MOLECULAR EVOLUTION OF FOUR SALIVARY PROTEINS WITHIN SPECIES OF THE *ANOPHELES GAMBIAE* COMPLEX

An Undergraduate Research Scholars Thesis

by

DIANA CARILU RODRIGUEZ and MARYAM IDREES KHAN

Submitted to Honors and Undergraduate Research
Texas A&M University
In partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLARS

Approved by
Research Advisor:

May 2013

Major: Biomedical Sciences

Entomology

TABLE OF CONTENTS

Page

ABSTRACT

Molecular Evolution of 4 Salivary Proteins within Species of the *Anopheles gambiae* Complex. (May 2013)

Diana Carilu Rodriguez & Maryam Idrees Khan
Department of Veterinary Medicine & Biomedical Sciences
Department of Entomology
Texas A&M University

Research Advisor: Dr. Michel Slotman
Department of Entomology

Some of the primary vectors of human malaria include female mosquitoes from the *Anopheles gambiae* complex, which is comprised of at least six different species within the genus *Anopheles,* including *A. gambiae* (M/S forms), *A. arabiensis*, *A. melas*, *A. bwambae*, *A. merus*, and *A. quadriannulatus*. Salivary gland proteins within the *Anopheles gambiae* complex interact with a vertebrate host's immune system by controlling vasodilatation, inflammation, and platelet aggregation at the feeding site on the vertebrate host. The way certain salivary proteins are expressed within different mosquito species has been studied, but there is still a need for a comparison between species of close proximity, such as those in the *A. gambiae* complex. This comparison could reveal genes that may interact with a host's immune system or with malaria parasites and hence may be under selection. Such genes may have crucial roles in the adaptation to specific hosts. For example, an excess of non-synonymous fixed differences in the gene would mean directional or positive selection, which may have resulted from interaction with various hosts. To gain further insight into 4 specific salivary gland proteins (Anophelin, Ichit, Glycosidase,

and Lysozyme), their patterns of polymorphism were analyzed in 3 species of the *An. gambiae* complex (*Anopheles gambiae M* and *S* forms, *Anopheles melas,* and *Anopheles arabiensis*). After analyzing these genes using several statistical tests, the comparison showed that three of the four genes, Anophelin, Ichit, and Glycosidase are highly conserved with no signs of positive selection or fixed differences between *A. gambiae*, *A. arabiensis*, and *A. melas* species. Further research exploring the genetic variation of other salivary proteins within the *A. gambiae* complex may identify proteins that are undergoing positive selection. This could locate genes involved in vector competence, either preventing or enhancing disease transmission in *Anopheles* mosquitoes.

DEDICATION

We would like to dedicate this thesis to Dr. Giri Athrey, who not only supervised our research, but also provided academic guidance and friendship outside of the laboratory. His genuine interest in our success made a profound impact in our experience as undergraduate research scholars.

ACKNOWLEDGEMENTS

Everyone in Dr. Slotman's laboratory offered their help and guidance from the moment we began working with them. Dr. Giri Athrey was patient in teaching us basic laboratory protocols and cloning techniques, while Dr. Theresa Hodges supplied the mosquito samples we needed from Equatorial Guinea. We could always count on Dr. Sharmilla Pathikonda to order our PCR and cloning supplies whenever we were running low and graduate students Luciano Cosme and Kevin Deitz came to the rescue when we needed help running different data analysis software. We thank the entire lab group for pitching in their comments and suggestions when it came to editing this thesis. It was nice knowing we could count on their support from the beginning. We would also like to acknowledge our friends and family for providing moral support.

NOMENCLATURE

PCR	Polymerase Chain Reaction
T	Temperature
PEG	Polyethylene glycine
A.	*Anopheles*
A. gambiae	*Anopheles gambiae*
A. arabiensis	*Anopheles arabiensis*
A. melas	*Anopheles melas*
A. merus	*Anopheles merus*
A. bwambae	*Anopheles bwambae*
A. quadriannulatus	*Anopheles quadriannulatus*

CHAPTER I

INTRODUCTION

Malaria is a vector transmitted disease that is a global health issue killing close to 1 million people annually, the majority of whom are children in sub-Saharan Africa. Several important vectors of human malaria are female mosquitoes of the *Anopheles gambiae* complex, which is comprised of at least six different species within the genus *Anopheles*, including *A. gambiae* (M/S forms), *A. arabiensis, A. melas, A. bwambae, A. merus*, and *A. quadriannulatus*. These species were first identified in the late 1960s and later X-linked sequences and fixed inversion differences aided in the discernment of an evolutionary relationship among them, as shown in Figure 1.1 (White *et al.* 2011). The members of this complex, although morphologically similar, differ in their degrees of anthropophilic (affinity for human hosts) tendencies and geographic distributions.

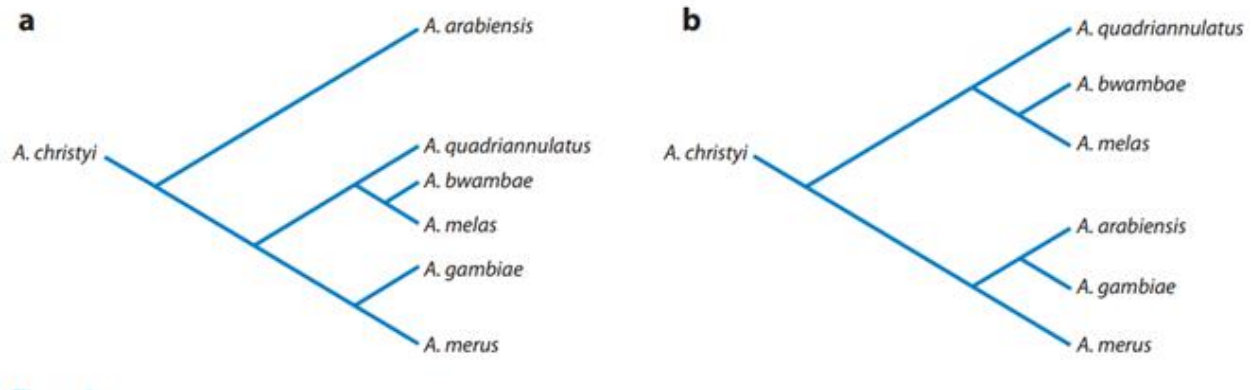


Figure 1.1. Evolutionary relationships for the A. gambiae complex based on (a) x-linked sequences and fixed inversions or (b) mtDNA and autosomal sequences. (White et al. 2011)

The two most medically important species are *A. gambiae* and *A. arabiensis*. They have the widest distribution and are the most anthropophilic species in the complex (Coluzzi *et al.* 1979). In contrast, *A. melas* has been documented with zoophilic and exophilic habits which are even more

consistently observed in *A. merus* and *A. quadriannulatus* (Coluzzi *et al.* 1964). *A. merus* and *A. quadriannulatus*, being zoophilic and/or exophilic are less efficient at transmitting malaria to humans than *A. arabiensis* and *A. gambiae* (Service *et al.* 1993).

A varied range of geographic distribution can also be seen across the *A. gambiae* complex as shown in Figure 1.2. (Coetzee *et al.* 2000). *A. arabiensis* is concentrated in the drier savannah and steppe areas of Africa, where the annual rainfall is <1000 mm, and in equatorial rainforest regions where there is a history of extensive land clearance (Coetzee *et al.* 2000). *A.quadriannulatus* is found mainly in southeast Africa, while *A. bwambae* inhabits the mineral springs of Semliki Forest in Uganda. *A. melas* is a brackish-water breeding species, occurring along the western coasts of Africa. *A. gambiae* has a cosmopolitan distribution throughout Africa.

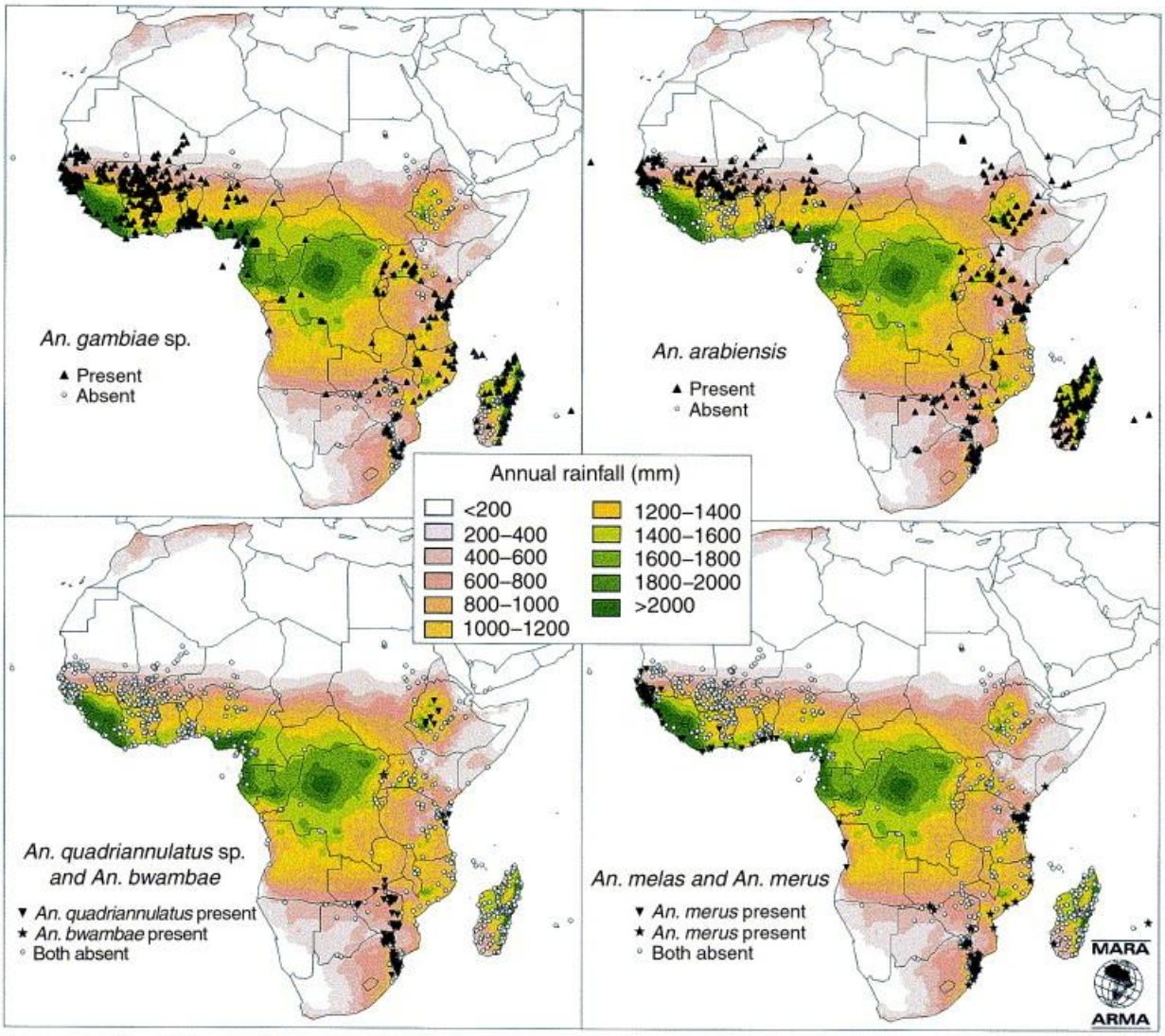


Figure 1.2. The geographical distribution of the *A. gambiae* complex against rainfall in mm. (Coetzee *et al*. 2000).

These differences in host preference and habitat distribution are strongly influenced by human activities. According to a study by Kamdem (2012), the ongoing diversification within the *A. gambiae* species involve speciation driven by environmental change and exposure to humans. Agricultural practices, deforestation, insecticide use, and other changes to the environment could

be responsible for changes in species composition in an area. Deforestation alone has led to an increase in local temperatures in African highlands, expanding the ambient temperature territory available for *Anopheles* mosquitoes further south into areas that were once too cold for larval development (Afrane *et al.* 2012). This demonstrates how human behavior and anthropogenic changes to the environment can lead to changes in a species' geographic distribution and feeding behaviors.

The way mosquito salivary proteins have coevolved with hosts is of particular importance when understanding host and vector evolution. Salivary glands of the *A. gambiae* complex express genes whose products allow them to efficiently take blood-meals from hosts. The saliva interferes with vasodilation, platelet aggregation, and coagulation that improve the mosquito's ability to take a blood meal efficiently (Arca *et al.* 1999). They also aid in the stimulation of certain biological pathways for the digestion of the blood (Arca *et al.* 2002). In addition to these functions in blood feeding, they can transmit the human malaria parasite [*Plasmodium spp.*] which completes part of its lifecycle in the mosquito's salivary glands. Once a mosquito ingests a sexual stage gametocyte from an infected host during blood feeding, female and male gametes are released in the midgut and eventually travel to the median and distal lobes of the salivary glands the (Dimopoulous *et al.* 1998). Salivary gland genes are important in other arthropod vectors as well. In sandflies, salivary proteins have the potential to be a vaccine against *Leishmania* infection, while in ticks, serine proteinase inhibitors, also known as serpins, are salivary proteins that may manipulate host defenses during feeding (Kato *et al.* 2013, Mulenga *et al.* 2009).

In this study, we examined the molecular evolution of four genes found in the salivary glands of species within the *A. gambiae* complex. The genes chosen were Anophelin, Ichit, Glycosidase, and Lysozyme. Discerning if certain salivary genes in the *A. gambiae* complex show signs of positive selection would identify potential candidates for further functional assay studies for how salivary proteins mediate the interaction between the malaria parasite and host.

Anophelin [ENSANGG00000009826, www.ensembl.org/Anopheles_gambiae] is a competitive alpha-thrombin inhibitor found in the salivary gland of mosquitoes within the *A. gambiae* complex. When the serine protease alpha-thrombin is inhibited by Anophelin, it can no longer convert soluble fibrinogen into insoluble strands of fibrin (Francischetti *et al.* 2010), which is a crucial step in the hemostatic (blood clotting) process. Therefore, the presence of Anophelin in the mosquito's salivary glands improves its ability to blood feed.

Ichit [EMBLAAR05803.1, http://www.ebi.ac.uk] is an antimicrobial protein that participates in the formation of the extracellular matrix (Patil *et al.* 2009). It contains two putative chitin-binding domains as well as mucin domains that play a role in trapping microbial pathogens passing through this matrix (Dimopoulos *et al.* 1998). This protein, found both in the salivary gland and midgut of the mosquito, is of interest because of its increase in production in *A. gambiae* mosquitoes that have been exposed to the malaria parasite (Patil *et al.* 2009). A study by Dimopoulous (1998) shows that Ichit is activated when the *Plasmodium* sporozoites are released from the midgut into the haemolymph to invade the salivary glands. After blood feeding, the malaria parasite elicits immune responses in the mosquito's various tissues. Ichit was thought to only be transcriptionally activated by bacteria, but this study shows that the malaria parasite can have the same effect on

the protein, possibly recognizing the malaria parasite as a foreign invader (Dimopoulos *et al.* 1998). This positive association with the malaria parasite could cause Ichit to be undergoing genetic changes.

Glycosidase is an enzyme present in saliva and aids in degrading complex carbohydrates. Red blood cell membranes are composed of a variety of glycoproteins which entails the presence of carbohydrates (Dana *et al.* 2005). Enzymes, such as Glycosidase, that are usually associated with metabolism of carbohydrates in nectar meals are also involved in blood digestion (Dana *et al.* 2005). Glycosidase has been noted to be used by *Anopheles* mosquitoes in the digestion of non-proteinaceous blood constituents. Positive selection for Glycosidase may be present in species that have a higher level of intake of complex carbohydrates which would require an elevated level of Glycosidase.

Lastly, Lysozyme [AGAP007386, www.vectorbase.com/Anopheles_gambiae] is a gene that is expressed in the salivary glands and midgut of larval and adult female mosquitoes in *A. gambiae* (Das *et al.* 2010). Lysozyme is thought to help prevent the growth of microbes in ingested sugar and blood meals that are stored in the mosquito crop (Das *et al.* 2010). Three separate transcripts for Lysozyme have been detected in *A. gambiae.* Of these three transcripts, two were found to be down-regulated in female *A. gambiae* mosquitoes 24 hours after blood feeding (Das *et al.* 2010). *A. gambiae* takes in fungi and bacteria while feeding on nectar, but is less likely to do so with blood because it is usually sterile (Kang *et al.* 1996). In a study conducted by Kang *et al.* it was observed that Lysozyme produced a much more evident signal in sugar-fed *A. gambiae*

mosquitoes (1996). This could either mean that taking a blood-meal suppresses the Lysozyme gene or sugar-meals induce its transcription (Kang *et al.* 1996).

The way salivary proteins are regulated within different mosquito species has been studied in previous research, but there is still a need for a comparison between species of close proximity, such as those in the *A. gambiae* complex (Arca *et al.* 2002, Drame *et al.* 2010, Mans *et al.* 2007, Ye *et al.* 2011). This comparison could reveal genes that are undergoing positive selection. If these positively selected genes were identified, the effects of silencing or altering them could be studied in attempts to decrease mosquito fitness. By analyzing data from the 4 genes, Anophelin, Ichit, Lysozyme, and Glycosidase within the *A. gambiae* complex, we will determine if these genes have evolved under positive selection.

CHAPTER II

MATERIALS AND METHODS

Samples

The mosquito DNA used in this experiment came from samples obtained from Equatorial Guinea as follows:

- 4 *A. gambiae* S form samples collected from Yengue, Ne in 2007
- 4 *A. gambiae* M form samples collected from Ukomba, Ne in 2007
- 4 *A. melas* samples collected from Cogo, Ne in 2009
- 4 *A. melas* samples collected from Luba, Ne in 2009
- 8 *A. arabiensis* samples collected from Riao, Cameroon in 2007

PCR along with TA (sticky end) cloning and Sanger sequencing was used to determine the DNA sequence of the Anophelin, Ichit, Glycosidase and Lysozyme genes in three main malaria vectors: *A. gambiae*, *A. arabiensis*, and *A. melas*. Primer 3 (Rosen *et al.* 2000) program was used to select appropriate primers based on the An. gambiae genome. The primer pairs chosen for Anophelin, Ichit, Lysozyme, and Glycosidase are listed in Table 2.1.

Table 2.1. Primer pairs chosen for Anophelin, Ichit, Lysozyme, and Glycosidase.

Gene Name	Primer Sequence
Anophelin – Forward 1	AGCAACAGCAAAGGCAAAGT
Anophelin – Forward 2	TGTCTTCCAAATCGATTGCT
Anophelin – Reverse 1	GCAATTATTCGTCCGACTCC
Anophelin – Reverse 2	ATAAATACGGCGCGAAGTTG
Ichit – Forward 1	GGCTGGAGCGTTAGTATTGG
Ichit – Forward 2	CCCGAGTGGCAAATAGAAAC
Ichit – Forward 3	CATTCCCGAGTGGCAAATAG
Ichit – Reverse 1	TACTGGTTGGGCTCTCTCGT
Lysozyme – Forward 1	ACCTCAACTACAGCGCCACT
Lysozyme – Forward 2	TCAACTACAGCGCCACTCTG
Lysozyme – Reverse 1	GCAACGAGCGTTATGGGTAT
Lysozyme – Reverse 2	CTGATCGGTCGGATTTCATT
Glycosidase – Forward 1	GACGGATGATCCCGAAGATA
Glycosidase – Forward 2	ACCTGCGACACATCTACACG
Glycosidase – Reverse 1	CCGTGCTATAAACCGTTGCT
Glycosidase – Reverse 2	CGATAGTTTTCGTACAGCTTGG

Primers targeted the non-transcribed regions of the genes on both sides of the coding sequence and the PCR products were analyzed using ethidium bromide gel electrophoresis to ensure that the product length was in the range of the expected base pair length of each protein. Four different primer pair combinations, F1R1, F1R2, F2R1, and F2R2, were used in PCR reactions

with the DNA samples for Anophelin, Lysozyme, and Glycosidase. For Ichit, F1R1, F2R1, F3R1 combinations were used. The primer pairs whose end products provided the brightest PCR product were chosen. GoTaq polymerase (Promega) was used first with a PCR cycle that included 2 minutes at 95°C, 35 cycles of 40 seconds at 95°C, 30 seconds at 54°C and 60 seconds at 72°C, finally ending with 15 minutes at 72°C. The PCR product of Anophelin was expected to be 467 base pairs long, while VectorBase reported Ichit at 1330 base pairs, Glycosidase at 981 base pairs in length and Lysozyme at 1053 base pairs. The primer pair combination that showed the brightest PCR product of expected length for Anophelin was ANPH-F1 and ANPH-F2. Ichit did not show any bands using the GoTaq polymerase, but did show bands when OneTaq 2X (New England Biolabs) and GoTaq (Promega) were used instead. The primer combination chosen for Ichit was ICHIT-F1 and ICHIT-R1. For Lysozyme, Phusion HD Polymerase (New England Biolabs) worked best, showing the optimal primer combination to be LYS-F2 and LYS-R1. Glycosidase showed the best results with the primer combination GLYC-F1 and GLYC-R1. These primers were then used to obtain PCR products using the mosquito DNA samples from Equatorial Guinea.

The PCR amplification program for Anophelin used Promega's GoTaq polymerase and a cycle that included 2 minutes at 95°C, 35 cycles of 40 seconds at 95°C, 30 seconds at 54°C and 60 seconds at 72°C, followed by 15 minutes at 72°C.

For Ichit, the two different amplification programs used for the different samples were as follows: *A. gambiae* and *A. melas* island samples were amplified using OneTaq 2X polymerase and a PCR cycle that included 30 seconds at 94°C, 30 cycles of 30 seconds at 94°C, 60 seconds

at 62°C, and 90 seconds at 68°C, followed by 5 minutes at 68°C. *A. melas* mainland and *A. arabiensis* samples were amplified using Promega's GoTaq polymerase and a PCR cycle that began with 2 minutes at 95°C, 35 cycles of 45 seconds at 95°C, 45 seconds at 60°C, and 60 seconds at 72°C, followed by 5 minutes at 72°C.

Lysozyme was amplified using PCR with Phusion polymerase and the primers LYS-F2 and LYS-R1. The PCR program ran for 2 minutes at 95°C, 35 cycles of 40 seconds at 95°C, 30 seconds at 54°C, and 60 seconds at 72°C, finally ending with 15 minutes at 72°C. *A. gambiae*, *A. melas,* and *A. arabiensis* showed the brightest PCR product for the primer combination with Phusion polymerase.

Amplification for Glycosidase was done using PCR with GoTaq polymerase and primers GLYC-F1 and GLYC-R1. The PCR program ran for 2 minutes at 95°C, 35 cycles of 40 seconds at 95°C, 30 seconds at 54°C and 60 seconds at 72°C, finally ending with 15 minutes at 72°C. Though the F1R1 combination displayed the brightest bands, it did so only for *A. gambiae* and *A. melas* samples. There were no apparent bands for *A. arabiensis* with the F1R1 combination, or any of the other combinations.

Cloning

PCR products were ligated into plasmids using the pGEM-T Easy Vector Systems cloning kit (Promega, Madison, WI, USA) and then cultured onto LB/Ampicillin agar plates. LB/ampicillin plates were poured and prepared with IPTG and X-Gal. The transformation culture was plated on the LB/ampicillin/IPTG/X-Gal plates. One plate was selected for each sample and white colonies

were isolated and picked using pipette tips. A standard PCR reaction using GoTaq polymerase was conducted using the DNA from the selected white colonies to check if the transformation was successful. [Technical Manual – Easy Vector Systems]

DNA cleanup

The final PCR product was first precipitated and purified using Polyethylene glycol (PEG) precipitation to isolate the DNA and remove impurities. The post-PCR cleaned-up DNA was then resuspended and cycle-sequenced using BigDye Direct Cycle Sequencing Kit (Applied Biosystems). Samples were then Sanger sequenced at Yale University's DNA Sequencing Facility.

Statistical Analysis

The raw sequences were analyzed using the program Sequencher 5.0 (Gene Codes Corporation) in order to delete the primer sequences and the T7 and SP6 sequences used in the pGEM-T easy Vector System (Librado *et al.* 2009). The introns were deleted and the original protein sequences were used to align them to the new clean sequences from Sequencher 4.9. The data was cleaned by deleting primer sequences and introns because they were not part of the chosen gene coding sequence for the genes we were analyzing. The initial gene sequences were used to align the cleaned samples.

DNA Sequence Polymorphism v5 (DnaSP) was used (Librado *et al.* 2009) to test if any of the species showed an excess of amino acid substitutions between species or evidence of a selective sweep. One of these tests is the McDonald-Kreitman Test which is based on the divergence

between species and compares the rates of synonymous to non-synonymous fixed differences between species to polymorphisms. A table is constructed with the number of synonymous and nonsynonymous substitutions that are fixed between species (D_S and D_N) and are polymorphic within a species (P_S and P_N) (refer to Table 2.2). Under positive selection, there is an increase in the non-synonymous divergence ($D_N/D_S > P_N/P_S$) while neutral conditions give values of $D_N/D_S = P_N/P_S$.

Table 2.2 McDonald-Kreitman Categories

	Fixed	**Polymorphic**
Synonymous	D_S	P_S
Nonsynonymous	D_N	D_N

Another method for determining if a DNA sequence is evolving is a sample variance test known as Tajima's D, which is based on the frequency distributions of polymorphisms within species. Sequences with the introns were used for this analysis. This statistic provides a value based on mean pairwise differences and the number of segregating sites. This value increases as the data deviates from a neutral equilibrium. A positive Tajima's D results from an excess of intermediate frequency alleles relative to what is expected and can result from balancing selection or a declining population size. Under equilibrium conditions, a negative Tajima's D value occurs when there is an excess of low frequency alleles relative to what is expected. These low frequency polymorphisms could be due to population size expansion following a bottleneck, from purifying selection, or even from selective sweep. All of these would cause the number of low frequency alleles to increase when compared to a neutral model.

Neighbor-joining trees

Neighbor-joining trees were created using the program MEGA v. 5 using 100 bootstrap replicates (Tamura *et al.* 2011). The contigs, obtained from Sequencher 5.0, containing the cleaned and aligned sequences and reference sequence, were used to create the trees. The trees were visualized in the program FigTree v. 1.3.1 (Rambaut 2009).

CHAPTER III

RESULTS

Anophelin results

From the initial set of DNA samples used for the study, 57 samples provided sufficient sequence data to be used for statistical analysis of the Anophelin gene. 19 *A. arabiensis*, 18 *A. gambiae* and 20 *A. melas* samples were used for the McDonald-Kreitman and Tajima's D tests. After removing introns, 272 base pairs, were compared among sequences. The results for the McDonald-Kreitman test, displayed in Table 3.1, revealed that there are no fixed differences between species (synonymous or nonsynonymous). The Tajima's D values, displayed in Table 3.2, for *A. gambiae, A. melas,* and *A. arabiensis* are all negative that show no significance. The results show that Anophelin is a highly conserved salivary gene with no signs of positive selection or fixed differences between *A. gambiae, A. arabiensis*, and *An. melas* species. The neighbor-joining tree, which is based on a distance matrix, in Figure 3.1 presents the taxonomical relationship of the Anophelin gene within *A. gambiae, A. melas,* and *A. arabiensis.*

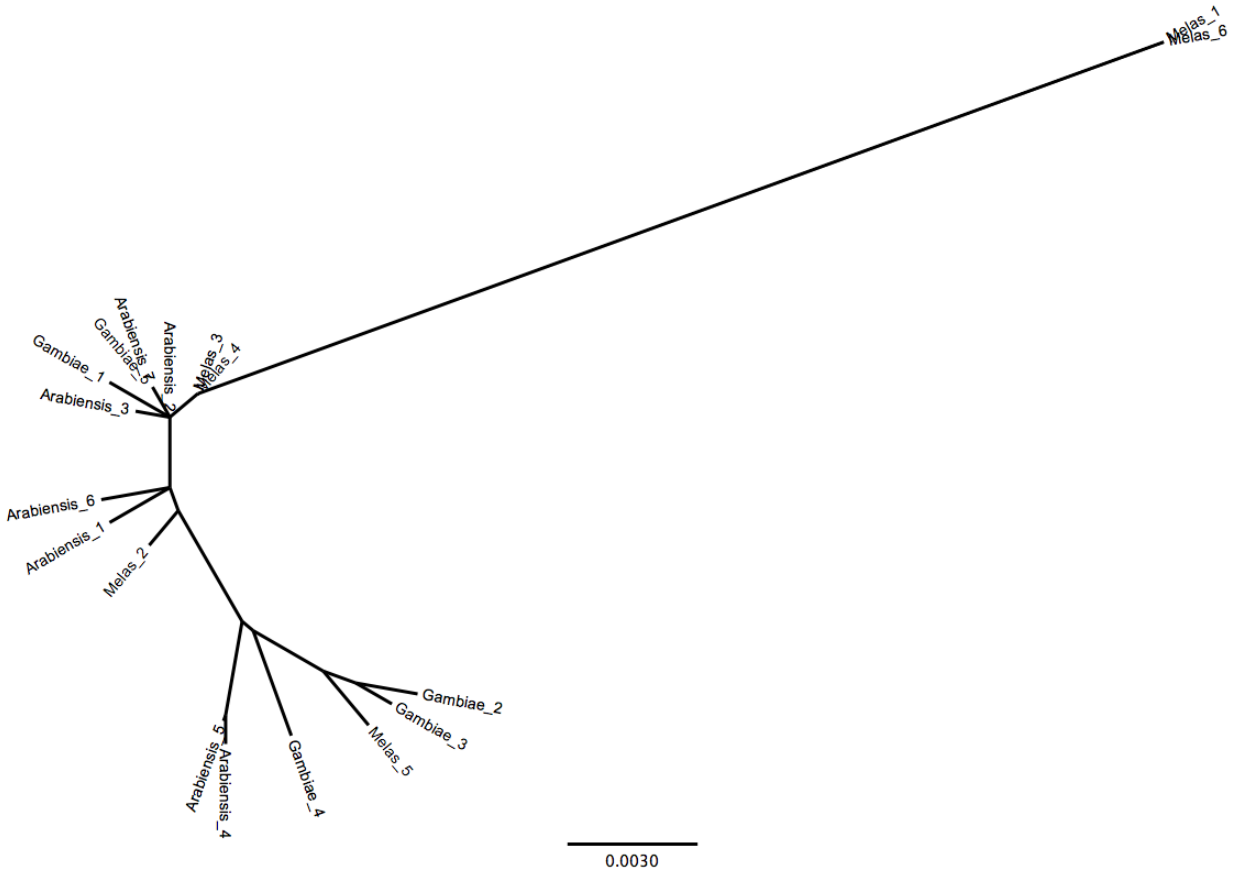
Table 3.1 McDonald-Kreitman Results for Anophelin

Species Combination	Synonymous	Nonsynonymous	Synonymous Fixed Differences	Nonsynonymous Fixed Differences
A. gambiae vs. A. melas	19	12	0	0
A. gambiae vs. A. arabiensis	16	14	0	0
A. melas vs. A. arabiensis	21	16	0	0

Table 3.2 Population Genetics Parameters for Anophelin

Species	Number of Sequences	Tajima's D Value	P-Value	Nucleotide Diversity
A. arabiensis	19	-1.168	.1159	.02079
A. gambiae	18	-.428	.374	.02009
A. melas	20	-.251	.449	.02012

Figure 3.1 Neighbor-Joining Tree for Anophelin



Ichit results

From the initial set of DNA samples used for the study, 36 samples provided sufficient sequence data to be used for statistical analysis of the Ichit gene. 14 *A. arabiensis,* 12 *A. gambiae* and 10 *A. melas* samples were used for the McDonald-Kreitman and Tajima's D tests. After sequencing

the raw data and removing introns, 857 base pairs, which included exon regions, were compared among sequences. The results for the McDonald-Kreitman test, displayed in Table 3.3, revealed that there are no fixed differences between species (synonymous or nonsynonymous). The Tajima's D values, displayed in Table 3.4, for *A. gambiae, A. melas,* and *A. arabiensis* are all negative that show no significance. The results show that Ichit is a highly conserved salivary gene with no signs of positive selection or fixed differences between *A. gambiae, A. arabiensis,* and *An. melas* species. The neighbor-joining tree, in Figure 3.2 presents the taxonomical relationship of the Ichit gene within *A. gambiae, A. melas,* and *A. arabiensis.*

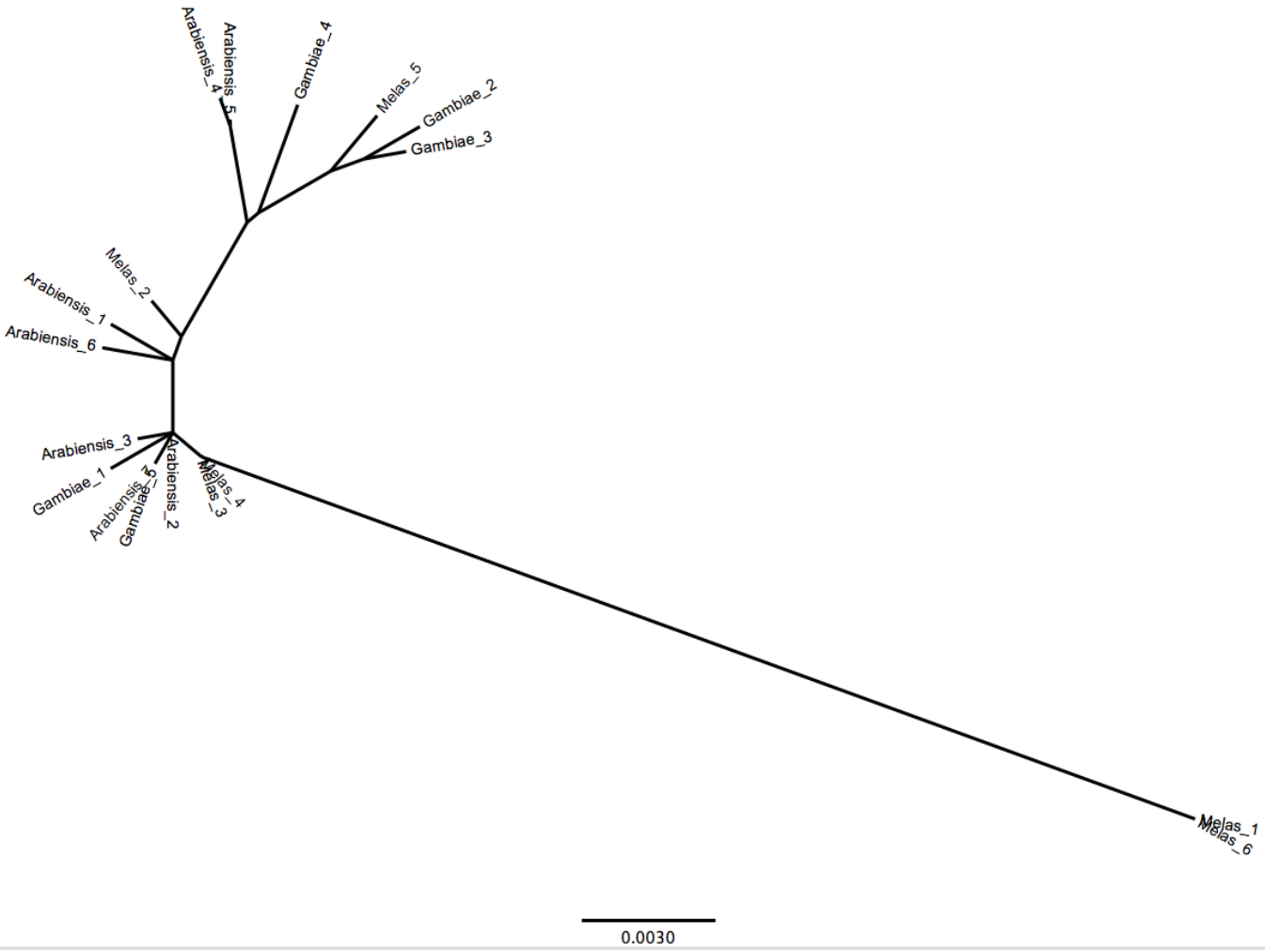
Table 3.3 McDonald-Kreitman Results for Ichit

Species Combination	Synonymous	Nonsynonymous	Synonymous Fixed Differences	Nonsynonymous Fixed Differences
A. gambiae vs. A. melas	12	8	0	0
A. gambiae vs. A. arabiensis	18	14	0	0
A. melas vs. A. arabiensis	14	12	0	0

Table 3.4 Population Genetics Parameters for Ichit

Species	Number of Sequences	Tajima's D Value	P-Value	Nucleotide Diversity
A. arabiensis	14	-1.19	.109	.01078
A. gambiae	12	-.869	.193	.02039
A. melas	10	-.840	.205	.01013

Figure 3.2 Neighbor-Joining Tree for Ichit



Lysozyme results

The DNA fragments were present after cloning of the gene, and confirmation by the insertion checks, but after the DNA cleanup step the PCR product was very poor. Hence, no results are available for this gene.

Glycosidase results

From the initial set of DNA samples used for the study, 10 samples provided sufficient sequence data to be used for statistical analysis of the Glycosidase gene. 4 *A. gambiae* and 6 *A. melas*

samples were used for the McDonald-Kreitman and Tajima D tests. After sequencing and removing introns, 976 base pairs were included in the analysis. The *A. arabiensis* samples were not able to be used because of poor PCR product. The results for the McDonald-Kreitman test, displayed in Table 3.5, revealed that there are no fixed differences between species (synonymous or nonsynonymous). Fisher's exact test cannot be performed when the cell values are zero.

Table 3.5 McDonald-Kreitman Results for Glycosidase

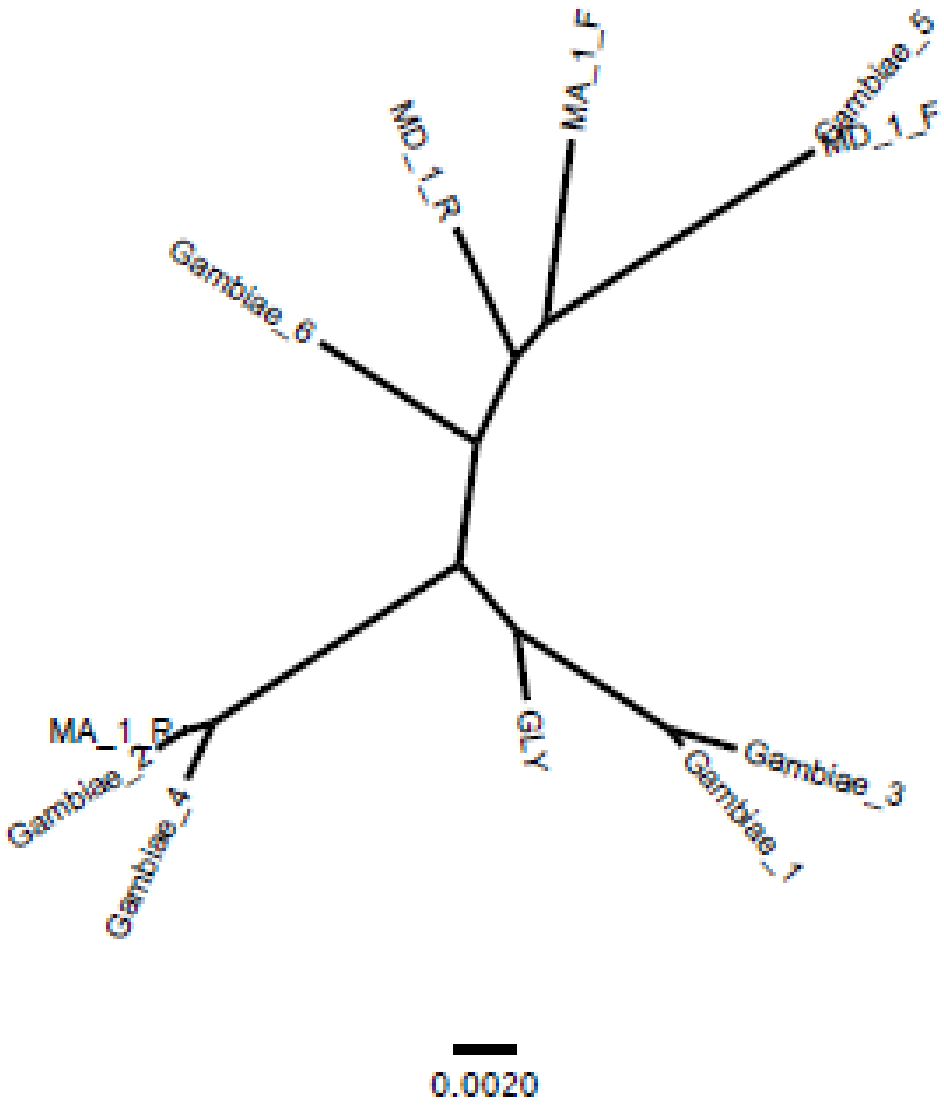
Species Combination	Synonymous	Nonsynonymous	Synonymous Fixed Differences	Nonsynonymous Fixed Differences
A. gambiae vs. A. melas	41	13	0	0

The Tajima's D statistic, displayed in Table 3.6, for *A. gambiae* and *A. melas* are both positive and thus indicate stable populations and do not provide evidence of these genes evolving under positive selection. The neighbor-joining tree in Figure 3.3 presents the taxonomical relationship of the Glycosidase gene within *A. gambiae* and *A. melas.*

Table 3.6 Population Genetics Parameters for Glycosidase

Species	Number of Sequences	Tajima's D Value	P-Value	Nucleotide Diversity
A. gambiae	6	.029	.519	.02111
A. melas	4	.048	.561	.02015

Figure 3.3 Neighbor-Joining Tree for Glycosidase. GLY indicates the reference sequence for glycosidase.



Discussion

The McDonald-Kreitman Test revealed no fixed differences for Anophelin, Ichit, or Glycosidase between any of the species included in the study. Tables 3.1, 3.3, and 3.5 present the ratio of synonymous polymorphic sites over nonsynonymous polymorphic sites for each species

combination. There was only variation within species. When analyzing the D_N/D_S values, those that are greater than one suggest the samples are undergoing positive Darwinian selection, while those values less than one imply stabilizing selection, and a ratio of exactly one indicating neutral selection or a possible accumulation of positive and purifying points that cancel each other out. The positive results of the Tajima's D statistics and no fixed differences found between any of the species through the McDonald-Kreitman Test indicate that Anophelin, Ichit, and Glycosidase are highly conserved genes across species. Introgression or recent evolution may contribute to the lack of fixed differences. For example, the rise in agricultural settlements may have caused forest populations of *A. gambiae* to come into secondary contact with *A. arabiensis*, resulting in hybridization and introgression from *A. arabiensis* into *A. gambiae* of chromosomal inversions beneficial in arid environments and helped *A. gambiae* populations spread to the drier savannas of Africa (White *et al.* 2011). Alternatively, when species share polymorphisms due to a recent ancestry, there would not be enough time for drift or selection to produce fixed differences.

The neighbor-joining trees present a way to visualize the distance matrix between different taxa. The neighbor-joining trees for Anophelin and Ichit both indicated introgression between two *A. gambiae* and three *A. arabiensis* samples. While the tree for Glycosidase showed introgression between one *A. melas* and two *A. gambiae* samples. The introgression between *A. gambiae* and *A. arabiensis* was not a surprise because of evidence of introgression provided by crossing experiments and genetic studies (della Torre *et al.* 1997). The introgression between *A. gambiae* and *A. melas* was unexpected because of the decreased likelihood of introgression between the two species (White *et al.* 2011). The introgression presented in the trees confirms positive results of the Tajima's D test and lack of fixed differences found through the McDonald-Kreitman Test.

As presented by Morlais *et al.* (2004), the average nucleotide diversity per gene for 35 conserved genes from *A. gambiae* was .0079, ranging from .0008 to .0266. The nucleotide diversity values of *A. gambiae* from Anophelin, Ichit, and Glycosidase fall within this range. This implies that these genes are also conserved. The lack of fixed differences obtained in this study demonstrates that Anophelin and Ichit are highly conserved genes between *A. gambiae, A. melas,* and *A. arabiensis* and Glycosidase is highly conserved between *A. gambiae* and *A. melas.* A high level of synonymous polymorphism and no fixed differences (synonymous or nonsynonymous) indicates polymorphisms between species.

The proteins we chose to study are known for their role in blood feeding. For example; Anophelin prevents coagulation during feeding, Ichit's antimicrobial properties protect the mosquito against foreign bacteria on the host, Lysozyme prevents growth of microbes in ingested sugar and blood meals, and Glycosidase breaks down complex carbohydrates that are present in nectar and blood meals. Changes in the amino acid sequences' of these genes may possibly alter a mosquito's ability to obtain and digest complex carbohydrates and nutrients in blood. We expect salivary genes to interact with the host's immune system since there is an exchange of saliva and blood between them. Prolonged contact between mosquitoes and humans can lead to co-evolution and the development of an immune response to mosquito salivary products, as well as a natural desensitization to the allergic reaction caused by them (Tosta, 2007). Highly conserved genes can be signs of functionally constrained sequences, indicating change would not be advantageous to the organism (Casillas *et al.* 2007).

Chapter IV

CONCLUSION

By comparing the patterns of polymorphism of these salivary genes within members of the *A. gambiae* complex, we determined that they have not evolved under positive selection. Instead, these genes did not have any fixed differences and proved to be highly conserved after running the McDonald-Kreitman test. This could be due to the important roles that these genes play in helping a mosquito acquire a successful blood meal. For example, Anophelin prevents coagulation during feeding, Ichit's antimicrobial properties protect the mosquito against foreign bacteria on the host, Lysozyme prevents growth of microbes in ingested sugar and blood meals, and Glycosidase breaks down complex carbohydrates that are present in nectar and blood meals. However, when the genetic diversity of mosquito populations does show positive selection, these salivary genes could be studied in an attempt to target those that are adapting to a host's immune system or exposure to the human malaria parasite [*Plasmodium spp*.]. These genes could lead to new ideas in vector control and population replacement techniques (Drame *et al.* 2010). With existing vector control programs being restricted due to insecticide and drug resistance, it is essential to identify genes that aid and abet in the transmission of malaria through the salivary glands of the A. gambiae complex (Slotman et al. 2007).

REFERENCES

Afrane, Y. A., Githeko, A. K., & Yan, G. (2012). The ecology of *Anopheles* mosquitoes under climate change: case studies from the effects of deforestation in East African highlands. Annuals of the New York Academy of Sciences, 1249, 204-210.

Arca, B., Lomardo, F., Lanfrancotti, A., Spanos, L., Veneri, M., Loius, C., Coluzzi, M. (2002). A cluster of D7-related genes is expressed in the salivary glands of the African malaria vector *Anopheles gambiae.* Insect Molecular Biology, 11, 47-55.

Ayala, F.J., Coluzzi, M. (2005). Chromosome speciation: humans, *Drosophila*, and mosquitoes. Proceedings of the National Academy of Sciences, 102, 6535–42.

Besansky, N. J., Powell, J. R., Caccone, A., Hamm, D. M., Scott, J. A., & Collins, F. H. (1994). Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between principal malaria vectors. Proceedings of the National Academy of Sciences, 91, 6885-6888.

Casillas, S., Barbadilla, A., Bergman, C. (2007). Purifying Selection Maintains Highly Conserved Noncoding Sequences in Drosophila. Molecular Biology Evolution. 24(10):2222-2234.

Coetzee, M., Craig, M., & Le Sueur, D. (2000). Distribution of African Malaria Mosquitoes Belonging to the *Anopheles gambiae*. Complex. Parasitology today, 73, 483-497.

Coluzzi, M. (1964). Morphological divergences in the Anopheles gambiae complex. Riv Malariol, 43, 197-232.

Coluzzi, M., Sabatini, A., Petrarca, V., & Deco, M. (1979). Chromosomal differentiation and adaptation to the human environments in the *Anopheles gambiae* complex. Transactions of the Royal Society of Tropical Medicine and Hygiene, 16, 74-77.

Dana, A., Hong, Y., Kern, M., Hillenmeyer, M., Harker, B., Lobo, N., Hogan, J., Romans, P., and Collins, F. (2005). Gene expression patterns associated with blood-feeding in the malaria mosquito *Anopheles gambiae*. BMC Genomics. 6:5.

Danalev, D. (2012). Inhibitors of Serine Proteinases from Blood Coagulation Cascade-View on Current Developments. Mini Reviews in Medicinal Chemistry, 12, 721-730.

Das, S., Radtke, A., Choi, Y., Mendes, A., Valenzuela, J., Dimopoulos, G. (2010). Transcriptomic and functional analysis of the *Anopheles gambiae* salivary gland in relation to blood feeding. BMC Genomics, 11, 566.

Della Torre, A. (1997). Polytene chromosome preparation from anopheline mosquitoes, pp. 329–336 inMolecular Biology of Disease Vectors: A Methods Manual, edited by J. M. Cr

Dimopoulos, G., Seeley, D., Wolf, A., & Kafatos, F. C. (1998). Malaria infection of the mosquito *Anopheles gambiae* activates immune-responsive genes during critical transition stages of the parasite life cycle. The EMBO Journal, 17, 6115-6123.

Drame, P.M., Poinsignon, A., Besnard, P., Cornelie, S., Le Mire, J. (2010). Human antibody responses to the *Anopheles* salivary gSG6-P1 peptide: A novel tool for evaluating the efficacy of ITNs in malaria vector control. PLoS ONE 5, e15596.

Fang, G., Rocha, E., Danche, A. (2005). How essential are nonessential genes. Molecular Biology Evolution, 22, 2147-2156.

Francischetti, I.M., Valenzuela, J.G., & Ribeiro, J.M. (2010). Anophelin: kinetics and mechanism of thrombin inhibition. Biochemistry, 38, 16678-16685.

Ghosh AK, Devenport M, Jethwaney D, Kalume DE, Pandey A, et al. (2009) Malaria Parasite Invasion of the Mosquito Salivary Gland Requires Interaction between the Plasmodium TRAP and the Anopheles Saglin Proteins. PLoS Pathog 5(1): e1000265. doi:10.1371/journal.ppat.1000265

Kamdem, C., Fossog, B. T., Simard, F., Etouna, J., Ndo, C., Kengne, P., & Costantini, C. (2012). Anthropogenic Habitat Disturbance and Ecological Divergence between Incipient Species of the Malaria Mosquito *Anopheles gambiae.* PloS one, 7, e39453.

Kang, D., Romans, P., Lee J. (1996). Analysis of a lysozyme gene from the malaria vector *Anopheles gambiae*. Gene, 174, 239-244.

Kato, H., Jochim, R. C., Gomez, E. A., Uezato, H., Mimori, T., Korenaga, M., & Hashiguchi, Y. (2012). Analysis of salivary gland transcripts of the sand fly *Lutzomyia ayacuchensis*, a vector of Andean-type cutaneous leishmaniasis. Infection, Genetics and Evolution, 13, 56-66.

Li, B., Calvo, E., Marinotti, O., James, A., Paskewitz, S. (2005). Characterization of the c-type lysozyme gene family in *Anopheles gambiae*. Genetics doi:10.1010.16.

Librado, P. and Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452.

Mans, B. J., Calvo, E., Ribeiro, J. M., & Andersen, J. F. (2007). The crystal structure of D7r4, a salivary biogenic amine-binding protein from the malaria mosquito *Anopheles gambiae.* Journal of Biological Chemistry, 282, 36626-36633.

Morlais I, Ponçon N, Simard F, Cohuet A, Fontenille D. (2004). Intraspecific nucleotide variation in Anopheles gambiae: new insights into the biology of malaria vectors. Am J Trop Med Hyg. (6):795-802. PubMed PMID:15642974.

Mendes, A. M., Awono-Ambene, P. H., Nsango, S. E., Cohuet, A., Fontenille, D., Kafatos, F. C., & Vlachou, D. (2011). Infection intensity-dependent responses of *Anopheles gambiae* to the African malaria parasite *Plasmodium falciparum*. Infection and immunity, 79, 4708-4715.

Mueller AK, Kohlhepp F, Hammerschmidt C, Michel K. (2010) Invasion of mosquito salivary glands by malaria parasites: prerequisites and defense strategies. Int J Parasitol. 40(11):1229-3

Mulenga, A., Khumthong, R., & Chalaire, K. C. (2009). *Ixodes scapularis* tick serine proteinase inhibitor (serpin) gene family; annotation and transcriptional analysis. BMC Genomics, 10, 217.

Okara, R. M., Sinka, M. E., Minakawa, N., Mbogo, C. M., Hay, S. I., & Snow, R. W. (2010). Distribution of the main malaria vectors in Kenya. Malar J, 9, 69.

Patil, D. P., Atanur, S., Dhotre, D. P., Anantharam, D., Mahajan, V. S., Walujkar, S. A., & Shouche, Y. S. (2009). Generation, annotation, and analysis of ESTs from midgut tissue of adult female *Anopheles stephensi* mosquitoes. BMC Genomics, 10, 386.

Rambaut A (2009) FigTree: Tree figure drawing tool, ver 1.3.1. http://tree.bio.ed.ac.uk/software/figtree/. Institute of Evolutionary Biology, University of Edinburgh.

Rozen, Steve, Skaletsky, Helen. (2000). Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386

Service, M.W., (1993). The *Anopheles* Vector: Bruce-Chwatt's Essential Malariology. pp. 96-123. London, Boston, Melbourne, Auckland.

Sinka, M. E., Bangs, M. J., Manguin, S., Coetzee, M., Mbogo, C. M., Hemingway, J., & Hay, S. I. (2010). The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. Parasitology Vectors, 3, 117.

Slotman, M., Parmakelis, A., Marshall, J., Awono-Ambene, P., Antonio-Nkondjo, C., Simard, F., Caccone, A., & Powell, J. (2007). Patterns of Selection in Anti-Malarial Immune Genes in Malaria Vectors: Evidence for Adaptive Evolution in LRIM1 in *Anopheles arabiensis*. PLoS ONE, 8, 793.

Takken, W., Eling, W., Hooghof, J., Dekker, T., Hunt, R., & Coetzee, M. (1999). Susceptibility of *Anopheles quadriannulatus theobald* (Diptera: Culicidae) to *Plasmodium falciparum*. Transactions of the Royal Society of Tropical Medicine and Hygiene, 93, 578-580.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28: 2731-2739.

Tosta, C. E.(2007). Coadaptation and malaria control. Memorias do Instituto Oswaldo Cruz. 102, 385-404

White, B., Collins, F., & Besansky, N. (2011). Evolution of *Anopheles gambiae* in relation to Humans and Malaria. Annual Review of Ecology, Evolution, and Systematics, 42, 111-32.

Ye, Y.G., Zhang, R.L., Zhao, Y.R., Geng, Y.J., Huang, D.N., Gao, S.T., Li, X.G., Liu, C.Y., Zhu, X.Q., Lin, R.Q. (2011). Identification of one putative D7 related gene specifically expressed in female *Anopheles anthropophagus.* Journal of Animal and Veterinary Advances, 10, 2618-2623.