CG* **11**: 91–102.

Bossart JL, Pashley Prowell D. 1998. Genetic estimates of population structure and gene flow: Limitations, lessons and new directions. *Trends Ecol Evol (Amst)* **13**: 202–206.

Collins FH, Paskewitz SM. 1995. Malaria: current and future prospects for control. *Annu Rev Entomol* **40**: 195–219.

Coluzzi M. 1999. The clay feet of the malaria giant and its African roots: hypotheses and inferences about origin, spread and control of Plasmodium falciparum. *Parassitologia* **41**: 277–283.

Connallon T, Knowles LL. 2007. Recombination rate and protein evolution in yeast. *BMC Evol Biol* **7**: 235.

Coop G, Wen X, Ober C, Pritchard JK, Przeworski M. 2008. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science* **319**: 1395–1398.

Costantini C, Ayala D, Guelbeogo WM, Pombi M, Some CY, Bassole IH, Ose K, Fotsing J-M, Sagnon N, Fontenille D, et al. 2009. Living at the edge: biogeographic patterns of habitat segregation conform to speciation by niche expansion in Anopheles gambiae. *BMC Ecol* **9**: 16.

Crawford JE, Lazzaro BP. 2010. The Demographic Histories of the M and S Molecular Forms of Anopheles gambiae s.s. *Mol Biol Evol* **27**: 1739–1744.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**: 491–498.

Elliott RJRJ1, Aggoun L, Moore JBJB1. 1995. *Hidden Markov models: estimation and control*. New York: Springer-Verlag.

Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* **78**: 737–756.

Garrett-Jones C, Shidrawi GR. 1969. Malaria Vectorial Capacity of a Population of Anopheles Gambiae - an Exercise in Epidemiological Entomology. *Bulletin of the World Health Organization* **40**: 531–.

Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the Drosophila genome that lack crossing over. *Genome Biol* **8**: R18.

Haldane J. 1922. Sex ratio and unisexual sterility in hybrid animals. *Journal of genetics*.

Hall AB, Papathanos PA, Sharma A, Cheng C, Akbari OS, Assour L, Bergman NH, Cagnetti A, Crisanti A, Dottorini T, et al. 2016. Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes. *Proc Natl Acad Sci U S A* **113**: E2114–23.

Hartl DL. 1980. *Principles of population genetics*. Sunderland, Mass.: Sinauer Associates.

Hillers KJ. 2004. Crossover interference. *Curr Biol* **14**: R1036–7.

Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.

Huxley JS. 1928. Sexual difference of linkage in Gammarus chevreuxi. *Journal of genetics*.

James AA. 2005. Gene drive systems in mosquitoes: rules of the road. *Trends in Parasitology* **21**: 64–67.

Joyce EF, Mckim KS. 2010. Chromosome Axis Defects Induce a Checkpoint-Mediated Delay and Interchromosomal Effect on Crossing Over during Drosophila Meiosis ed. G.P. Copenhaver. *PLoS Genet* **6**.

Kim Y, Stephan W. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160**: 765–777.

Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, Walters GB, Jonasdottir A, Gylfason A, Kristinsson KT, et al. 2010. Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**: 1099–1103.

Lenormand T. 2003. The evolution of sex dimorphism in recombination. *Genetics* **163**: 811–822.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

McGaugh SE, Heil CSS, Manzano-Winkler B, Loewe L, Goldstein S, Himmel TL, Noor MAF. 2012. Recombination Modulates How Selection Affects Linked Sites in Drosophila ed. N.H. Barton. *PLoS Biol* **10**: e1001422.

Nachman MW, Bauer VL, Crowell SL, Aquadro CF. 1998. DNA Variability and Recombination Rates at X-Linked Loci in Humans. *Genetics* **150**: 1133–1141.

Nene, V., Wortman, J. R., Lawson, D., Haas, B., Kodira, C., Tu, Z. J., et al. (2007). Genome Sequence of Aedes aegypti, a Major Arbovirus Vector. *Science*, *316*(5832), 1718–1723.

Nielsen R. 2005. Molecular signatures of natural selection. *Annu Rev Genet* **39**: 197–218.

Obbard DJ, Welch JJ, Little TJ. 2009. Inferring selection in the Anopheles gambiae species complex: an example from immune-related serine protease inhibitors. *Malar J* **8**: 117–117.

Presgraves DC. 2005. Recombination enhances protein adaptation in Drosophila melanogaster. *Curr Biol* **15**: 1651–1656.

Przeworski M, Coop G, Wall JD. 2005. The signature of positive selection on standing genetic variation. *Evolution* **59**: 2312–2323.

R Core Team. 2016. *R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014*.

Roze D, Barton NH. 2006. The Hill-Robertson effect and the evolution of recombination. *Genetics* **173**: 1793–1811.

Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**: 832–837.

Schultz J, Redfield H. 1951. Interchromosomal Effects on Crossing Over in Drosophila. *Cold Spring Harb Symp Quant Biol* **16**: 175–197.

Simard F, Ayala D, Kamdem GC, Pombi M, Etouna J, Ose K, Fotsing J-M, Fontenille D, Besansky NJ, Costantini C. 2009. Ecological niche partitioning between Anopheles gambiae molecular forms in Cameroon: the ecological side of speciation. *BMC Ecol* **9**: 17.

Thomas DD, Donnelly CA, Wood RJ, Alphey LS. 2000. Insect population control using a dominant, repressible, lethal genetic system. *Science* **287**: 2474–2476.

Turissini DA, Gamez S, White BJ. 2014. Genome-Wide Patterns of Polymorphism in an Inbred Line of the African Malaria Mosquito Anopheles gambiae. *Genome Biol Evol* **6**: 3094–3104.

Washington ClO, Bridges CB. 1919. *Contributions to the Genetics of Drosophila Melanogaster...*

Weetman D, Wilding CS, Steen K, Pinto J, Donnelly MJ. 2012. Gene Flow–Dependent Genomic Divergence between Anopheles gambiae M and S Forms. *Mol Biol Evo* **29**: 279-291.

White BJ, Kundert PN, Turissini DA, Van Ekeris L, Linser PJ, Besansky NJ. 2013. Dose and developmental responses of Anopheles merus larvae to salinity. *Journal of Experimental Biology* **216**: 3433–3441.

White BJ, Lawniczak MKN, Cheng C, Coulibaly MB, Wilson MD, Sagnon N, Costantini C, Simard F, Christophides GK, Besansky NJ, et al. 2011. Adaptive divergence between incipient species of Anopheles gambiae increases resistance to Plasmodium. *Proc Natl Acad Sci U S A* **108**: 244–249.

Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, Ulge UY, Hovde BT, Baker D, Monnat RJ, Burt A, et al. 2011. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* **473**: 212–215.

WHO. 2016. World Malaria Report.

CHAPTER 3

Structural Variation at the TEP Anti-Pathogen Locus in African Malaria Mosquitoes

ABSTRACT

Several species of mosquito in the *Anopheles gambiae* species complex are highly efficient vectors of malaria in sub-Saharan Africa. Two of these species, *An. coluzii* and *An. gambiae*, have only recently been elevated to the level of separate species, and were previously thought to be separate forms of the same species. Genome scans have revealed a region of high divergence on the left arm of the third chromosome (3L), containing at least 6 genes of the thioester-containing protein (TEP) family, which have been implicated in the innate immune response. Interestingly, *An. coluzzii* populations from Mali and Burkina Faso are fixed for an allele at *TEP1* – r^B, not present in sympatric *An. gambiae* populations – that confers additional resistance to the malaria parasite. To annotate the TEP Anti-Pathogen Locus (TAPL) and identify structural variation among the three known haplotypes – r^B, r^A, and S – in the TAPL region, BAC clones of each of the three haplotypes have been sequenced and aligned them to each other. These alignments indicate several significant structural differences between the three haplotypes, including a duplication in one haplotype that has created a chimeric *TEP* gene.

INTRODUCTION

There were an estimated 212 million malaria cases and 429,000 deaths in 2015 (WHO, 2016). Sub-Saharan Africa accounted for nearly 83% of cases and >90% of deaths, largely due to the presence of vectors in the *Anopheles gambiae* species complex. This species complex consists of at least eight morphologically indistinguishable species. The genomic differentiation between two of these species, *An. gambiae* and *Anopheles coluzzii*, has been studied in depth (cf. Fontaine et al. 2015; Crawford et al. 2012; White et al. 2011; Lawniczak et al. 2010; White et al. 2010), partially owing to the fact that they were originally considered two molecular forms of *An. gambiae sensu stricto*, having only recently been raised to the status of separate species (Coetzee et al. 2013). One of these studies (White et al. 2010), revealed four regions of high divergence between *An. coluzzii* and *An. gambiae* (previously termed the M and S forms, respectively): three pericentromeric "speciation islands" and a region on the left arm of chromosome three (3L). Interestingly, the region on 3L contains *TEP1*, a gene that has been shown to play a key role in mediating the mosquito's response to infection with the *Plasmodium* parasite (Blandin et al. 2004). This region also contains five other genes in the thioester-containing protein (TEP) family; because of the number of *TEP* genes and the role that *TEP1* plays in mosquito immunity, I have termed this region the TEP anti-pathogen locus (TAPL). Population sampling in across Africa revealed the presence of three alleles of *TEP1* in wild populations; two increase resistance to *Plasmodium* (*TEP1r*^{A} and *TEP1r*^{B}), while mosquitoes harboring the third allele remain susceptible to

infection by the parasite (*TEP1s*) (White et al. 2011). The *TEP1r*^{B} allele has been swept to fixation in *An. coluzzii* populations in Mali and Burkina Faso, while sympatric *An. gambiae* populations show no trace of the allele, despite appreciable rates of hybridization between the two species (Choi and Townson 2012; Etang et al. 2009; Djogbenou et al. 2008). The selective sweep for *TEP1r*^{B} in Mali and Burkina Faso likely affected the makeup of the flanking regions via genetic hitchhiking. To determine the factors affecting evolution of this region, as well as structural differences between haplotypes bearing the three *TEP1* alleles, I sequenced bacterial artificial chromosomes (BACs) consisting of the three haplotypes. I identified structural (insertions/deletions) and sequence differences (SNPs) between the haplotypes. I also mapped transposable element (TE) sequences onto them, as the movement of TEs may play an important role in genome evolution.

MATERIALS AND METHODS

Screening of BAC libraries

A 334 bp sequence from *TEP3* that showed 100% homology to Mali-NIH strain scaffolds and 99.7% homology to Pimperena strain scaffolds and was unique in the genome was sent to the Clemson University Genomics Institute for probe design and screening of BAC libraries AGMCBa and AGSCBb. Liquid stocks of clones that the probe successfully hybridized to were then ordered. Clones of the ND-1 and ND-TAM libraries were

identified via their end-sequencing mapping locations on VectorBase and subsequently ordered from Notre Dame.

BAC isolation and sequencing library preparation

Clones were grown in liquid LB media with chloramphenicol antibiotic to select for cells carrying the BAC. These were subsequently streaked onto LB-agar plates, again with chloramphenicol selection. To reduce variation in the sequencing populations, single colonies were picked from these plates and then grown in liquid LB media overnight. Following overnight growth, BACs were isolated from the *E. coli* cells with the Zymo ZR BAC DNA Miniprep Kit (Zymo Research, Irvine, CA, USA) using standard protocols. Library preparation and multiplexing for the Illumina MiSeq runs was carried out using the Illumina Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA). DNA to be sequenced on the PacBio RS II system was sheared to ~10kb using a G-Tube (Covaris, Woburn, MA, USA), after which blunt adapters, then sequencing primers were ligated to create SMRTbell templates for sequencing.

Sequence assembly and analysis

Contaminating bacterial and vector sequence was filtered out of the resulting sequence reads by mapping the reads to the vector and genome sequence using BWA (Li and Durbin 2009). Preliminary assembly of the resulting Illumina reads was carried out using SGA (Simpson and Durbin 2012). The PacBio reads were corrected by using Proovread

(Hackl et al. 2014) to align Illumina reads to the PacBio reads, which were split into smaller blocks to allow for higher coverage of Illumina reads, increasing the accuracy of correction. The resulting corrected PacBio reads, together with the Illumina reads, were then assembled using PBcR (Koren et al. 2012). The resulting handful of scaffolds for each haplotype were stitched together and aligned by hand, based on colinearity and sequence similarity. Polymorphisms and indels in the alignment were called using a custom Perl script (script can be found at: https://github.com/esmith1032/tep_assembly). Transposable element sequences were then downloaded from RepBase and mapped to the alignments using BLAST. To determine the prevalence of the chimeric gene in wild populations, the .vcf file of chromosome 3L was downloaded from the *Anopheles gambiae* 1000 genomes project (The *Anopheles gambiae* 1000 Genomes Consortium, 2015) and parsed for coverage and informative SNPs using custom Perl scripts (scripts can be found at: https://github.com/esmith1032/tep_assembly). d_S between the chimeric *TEP1*/*TEP3*, *TEP1*, and *TEP3* was calculated using the Yang-Nielsen algorithm, as implemented in PAML (Yang 2007) . All other statistical analyses were carried out in R.

RESULTS

The screen of two BAC libraries at the Clemson University Genomics Institute (CUGI) resulted in 60 BAC clones that were positive for the probe, designed to target *TEP3*. Of these 60 clones, 45 were from the *An. coluzzii* Mali-NIH library (AGMCBa) and the other

15 were from the *An. gambiae* Pimperena library (AGSCBb). An additional 24 BACs from libraries housed at Notre Dame (ND-TAM and ND-1) were identified as potentially carrying sequence from the TAPL region as a result of them being used for BAC end sequencing and subsequent mapping (Hong et al. 2003). These BACs were then sequenced using 2x250 reads on the Illumina MiSeq (Table 3.1). After initial assembly with SGA (Simpson and Durbin 2012), it was determined that all BAC clones from the CUGI *An. coluzzii* library contained the *TEP1r*^{B} allele, while all BACs from the CUGI *An. gambiae* library contained the *TEP1r*^{A} allele. Of the BACs from the Notre Dame libraries that I was able to positively genotype at *TEP1*, one contained the *TEP1s* allele, while the rest were *TEP1r*^{A} or *TEP1r*^{B}. My initial BAC assemblies using short reads only proved to be highly fragmented, so I chose the two best assembled BACs of each allele type – except in the case of *TEP1s*, where I only had one clone – and sequenced those on the PacBio RSII system (Table 3.2). I corrected the PacBio reads using Proovread (Hackl et al. 2014) and then performed a hybrid assembly using PBcR (Koren et al. 2012) to incorporate both the Illumina short reads and the PacBio long reads to make a consensus sequence for each of the three haplotypes (hereafter r^{A} for *TEP1r*^{A}, r^{B} for *TEP1r*^{B}, and s for *TEP1s*). The consensus sequences for the r^{A} and r^{B} haplotypes extend across all of the genes in the TAPL region, while the s haplotype consensus sequence extends only through *TEP1* and stops just upstream of *TEP3*, missing sequence for *TEP3, AGAP010817, TEP11,* and *TEP10*. The r^{A} sequence is 132,286 bp, the r^{B} sequence is 191,169 bp, and the s sequence is 136,210 bp. I was able to determine the endpoint of

TEP18, previously masked by Ns in the PEST reference sequence. I also used this multi-sequence alignment to call insertions/deletions (indels) and polymorphisms between the three haplotypes (Figure 3.1). Most indels were relatively small – 88% were less than 100bp and 71% were less than 10bp – but several were substantial in size. Two of these are sequences present in the s haplotype that are not present in either of the other two haplotypes or the reference, spanning 6,387 bp and 4,399 bp. The 3' half of the smaller of these two insertions shows some homology to *An. gambiae CLIPC9*, encoded on chromosome 2L, as well as some other, unannotated, genes throughout the reference genome. Aside from this genic homology, these two sequences are homologous to several non-protein coding regions scattered about the reference genome. Additionally, there is a 2,152 bp insertion present in all three haplotype consensus sequences that is not present in the reference and a 2,135 bp sequence present in the reference genome that is not present in any of the three consensus sequences. There are another two sequences that are present in only the r^B haplotype, but not in r^A or PEST (I don't have sequence for the s haplotype in this region), totaling 5,204 bp. The remainder of the large (>1kb) indels are shared by the reference and at least one other haplotype, save for one 10,537 bp insertion into the r^A haplotype that will be discussed later.

Several of the indels correspond to transposable element (TE) sequence, so I took the list of known *An. gambiae* TEs from RepBase and mapped those onto the sequence alignment. The most abundant TE sequences were fragments of non-long terminal

repeat retrotransposons in the RTE-1 family. These fragments ranged from 68 to 851 bp and showed >88% nucleotide identity to the haplotype consensus sequences. There are five fragments present in both the r^A and r^B haplotypes, and four fragments present in the s haplotype. The most abundant complete – or nearly so – TE sequences were short, interspersed nuclear element (SINE) X1, though the number per haplotype varied quite dramatically. There were seven copies of SINE-X1 in the r^A haplotype (three more than are in the reference in the TAPL region), four copies in the r^B haplotype, and just two in the s haplotype (this could result from the fact that the s sequence doesn't cover all of TAPL). The region is also littered with fragments of TEs from the chicken repeat 1 family of retrotransposons. These fragments range in size from just 72 bp to 4,323 bp. There are eight of these fragments in the r^B haplotype, six in the s haplotype, and, interestingly, just one in the r^A haplotype. Fragments of TEs from several other families are scattered throughout the region, as well (Table 3.3).

Perhaps the most interesting finding of the sequence comparisons was the presence of a novel gene between *TEP1* and *TEP3* in the r^A haplotype that appears to be a chimera of those two genes and represents ~50% of a ~10kb insertion in the r^A haplotype. This chimera is not present in the r^B haplotype or the reference, and I don't have sequence from the s haplotype to determine whether it is present in that consensus. The chimera is homologous to *TEP1* on the 5' half of the coding sequence and homologous to *TEP3* on the 3' half (the gene is encoded on the negative strand). Between the two BACs that

were pooled to call the r^A consensus sequence, more than 300 PacBio reads contain the entire sequence of the chimera, ruling out the possibility that the chimera resulted from a few chimeric reads being assembled in a low-coverage assembly. The junction between *TEP3* and *TEP1* in the chimera occurs in the 8^{th} exon of each gene, where there is a 54 bp deletion, resulting in a loss of 18 amino acids. To determine how divergent the chimera is, I calculated d_S for 21 non-overlapping windows spanning the alignments between *TEP1*, *TEP3*, and the chimera (Figure 3.2). Average d_S between the homologous regions of the *TEP1* and *TEP3* and the chimera were 0.06 and 0.05, respectively, while d_S in the non-homologous regions averaged >3 and 1.4 for *TEP1* and *TEP3*, respectively. For comparison, d_S between *TEP1* and *TEP3* averaged 1.6 across the entire region. To further validate the existence of this chimera, I called polymorphic sites between the chimera and homologous regions of *TEP1* and *TEP3* and used that information to preliminarily classify individuals from the *Anopheles gambiae* 1000 Genomes Project (The *Anopheles gambiae* 1000 Genomes Consortium, 2015) as either containing the chimera or not. I then looked at the ratio of coverage in the regions of *TEP1* and *TEP3* that are homologous to the chimera to coverage in the non-homologous regions of these genes. Since the chimera essentially represents a second copy of the homologous portions of these genes, I would expect twice as many reads to map to the homologous portions than map to the non-homologous portions. I do, in fact, see significantly elevated coverage in the *TEP1*/*TEP3* homologous portions of these genes in individuals that polymorphism data indicated may be harboring the chimera (Mann-Whitney U, W =

32923, p < 2.2e-16, Figure 3.3). Furthermore, individuals were classified according to their *TEP1* genotype; polymorphism and coverage data indicate that the chimera is present in some individuals of the s haplotype.

DISCUSSION

Despite the high sequencing coverage obtained, TAPL proved to be a difficult region to assemble, possibly due to the presence of numerous TEs containing repetitive sequence. Large structural differences exist between the three haplotypes, which may have played an important role in the past evolution of the region and may play an important role going forward. Insertions between genes may act to increase the recombination rate between those two genes, which could result in a higher probability that advantageous alleles at each gene get inherited together, providing an increase in fitness greater than either of the two alleles being inherited alone. If the proteins manufactured from those genes interact, recombination may result in incompatible alleles at the two loci, decreasing the fitness of the individual harboring them. An increase in the amount of sequence due to insertion may also provide the raw material necessary for regulatory regions to arise, or for existing regulatory regions to get duplicated and modified, resulting in finer scale regulatory control and more fine-tuning of gene expression, which may also increase fitness via less energy waste manufacturing proteins that aren't needed.

The large number of TE sequences present in the region also likely played an important role in its evolution. Movement of TEs can be very detrimental, particularly if they insert in genes or regulatory regions and disrupt function of those genes. Transposases may also target the wrong sequence and move genes to areas under different regulatory control. Given the high number of TE sequences in TAPL, it is possible that some of the genes in the region originated elsewhere in the genome and were moved via this mechanism, or that genes previously in this region were exported to a different region. The discovery of the *TEP1*/*TEP3* chimera illustrates the importance or long read sequencing technologies going forward. Because of the homology of regions of this gene to both *TEP1* and *TEP3*, short reads map to either of those two genes, or reads that span the junction between the two may be thrown out as mapping error. The chimera itself could prove to be important in immune function of mosquitoes harboring it. The chimera contains the active site (thioester domain, or TED region) of *TEP3*, but a majority of the rest of the gene is homologous to *TEP1*, indicating that it may target pathogens similar to *TEP3* (Mitri et al. 2015) but interact with other proteins in the manner of *TEP1* (Fraiture et al. 2009). It remains unclear whether this gene is expressed, however, so assessing its impact on the mosquito is difficult at this time. It is also unclear how this gene arose, but two possibilities initially come to mind: 1) asymmetric recombination such that the breakpoint on one chromosome was in the middle of *TEP1* and the breakpoint on the other chromosome was in the middle of *TEP3*. This would result in the one chromosome having functional copies of *TEP1*, *TEP3*, and the

TEP1/*TEP3* chimera (as seen in the r^A haplotype) and one chromosome having only the chimera and no *TEP1* or *TEP3*; 2) retrotransposition of the region spanning from the middle of *TEP1* to the middle of *TEP3*, with insertion into the middle of *TEP3*, resulting the pattern seen in the r^A haplotype. Regardless of how this gene arose, there is evidence that *TEP1* arose as a chimera of two other *TEP* genes (*TEP5* and *TEP6*) (Obbard et al. 2008), indicating that this may be a more common mechanism for gene creation than previously thought.

This study provides the foundation for further functional study of the genes in TAPL to elucidate their importance in *Anopheles* immunity. Using the data generated here, it would be possible to develop a PCR screen to determine whether a lab colony of *An. gambiae* contains the chimeric *TEP1/TEP3* gene. Once a colony harboring this gene is identified, further study could be done to determine whether or not the gene is expressed, and if it is expressed, determine its effect of mosquito biology via knockout or knockdown. Likewise, the remaining TEP genes in this region that have not been functionally characterized can be interrogated in a similar manner to determine their function. Additionally, it should be possible to determine the age of the chimeric gene to determine when it arose and if that corresponds to anything important in the demographic history of this species (i.e. *gambiae*/*coluzzii* split, migration into a new environment, etc.) Gaining a fuller picture of mosquito immunity and response to *Plasmodium* will be important in developing malaria control strategies, as will

understanding the factors affecting the evolution of these important immune genes. This study also demonstrates the key role that long read sequencing technologies will play going forward. Had I not used long read technology, the chimeric *TEP1*/*TEP3* would have remained undiscovered.

TABLES AND FIGURES

Table 3.1. Illumina MiSeq sequencing statistics

Library	Estimated BAC size (kb)	Average number of reads	Average sequence (kb)	Estimated average coverage
AGMCBa	110	33,100	8,275	75x
AGSCBb	100	55,520	13,880	138x
ND-TAM	133	467,532	116,883	879x
ND-1	120	393,011	98,253	819x

Table 3.1 Illumina MiSeq Sequencing Statistics Estimated BAC size (kb), average number of reads, average total sequence (kb), and estimated average coverage for the four libraries sequenced on the Illumina MiSeq.

Table 3.2. PacBio RSII sequencing statistics

Sample	Estimated size (kb)	Number of reads	Average read length (bp)	Maximum read length (bp)	Estimated coverage
r^A1	100	60,990	6,283	39,760	3,832x
r^A2	100	87,439	6,321	41,205	5,527x
r^B1	110	107,855	5,866	38,423	5,752x
r^B2	110	71,350	5,524	40,883	3,583x
s	120	121,165	6,938	53,953	7,005x

Table 3.2 PacBio RSII Sequencing Statistics Estimated size (kb), number of reads, average read length (bp), maximum read length (bp), and estimated coverage for the five samples that were sequenced on the PacBio RSII.

Table 3.3 Percentage of haplotype sequence to which known transposable element sequences map			
	Percentage of haplotype sequence		
Transposable Element	r^A	r^B	s
Ag-CR1-7_CR1_Anopheles_gambiae_str._PEST	0.0000	2.5182	0.0000
Ag-Jock-12_Jockey_Anopheles_gambiae	0.0461	0.0319	0.0000
AgaP8_P_Anopheles_gambiae_str._PEST	0.4135	0.2469	0.4354
Clu-111_AG_DNA_transposon_Anopheles_gambiae_str._PEST	1.8505	0.0000	0.0000
Clu-13_AG_DNA_transposon_Anopheles_gambiae_str._PEST	0.6690	0.8882	0.0000
Clu-15B_AG_DNA_transposon_Anopheles_gambiae_str._PEST	0.0000	0.0000	0.3370
Clu-15_AG_DNA_transposon_Anopheles_gambiae_str._PEST	0.0000	0.0000	0.5425
Clu-47C_AG_DNA_transposon_Anopheles_gambiae_str._PEST	0.6940	0.0000	0.0000
Clu-87A_AG_DNA_transposon_Anopheles_gambiae_str._PEST	0.0000	0.0000	0.7063
Clu-87B_AG_DNA_transposon_Anopheles_gambiae_str._PEST	0.0000	0.0000	0.5242
CR1-2_AG_CR1_Anopheles_gambiae_str._PEST	0.0000	0.0000	2.1269
CR1-6_AG_CR1_Anopheles_gambiae_str._PEST	0.0000	0.1742	0.1916
CR1-7_AG_CR1_Anopheles_gambiae_str._PEST	0.3379	1.6075	0.1351
HATN2_AG_hAT_Anopheles_gambiae_str._PEST	0.2139	0.0361	0.2114
HATN3_AG_hAT_Anopheles_gambiae_str._PEST	0.1580	0.0000	0.0000
HELITRON2N_AG_Helitron_Anopheles_gambiae_str._PEST	0.2910	0.2218	0.1424
HELITRON2_AG_Helitron_Anopheles_gambiae_str._PEST	0.3402	0.2558	0.1424
INVADER1-LTR_AG_Gypsy_Anopheles_gambiae_str._PEST	0.0000	0.0000	0.1490
Mariner-N16_AG_Mariner/Tc1_Anopheles_gambiae_str._PEST	0.7454	0.3819	0.5257
MARINERN7_AG_Mariner/Tc1_Anopheles_gambiae_str._PEST	0.2941	0.2035	0.0000
P4_AG_hAT_Anopheles_gambiae_str._PEST	0.1406	0.0000	0.1402
RETRO18_AG_LTR_BEL_Anopheles_gambiae	0.0000	0.0000	0.3950
RETRO33_AG_LTR_BEL_Anopheles_gambiae	0.0000	0.2427	0.0000
RTE-1_AG_RTE_Anopheles_gambiae_str._PEST	1.6321	1.0818	1.7752
RTE-2_AG_RTE_Anopheles_gambiae_str._PEST	0.3961	0.2741	0.0000
SINEX-1_AG_SINE_Anopheles_gambiae_str._PEST	0.9525	0.3604	0.2875
SINEX-2_AG_SINE2/tRNA_Anopheles_gambiae_str._PEST	0.0000	0.1004	0.1410
TC1N-1_AG_Mariner/Tc1_Anopheles_gambiae_str._PEST	0.0000	0.0000	0.2481
Total sequence to which known TEs map	**9.1749**	**8.6254**	**9.1569**

Table 3.3 Percentage of haplotype sequence to which known transposable element sequences map The names of the transposable elements are presented as they are in the RepBase database, and the percentages represent how much of the haplotype sequence is covered by the given TE, regardless of how much of the TE sequence is present in the haplotype.

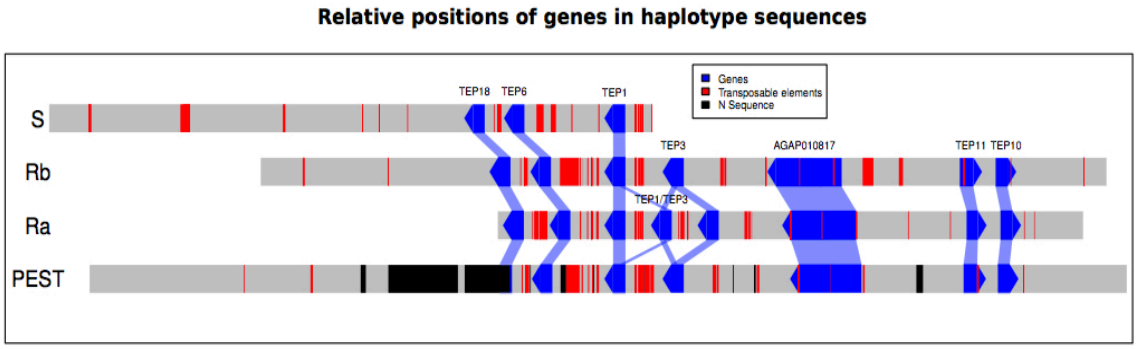

Figure 3.1 Relative positions of TAPL genes in haplotype sequences Grey bars represent the totality of sequence for each haplotype, blue bars and the bands going between them represent genes in the TAPL region, red bars on each haplotype correspond to regions to which TEs mapped, and black bars indicate regions in the reference that are masked by Ns.

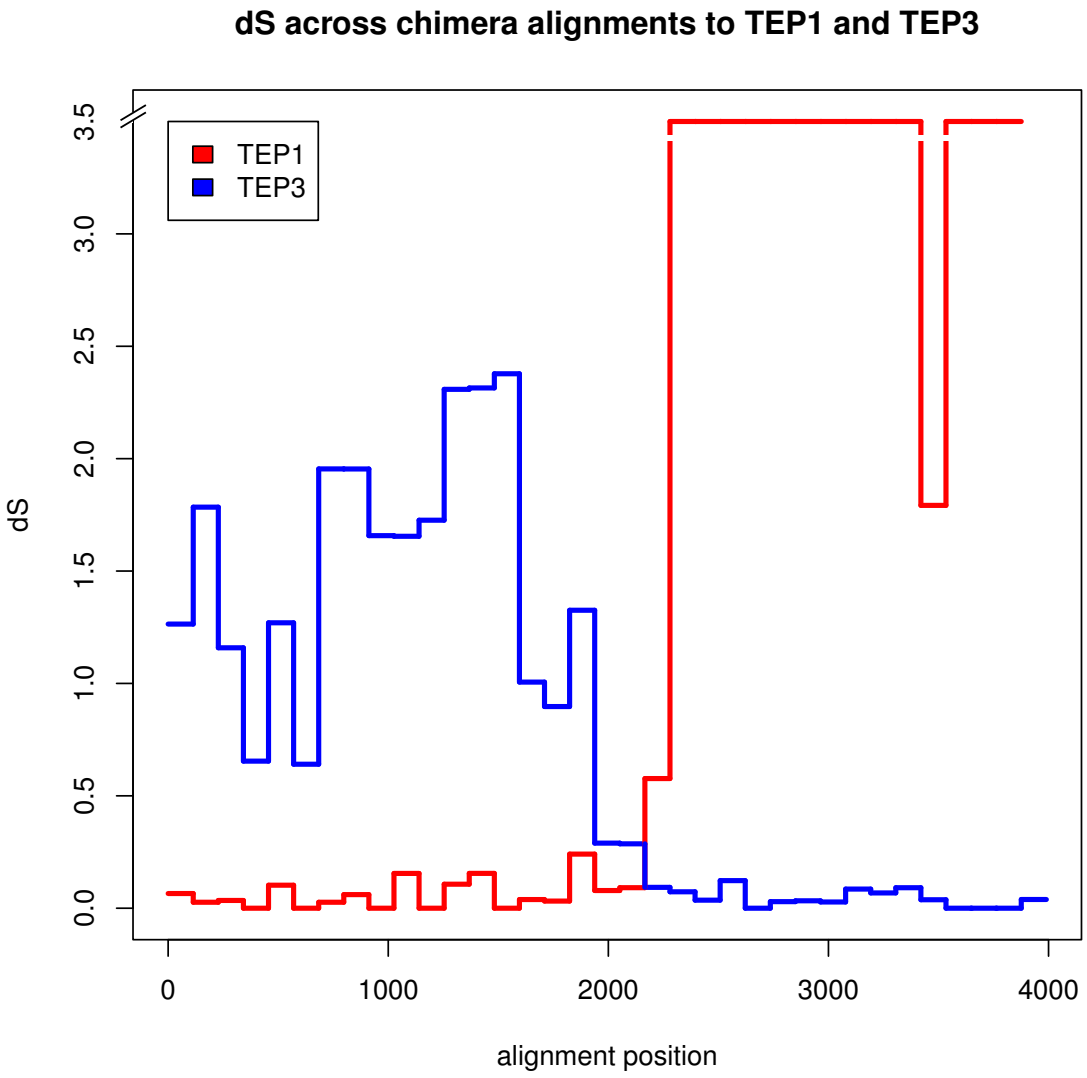

Figure 3.2 dS across chimera alignments to TEP1 and TEP3 dS across 21 non-overlapping windows across alignments of the chimeric *TEP1*/*TEP3* gene and *TEP1* (red) and *TEP3* (blue). The graph has been truncated when dS > 3.4.

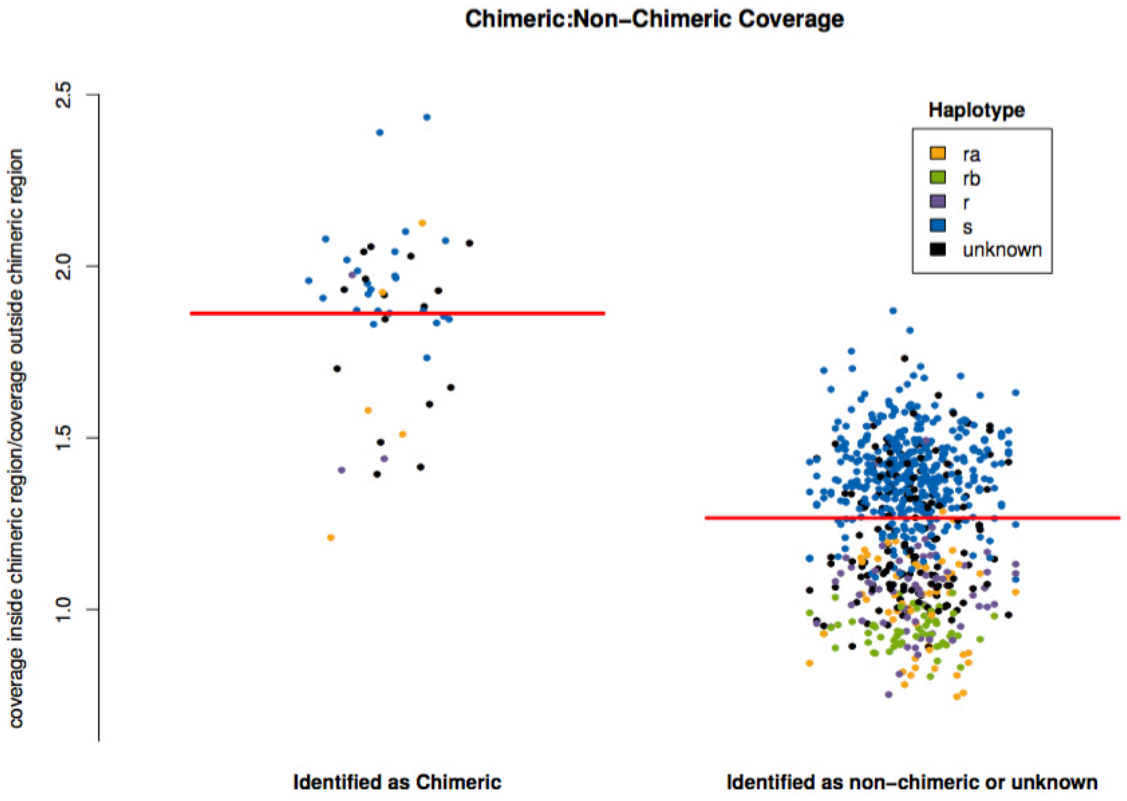

Figure 3.3 Chimeric:non-chimeric coverage Ratio of coverage of regions of *TEP1* and *TEP3* present in the chimera to regions of *TEP1* and *TEP3* not present in the chimera of samples in the *Anopheles gambiae* 1000 genomes project identified as chimeric by SNP data (left) and identified as non-chimeric or unknown by SNP data (right). Each point represents one sample and points are color-coded according to haplotype (r^A, yellow; r^B, green; r, purple; s, blue; and undetermined, black). Samples that were determined to be r, but were unable to be class as r^A or r^B are simply labeled as r. Red bars indicate the mean ratio for each group. The difference between the means is significant, as determined by a Mann-Whitney U test ($W = 32923$, $p < 2.2e\text{-}16$)

REFERENCES

Blandin S, Shiao S-H, Moita LF, Janse CJ, Waters AP, Kafatos FC, Levashina EA. 2004. Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector Anopheles gambiae. *Cell* **116**: 661–670.

Choi KS, Townson H. 2012. Evidence for X-linked introgression between molecular forms of Anopheles gambiae from Angola. *Med Vet Entomol* **26**: 218–227.

Coetzee M, Hunt RH, Wilkerson R, Torre della A, Coulibaly MB, Besansky NJ. 2013. Anopheles coluzzii and Anopheles amharicus, new members of the Anopheles gambiae complex. *Zootaxa* **3619**: 246–274.

Crawford JE, Bischoff E, Garnier T, Gneme A, Eiglmeier K, Holm I, Riehle MM, Guelbeogo WM, Sagnon N, Lazzaro BP, et al. 2012. Evidence for Population-Specific Positive Selection on Immune Genes of Anopheles gambiae. *G3 (Bethesda)* **2**: 1505–1519.

Djogbenou L, Chandre F, Berthomieu A, Dabire R, Koffi A, Alout H, Weill M. 2008. Evidence of Introgression of the ace-1(R) Mutation and of the ace-1 Duplication in West African Anopheles gambiae s. s. ed. D.A. Carter. *PLoS ONE* **3**.

Etang J, Vicente JL, Nwane P, Chouaibou M, Morlais I, do Rosário VE, Simard F, Awono-Ambene P, Toto JC, Pinto J. 2009. Polymorphism of intron-1 in the voltage-gated sodium channel gene of Anopheles gambiae s.s. populations from Cameroon with emphasis on insecticide knockdown resistance mutations. *Mol Ecol* **18**: 3076–3086.

Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, Jiang X, Hall AB, Catteruccia F, Kakani E, et al. 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* **347**: 42–.

Fraiture M, Baxter RHG, Steinert S, Chelliah Y, Frolet C, Quispe-Tintaya W, Hoffmann JA, Blandin SA, Levashina EA. 2009. Two Mosquito LRR Proteins Function as Complement Control Factors in the TEP1-Mediated Killing of Plasmodium. *Cell Host Microbe* **5**: 273–284.

Hackl T, Hedrich R, Schultz J, Foester F. 2014. proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* **30**: 3004–3011.

Hong YS, Hogan JR, Wang X, Sarkar A, Sim C, Loftus BJ, Ren C, Huff ER, Carlile JL, Black K, et al. 2003. Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito Anopheles gambiae. *Mol Gen Genomics* **268**: 720–728.

Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, et al. 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat Biotechnol* **30**: 692–.

Lawniczak MKN, Emrich SJ, Holloway AK, Regier AP, Olson M, White B, Redmond S, Fulton L, Appelbaum E, Godfrey J, et al. 2010. Widespread divergence between incipient Anopheles gambiae species revealed by whole genome sequences. *Science* **330**: 512–514.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Mitri C, Bischoff E, Takashima E, Williams M, Eiglmeier K, Pain A, Guelbeogo WM, Gneme A, Brito-Fravallo E, Holm I, et al. 2015. An Evolution-Based Screen for Genetic Differentiation between Anopheles Sister Taxa Enriches for Detection of Functional Immune Factors. ed. G. Lanzaro. *PLoS Pathog* **11**: e1005306.

Obbard DJ, Callister DM, Jiggins FM, Soares DC, Yan G, Little TJ. 2008. The evolution of TEP1, an exceptionally polymorphic immunity gene in Anopheles gambiae. *BMC Evol Biol* **8**: 274.

Simpson JT, Durbin R. 2012. Efficient de novo assembly of large genomes using compressed data structures. *Genome Res* **22**: 549–556.

The Anopheles gambiae 1000 Genomes Consortium (2015): Ag1000G phase 1 AR3 data release. MalariaGEN. http://www.malariagen.net/data/ag1000g-phase1-AR3

White BJ, Cheng C, Simard F, Costantini C, Besansky NJ. 2010. Genetic association of physically unlinked islands of genomic divergence in incipient species of Anopheles gambiae. *Mol Ecol* **19**: 925–939.

White BJ, Lawniczak MKN, Cheng C, Coulibaly MB, Wilson MD, Sagnon N, Costantini C, Simard F, Christophides GK, Besansky NJ. 2011. Adaptive divergence between incipient species of Anopheles gambiae increases resistance to Plasmodium. *Proc Natl Acad Sci U S A* **108**: 244–249.

WHO. 2016. World Malaria Report.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591.

CONCLUSION

Despite the continued progress of vector control in the fight against mosquito-borne diseases, mosquitoes continue to be the deadliest animal on the planet, with malaria and dengue fever combined accounting for roughly 450,000 deaths and 600 million cases annually (WHO, 2016; Bhatt et al. 2013). Mosquitoes also vector pathogens such as Chickungunya virus, West Nile virus, yellow fever virus, Zika virus, and a number of encephalitis viruses, among others, which only add to the burden of mosquito-borne diseases. Past and current vector control strategies continue to rely heavily on the use of insecticides, either via fogging, indoor residual spraying, or insecticide-treated bed nets. While these strategies have proven effective at reducing mosquito populations and eliminating disease at a small scale in some instances (Wakabi 2007; Batra et al. 2005; Mabaso et al. 2004; Shiff 2002; Sampath et al. 1998; Trigg and Kondrachine 1998; Jana-Kara et al. 1995; Sharma and Yadav 1995), the burden of mosquito-borne diseases is still tremendous. In addition to insecticide use, the use of ovitraps has been a large component in vector control strategies targeted towards *Aedes* mosquitoes (Long et al. 2015; Barrera et al. 2014; 2013a; 2013b; Norzahira et al. 2011; Reiter and Colon 1991; Cheng et al. 1982). Recently, the development of a number of genetic strategies for vector control have shown promise, including transposons (James 2005), homing endonucleases (Windbichler et al. 2011), Medea (Akbari et al. 2014), RIDL (Thomas et al. 2000), and the introduction of *Wolbachia* (Bian et al. 2010; Kambris et al. 2009; McMeniman et al. 2009; Moreira et al. 2009).

While a lot of these strategies have been successful and show promise, there are still a number of limitations and gaps in our knowledge that need to be addressed. First, insecticide-based controls suffer from the evolution of both physiological and behavioral resistance (Riveron et al. 2014; Ffrench-Constant 2013; Gatton et al. 2013; Jones et al. 2012). Second, cryptic population structure and a lack of knowledge of recombination rate variation limits our ability to accurately predict and model how genetic constructs will spread within and between populations (cf. Crawford et al. 2017; Kamdem et al. 2017; Riehle et al. 2011; Connallon and Knowles 2007; Haddrill et al. 2007). Third, cryptic habitats provide a potential reservoir for mosquitoes to avoid vector control strategies (Arana-Guardia et al. 2014; Gustave et al. 2012; Pilger et al. 2011; Barrera et al. 2008).

This dissertation fills in some of the knowledge gaps relating to mosquito evolution and population genetics. In chapter 1, I modeled effective population size changes of *Aedes aegypti* in response to a novel trap. The modeling method used has been successful in detecting changes in *Anopheles gambiae* effective population size (Hodges et al. 2013; Athrey et al. 2012). However, my application of this modeling method to populations of *Ae. aegypti* was not as successful as previous reports. This could be due to the fact that *Aedes* generally have a larger effective population size than *Anopheles*, a population bottleneck in *Ae. aegypti* in the new world (Crawford et al. 2017), an insufficient sample

size (Wang 2001), or a number of other explanations including that *Ae. aegypti* effective population size does not respond in the same way to vector control as *An. gambiae* does.

In chapter 2, I generated fine-scale recombination rate maps including data from thousands of individuals. I showed that recombination rate is variable across the genome, and some of the variation is dependent upon inversion karyotype and sex. On chromosomes where no inversion heterokaryotype is present, sex seems to be the main determinant of the pattern of recombination, whereas on chromosomes containing inversions, the inversions seem to be the main driver of recombination rate patterns. Overall, there was a general pattern of females having higher recombination rates than males, in agreement with previous findings (Kong et al. 2010; Coop et al. 2008), though the magnitude of the disparity between the sexes varies between chromosomes.

In chapter 3, I used bacterial artificial chromosomes to reconstruct a genomic region of high divergence between *An. gambiae* and *An. coluzzii* (White et al. 2011; Lawniczak et al. 2010). I have demonstrated several structural differences between the three dominant haplotypes in this region, some of which are the result of transposable elements. I also showed that there is a previously unannotated gene in this region that arose as a chimera of two neighboring genes. This mechanism of gene creation has been noted before in this region and may represent a previously underappreciated mode of

genome evolution (Obbard et al. 2008). Using data from the *Anopheles gambiae* 1000 Genomes Project, I have shown that this new gene appears to be present in wild populations, though its function is not yet known. This demonstrates the need for long read sequencing technologies, as the short read sequences used for assembling the reference *An. gambiae* genome and further sequencing projects fail to detect this gene as a result of its homology to nearby genes.

Future directions

The projects presented here fill in some important gaps in our current understanding of mosquito vector population genetics and evolution. Using these data will allow the vector control community to more adequately monitor and adapt control strategies and hopefully continue reducing the burden of mosquito-borne diseases. However, while these projects address gaps in our current knowledge, they also raise additional questions. Going forward, it will be important to understand how *Aedes* effective population size reacts to vector control interventions. The data from chapter 1 have the potential to address this. It is possible that the ddRADseq data generated in chapter 1 contain microsatellite loci that could be used to re-run simulations and determine whether microsatellite data is more appropriate for population size modelling in *Aedes*. It might prove fruitful to design an experiment similar to the one carried out in chapter 1, but sample more frequently and deeply (i.e. collect more samples per site, per time point) and run simulations from those data to see if they correspond more closely to the

census population size numbers collected. Additionally, further refinement of the ddRADseq protocol and optimization for the *Ae. aegypti* genome – which is nearly five times the size of the *An. gambiae* genome that the ddRAD protocol I used was optimized for – would produce results more in line with what would be expected. Whole genome sequencing of collected samples may also provide a more accurate result; though it may be cost prohibitive currently, sequencing costs are rapidly coming down and it may be feasible in the near future to sequence whole genomes of many individuals.

Chapter 2 provided a comprehensive overview of recombination in lab crosses of *Anopheles gambiae*, but it is important to bear in mind that this is one set of crosses using three lab colonies and the results are not universally applicable. It is possible that similar patterns and trends will be seen in wild populations of *An. gambiae*, but there are a number of other *Anopheline* species and recombination likely looks different in all of them. Additionally, cryptic subpopulation structure may alter population-level recombination rates. Recombination in other vector genera is certainly vastly different, as well. *Ae. aegypti* has homomorphic sex chromosomes and it remains to be seen what effect that has on local and global patterns of recombination. The methods that have been developed in chapter 2 can be used to look at recombination in other species, which would provide a basis for making more generalized claims about recombination in mosquito species and would also allow for further dissection of the factors influencing

recombination (e.g. sex, inversions, environmental factors). The genome of *Ae. aegypti* is not yet resolved to the chromosome level – a necessity if a fine-scale measure of recombination is desired – however, it is still possible to get a coarse picture of recombination at the scaffold level, which could prove beneficial in the race to control this species. Regarding the data generated in chapter 2, further exploration to determine where the largest discrepancies between males and females occur and the genes present in those regions may help elucidate some of the selective pressures driving the evolution of recombination differences between males and females. Additionally, there is an inversion on chromosome 2R that appears to be polymorphic in the crosses used in this study and may provide an avenue for exploring the effect of inversion size on recombination rate. It may be the case that smaller inversions reduce recombination more drastically, allowing for the potential to engineer gene drive systems to introduce inversions and lessen the chance of recombination breaking them down.

While I was able to determine the presence of a new gene in chapter 3, the function of this gene – and whether it is expressed – remains to be elucidated. A prudent first step in determining if this gene is expressed and what it does would be to screen colonies of *An. gambiae* for the presence of the genic sequence. This could be done by designing PCR primers to amplify the junction between *TEP1* and *TEP3*, for instance. Once a colony containing the sequence has been identified, a standard RNA sequencing project should

be able to determine whether or not it is expressed. If it is expressed, it then becomes a matter of knocking down expression under various challenges to determine function. This final step should be carried out with the remainder of the genes in the TAPL region, as well, as functional characterization of the remaining *TEP* genes in the region has yet to be performed. It is also necessary to determine the age of the chimeric gene so that it may be compared to important events in the demographic history of *An. gambiae* (e.g. spitting from *An. coluzzii*, colonizing a new region or niche, etc.). This will provide a clearer picture of the selective forces that gave rise to this potentially important gene. The selective sweep for the $TEP1r^{B}$ allele in *An. coluzzii* populations in west Africa provides some insight into evolutionary forces acting on important immune genes and elucidating the forces that have given rise to other genes in the region will provide further insight into how this region of high divergence has evolved.

Mosquito vector control has made great strides in recent decades, but the fight to eliminate mosquito-borne diseases still has a long way to go and there are still many questions to be answered. The projects presented here address some important concerns for vector control and lay the foundation for further exploration into important aspects of mosquito biology. Answering some of the new questions that have arisen from this work will aid the continuing battle against mosquitoes and mosquito-borne diseases and push us further towards finally eliminating the tremendous burden they place on human populations.

REFERENCES

Akbari OS, Chen C-H, Marshall JM, Huang H, Antoshechkin I, Hay BA. 2014. Novel Synthetic Medea Selfish Genetic Elements Drive Population Replacement in Drosophila; a Theoretical Exploration of Medea-Dependent Population Suppression. *ACS Synth Biol* **3**: 915–928.

Arana-Guardia R, Baak-Baak CM, Loroño-Pino MA, Machain-Williams C, Beaty BJ, Eisen L, García-Rejón JE. 2014. Stormwater drains and catch basins as sources for production of Aedes aegypti and Culex quinquefasciatus. *Acta Tropica* **134**: 33–42.

Athrey G, Hodges TK, Reddy MR, Overgaard HJ, Matias A, Ridl FC, Kleinschmidt I, Caccone A, Slotman MA. 2012. The Effective Population Size of Malaria Mosquitoes: Large Impact of Vector Control ed. D.A. Petrov. *PLoS Genet* **8**: e1003097.

Barrera R, Amador M, Acevedo V, Caban B, Félix G, Mackay AJ. 2013a. Use of the CDC Autocidal Gravid Ovitrap to Control and Prevent Outbreaks of Aedes aegypti (Diptera: Culicidae). *J Med Entomol* **51**: 145–154.

Barrera R, Amador M, Acevedo V, Hemme RR, Félix G. 2014. Sustained, area-wide control of Aedes aegypti using CDC autocidal gravid ovitraps. *Am J Trop Med Hyg* **91**: 1269–1276.

Barrera R, Amador M, Diaz A, Smith J, Munoz-Jordan JL, Rosario Y. 2008. Unusual productivity of Aedes aegypti in septic tanks and its implications for dengue control. *Med Vet Entomol* **22**: 62–69.

Barrera R, Mackay AJ, Amador M. 2013b. An Improved Trap to Capture Adult Container-Inhabiting Mosquitoes. *http://dxdoiorg/102987/13-63431* **29**: 358–368.

Batra CP, Raghavendra K, Adak T. 2005. Evaluation of bifenthrin treated mosquito nets against anopheline & culicine mosquitoes. *Indian Journal of …*.

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, et al. 2013. The global distribution and burden of dengue. *Nature* **496**: 504–507.

Bian G, Xu Y, Lu P, Xie Y, Xi Z. 2010. The endosymbiotic bacterium Wolbachia induces resistance to dengue virus in Aedes aegypti. ed. D.S. Schneider. *PLoS Pathog* **6**: e1000833.

Cheng ML, Ho BC, Bartnett RE, Goodwin N. 1982. Role of a modified ovitrap in the control of Aedes aegypti in Houston, Texas, USA. *Bull World Health Organ* **60**: 291–296.

Connallon T, Knowles LL. 2007. Recombination rate and protein evolution in yeast. *BMC Evol Biol* **7**: 235.

Coop G, Wen X, Ober C, Pritchard JK, Przeworski M. 2008. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science* **319**: 1395–1398.

Crawford JE, Alves JM, Palmer WJ, Day JP, Sylla M, Ramasamy R, Surendran SN, Black WC, Pain A, Jiggins FM. 2017. Population genomics reveals that an anthropophilic population of Aedes aegypti mosquitoes in West Africa recently gave rise to American and Asian populations of this major disease vector. *BMC Biology 2017 15:1* **15**: 16.

Ffrench-Constant RH. 2013. The molecular genetics of insecticide resistance. *Genetics* **194**: 807–815.

Gatton ML, Chitnis N, Churcher T, Donnelly MJ, Ghani AC, Godfray HCJ, Gould F, Hastings I, Marshall J, Ranson H, et al. 2013. The importance of mosquito behavioural adaptations to malaria control in Africa. *Evolution* **67**: 1218–1230.

Gustave J, Fouque F, Cassadou S, Leon L, Anicet G, Ramdini C, Sonor F. 2012. Increasing Role of Roof Gutters as Aedes aegypti (Diptera: Culicidae) Breeding Sites in Guadeloupe (French West Indies) and Consequences on Dengue Transmission and Vector Control. *Journal of Tropical Medicine* **2012**: 1–3.

Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the Drosophila genome that lack crossing over. *Genome Biol* **8**: R18.

Hodges TK, Athrey G, Deitz KC, Overgaard HJ, Matias A, Caccone A, Slotman MA. 2013. Large fluctuations in the effective population size of the malaria mosquito Anopheles gambiae s.s. during vector control cycle. *Evolutionary Applications* **6**: 1171–1183.

James AA. 2005. Gene drive systems in mosquitoes: rules of the road. *Trends in Parasitology* **21**: 64–67.

Jana-Kara BR, Jihullah WA, Shahi B, Dev V, Curtis CF, Sharma VP. 1995. Deltamethrin impregnated bednets against Anopheles minimus transmitted malaria in Assam, India. *J Trop Med Hyg* **98**: 73–83.

Jones CM, Liyanapathirana M, Agossa FR, Weetman D, Ranson H, Donnelly MJ, Wilding CS. 2012. Footprints of positive selection associated with a mutation (N1575Y) in the voltage-gated sodium channel of Anopheles gambiae. *Proc Natl Acad Sci U S A* **109**: 6614–6619.

Kambris Z, Cook PE, Phuc HK, Sinkins SP. 2009. Immune Activation by Life-Shortening Wolbachia and Reduced Filarial Competence in Mosquitoes. *Science* **326**: 134–136.

Kamdem C, Fouet C, Gamez S, White BJ. 2017. Pollutants and Insecticides Drive Local Adaptation in African Malaria Mosquitoes. *Mol Biol Evol* **34**: 1261–1275.

Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, Walters GB, Jonasdottir A, Gylfason A, Kristinsson KT, et al. 2010. Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**: 1099–1103.

Lawniczak MKN, Emrich SJ, Holloway AK, Regier AP, Olson M, White B, Redmond S, Fulton L, Appelbaum E, Godfrey J, et al. 2010. Widespread divergence between incipient Anopheles gambiae species revealed by whole genome sequences. *Science* **330**: 512–514.

Long SA, Jacups SP, Ritchie SA. 2015. Lethal ovitrap deployment for Aedes aegypti control: potential implications for non-target organisms. *Journal of Vector Ecology* **40**: 139–145.

Mabaso MLH, Sharp B, Lengeler C. 2004. Historical review of malarial control in southern African with emphasis on the use of indoor residual house-spraying. *Trop Med Int Health* **9**: 846–856.

McMeniman CJ, Lane RV, Cass BN, Fong AWC, Sidhu M, Wang Y-F, O'Neill SL. 2009. Stable Introduction of a Life-Shortening Wolbachia Infection into the Mosquito Aedes aegypti. *Science* **323**: 141–144.

Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT, Hedges LM, Rocha BC, Hall-Mendelin S, Day A, Riegler M, et al. 2009. A Wolbachia Symbiont in Aedes aegypti Limits Infection with Dengue, Chikungunya, and Plasmodium. *Cell* **139**: 1268–1278.

Norzahira R, Hidayatulfathi O, Wong HM, Cheryl A. 2011. Ovitrap surveillance of the dengue vectors, Aedes (Stegomyia) aegypti (L.) and Aedes (Stegomyia) albopictus Skuse in selected areas in Bentong, Pahang, …. *Trop Biomed*.

Obbard DJ, Callister DM, Jiggins FM, Soares DC, Yan G, Little TJ. 2008. The evolution of TEP1, an exceptionally polymorphic immunity gene in Anopheles gambiae. *BMC Evol Biol* **8**: 274.

Pilger D, Lenhart A, Manrique Saide P, Siqueira JB, da Rocha WT, Kroeger A. 2011. Is routine dengue vector surveillance in central Brazil able to accurately monitor the Aedes aegypti population? Results from a pupal productivity survey. *Tropical Medicine & International Health* **16**: 1143–1150.

Reiter P, Colon M. 1991. Enhancement of the CDC Ovitrap with hay Infusions for Daily Monitoring of Aedes aegypti Populations. *J of Am Mosquito Cont Assc* **7**: 52-55.

Riehle MM, Guelbeogo WM, Gneme A, Eiglmeier K, Holm I, Bischoff E, Garnier T, Snyder GM, Li X, Markianos K, et al. 2011. A Cryptic Subgroup of Anopheles gambiae Is Highly Susceptible to Human Malaria Parasites. *Science* **331**: 596–598.

Riveron JM, Yunta C, Ibrahim SS, Djouaka R, Irving H, Menze BD, Ismail HM, Hemingway J, Ranson H, Albert A, et al. 2014. A single mutation in the GSTe2 gene allows tracking of metabolically based insecticide resistance in a major malaria vector. *Genome Biol* **15**: R27.

Sampath TR, Yadav RS, Sharma VP, Adak T. 1998. Evaluation of lambdacyhalothrin-impregnated bednets in a malaria endemic area of India. Part 2. Impact on malaria vectors. *J Am Mosq Control Assoc* **14**: 437–443.

Sharma VP, Yadav RS. 1995. *Cyfluthrin impregnated mosquito nets to control malaria in the mining settlement of Orissa, India*. Public Health.

Shiff C. 2002. Integrated approach to malaria control. *Clin Microbiol Rev* **15**: 278–293.

Thomas DD, Donnelly CA, Wood RJ, Alphey LS. 2000. Insect population control using a dominant, repressible, lethal genetic system. *Science* **287**: 2474–2476.

Trigg PI, Kondrachine AV. 1998. Commentary: malaria control in the 1990s. *Bull World Health Organ* **76**: 11–16.

Wakabi W. 2007. Africa counts greater successes against malaria. *Lancet*, December 8.

Wang J. 2001. A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genetics Research* **78**: 243–257.

White BJ, Lawniczak MKN, Cheng C, Coulibaly MB, Wilson MD, Sagnon N, Costantini C, Simard F, Christophides GK, Besansky NJ, et al. 2011. Adaptive divergence between incipient species of Anopheles gambiae increases resistance to Plasmodium. *Proc Natl Acad Sci U S A* **108**: 244–249.

WHO. 2016. World Malaria Report.

Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, Ulge UY, Hovde BT, Baker D, Monnat RJ, Burt A, et al. 2011. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* **473**: 212–215.