MOLECULAR CHARACTERISTICS AND PUTATIVE FUNCTIONS OF REPRODUCTION-ASSOCIATED PROTEINS OF*ANOPHELES GAMBIAE* MOSQUITO

ACHINKO DANIEL AYEHNGWO

A Thesis Submitted to the Graduate School in Partial Fulfillment of the Requirements for the Award of the Doctor of Philosophy Degree in Biochemistry of Egerton University

EGERTON UNIVERSITY

May, 2013

DECLARATION AND RECOMMENDATION

DECLARATION

This thesis is my original work and has not been submitted in part or whole for an award in any other institution.

Mr. D. A. Achinko

SD14/0374/10

Signature: ……..... Date: ……

RECOMMENDATION

This thesis has been submitted for examination with our approval as university supervisors

Dr. P. Mireji,

Egerton University, Kenya

Signature: …….. Date: …….

Dr. D. Masiga

International Center for Insect Physiology and Ecology (ICIPE), Kenya

Signature: ...……. Date: …..……

Pr. F. Catteruccia

Universita` degli Studi di Perugia, Italy.

Signature: …… ……..….. Date:…….……....

COPYRIGHT

DEDICATION

I dedicate this work to my lovely and God given parents Mr. Enjuweson Philip Achinko and Mrs. Achinko Theresia Ajala, to my brothers; Achinko Godlove Egoh and Anyangwa Elias and to my sisters Achinko Elisabeth Ajabia, Achinko Ambangkoh Mary, Achinko Estella Endah, Enjuweh Miscline Abla and Enjuweh Joycelyn Mbe; for their endless support and prayers to see me succeed in the things I do and become a responsible man.

ACKNOWLEDGEMENT

I thank Dr. Paul O. Mireji my university supervisor for tirelessly going through my work, correcting it and advising me on how to face the scientific future. I wish to thank Dr. Daniel Masiga, my *icipe* supervisor for accepting to work with me as his student and equally advising me on research methods despite my short comings. I wish to thank Prof. Flaminia Catteruccia for accepting me into her laboratory in Terni, Italy, advising me all the time on my work progress, supporting me financially during my difficult moments and accepting me as part of her team. In the same vain I also thank the Terni lab members for being there for me all the time when I needed them and for supporting me through the language difficulties. I would like to thank *icipe* scholars association (IScA) and the *icipe* family especially Ms. Lilian Igweta (Coordinator of student's affairs) who stood by me all through my research at icipe and always shared my joyful and sad moments. I am grateful to my friends at icipe Mbita who accepted me and worked with me to see me succeed. I would like to thank my friends Ayuka Teboh Fombong and Tchouassi Poumo David for their endless brotherly love and support through my difficult moments in Nairobi. They always read through my work, guided me on my PowerPoint presentations and cooked delicious meals. With them, I will also thank Janis Thailayil, Muangala Robert and my friends in the UK who always cared about my work and prayed for me. I wish to give special thanks to my fiancé Nsang Solange, for loving me through all these difficult moments of my studies. I would like to appreciate my family in the USA especially my aunts Lydia Anagho and Enjuweh Isabel and their husbands Mr. Anagho Philip and Mr. Pougoue Martin respectively. They always checked on me ensuring that I had the moral support I needed to keep going with my work. I thank my cousins Gloria Abo and Sherade Anagho for supporting me. I thank my parents, brothers and sisters who always prayed for me, encouraged me, advised me on how to face life far away from home and above all for loving me so much. I wish to thank *icipe* headed by Prof. Christian Borgemeister and DAAD for providing the facilities and infrastructure needed for my work and the scholarship, respectively. I would like to thank EU-FP7 MALVECBLOK grant which assisted me in the realization of these thesis by organizing workshops for Ph.D students and facilitating capacity building through the acquiring of research skills in the different labs involved with the project. Above all, I thank God Almighty for always having mercy on me, guiding me in my research, watching and protecting over me from the hands of the evil one. May He take the glory.

ABSTRACT

In *Anopheles gambiae* mosquito vector of malaria, mating is mediated by transfer of a plug (mating plug protein) from males to females. In the females, the constituents of the plug include male accessory glands (MAGs) proteins and female specific proteins that influence ovulation, oviposition, lifetime refractoriness to mating and other post-mating behaviors in the females. Manipulation of the plug can significantly affect the reproduction potential and concomitant population density of the vector. Investigations were conducted to establish in the mosquito 1) putative structural properties of MAGs proteins (transglutaminase and plugin) mediating successful interaction for mating-plug formation in the male *An. gambiae* 2) putative interacting proteins potentially mediating the mating-plug formation and post-mating phenotype, 3) mating responsive molecular factors in the proteins, 4) influence of the interactions among the proteins on mating responses and 5) patterns and roles of swarms in mating responses in females. Structural properties of *An. gambiae* MAGs (transglutaminase and plugin) proteins were identified using Protein-Predict software and type of selection pressure among related orthologs in different taxa was determined through phylogenetic analysis. Orthologs of the *An. gambiae* plug proteins in *Drosophila melanogaster* were obtained from STRING database and their interactions (protein-protein) predicted using CYTOScape software. Phylogeny among key *D. melanogaster* interacting protein homolog equivalents in *An. gambiae* (interologs) was determined by maximum likelihood phylogenetic method. Pathway analyses of the interologs in Reactome database was used to identify proteins involed in plug formation and post-mating phenotype. Expression of mRNA for genes of the proteins in males (MAGs and testes) and females (atria, ovary and spermatheca) reproductive tissues was determined by RT-qPCR, and identification of the genes verified through sequencing of the amplicons and western blot analyses. Role of the genes in mating responses in the mosquito was verified through RNAi technique. Patterns and roles of swarms in mating responses in females were determined through cage releases and observations in a semi–field setting. Plugin was predicted to be non-structural and under positive selection among the MAGs proteins, and also identified as the main protein in mating plug formation on the protein-protein interaction network. Plugin was also shown to putatively interact with TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) in the female atria after mating through a putative two response regulatory pathway. Plugin (AGAP009368), TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) and CATHEPSIN B (AGAP004533) appeared to interact and influence postmating effects in the females. CATHEPSIN B (AGAP004533) was co-expressed and regulated by TRYPSIN-LIKE

SERINE PROTEASE (AGAP005195) in the female. TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) specifically interacted with plugin on the mating plug and was transported to the spermatheca after mating, implicating it in postmating events within the females. Knockdown of CATHEPSIN B suggested that plugin was alternatively spliced in the male. The knockdown implicates CATHEPSIN B and TRYPSIN-LIKE SERINE PROTEASE in a common reproductive pathway in females, supported by mating propensity phenotype observed in the females. A swarming window of about 30 minutes was identified for *An. gambiae* mosquito where 41.3% of females were successfully inseminated. Interaction between CATHEPSIN B and plugin in the male then CATHEPSIN B and TRYPSIN-LIKE SERINE PROTEASE in the female produces successful interactions which are involved in reproductive pathways within the male and the female mosquitoes hence potentially and significantly affecting male and female reproductive processes. Detailed molecular analysis through high throughput interventions can potentially elucidate actual binding sites between CATHEPSIN B and its partner in the male (plugin) and female TRYPSIN-LIKE SERINE PROTEASE mosquitoes which can be exploited for vector control interventions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

aLRT	Approximate Likelihood-Ratio Test
dsRNA	double Stranded RNA
HMM	Hidden Markov Model
IHC	Immunohistochemistry
IMC	Isolation of mating couples" technique
IPT	Intermittent presumptive treatment
LRT	Lower reproductive tract
M and S	Molecular forms of *Anopheles gambiae* –Mopti and Savannah
MAGs	Male Accessory Glands
mtDNA	Mitochondrial DNA
PHYML	Maximum likelihood phylogeny
rDNA	Ribosomal DNA
RIDL	Release of Insects carrying Dominant Lethals
S.O.C	Super Optimal broth medium
SDS-PAGE	Sodium dodecyl sulphate Polyacrylamide gel electrophoresis
SIT	Sterile Insect Technique
SXL	Sex lethal
TAE	Tris-acetate-ethylenedianinetetraacetic acid
YPPs	Yolk Protein Precursors
lacZ	lac operon (beta galactosidase)
RNAi	RNA interference
BLAST	Basic Local Alignment Search Tool
BLASTp	Protein BLAST
BLASTt	Translational Blast
BLASTn	Nucleotide Blast
GABA	Gamma Amino acid Butyric Acid
PTHR	PANTHER FAMILY Information domain database

APPENDICES

CHAPTER 1
INTRODUCTION

1.1 Background information

Anopheles gambiae being the most important malaria vector mosquito in Africa (Collins and Paskewitz, 1995) is highly anthropophilic; hence, its history of evolution may have been structured mainly by the demography of human populations in Africa (Coluzzi *et al*, 1979; Coluzzi *et al*, 1985). This therefore makes vector Control a primary concern in the control of malaria transmission (Macdonald, 1957). Malaria transmission can easily be curtailed if the vector population density is reduced below a critical level. This can be attained using Sterile Insect Technique (SIT) (Knipling, 1955). Hypothetically it has been shown that if sterile males are initially released at a sufficient rate capable to cause a decline in wild populations of mosquitoes, then releasing a constant number of sterile males into wild mosquito populations per generation while sustaining it, will progressively achieve a higher degree of suppression of this wild population in each successive generation (Knipling, 1955). A similar research has shown that mosquito vectors with high reproductive rates play very important roles in their vectorial capacity (Tripet *et al*, 2003). Mosquitoes have been shown to copulate only once in the course of their lives (Tripet *et al*, 2003), therefore interfering with the mating process remains a promising area for research into vector control.

Studying mechanisms that regulate reproduction in the mosquito could possibly identify new targets which can be used to develop tools to break the reproduction cycle. However the knowledge behind the molecular or physiological events related to insemination and mating, known as; sperm storage, oviposition, and inhibition of re-mating in thefemale vector of malaria is still poorly understood (Rogers *et al*, 2009). For ovulation and oviposition to take place in the female mosquito successful copulation is required (Klowden, 2001). At post-copulation, the female *An. gambiae* experiences a significant change in behaviour including egg-laying induction in blood fed females accompanied by a serious reduction in mating-receptivity to males for further insemination (Klowden, 2001). Prominent transcriptional changes takes place in the female mosquito after mating, which unveils a dedicated ‘‘mating machinery’’ within the reproductive tract. It comprises of molecular and structural parameters that are switched off or altered intensely after copulation (Rogers *et al*, 2008).

Seminal fluid proteins derived from male accessory glands (MAGs) are known so far to be transferred to females during copulation and known to regulate storage of sperms and viability, trigger oviposition and reduce the female’s receptivity to re-mating in *D.*

melanogaster insect species (Ravi *et al,* 2007). Seminal fluid proteins and their importance to mating biology in *An. gambiae* are quite limited, and their proteomic role in reproduction is based mainly on similar functional class predictions between *Anopheles* and *Drosophila* Acps (Dottorini *et al,* 2007). MAGs secretions, besides egg laying stimulation and also reducing sexual reception have also been shown to initiate the expression of immune peptides and likewise reducing the female's lifespan (Wolfner, 1997; Kubli, 2003; Chapman *et al,* 2004). Therefore MAGs products essentially modulate female behaviour and identifying vital targets for use in biological and genetic control of insect pests (Chapman *et al,* 2004) will be of significant importance. A gelatinous mating plug is produced by the MAG secretions during mating in *An. gambiae* and it is deposited in the atrium of the female reproductive tract (Gillies, 1956; Giglio *et al,* 1966). In invertebrates, reptiles, and mammals mating plugs are very common and important for reproduction (Birkhead *et al,* 1998), however, among mosquitoes; they have been shown to be exclusive to *Anopheline* species (Yuval, 2006). The *An. gambiae* plug is formed entirely within the male and is digested by the female over a period of 24 - 48h. However, the genotypes and distinct components and functions of most of the *An. gambiae* MAGs are poorly understood.

The availability of *An. gambiae* genome (Holt *et al,* 2002) in the databank, multiplicity of bioinformatics tools and development of novel technologies for gene silencing and transfer in the vector (Catteruccia *et al,* 2000; Grossman *et al,* 2001) has facilitated the manipulation of mosquitoes genetically (Kim *et al,* 2004; Abraham *et al,* 2005). This has potentially helped in interrogating the novel properties of MAGs in reproductive modulation in the mosquito.

1.2 Statement of the problem

Anopheles gambiae, the principle vector of malaria is responsible for more than 2.5 million cases of morbidity and mortality related to the disease. Most vector control techniques applied this far have met serious draw-backs like insecticide resistance and failure in transgenic techniques related to uncompetitive male mosquitoes in the wild. The failure in most recently used transgenic techniques for reproductive control resulted from the paucity in knowledge on the reproductive biology of the Anopheles mosquito species. Reproductive efficiency within the *gambiae* complex is linked to spatio-temporal swarm formation with successful insemination comprising of sperm transfer to the female and consequent formation of the mating plug. Understanding mosquito swarming patterns in semi-field scenarios and swarm related genes will be a good guide to dissect the competitive nature of these mosquito species in the wild and also, elucidating the molecular properties of proteins responsible for

the successful formation of the plug and post-mating interactions within the female, will provide novel targets for use in mosquito transgenesis and its application in reproductive control of this mosquito species.

1.3 Justification of the study

Population dynamics in *An. gambiae* has a significant influence on malaria transmission in endemic areas. The use of biotic (larvivorous fish) and abiotic (insecticides) vector control interventions has successfuly reduced transmission hence reducing the populations of the mosquito. However, the vector has been able to develop mechanisms to surmount these intervention approaches, necessitating a need to prospect for other alternative methods to assuage the existing tools. Current transgenic techniques like the Sterile Insect Technique (SIT) have potential application in reducing the population of the vector. Male accessory glands (MAGs) proteins mediate reproduction success in the mosquito, and are a viable target for disrupting mating in the mosquito. The MAGs proteins produce a mating plug mainly consisting of Plugin and Transglutaminase enzymes that is deposited in the female atria after mating. The plug is rapidly digested (within 24-48hrs) in atria, simultaneously imparting in the mosquito a life-long refractory phenotype to subsequent mating. It is postulated that the digestion also influences oviposition, ovulation and blood-meal search post-mating behaviour in the mosquito. These changes in phenotypes of the mosquito are accompanied by differential genetic responses within the atria and spermatheca of the female. Specific trypsin proteases (Chymotrypsin-Like Serine Protease and Trypsin-Like Serine Protease) have been identified in the plug and is down-regulated 24h post-mating. Equivalent of MAGs proteins (Acp) in *D. melanogaster* with similar post mating phenotype in the females with orthologs in *An. gambiae* have been identified. However, their molecular characteristics and specific functional role in the mosquito have not been elucidated, and yet hold key to genetic manipulation of the mosquito reproduction, with potential application in control of population of the vector and malaria transmission in nature.

1.4 General hypothesis

Mating-associated-proteins do not possess molecular and functional characteristics that can potentially be targeted to disrupt mating in *An. gambiae* to reduce populations of the malaria vector in nature.

1.5.1 Specific hypothesis

1. Putative protein structural properties of MAGs proteins (transglutaminase and plugin) do not favour their successful interaction for mating plug formation in *An. gambiae.*
2. Putative Protein-protein interactions neither mediate mating-plug formation nor influence post-mating phenotype in *An. gambiae.*
3. Molecular factors among the interacting proteins do not mediate mating responses in *An. gambiae.*
4. Interactions among the interacting proteins do not influence mating responses in *An. gambiae*
5. Swarming patterns in male *An. gambiae* mosquitoes do not influence mating responses in respective females.

1.5 General objective

To determine molecular characteristics and functional roles of mating-associated-proteins that can potentially be targeted to disrupt mating in *An. gambiae* to reduce populations of the malaria vector in nature.

1.5.1 Specific objectives

1. To determine putative protein structural properties of MAGs proteins (transglutaminase and plugin) mediating successful interactions for mating plug formation in *An. gambiae*
2. To determine putative protein-protein interactions mediating mating-plug formation and post-mating phenotype in *An. gambiae.*
3. To document molecular factors among the interacting proteins mediating mating responses in *An. gambiae.*
4. To determine influence of the interactions among the interacting proteins on mating responses in *An. gambiae.*
5. To describe swarming patterns in male *An. gambiae* mosquitoes and their role on mating responses in respective females.

CHAPTER 2
LITERATURE REVIEW

2.1 Malaria in Sub Sahara Africa

Half of the world population is approximately at risk of malaria, and estimatively 243 million cases of infection resulted in about 863,000 deaths in 2008 (WHO, 2009). Ninety-one percent of all malaria-related deaths take place in sub-Saharan Africa (SSA), where, an estimated annual loss of 35.4 million Disability Adjusted Life Years with 85% of deaths occurring in children below five years of age is said to result from malaria. Approximately 40% of all public health spending in SSA is malaria oriented (WHO, 2010).

Given the distressing records observed, lowering the number of malaria cases and related deaths to about 50% over the last 10 years have been documented in several high burden African countries (O'meara *et al,* 2010) including Eritrea, Rwanda, Zanzibar (WHO, 2009), Pemba (Jaenisch *et al,* 2010), Tanzania mainland (Mmbando *et al,* 2010), Kenya (O'meara *et al,* 2008) and Zambia (Chizema-Kawesha *et al,* 2010). The reduction in infection rates and disease burdens in general as well as declining asymptomatic carriers are seriously considered a consequence of improved health systems quality, alongside improved case management, such as improved diagnosis and use of highly effective drugs against malaria. The burden of malaria has been reduced significantly through increased expenditures in intervention programmes aimed at specifically achieving high coverage using; bed nets, indoor residual spraying (IRS) campaigns and intermittent presumptive treatment (IPT) implementation in vulnerable groups (WHO, 2009).

Malaria control interventions have been scaled up in high-endemic countries and this has without doubt contributed a great deal to the observed reduction in malaria cases and deaths. Other factors however not known to be involved in intervention on mosquito vectors could have a potential impact on them, hence reducing transmission, which will subsequently reduce the number of infected cases. Some of these factors include; urbanization, land use and variations in agricultural practices, and economic development which results in for example better housing construction (O'Meara *et al,* 2008). Global climatic Changes resulting in rainfall, humidity and temperature known as global effects could also affect malaria vectors (Rogers *et al,* 2006; Stresman, 2010; Parham *et al,* 2010). Given the importance of these multiple determining factors and their important roles in reducing the malaria burden, majority of studies carried out in the field report on changing the focus of malaria epidemiology mainly to the change in emerging human infections and diseases. In this light,

limited efforts in research have also been invested in systematic collection of data and also mapping trends in vector densities with long-term effects and other related parameters to vector transmission areas where reduction in malaria infections and the disease burden have been reported (Meyrowitsch *et al*, 2011).

2.2 Classification of *Anopheles gambiae* mosquitoes

The principal malaria vectors in the *An. gambiae* complex are *An. gambiae sensu stricto* (s.s.) and *An. arabiensis*. For the other members of the complex, *An. quadriannulatus* species A, is widespread in southern Africa, and *An. quadriannulatus* species B is found in Ethiopia and they are both considered to be non-malaria vectors and zoophilic (Coetzee *et al*, 2000; Coetzee, 2004). *Anopheles melas* and *An. merus* both breed in salt water and are therefore considered important vectors only in coastal regions (Moreno *et al*, 2004; Tsy *et al*, 2003). The last member of the complex named *An. bwambae* is found mainly in a region close to the Buranga hot springs in Uganda (White, 1985). The proper identification of the *An. gambiae* complex and each of its sibling species has far-reaching consequences practically, and this is mainly due to their underlying behavioral and bionomic differences (Van Bortel *et al*, 2001). Classical mosquito taxonomy cannot be applied to these species because they cannot be distinguished morphologically (Krzywinski *et al*, 2003). Therefore the implementation of a multidisciplinary approach including morphological, molecular, distribution, and bionomic data is required to resolve these specific taxonomic differences regarding this species (Nguyen, 2000).

The subfamily *Anophelinae* is commonly subdivided into *An. Bironella*, and *Chagasia* genera. *Anopheles* is again subdivided into *An. Cellia*, *Kerteszia*, *Lophopodomyia*, *Nyssorhynchus*, and *Stethomyia* subgenera, and infra-subgeneric informal groupings. Despite the fact that Ross hypothetically first explained the evolutionary relationships 50 years ago within the Culicidae family, a hand full of phylogenetic studies including the genus *Anopheles* was considered until relatively recently (Ross, 1952). The use of rDNA, mitochondrial DNA (mtDNA), and single-copy nuclear genes were the modern morphological and molecular approaches used to strongly support very close relationships, including monophyly of the subfamily *Anophelinae* (Krzywinski *et al*, 2003) which includes the genus *Anopheles*.

Anopheles genus, subdivided into seven subgenera was based on the positions of specialized setae and number on the male genitalia gonocoxites. They were *An.Myzomyia*, *Nyssorhynchus*, *Stethomyia*, *Baimaia* and *Lophopodomyia* subgenera (Foley, 1998;

Krzywinski, 2003). Mosquitoe's classification into species is based on morphological characteristics: - wing spots, anatomy of the head, larval and anatomy of the pupa, structure of the chromosome, and more recently using DNA sequences. The current number of species identified within the subgenera includes 206 *Anopheles* species.

In Kenya, *An. gambiae* complex consist of *An. gambiae sensu stricto, An. arabiensis* and *An. Merus* (Coluzzi *et al,* 1985; Collins *et al,* 1988; Petrarca and Beier, 1992). *Anopheles gambiae s.s* and *An. arabiensis* within the complex are closer to humans and are considered the major vectors of malaria in the country (Muirhead-Thomson, 1951; Highton *et al,* 1979). These two species in Kenya have a distribution which overlaps and they occur in sympatry in large areas of tropical Africa. Predominantly *An. gambiae s.s* and *An. arabiensis* mosquitoes are present along the coastal, western and parts of central Kenya (Githeko *et al,* 1993; Mbogo *et al,* 1995). Their spatio temporal distribution in the wild in not currently known and therefore understanding their swarming patterns in real time and mating competitive behaviour will help in understanding species conservation through several generations.

2.3 Life cycle and control of *Anopheles gambiae* mosquito

The summarized mosquito life cycle as described by Service, (1980) and Lindsay *et al,* (2000), can be as follows: After mating and obtaining a blood meal, the female *Anopheles* lays approximately 50-200 small (1mm long) brown or blackish boat-shaped eggs on the surface of water. Usually, the eggs are white when laid freshly, and later turn to brown, then finally black upon maturing. Viable eggs hatch into larvae in about 2-3 days in the tropics, but in temperate regions which are cooler, the eggs may hatch only after 4-7 days or longer. On the water surface, the larvae, lies parallel to it so as to allow intake of air and surface feeding. When the mean temperatures of water is 25-28^{o}C, larvae molts four times within 6-9 days reaching the pupal stage, and this stage lasts for 2-3 days depending on the temperature conditions. Hence, the immature developmental stage per generation may last for 10-11 days. The males are known to emerge first as adultsfor every progeny of any one egg batch. Within 24 hrs post-emergence the males become active for mating so that when the females emerge; the males are competent to start mating. Swarming as a process takes place prior to mating and during this process the males are known to associate over a marker, flying in a particular manner. Most male mosquitoes are usually seen dead after mating. The successful mating age for males still remains an area of concers and understanding it gives a guide on what age to use for SIT. For the development of ovaries and maturation of eggs, the female *Anopheles* needs a blood meal, which is followed by oviposition of batches of eggs (Gillies, 1955).

Blood meal search occurs usually after mating and therefore understanding swarm drives in females and post-mating events that occur will open new methods of targeting the female before a blood meal. The final percentage of the eggs leading to adult formation is unknown, but mortality is usually heavy, especially among larvae and this is due to; predators, disease, drought, and flushing (Gillies, 1955). Predators effect in mosquito control is poorly understood and bridging this gap by identifying suitable predators able to control mosquito populations in swarms can be a promising area of vector control intervention. The loss of Larvae due to predation is one of the main factors reducing the numbers of larvae which actually develop into adults. Predation of larvae in established pools has been recognized to be an important factor in limiting larval numbers (Lindsay *et al*, 2000). In *C. tigripes* for example, alongside other predators which colonize the same pools as *An. gambiae*, have been shown to predate on *An. gambiae* larvae (Haddow, 1942; Christie, 1958). There is great possibility that similar pressures should exist in other types of permanent waters, hence limiting the productivity for *An. gambiae* larvae. Water movements generated by *An. gambiae* larvae; also favorably increases their vulnerability to predator attack (Service, 1980). Larval predation also provides a good area in preventing mosquitoes from attaining adult stages and going to the swarms to mate and also mating responsive genes and their expression in the larvae through adulthood is poorly understood. This study will help in providing solutions to such areas which can be exploited for better understanding the reproductive biology of the larvae and the adult

Vector control of *An. gambiae* mosquitoes is essential for the reduction of malaria transmission (Molyneux, 1999). It is now shown that Insecticide Treated Nets (ITNs) possess efficiencies operating in a similar manner to indoor spraying programs, and so are more preferred (Curtis, 1998; Armstrong *et al*, 1999). There is an undergoing renaissance to target the genome of pest insects to reduce their deleterious effects and the focus is especially in mosquitoes, with much attention on using transgenesis to modulate vector competence (Aksoy, 2008). The use of transgenic techniques may contribute to the developing of male-specific strains (Papathanos *et al*, 2009) hence providing alternative methods to sterilize mosquitoes (Catteruccia *et al*, 2009). Genes involved in reproduction have been identified but their roles in reproduction are not yet known. Transgenic techniques could be more potent over Sterilizing techniques which have failed due to the uncompetitive nature of such mosquitoes. Transgenesis requires the identification of specific ineracting molecules between the males and females during mating and understanding how swarm competition affects the

regulation of these genes before and after mating. This study attemps to elucidate such molecular targets and their roles in mating

2.4 Vectorial capacity, feeding and resting behavior in *Anopheles gambiae* mosquitoes

Female *Anopheles* mosquitoes obtain gametocytes with the blood meal after an infective bite of the mosquito (Kettle, 1992). Microgametes and macrogametes of the *Plasmodium* parasite fuse to form zygotes, which remain motionless for about 18-24 hours after which they elongate to form motile ookinetes invading and infecting the mosquito vector (Aikawa and Seed, 1980; Kettle, 1992). Several studies related vector populations dynamics to the *P. falciparum* infection incidences and disease in human population (Mbogo *et al,* 1993; Beier *et al,* 1994; Mbogo *et al,* 1995; Kabiru *et al,* 1997). Such studies indicate that for malaria to be transmitted, an individual *Anopheles* must at least feed twice, successively on humans within a week. During the first feeding it acquires the infection and in the second it transmits the parasite to another human individual.

Eighty percent of *Anopheles* feed on any large mammals (Gillies, 1972), and of the known species of *An. gambiae* complex, *An. arabiensis*, *An. merus*, *An. melas* and *An. bwambae* are partly zoophilic and endophilic (White, 1974; Mosha and Petrarca, 1983; Mutero *et al,* 1984). Blood meal studies undertaken in western Kenya on *An. arabiensis* showed that it has a lower proportion of human meals, which is a reflection of exophily at a higher degree (Joshi *et al,* 1975; Highton *et al,* 1979; Petrarca *et al,* 1991; Githeko *et al,* 1994). Cattle feeding have been shown to be generally more preferred by *An. arabiensis* compared to *An. gambiae s.s.* (Githeko *et al,* 1994) known to be primarily endophilic and endophagic, whereas *An. arabiensis* and *An. merus* show some partial degrees of exophily and zoophagy (White, 1974; Coluzzi *et al,* 1979; Gillies and Coetzee, 1987).

Presently in Western Kenya, *An. arabiensis* is the dorminant Anopheles species in most localities and malaria is equally still endemic across the region, therefore the feeding of this mosquito species on humans outdoors still remains a great challenge. Understanding their swarming patterns within the area will be of great importance as this can help in the implementation of malaria control programs within the region.

2.5 Sex differentiation and sexual maturity in *Anopheles gambiae*

Naturally, the achievement of sexual differentiation occurs through several mechanisms, resulting in the determination of morphological, physiological and behavioural traits in a good number of living organisms. Despite these clear differences in the sex

determining machinery and in the individual genes involved, common patterns could be recognized among well distinct taxonomic groups: a primary signal, a main gene and a gene regulating controlwhich leads to a double-switch gene, selecting between sexual programmes alternatively (Scali *et al,* 2005). In *D. melanogaster*, as one of the most genetically and molecularly characterized organisms, the primary signal of sex determination in somatic cells is given as the ratio of X chromosomes to autosomes (X:A ratio). When the X: A ratio= 1.0 (2X: 2A) it defines female development while when the X: A ratio = 0.5 (1X: 2A) dictates male development (Cline, 1993). The key gene sex-lethal (*sxl*) is the only target of the primary signal and it becomes active in females but inactive in males. *Sxl* gene later regulates the splicing of *transformer (tra)* pre-mRNA (Boggs *et al,* 1987). Despite the transcription of *tra* in both sexes, a complete open reading frame (ORF) of its spliced transcript occurs solely in females. The ultimate double-switch gene in the cascade somatic sex determination is the *doublesex (dsx)* gene. Two sex-specific transcription factors (DSX^F in females and DSX^M in males) coded for by *dsx* activates or represses the last genes of the sexually dimorphic traits necessary differentiation. In *An. gambiae*, the *dsx* gene covers an 85 kb region on chromosome 2R composed of several exons, alternatively spliced to produce transcripts specific to females and males (Scali *et al,* 2005). The gap here relates clearly to the fact that the relationship and molecular interaction between *dsx*, male and female reproductive genes is not yet elucidated. Such interactions can be understood through Protein networks *in silico* and RNAi studies which can open up novel targets for transgenesis both at the larval and the adult stages.

Male mosquitoes are not immediately sexually competent post-emergence in many Diptera species, though maturation is known mainly to be a permanent 180° rotation of their genitalia (Marshall, 1938). The male mosquito genitalia comprise 8 to10 segments located abdominally. The claws located at the tip of their claspers found on the 10^{th} segment enables the male grasp the female during copulation (Rees *et al,* 1951). After emergence these male claws are rotated dorsally, preventing them from copulating until the rotation has taken place. Two sets of opposed and crossed muscles drive this rotation (Chevone *et al,* 1976) and can equally happen frequently either counter-clockwise or clockwise (Chevone *et al,* 1976; Roth, 1948). The time required to complete this event is species-specific. For anopheline mosquitoes, the rate of genitalia rotation is not yet known which can be done by observing their behaviours in swarms correlated to their mating age. Based on observations of the insemination status, it was shown that 24 h post-emergence are required at least for mating in *An. arabiensis, An. gambiae* s.s (Verhoek *et al,* 1994) and *An. stephensi* (Mahmood *et al,*

1982). Despite the requirements for maturation in males, the females of a majority of mosquito species are unreceptive within the first 30-60 h after emergence; even though they can allow copulation, they may not become inseminated (Clements, 1999). According to Mahmood and Reisen (1982) females of *An. stephensi* became sexually mature by the 2nd night of life, though the proportion of females inseminated by older males less than 12 h after their emergence was very low. The mechanism underlying the unreceptive nature of female mosquitoes is still poorly understood and this can be related to mating expressed genes at a given age of the mosquito. Also looking at the regulation of such genes before and after swarm formation where these females are mated could provide information on swarm patterns in the field and when mating drives actually occur in the female.

2.6 Reproduction in *Anopheles gambiae* mosquitoes

Swarming by *An. gambiae* males is known to entice females into insemination psychology and improves the success of mating/copulation in males (Thornhill 1976; Steele 1986; Eggert and Sakaluk, 1994). The likely selective context leading to evolution in courtship feeding must have been sexual conflict over paternity for the offspring (Sakaluk and Eggert 1996). A gelatinous mating plug formed by the MAG secretions in *An. gambiae* is deposited into the female reproductive tract (atrium) of the female during mating (Gillies, 1956; Giglioli, 1966). Though mating plugs are common features in the reproductive process of many organisms (Birkhead, 1998), only *Anophelines* among mosquitoes are known to produce the plug (Yuval, 2006). The *An. gambiae* plug is formed entirely within the male and is digested by the female within 24 hrs. Transglutaminase - Plugin protein interactions are responsible for mating plug formation (Rogers *et al*, 2009). Given the fact that the role of the plug has unequally not been functionally elucidated, hypothetically it is postulated that the mating plug of *An. gambiae* blocks access to the spermatheca by serving as a physical barrier to re-insemination (Parker, 1970; Yuval, 2006). Rarely observations of females with two plugs and sperm trapped between them has previously been made (Gillies, 1956), though an alternative hypothesis suggests that another function of the mating plug might be to prevent loss of sperm from the female spermatheca, favoring reproductive success by the males (Gerber, 1970). The particularity of the plug to Anopheles species, Vectors of malria remains an area of interest given that this plug is produced in the male and transferred to the females during mating. Therefore understanding interactions favouring plug formation in the male could identify potential targets which if molecularly compromised could disrupt plug formation in the male and hence interfering with proper female insemination during mating.

2.7 Organogenesis in the male and female *Anopheles gambiae* mosquito

2.7.1 The testes and spermatogenesis

Spermatogenesis in *An. gambiae* is initiated at the early larval stage with acceleration in development. With fluorescent containing sperm in transgenic mosquitoes, testes can be observed at the early larval L2 stage (Papathanos *et al,* 2009). The testes are fully developed in the pupae, while mature sperm are observed as early as 24 hours after pupation. A 'hub' known to be present in the testes is present at the anterior end where the primordial germ cells are formed. Each cyst containing a sperm cell goes through meiotic and mitotic divisions, creating more cells of the same nature. Each of the testicles of a young male contains three to five cysts which decrease in number with age. These cysts migrate to the opposite end of the testes where the development of spermatozoa takes place. After maturation, the cysts lyse then release the sperm cell into the reservoir of the testes located at the posterior pole. At this point, the sperm is capable of traveling via the vas deferens to the ejaculatory duct to be released by ejaculation (Huho *et al,* 2006). Sperm activation at the adult stages is still not elucidated but is believed to be activated once in the female spermatheca, therefore understanding the molecular mechanisms leading to sperm activation in adult males is of great importance for this study.

2.7.2 The male accessory glands (MAGs) and their secretions

The MAGs are comparable to the prostate in humans. The formation of the MAGs at the last abdominal segment occurs during larval development as an imaginal disk which significantly grows at the pupal stage. The MAGs are already formed when males emerge, but they require 2-3 days to fully contain secretions. Maturity of the MAGs and their successful mating age is not yet clearly understood given that it is vital for mating success in swarms. These secretions are made up of peptides and proteins received by the female during copulation. A recent bioinformatics study has identified 46 MAGs-specific genes in *An. gambiae* (Dottorini *et al,* 2007) and more genes have been identified by mass spectrometry analysis (Rogers *et al,* 2009). How these genes interact in the males prior to mating and plug formation is still not clearly known and this study will attempt to elucidate these pathways using *in silico* analysis. During copulation in *An. gambiae,* a coagulated mass called a mating plug formed by the MAGs secretions are transferred to the female atrium (Figure 2). 27 proteins were identified on the mating plug, 15 of which were originated mainly from the MAGs and six solely from the atrium. The rest of 6 proteins belonged to both the male and female reproductive tissues. Analysis of plug proteins for functions revealed the mechanism of plug formation, which is based on the cross-linking activity of a transglutaminase enzyme

(MAG-specific) on seminal fluid proteins, with the most abundant named Plugin (AGAP009368), identified as the major substrate (Rogers *et al,* 2009). Plugin is specifically expressed in the MAGs and through immunofluorescence techniques it was located to the anterior part of the secretory epithelium, within channels composed of an actin-rich muscle network lined on the outward part of the glands (Rogers *et al,* 2009). No characteristic protein domains have been identified on plugin but composition wise it is rich in glutamine residues (134/557 amino acids) often proceded by a lysine at the +2 position, which is the amino acid conformation known for Transglutaminase-mediated cross-linking. On another interesting note, the authors found a Transglutaminase enzyme (TRANSGLUTAMINASE (AGAP009099)) in the mating plug samples which were analyzed by mass spectrometry. Transglutaminase was shown to be localized at the posterior end of the MAGs and also involved in the cross-linking of the plugin substrate for plug formation (Rogers *et al,* 2009). This far, transglutaminase remains the only male protein whose functional role has been evaluated using RNAi studies. Protein-protein interactions using *in silico* networks will elucidate several tagets involved in plug formation which can be evaluated by RNAi studies and further evaluated for transgenesis

2.7.3 The ovaries and oogenesis

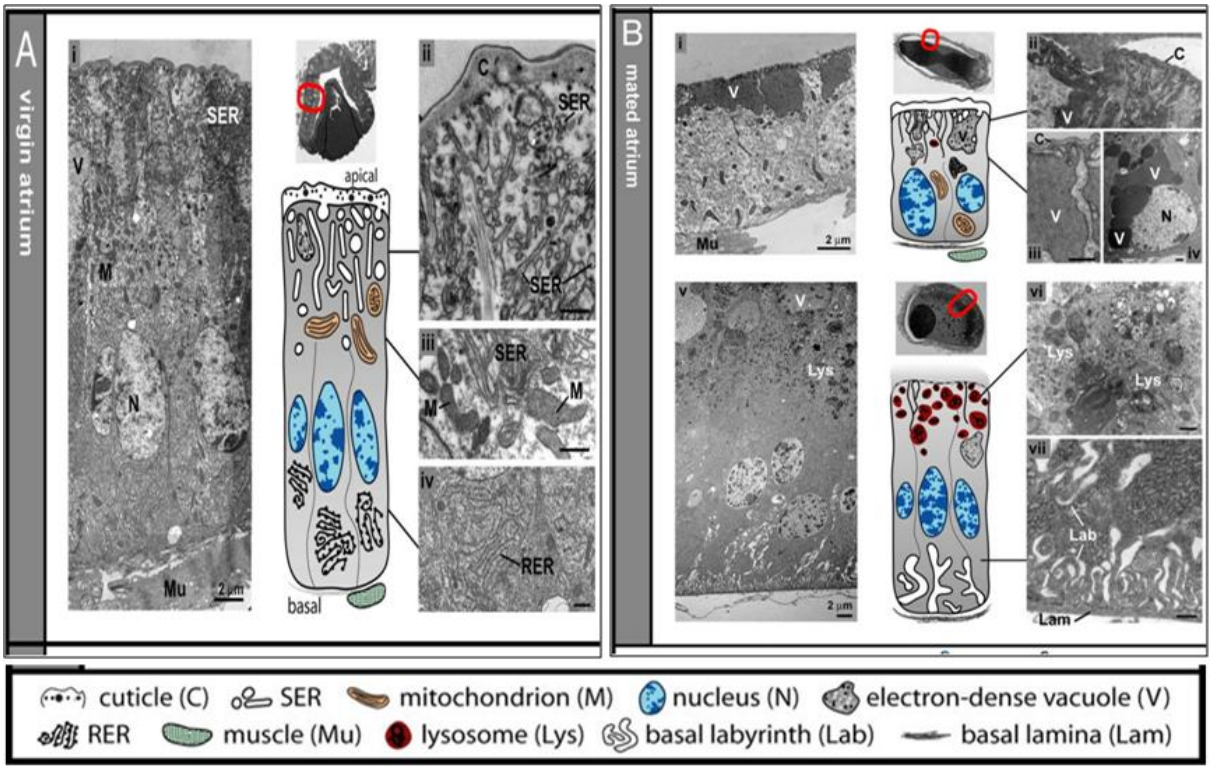
A transgenic line expressing a gonad-specific eGFP in the female *An. gambiae* showed that gonads detection could be possible from the L2 larval stage and onwards (Papathanos *et al,* 2009). The ovarian elongated shape at the pupal stage could be seen in the 6^{th} abdominal segment. The ovaries in *Anopheles*, contains 90-120 ovarian follicles, containing each an oocyte and 7 nurse cells derived from the germ cells. These follicles are formed in the germarium, where the oogonium (female stem line) is located. A cytoblast generated by the oogonium undergoes 3 mitotic divisions forming a cluster of eight cells surrounded by somatically cells. Nurse cells are formed by 7 of these cells while the eight cell becomes an oocyte whch begins meiosis. Repeated rounds of DNA replication are undergone by the nurse cells without cell division and which makes them transcriptionally very active. The follicles separate from the germanium though they remain connected through the follicular stalk. This process takes place before emergence in *Anopheles*, and then the establishment of complete separation occurs 1-2 days after eclosion. While the primary follicles undergo maturation, Secondary follicles are seen in the germarium. The development of the ovary at this point enters a previtellogenic arrest until the female finds a host. The transcription of the yolk protein precursors (YPPs) is induced by follicular growth

triggered by a blood meal from its host. Transcription gets to its peak at approximately 36 hours after a blood meal, and then YPPs are absorbed into the rapidly growing oocytes. The chorion is formed by the deposition of structural proteins on the follicle surface. At this point the oocyte is ready for release by the ovariole directly into the oviduct under the action of muscular contractions (ovulation), controlled by myotropin hormones. It is expected that females will not lay eggs until mating occurs and if they do, then the eggs will not be viable. This could mean that ovulation does not depend solely on a blood meal but equally on mating. This could also mean that, factors driving females to swarms could also be related to egg development; therefore elucidating such molecular interactions will be a point of focus in this study.

2.7.4 The female lower reproductive tract: atrium and spermatheca

The ovaries are linked to the atrium through the common oviduct together with the vagina (Figure 2). The dorsal and ventral side of the atrium form from pockets within the eighth and ninth segments (Clements, 1999). The atrium is ventrally attached to the lining within the ventral cuticle. The atrium connects the spermathecae and the female accessory glands (parovarium) through individual ducts (Figure 2). Based on transmission electron microscopy it has been revealed morphologically that the virgin atrium undergoes dramatic changes from 8 hours post-mating (Rogers *et al,* 2008). What drives these changes have not been established but are believed to be related to MAGs secretions deposited within the female. Highly polarised cells in virgins could be seen with the luminal surface lined with a cuticle presenting without channels and pores (Figure 1). Within the cytoplasm, smooth endoplasmic reticulum (SER) and many mitochondria could be seen present in the apical side. Basally this end contained extended complex of rough endoplasmic reticulum (RER), which could suggest high translational activity. Comparatively, the mated atria lost the SER or it was fused to cuticle while the RER was abundantly reduced (Figure 1). The cytoplasm showed a distribution of mitochondrial organelles while cells of the epithelium became cuboidal. Additionally, the luminal side saw the appearance of large electron dense vacuoles which presented with movements towards the basal end of the cell, and could suggest, transport of molecules (probably derived from processing of the mating plug) from the lumen of the atria to the hemolymph externally. The accumulation of endosomes and lysosomes of high density could suggest intensive serious exchange between the atrial epithelium and the plug, while expansion observed at the basal labyrinth and lamina suggested an active uptake of mating plug materials by cells of the atria. Molecular interactions between the plug and the

atria have not been elucidated and could form the basis behind the drastic changes observed in the tissue. This study will attempt to elucidate these interactions and evaluate their roles through RNAi technigue. These features taken together are indicative of the fact that the epithelium of the atria observes a drastic transition from a highly protein synthesis dedicated tissue into a tissue dedicated to move material into the hemolymph. The reception of sperm during mating is stored and maintained in the spermathecal capsule (Figure 2). The number of spermathecae shows a variation between culicidae: *Aedes* and *Culex* mosquitoes present three while *Anopheles* has only one. The only spermatheca in *An. gambiae* females derives from imaginal disks in the 8^{th} abdominal segments (Clements, 1999) enclosed in a cuticular shell, the capsule, is known to be formed by secretions of the epithelial cell layer of the spermatheca (Figure 1). The plaque surrounding the capsule is composed of large glandular cells connected through perforations to the spermathecal lumen present in the shell which helps in facilitating fluid transport. It will be wise to think that the glandular cells could release components meant to 'protect' the viability of sperms and also ensuring sperm related functions, like those factors involved in modulating oxidative stress



Rogers *et al,* 2008

Figure 1. Transmission electron micrographs of atrium cells. In virgin mosquitoes; (*A*), and in mated mosquitoes (8h) (*B*). (*Ai*) virgin atrial cell, showing SER and storage vacuoles filling the apical cytoplasm (*ii*, detail), mitochondria together with SER (*iii*, detail), and many RER located at the basal pole (*iv*, detail). (*Bi*): Atrial cell observed at post-mating, presenting electron-dense vacuoles located apically with collapsed SER membranes (*ii* and *iii*). Vacuoles of varying composition are contained in other cells that cluster at the base (*iv*, detail). Cell shown in contact directly with the mating plug in another mated female (*v*). Endosomes and lysosomes densely occur apicallly (*vi*, detail), expansion of the basal labyrinth is observed and the basal lamina is quite distinct (*vii*, detail). Some vacuoles can be clearly seen (*v*). Thumbnail images seen above the diagrams present low-magnification sections of the entire atrium showing the micrograph locations. Apical surface of images are orientated uppermost. (Scale bar, 500 nm unless otherwise stated.)

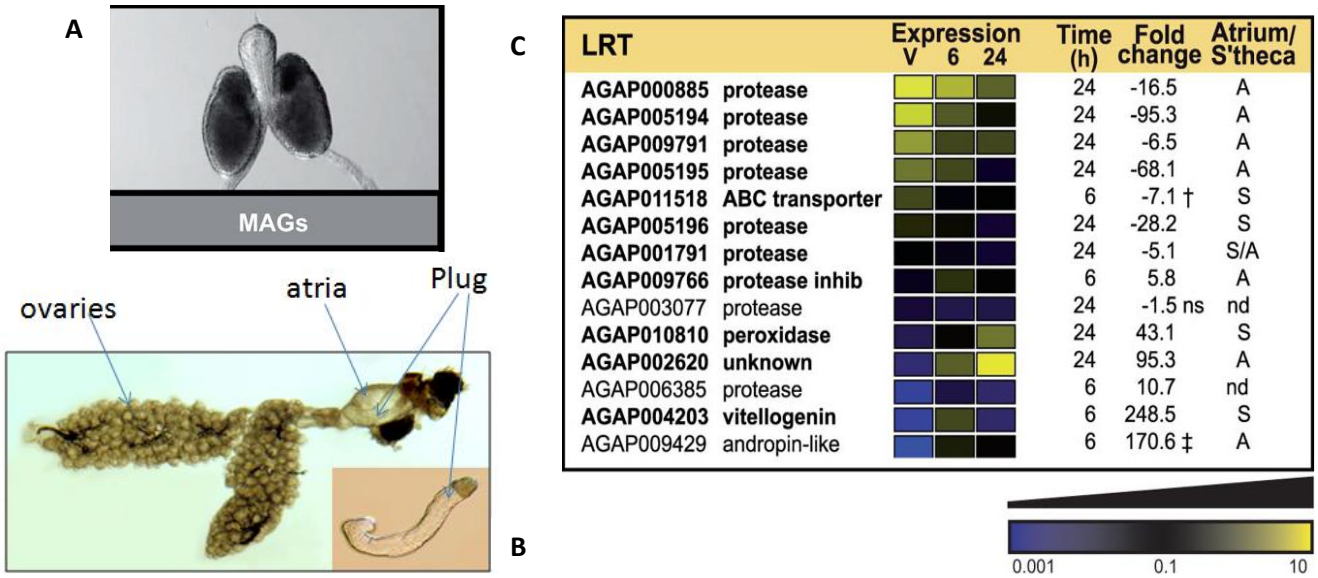
2.8 Mating Process in the mosquito *Anopheles gambiae* mosquitoes

Anopheles gambiae males usually become sexually active 48 h after emergence with mating activity peaking at 3-7 days after emergence (Mahmood *et al,* 1982). Despite the observation that males of *An. gambiae* and *An. arabiensis* were sexually active 48 hrs after emergence, low rates of insemination was observed in males younger than 3 days old (Charlwood *et al,* 1979) and this rate peaked at 7 days post emergence (Verhoek *et al,* 1994; Chambers *et al,* 2001). A period of 72-100 h is required for the accumulation of male accessory glands (MAG) proteins (Figure 2A) (Mahmood, 1997; Chambers *et al,* 2001) and it was shown to coincide with intense mating activity in male anophelines (Mahmood *et al,* 1982; Mahmood *et al,* 1994; Chambers *et al,* 2001). Additionally, the male antennal fibrillae shows a full functional activity level of essential for mating and occurring only around 12 h post emergence. When erect, fibrillae should respond to female flight tones which are also required for localizing females (Nijhout *et al,* 1979; Clements, 1999). The perception of female wing beat tones, should therefore, correspond apparently with sexual maturation in males.The relation between antennal perception and mating is not yet known and how this interaction will affect mating drive related events has not been established, therefore this study will struggle to find out the role played by these two physiological events during mating for both males and females. Swarms have been shown to be the only location for mating in most anopheline mosquitoes (Downess, 1968; Yuval, 2006) event though other reports propose indoor mating in African anophelines (Dao *et al,* 2008). The decrease in light intensity doesn't only define the onset of mating but is also coupled to the built-in circadian rhythm in anophelines. Charlwood *et al* (1979) showed that an alteration in male circadian rythms resulted in a reduction in inseminations. The formations of swarms at varying heights have been reported as well, which could possibly result from species-appropriate swarm markers. Markers are usually visual objects or contrasting areas occurring either on the ground or horizon which is used for swarming orientation by mosquitoes (Downess, 1968). Swarming in *An. gambiae* has been observed between 2-3 mabove no specific marker (Diabate *et al,* 2003) and 1to 4 m (Charlwood *et al,* 2002) over ground markers. No discernable markers were observed for *An. gambiae s.l.* swarms at heights between 0.5-2 m (Marchand, 1984). *Anopheles funestus* males converged in swarms of 2-4 m high using horizon markers (Charlwood *et al,* 2003). Populations of *An. philippenensis* and *An.hyrcanus* swarmed in Sympatry at a height of 0.3 to 0.6 m over small flowers and shrubs (ground markers) and 1.5 to 2.5 m in big open spaces (horizon markers) (Wharton, 1953). Swarm markers may play a serious role in defining Swarming height orientation in mosquitoes.

Males of *An. gambiae* S-form, swarm close to female feeding sites using a clear sky view (Marchand, 1984) while *An. gambiae* M-form place themselves in doorways and close to eaves during mating, irrespective of any markings on the ground (Dao *et al,* 2008). The variation in Swarming heights, may therefore act as another inter-species barrier to prevent mating. Mosquitoes locate themselves in space and time ensuring their availability to mate. At the approach of dusk, males start flying over the swarming location (Reisen *et al,* 1976; Charlwood *et al,* 1980) then more males gradually join them forming large and loose aggregations. Then the males evolve from a distended-extended, non-specific circular motion to a more rigid and condensed group (Reisen *et al,* 1976; Charlwood *et al,* 1980). Males mainly make up the swarms with the number of individuals varying from about 5 to 5000. Males become active before the females supporting the fact that females only get to already formed swarms by males. Until 5-20 min post swarm initiation, then can couples be observed. Females approaching the swarm may be detected from the frequencies of their wing beat, which is normally lower than that of males. What drives the females to identify these swarms still remains unknown because a swarm formed a distance away from where the female emerges will pose serious challenges to mating success therefore implying males could also release a possible chemical which facilitates females in identifying swarms.

The actual period of observed mating couples during swarming is quite brief lasting 10-30mins once mating has been initiated, copulating pairs are not usually seen anymore as darkness approaches (Diabate *et al,* 2003; Charlwood *et al,* 2003; Diabate *et al,* 2006). This period of mating may be shortened relative to the length of time spent by the mosquitoes *in copula*. Copulation in *An. gambiae* usually lasts 15-20 s (Charlwood *et al,* 1979) and lasts for about 27 s in *An. culicifacies* (Reisen *et al,* 1976). Intensive studies based on copulation in *Anopheles* have shown that as a male gets closer to the female, he grasps her using his tarsal claw on his first pair of legs and then moves his abdomen up to grasp her genitalia. Once the hook is formed, the male releases his tarsal grip on the female and they adopt the venter-to-venter position resuming flight (Downess, 1968; Charlwood *et al,* 1979; Omar *et al,* 1987). Copulating pairs then leave the swarm, flying off to complete the mating process after which the male is believed to join the swarm again but it is not clear if males will mate again on that same night. In any case, observations have been documented of anopheline males going back to the swarm after mating (Yuval *et al,* 1993). Males are believed to have limited opportunities to mate due to availability of few females, reported high risk of predation in the course of swarming, and the high energy budget needed for smooth swarming process (Yuval, 2006). This implies, males would benefit if they invest in high quality inseminations

hence forcing monogamy on females through the transfer of large volumes of sperm and a mating plug. At the completion of mating, the female is considered refractory to further mating due to the transfer of several compounds from the male. Males usually insert a mating plug after ejaculation, (Figure 2B) composed of MAG substances (Dottorini *et al*, 2007). MAGs substances in other culicids are known for causing a behavioural switch in the female from unmated to mated. Given that the functions of MAG proteins in anophelines have not been identified, a good number of specific male accessory gland proteins that have been shown to trigger post mating changes in behaviour in *Drosophila* (Dottorini *et al*, 2007) have been identified recently in *An. gambiae*. In a Similar manner, the transfer of the mating plug, and its related substances, has been identified to turn off the mating machinery in the females of *An. gambiae* resulting possibly in a remating physiological barrier (Rogers *et al*, 2008). Although contradictory in a superficial way to earlier work (Bryan, 1972; Dottorini *et al*, 2007), the digestion of MAG substances within the female's atrium could be necessary for stimulating post-mating events, which possibly could have been missed somehow in the original experiments (Klowden, 2001). Microarray analysis done by Rogers *et al* (2008) showed a well organised regulation of female genes (Figure 2C) in virgins and post-mated females (6h and 24h) during which the plug is known to have been digested.



Rogers *et al,* 2008

Figure 2: Mating events in *Anopheles gambiae*; A) Male Accessory glands (MAGs) wherein proteins involved in the formation of the mating plug are formed. B) Female atria containing the mating plug after mating. The atrium is attached to the ovaries showing egg development. C) Gene expression variations in the lower reproductive tract (LRT (atrium and spermatheca)) before mating then 6hrs and 24hrs after mating. The scale shows up and down regulation of genes at different time points in the female mosquito.

2.9 Oviposition in the female mosquito *Anopheles gambiae*

Initiation of oviposition (egg laying) in the female Anopheles depends on signals originating from the central nervous system (CNS) indicating that mating has occurred and the mature eggs are ready to be released. The triggering of these stimuli is not clear and their nature is unknown, but MAGs proteins and/or sperm have been postulated to play important roles in inducing egg laying (Dottorini *et al*, 2007). Mechanical stimulation of sensory receptors in the oviduct results from the distension created by the descending eggs. Motor neurons are known to activate the muscular walls of the spermathecal duct allowing the flow of sperm from the spermatheca into the egg via a small channel known as micropyle. The deposition of glycogen deposits could assist spermatozoa as an important energy source to penetrate the eggs. The initial meiotic division in the oocyte enters a state of arrest during the metaphase pending the penetration of a sperm cell into the egg. Only then can meiosis resume hence producing a haploid oocyte nucleus with three polar nuclei which develops in the periphery of the egg. The nucleus, known to be protected by a cytoplasmic layer around it, moves into the egg fusing together with the sperm (fertilisation). Oviposition known to be directly linked to ovulation in *Anopheles* occurs successively without a break. At post-fertilisation, the oocyte is forced down through peristaltic waves from muscular contractions and later released externally through the genital opening. The eggs can only move in one direction (towards the ovipositor) depending on a ratchet-like mechanism possibly favored by scales directed backward within the oviduct. The fact that oviposition depends on several environmental factors makes it a complex process. Most likely gravid females should experience an induced stimulus to find an oviposition site, though in the absence of mature eggs; they become more prone to select host stimuli for feeding. Such stimuli are known in in *D. melanogaster* to result from MAGs secretions and establishing such relations in *An. gambiae* will be a focus of this study. The mating status of *Anopheles* may not influence blood-feeding behavior either; given that most *An. gambiae* females obtain their first blood meal in the field one day post-eclosion and then mate one day later, and are more likely to take a full blood meal like females mated prior to a blood meal (Gillies, 1954; Detinova and Gillies, 1964; Brengues and Coz, 1973). Post blood-feeding, a batch of eggs undergoes maturation within the females as they rest indoors. They usually exit the house later and search for sites to oviposit (Huang *et al*, 2005). *Anopheles* females are known to lay eggs only on wet surfaces and in that light they need to search for favorable oviposition sites. Given that larval development and their survival depends alot on this selection, it is therefore important that gravid females should choose their oviposition sites very carefully. They are

well equipped with sensory receptors located on the antennae, on the tarsi and at the abdominal tip providing necessary information about the various surfaces, substrates and a favourable site for oviposition. Suggestions have been made that females are attracted to certain water bodies containig bacteria over others (Lindh *et al,* 2008) and to dark surfaced areas. Chemical cues may also provide insights as repellents or attractants. All sensory data collected are sent to the brain through the CNS where the information is analyzed before the female selects a suitable site for oviposition. Understanding where the female *An.gambiae* mosquito releases her eggs and the factors triggering oviposition are crucial for the understanding of the basic biology of this malaria vector and this will assist the progress towards effectively managing the vector (Otienoburu, 2007). Observing directly the oviposition of a mosquito in nature is not very possible because of being unable to trace its movement, its nocturnal activity, and very small size of mosquitoes (Chen *et al,* 2006). *An. gambiae* females will however in a general sense use non-permanent aquatic sites for oviposition, and the density of their larvae in a habitat is generally lower (less than 50 larvae) (Wagbatsoma *et al,* 1995). *An. gambiae* sensu lato gravid females will prefer water dishes with black-bottoms compared to white-bottomed ones relative to oviposition, and most of their eggs are laid in pools containing mud and lack of vegetation at the edges (McCrae, 1983; McCrae, 1984). Factors such as size of habitat, semiochemicals, substrate, predators, microbial fauna, vegetation, and types of land cover also affected the female's choice of aquatic breeding sites (Service, 1973; Munga *et al,* 2005).

2.10 Transgenesis and Bioinformatic applications in *Anopheles gambiae* mosquito

Developing of an elaborate technology for gene transfer in *An. stephensi* and *An. gambiae* mosquitoes (Catteruccia *et al,* 2000; Grossman *et al,* 2001), in combination with *An. gambiae* genome sequence project recently completed (Holt *et al,* 2002), has opened a new arena on reproductive genetic studies of these important disease vectors. However, progress towards the identification of reproductory molecules in males involved in the interaction with those in the female *An. gambiae* mosquito has been hampered by the lack of efficient technologies, such as mutagenesis screens and gene knock-out by homologous recombination, routinely used in other organisms to perform functional genomic studies. Silencing genes through RNA interference (RNAi), is a conserved evolutionary phenomenon driven by double-stranded RNA molecules (dsRNAs), and this has recently been looked upon as a powerful tool for studying the functions of genes in different model organisms (Fire *et al,* 1998). Levashina *et al,* (2001) used dsRNA knockout in hemocyte-like *An. gambiae*

cultured cells to demonstrate the conserved role of a complement-like protein in phagocytosis. Blandin *et al,* (2002) utilized direct injection of dsRNA in order to knock-down the *defensin* gene in *An. gambiae* and demonstrate its function. Rogers *et al,* (2009) also used dsRNA to knock-down transglutaminase (TRANSGLUTAMINASE (AGAP009099)) to ascertain its function in the crosslinking action of Plugin (AGAP009368) which is its main substrate for the formation of the mating plug. For the effective implementation of vector control by targeting reproduction in the mosquito *An. gambiae*, functional analysis using dsRNA technology on the known expressed genes in the males and females before and after mating will be very useful. Work done by Rogers *et al,* (2008) identified using microarray technology, a wide plethora of genes with transcript variations before and after mating. Understanding the molecular mechanisms for sex differentiation in the male and female reproductive systems that drive mating function, will be a powerful guide to implementation of the RNAi technology and vector control through development of transgenic mosquitoes

The use of genetic technology in the manipulation of mosquito species that transmit malaria to humans (Catteruccia *et al,* 2000; Grossman *et al,* 2001; Perera *et al,* 2002) has observed considerable progress hence expanding the range of possible methods needed to achieve eradication of vector populations locally or replacing them with mosquito populations refractory to malaria. Ethical concerns, environmental and safety issues have been raised relative to the risks of permanently varying the genetic makeup of the target vector population, with emphasis on horizontal gene transfer to untargetted species of insects (Alphey, 2002). The Sterile Insect Technique (SIT) could be employed on these mosquitoes (Knipling, 1959; Knipling *et al,* 1968) or its derivation, Release of Insects carrying Dominant Lethals (RIDL) (Alphey, 2002), to suppress and eradicate targeted insect pests populations. Actually transgenic techniques being applied include changing several aspects of the mosquito's biology that could help reduce vectorial capacity, such as the production of transgenic mosquitoes which are non-biting, reducing vector survival rate and/or human-biting rate, or variation of the rate of parasite development (Toure´ *et al,* 2004).

Sexual sterility induced in males comprising the SIT technology is based on the use of irradiation or chemical sterilants, and flooding natural populations with such sexually compromised males (Knipling, 1955; Knipling 1959) with the aim of reducing or eliminating vector populations. The SIT technique was successfully used to control the screwworm fly, the Medfly and tsetse (Knipling, 1955; Knipling 1959). Using dominant, repressible, female-specific lethal genetic systems forms the basis of the RIDL technology (Thomas *et al,* 2000; Alphey, 2002), and it has been proposed as a possible method for improving SIT. Transgenic

female insects reared on a tetracycline supplied medium grow while expressing the lacZ reporter gene. Breaking the tetracycline supply to the insects ready to be released will result in death of the female. Released RIDL males and mated with wild females will equally result in death of female offsprings and hence leading to reduction in the vector population.

In order to develop new, arthropod-borne strategies for disease control, several molecular biology and genomics tools are being used to investigate the interactions between vectors and parasites. Identifying new promoter sequences in the mosquito which are specifically regulated in reproductive tissues and at the right time are areas of great interest for the development of transgenic mosquitoes which when used in the field, they can be reproductively competitive. Furthermore, identification of *An. gambiae* male molecules involved in reproduction which can drive postmating responses in the female after mating will also provide novel insights to the development of male and female molecular transgenes for reproductive intervention. Information on this technology can be obtained by using genomics and microarray chip technologies and later analyzed by complex bioinformatic tools. Of recent, completion of *An.gambiae* genome, the most important vector of human malaria, was sequenced by an international consortium (Holt *et al,* 2002). Dottorini *et al* (2007) identified 46 MAG genes from *An. gambiae*, 25 of which are male reproductive tract-specific. This was achieved through a combination of bioinformatics searches and manual annotation confirmed by transcriptional profiling. To identify factors and mechanisms relevant to the fertility of *An. gambiae*, Rogers *et al* (2008) performed a comprehensive analysis of the molecular and cellular machinery associated with copulation in females. Initial whole-body microarray experiments comparing virgins with mated females after 2hrs, 6hrs, and 24hrs detected large transcriptional changes. Their results identified genes and mechanisms regulating the reproductive biology of *An. gambiae* females. Understanding the actual pathways in which these genes could be involved in at the protein level could be better appreciated by mapping these reproductive proteins in a protein-protein interaction network *in silco*. Protein-protein interactions have been shown to be largely mediated by structural domains interactions (Schuster-Bockler *et al,* 2007). For this to be achieved achieved, they usually search for (1) closest atomic distances between interacting domains, (2) defined bonding patterns between domains, and (3) conserved interacting sites. Upon determining and defining a 3D structural interface conserved between two domains, the databases for structural domain-domain interactions (sddi) uses it to predict all known protein interacting pairs where such structural domain pair has been analyzed. In this manner, the sddi datasets permit the inference of all possible and known protein interacting pairs that involve such

structural domains that have been identified to interact with at least one protein characterized in the Protein Data Bank (PDB). Uniprot accession numbers could be obtained and the used in; 1) CYTOSCAPE (Hernandez-toro *et al*, 2007) an open source bioinformatics software platform for visualizing protein networks and molecular interactions and complementing them with gene expression profiles and other necessary data for analysis and 2) Reactome database for viewing documented protein interactions hosted by European Bioinformatic Institute (EBI, www.ebi.ac.uk). These interactions with domain identifications can identify putative molecular targets which can be used in gene knockout and protein modification applications in the control of the mosquito *An. gambiae* as a malaria vector.

CHAPTER 3

MATERIALS AND METHODS

3.1 Determination of putative protein structural properties of MAGs proteins (transglutaminase and plugin) mediating successful interactions for mating plug formation in *Anopheles gambiae*

3.1.1 Identification and structure prediction of transglutaminase and plugin male accessory gland (MAG) protein orthologues in mammals and insects

Protein sequences of transglutaminase (AGAP009099) and plugin (AGAP009368) genes in *An. gambiae* were obtained from National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) database through bioinformatic searches (Altschul *et al*, 1990). Frequency of distribution of amino acid residues in the *An. gambiae* transglutaminase and plugin proteins were mapped using hidden markov model (HMM) routine in Genomics Workbench version 5.5 (CLC Bio, Aarhus, Denmark) and secondary structures predicted using PredictProtein server version 10.20.04 (Rost *et al*, 2004). Neural Network-Based-Residue-Residue Contact Map (NNcon) prediction server (Cheng *et al*, 2005) was used to map putative contacts among residues within *An. gambiae* transglutaminase or plugin proteins to better understand their predicted global structures. Conformations (3D) of the *An. gambiae* transglutaminase or plugin proteins and putative binding sites favouring the interaction of the proteins were predicted using iterative threading assembly refinement (I-TASSER) server (Ambrish *et al*, 2010). The transglutaminase or plugin conformations were superimposed on each other using MArkovian TRAnsition of protein structure (Matras) server (Kawabata *et al*, 2000) to validate established interacting surfaces among the proteins in mating plug (Rogers *et al*, 2009). Functional homologues of *An. gambiae* transglutaminase and plugin genes in *Homo sapiens* (human), *Mus musculus* (mouse), *Drosophilidae* (fruit fly), *Aedes aegypti*, *Culex quinquefasciatus*, *Caenorhabditis elegans*, *Pongo abelii* (Orangutan), *Apis mellifera*, *Plasmodium falciparum* were identified by BLASTp on non redundant protein database (Altschul *et al*, 1997).

3.1.2 Maximum likelihood and evolutionary relationships in transglutaminase and plugin gene sequences

Evolutionary phylogenies among the functional homologues of transglutaminase and plugin protein (retrieved 3.1.1) were determined through Genomics Workbench version 5.5 (CLC Bio, Aarhus, Denmark) based on gap open cost of 10.0 and a free end gap cost and Jukes and Cantor (JC69) and Kimura (K80) maximum likelihood models. Substitution rate matrices were estimated using parametrization method of Yang (1994a). Discrete gamma

model of Yang (1994b) was used to determine rate variation among sites and tree topology was also estimated based on the tree starting parameters (Yang, 1994b). In the absence of rate variation, estimation of substitution parameters and branch lengths were carried out according to expectation maximization algorithm of Dempster *et al* (1977). The topology space search was performed according to the PHYML method (Guindon *et al,* 2003) to efficiently search and estimate large phylogenies in our datasets.

3.1.3 Evaluation of population genetics parameters in transglutaminase and plugin gene sequences

Population genetics parameters in *An. gambiae* and the functional homologues in transglutaminase and plugin gene sequences were evaluated using routines in DnaSP version 5.10.01 (Librado and Rozas, 2009) software. Polymorphism analyses were conducted using locally aligned regions in all sequences. For Mc Donald and Kreitman test of neutrality, the groups were separated into mammals, insect (including *An. gambiae*) + others. For transglutaminase analysis, 16 and 9 sequences were used for mammals (nT1) and insects + others (nT2) respectively. Similarly, 18 and 17 sequences were used for mammals (nP1) and insects + others (nP2) respectively for plugin analyses. Protein coding regions were assigned to proteins in each group, with codon positions set at the first amino acid site. Synonymous and non-synonymous substitutions data was used to calculate ω, ratio of variation of non-synonymous to synonymous substitutions.

3.2 Determination of putative protein-protein interactions mediating mating-plug formation and post-mating phenotype in *Anopheles gambiae* mosquito

3.2.1 Identification of *Anopheles gambiae* mating plug proteins and their orthologs in *Drosophila melanogaster*

Information on *An. gambiae* mating plug proteins, with and without accession numbers, was obtained from Rogers *et al,* (2009). Sequences of the protein were retrived from VectorBase database (Lawson *et al,* 2009) and were subsequently applied in Search Tool for Retrieval of Interacting Genes (STRING) (Szklarczyk *et al,* 2010) database to predict orthologs of the proteins in *D. melanogaster*. The *An. gambiae* orthologs in *D. melanogaster* were applied in Artemis Comparison Tool (ACT) Version 9.0 (Carver *et al,* 2005) to confirm related proteins.

3.2.2 *In silico* determination of protein-protein interaction network among *Anopheles gambiae* orthologs in *Drosophila melanogaster*

Identity of *An. gambiae* protein orthologs in *D. melanogaster* were obtained from Uniprot Protein KnowledgeBase (UniProtKB) (Magrane *et al,* 2011) and subsequently

applied in Agile Protein Interaction DataAnalyzer (APID2NET) routine within Cytoscape version 2.8.0 (Hernandez-toro *et al,* 2007) to predict protein interaction among the orthologs, and complexes within the network identified using ClusterViz (Pavlopoulos *et al*, 2011) routine within the software. Main hubs (key nodes) within the network were detected using Cyto-Hubba (Lin *et al,* 2008) and Edge Percolation Component (EPC) algorithm (Chin *et al,* 2003) routines in the software. Pathways involving the complexes were identified using pathway identifier routine in Reactome database (Croft *et al,* 2011). Information from the hubs and pathways guided the extraction of a sub-network based on interactions on putative association with the *An. gambiae* plugin (Acession # AGAP009368), serine protease (Accession # TRYPSIN-LIKE SERINE PROTEASE (AGAP005195)) and *D. melanogaster* (male Accesory gland (Accession # ACP29AB)) proteins. Putative expressions of these proteins in reproductive tissues in *D. melanogaster* within the sub-network were determined from FlyATLAS (Chintapalli *et al,* 2007) and FlyBase (McQuilton *et al,* 2012). Comparative analysis between orthologs was done using PROSITE (Sigrist *et al*, 2012) a protein domain database for functional characterization and annotation and ProteinPredict softwares.

3.2.3 Maximum Likelihood phylogenetic analyses of network identified putative *Anopheles gambiae* mating plug proteins and related orthologs in *Drosophila melanogaser*

The orthologs of *An. gambiae* plug proteins in *D. melanogaster* via Genious version 5.5 created by Biomatters (Drummond *et al,* 2010) were subjected to a Multiple Sequence Alignment using Multiple Alignment Program for amino acid or nucleotide sequences (MAFFT) routine version 6.814b (Katoh and Toh, 2010). The alignments were subjected to a maximum likelihood phylogeny (PhyML) based on nearest neighbour interchanges (NNI) to explore space of tree topologies starting from a fast distance-based tree (Guindon and Gascuel 2003). Reversible mitochondrial substitution model (mtREV) (Adachi and Hasegawa, 1996) and Whelan and Goldman (WAG) (Whelan and Goldman, 2001) amino acid substitution models on *D. melanogaster* orthologs were evaluated. The best identified model for the tree based on a ML ratio test was used to build the tree with *An. gambiae* interologs.

3.3 Documentation of molecular factors among interacting proteins mediating mating responses in *Anopheles gambiae*

3.3.1 Experimental samples of *Anopheles gambiae* obtained

Anopheles gambiae mosquitoes, obtained as pupae from Terni Italy (G3 strain), Nairobi and Mbita-Point, Kenya (Mbita strain), were sexed as males and females into

seperate 200 ml glass bowls. The bowls were placed into separate 30cm long x 30cm wide x 30cm high cages and the mosquitoes left overnight to emerge as virgin males and females.

3.3.2 Mating experiments with *Anopheles gambiae* mosquitoes

Isolation of mating couples technique (IMC) Thailayil, (2011) was used to maximise (>95%) female insemination. Matured virgin females (same age as males) were collected with an electric or plastic glass aspirator and transferred to male cage. During swarming, mating pairs (couples) dropped to the bottom of respective cages, and were collected using Falcon tubes (Falcon Plastics, LA, USA). Males and females were subsequently separated, with successful insemination determined by 1) dissection and analysis of female spermatheca for the presence of sperm, and atria for mating plug and 2) qRT-PCR (section 3.2.7) of the atria for expression of insemination specfic serine proteases (TRYPSIN-LIKE SERINE PROTEASE (Accesion #AGAP005195)) (Rogers *et al*, 2008)

3.3.3 Dissection of *Anopheles gambiae* mosquitoes

Mosquitoes were placed on ice or CO_2 using INJECT+MATIC sleeper (Drummond Scientific, PA, USA) prior to dissection. Each mosquito was placed on a microscope slide with a drop of fresh 0.01M PBS (137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4 and 2mM KH_2PO_4). Whole mosquitoes and tissues of interest (MAGs, testes, ovaries, atria and spermatheca) were dissected and placed into a 1.5ml eppendorf tube containing 50µL of RNA later (for RNA extraction section 3.2.5) or 200µL grinding buffer (Sigma-Aldrich, NJ, USA) (DNA extraction, Appendix I). Both tubes were stored at -80°C for RNA or DNA extraction respectively.

3.3.4 DNA extraction from *Anopheles gambiae* tissues

The extraction was done using Phenol Chloroform (PCL) method (Appendix I). Tissues stored at -80°C were crushed using sterile pestles and 200µL extraction buffer added to the homogenate. One microlitre of proteinase K stock solution (10mg/ml) was added to each homogenate, inverted gently and incubated at 68°C for 15minutes after which 400µL of PCL was added to each tube, vortexed and centrifuged at 12,000 rotations per minute (rpm) using a micro-centrifuge (Eppendorf, Hamburg, Germany) for 5minutes. The resulting supernatant was transferred into a new 1.5ml eppendorf tube to which 400µL of 100% ethanol was added, vortexed and stored at -20°C for 30minutes to precipitate DNA. The tubes containing the precipitated DNA were then centrifuged at 12,000rpm for 30minutes at 4°C using a refrigerated centrifuge (Eppendorf, Hamburg, Germany). The Supernatant was then discarded and 400µL of 70% ethanol was added to each tube to wash the pellet. The tubes

were then inverted gently and then centrifuged at 12,000rpm for 5minutes at 4°C. The supernatant was discarded and the pellet formed was air dried at room temperature (RT). A volume of 100µL TE buffer was added to each tube and stored at -20°C for further analysis. DNA concentration was checked using a fluorometer (Eppendorf Biophotometer plus, Hamburg, Germany).

3.3.5 RNA extraction and DNase treatment of *Anopheles gambiae* tissues

Tissues stored at -80°C were crushed under a hood using sterile pestles. A volume of 150µL of TRI reagent (Sigma-Aldrich, NJ, USA) was added to each tube and incubated at RT for 3-5minutes. Fifty microlitres of chloroform (Molecular biology grade) was added to each tube, inverted to mix and incubated at RT for 3-5minutes. The samples were centrifuged at 12,000rpm for 15minutes at 4°C. The resulting supernatant was then transferred to a new 1.5ml eppendorf tube to which 60µL of Isopropanol (Molecular Biology grade) was added to each tube and incubated on ice for 30minutes then centrifuged at 12,000rpm for 15minutes at 4°C. The resulting supernatant was discarded and 200µL of 70% ethanol (Molecular biology grade) was added to each tube, inverted to mix and centrifuged at 7,500rpm for 5minutes at 4°C after which the supernatant was discarded carefully retaining the pellet. Tubes containing the pellet were blotted on tissue paper for 2-3 minutes and then air dried at RT for 5-10minutes. RNA was quantified according to section 3.4 above then resuspended in 10µL of DEPC water and normalized to 1µg/µL after which 1uL of DNAse (Invitrogen, UK) was added to each RNA sample. The RNA samples were then incubated at RT for 15minutes and 1µL EDTA (Invitrogen, UK) was added to each sample which was further incubated for 10minutes at 65°C in a heat block and afterwards stored at -80°C for cDNA synthesis.

3.3.6 cDNA synthesis of extracted RNA from *Anopheles gambiae*

Each extracted RNA sample was made up to a pre-mix of 100µL final volume with reagents as described by the manufacturer (Invitrogen, UK). The pre-mix contained 20µL of 5x first strand buffer (FBS), 5µL of 0.1M Dithiothreitol (DTT), 5µL of 10mM dNTPs, 5µL of 50µM random hexamers, 1µL of RNase inhibitor (out), 0.625µL of Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLVRT), 11µL of RNA template and 52.375µL DEPC water. The synthesis was completed by incubating the pre-mix at RT for 25°C for 10minutes, then at 37°C for 50minutes and finally 70°C for 15minutes. The synthesized cDNA was then stored at -20°C for further analysis.

3.3.7 qRT-PCR and expression profiling of *Anopheles gambiae* transcripts

Tissue specific expression was quantified using SYBR green detection dye. SYBR green binds to double-stranded DNA and hence the fluorescence increases with the amount of amplified product. Quantification of these samples were carried out in triplicate in a 10µl reactions using 1x Fast SYBR Green Master Mix (Applied Biosystems, CA, USA), HPLC-purified qRT-PCR primers and 5µl of 5x diluted cDNA and performed on Applied Biosystems StepOne plus RealTime PCR System. The following programme was used for amplification: 95ºC for 15minutes, then 40 cycles (95ºC for 15 seconds, 60ºC for 60 seconds) followed by a dissociation curve. Tissue specific expression was measured using the ΔCt method and standard curve for absolute quantification. The ribosomal gene RpL19 (AGAP004422) is a highly expressed 'housekeeping' gene. Its expression is very stable and does not change by mating. Therefore, RpL19 was used as a standard to normalise expression levels from other genes. All primers that were used for the qRT-PCR's are provided in Appendix II.

3.3.8 RT-PCR amplification of *Anopheles gambiae* transcripts

The cDNA was used for amplifying the genes of interest (Appendix II) prior to double stranded RNA preparations and cloning procedures. The PCR mix was prepared according to the manufacturers' instructions (QIAGEN, Hilden, Germany). A 25µL final volume reaction was made using 2.5µL of 10x reaction buffer, 1µL of 25mM $MgCl_2$, 1.25µL of 10mM dNTPs, 0.125µL of Taq (Hot start), 1µL (10pM) each for both forward and reverse primers and 16.125µL of nuclease/RNase free water. The PCR product, which was later ran on a gel for band size observations, was either cloned for sequencing or used for direct sequencing.

3.3.9 Cloning and Colony PCR of various amplicons obtained from RT-PCR

This procedure was undertaken with the TOPO TA cloning kit (Invitrogen, UK). The PCR product which contains a single deoxyadenosine (A) overhang at its 3' end produced with Taq polymerase was highly specific to the PCR 2.1-TOPO vector designed with single 3'-Thymidine overhangs. A 6µL of mix containing 4µL of the PCR product, 1µL of salt solution (200mM NaCl; 10mM $MgCl_2$) and 1µL of the vector, was composed and left at RT for 5–30minutes and then placed on ice prior to transformation. For transformation, 2µL of the TOPO cloning reaction was added to 1 vial of DH5α-T1 competent cells with ampicillin resistance. The cloning mixture was then placed on ice for 5–30 minutes then heat shocked for 30seconds at 42°C. The tubes containing cloning mixture were immediately placed on ice to which 250µL of super optimal broth (SOC) (Invitrogen, UK) medium preheated to 37°C

was added. These tubes were then thoroughly capped and shaken horizontally (Stuart, orbital incubator) at 200rpm for 1hour. Plates containing Lysogeny Broth (LB) Agar (for total Volume of 1L add; 10 g Tryptone, 5 g Yeast Extract, 5 g NaCl, 800 mL of distilled water) and 50µg/µL Ampicillin (Invitrogen, UK) were incubated at 37°C prior to plating the transformants. A solution of 40mg/ml X-gal in dimethylformamide (DMF) for blue and white screening was spread on the LB Agar plates before adding the transformants. A volume of 20µL of SOC medium was added to each LB Agar plate. Two different volumes of the transformants were plated for even spacing of colonies and incubated at 37°C for 12hours. Successful transformants produced white colonies which were tested by colony PCR. The plates with successful transformants were then sealed with parafilm and kept at 4°C for further analysis. White colonies were tested for the right band size and other variants through colony PCR using M13 primers as provided by the TOPO TA cloning kit. The PCR was set up according to section 3.2.8.

3.3.10 Gel electrophoresis analysis of amplicons

PCR product were analysed on a 1% agarose (Invitrogen, UK) gel run in TAE (Tris-acetate-ethylenedianinetetraacetic acid; 9.68g Tris, 4ml 0.5M EDTA (pH 8), 2.28ml Glacial acetic acid). 10µL of PCR product, homogenized in a loading buffer (Invitrogen, UK) was loaded per well. 2.5µL 1kb ladder (Invitrogen, UK) was separately loaded alongside, and run was at 70 volts (Thermoscientific, UK) for one hour visualised with a UV transilluminator (FujiFilm, UK).

3.3.11 Gel Purification of identified amplicons

The described bands on the gel were inscised using a new blade for each band and placed into a 1.5ml eppendorf tube. Purification was done using gel purification kit (QIAGEN, Hilden, Germany) according to the manufaturer's instructions. Three volumes of QG buffer was added to 1 volume of the excised gel. Samples were incubated for 10minutes at 50°C in a water bath while being vortexed at 2-3 minutes interval until completely dissolved. One gel volume of isopropanol was added to the mixture and inverted several times. The reaction mix was loaded onto a QIAquick column and centrifuged for 1minute at 13,000 rpm to collect the flow into a collection tube. A volume of 750µL PE buffer was added to the column and left to stand at RT for 2-5 minutes after which it was centrifuged for 1minute at 13,000 rpm to wash. The column was centrifuged for a further 1minute at 13,000rpm and was transferred to a 1.5ml eppendorf tube to which 30µL of elution buffer was added to the center of the membrane and left to stand for 1minute for DNA elution. The

tube was centrifuged for 1minute at 13,000rpm. Five microlitres of eluted DNA was viewed on agarose gel as per section 3.2.10. Samples were stored at -20°C for further analysis

3.3.12 Miniprep purification of identified positive clones through colony PCR

Clones previously confirmed by colony PCR (section 3.2.9) were placed in a falcon tube containing 5ml of LB medium (section 3.2.9) and 50µL (50µg/mL) of ampicillin (Invitrogen, UK), incubated at 37°C overnight and left shaking at 200 rpm for bacteria to grow. Overnight cultures were pelleted and completely resuspended in 250µL RNAse A containing Resuspension buffer (R3) ensuring no clumps formed. A volume of 250µL of lysis buffer (L7) was added to the cells, inverted several times, incubated at RT for 5minutes and 350µL of precipitation buffer (N4) added to the mix then inverted immediately several times to ensure the solution was homogenous. The mix was centrifuged at RT for 10minutes at 12,000rpm and supernatant loaded into a spin column for purification. The column was centrifuged at 12,000 rpm for 1minute and the flow through discarded. A volume of 500µL of wash buffer (W10) was added to the column and incubated for 1minute at RT then centrifuged at 12,000rpm for 1minute after which 700µL of wash solution (W9) was added to the column and centrifuged for 1minute at 12,000rpm. The flow through was discarded and the column placed in a sterile 1.5ml eppendorf tube to which 50µL of preheated TE buffer was added to the center of the column, incubated for 1minute at RT then centrifuged at 12,000rpm for 2minutes for elution. The purified plasmid was aliquoted for sequencing as per the manufacturer's instructions and the remaining aliquot was stored at -80°C.

3.3.13 Western blot analysis of mating plug proteins in *Anopheles gambiae*

Western blot analysis was done to observe actual protein size of the transcript of interest after translation. Three to four days old virgin mosquitoes sexed as pupae were dissected for tissues of interest (MAGs, testes, spermatheca and atria) into 1.5ml eppendorf tubes containing 15µL of extraction buffer (Appendix III) placed on ice. The tissues were homogenized with sterile pestles and placed on ice for 10 minutes to allow for lysis and denaturing of molecular bonds to obtain a homogenate which was centrifuged at maximum speed for 15 minutes at 4°C. The resulting supernatant was separated from the pellet and Bradford test was done (Appendix IV) to quantify protein concentrations per sample. Samples were diluted to the same concentrations to which 6µL of sample buffer 4x (Invitrogen, UK) and 2µL of DTT 10x (Invitrogen, UK) was added to each sample. The samples were incubated at 70°C for 10 minutes in a heating block and 20µL of each sample was loaded on a NuPAGE SDS PAGE gel (Invitrogen, UK) in a tank containing running

buffer according to the manufacturer's instructions. Ten microlitres of the Novex Sharp Protein Standard (Invitrogen, UK) was loaded alongside the samples. The samples were run for 1 hour and 30 minutes at 70volts to separate the protein bands. Transfer of protein bands from the gel to a nitrocellulose membrane was carried out according to the manufacturer's instructions (Invitrogen, UK). The protein transfer setup was placed in a tank containing transfer buffer which was left to run for 1hour at 30volts until transfer was complete. The nitrocellulose membrane was placed in blocking buffer (Appendix V) overnight. The nitrocellulose membrane was incubated for 1 hour in primary antibody (1:1000µL of blocking buffer, Abcam, Cambridge, MA) while shaking, then washed with PBS Tween (Appendix III) by shaking 3times at 10minutes intervals. The nitrocellulose membrane was then incubated in secondary antibody (1:10000µL of blocking buffer, Santa Cruz Biotechnologies: sc-2030 and sc-2314) for 1 hour while shaking to ensure effective fixing of the secondary antibody. The nitrocellulose membrane was then washed with PBS Tween by shaking 3times at 10 minutes intervals. Bands were visualized using ECL Western blotting detection reagents (GE Healthcare, UK) on an LAS-3000 imaging system (FujiFilm, UK).

3.3.14 Immunostaining of mating plug proteins in *Anopheles gambiae*

Immunostaining was used for localization of proteins expressed in various tissues of the mosquito *An. gambiae* with emphasis on reproductive tissues. Mosquitoes of different ages were dissected for tissues of interest (MAGs, testes, spermatheca and atria) into individual baskets to hold them and were later placed into each well of a 24well plate containing 4% paraformaldehyde solution and left to stand for 40 minutes. The tissues were washed 3 times at 15 minutes intervals in fresh PBS then incubated in 2% hydrogen peroxide (H_20_2) for 5minutes to remove auto-fluorescence from these tissues. The tissues were then washed 3 times at 5 minutes intervals in fresh PBS. The washed tissues were kept in 1% Permeabilization and blocking buffer (PBB) (Appendix V) for 1 hour then incubated with the primary antibody (1:300µL 1% PBB) and kept overnight at 4°C. Tissues were then washed 3 times at 20 minutes intervals in 1% PBB and incubated in a dark box with the secondary antibody [fluorophore (1:1000µL PBB)] at RT for 1 hour while shaking. The tissues were washed 3 times at 20 minutes interval with 1%PBB and mounted on DAPI-containing Vectashield medium (Vector Laboratories, Inc.) then visualized using Apotome (Carl Zeiss technology). Stacks were generated using 19 consecutive 0.5 mm optical sections.

3.4 Determination of influence of interactions among proteins on mating responses in *Anopheles gambiae*

3.4.1 Functional genomics analysis of CATHEPSIN B (AGAP004533) in *Anopheles gambiae*

The CATHEPSIN B (AGAP004533) (4533) gene in the mosquito and the LacZ controls were validated using RNA interference technique of Catterruccia *et al.* (2009). Transcript sequences were obtained from VectorBase (Lawson *et al,* 2009) and used in Primer 3 software (Rozen and Skaletsky 2000) to design primers specific to a 502bp stretch of CATHEPSIN B (AGAP004533). The sequence was amplified from cDNA obtained from the MAGs and atria and gel purified as perviously described (sections 3.2.11 and 3.2.12). *ds*4533 and *dsLacZ* were synthesised using pLL10 and pLL100 plasmids (Rogers *et al,* 2009) respectively. For control, β-galactosidase (β-gal) dsRNA was produced from a 500-bp amplicon of the β-gal gene from the commercial plasmid pcDNA3.1/His/LacZ (Invitrogen, UK). The plasmids were purified, with the insert sequenced and verified as previously described (section 3.2.8). XhoI and XbaI restriction digestion was conducted on the plasmid to recover the insert. *In vitro* transcription of the insert was conducted using The MEGAscript kit (Invitrogen, UK) according to the manufacturer's instructions. Briefly, 20µL mix of 8µL of PCR product, 2µL of 10x reaction buffer, 2µL each of 75mM ATP, CTP, GTP and UTP and 2µL of T7 enzyme mix was composed and incubated at 37°C for 12–16hours. 1µl of TURBO DNase (2U/µL) added and incubated for 15minutes at 37°C. Nuclease free water (115µl) was added, followed by incubation at 60°C for 10minutes. Ammonium acetate (15µl) was added to the mix to terminate the DNAse action. The product was purified by adding 150µL of Phenol/Chloroform (50/50 mix), vortexed and centrifuged for 5 minutes at RT. The supernatant was transferred into a new 1,5ml eppendorf tube wherein 150µL of Chlorform was added, mixed by vortexing, and centrifuged at 13,000 RPM (rotations per minute) for 5minutes at RT. The supernatant was collected into a new 1.5 ml eppendorf tube. A volume of 110µL isopropanol (laboratory grade) was added to the supernatant and incubated at for 15minutes at RT to precipitate. The solution was centrifuged at 13,000rpm for 30minutes at 4°C. Supernatant was discarded, the pellet air-dried, dissolved in 20µL RNAse free water and subsequently heated at 60 - 65°C in a water bath to dissolve pellet. The RNA concentration was determined using a UV spectrophotometer (eppendorf Biophotometer Plus) in a 1:59 dilution.

3.4.2 Injection of male and female *Anopheles gambiae* with dsRNA

Sense and antisense strands of AGAP004533 and LacZ were diluted to a final required concentration of 5µg/µL (1.9M) and 6µg/µL (2.3M) for injection into female and

male mosquitoes respectively. Equal concentrations of sense and antisense strands were mixed and boiled for 5minutes in a beaker of water then left to cool down to RT. All RNAs were analyzed on agarose 2% gel to confirm annealing was successful. 1µL of the single RNA strands were diluted in 5µL RNAse free water and loaded on the gel alongside 1µL of the double stranded RNA using RNA loading buffer. Double stranded RNA was stored at -80°C for mosquito injection process. Injections were done using NanojectII (Drummond Scientific, PA, USA) according to the manufacturer's instructions. The mosquitoes were injected at the thorax as previousely described (Garver *et al*, 2007) where in 5µg/µL and 6µg/µL of ds4533 and ds LacZ were injected into female and or male mosquitoes respectively.

3.4.3 Evaluation of gene knockdown in *Anopheles gambiae*

Male and female reproductive tissues were dissected 2-4 days post injection and RNA extracted as perviously described (sections 3.2.3 and 3.2.5). Gene specific cDNA was synthesized for AGAP004533 and gene silencing was evaluated using qRTPCR as previously described (section 3.2.7).

3.5 Determination of swarming patterns and roles in male *Anopheles gambiae* mosquitoes mating responses to females

3.5.1 Detection of swarming patterns and mating couples in semi-field screen houses

Eggs and larvae of *An.gambiae s.s* (Mbita strain) were held in a screen house of 7.1m long x 5.7m wide x 4.4m high. The screen house was made of all round netting. Eggs and larvae were placed in plastic bowls (10L) containing fresh filtered water from Lake Victoria. Eggs and larvae were exposed to natural weather conditions of rainfall, sunlight, night and day. Eggs hatched 2–3 days into larvae, which pupated at 9–10 days later. Adults were held in 30 x 30 x 30cm cubical cage placed in a 3 x 3 x 3m block walls room with window opening closed with a cloth, but with natural changes of night and day. The room was also exposed to natural weather conditions of sunlight and rainfall. Caged mosquitoes were arm-fed and an ethical concern for this work was obtained from the Kenyan government (Appendix VIII). Two semi field cages (screen houses) (Figure 3) were used for release of mosquitoes to analyse swarming patterns and mating, Malaria sphere 2 (7.1 x 5.7 x 4.4m), had internal netting with the same length and width but 2m high. The sphere contained a hut (3.2 x 2.8 x 1.7m) made of mud walls and thatched (dry grass) roof simulating human residence. The screen house contained varied vegetation within and mosquito predators (spiders, geckos, praying mantis and ants) to mimic a natural ecosystem. The screen house

was exposed to natural conditions of sunlight and rainfall. Daily temperature was ~30ºC in March then came the rainy season in April which extended into May 2012.

Malaria sphere 1 (11.4 x 7.1 x 3.4m) had no internal netting But with similar vegetation, predators and natural weather conditions to malaria sphere 2. These two screen houses were used to study natural swarm patterns and mating of mosquitoes in natural field conditions. Behavioral mating of the *An. gambiae s.s* (Mbita strain) was studied by collecting 500 pupae daily from the icipe Mbita insectary. The pupae were sexed into males and females, placed into separate 30 x 30 x 30cm high cubical cages and mated as 3 days old virgins using isolation of mating couples technique of Thailayil *et al,* (2011).

One thousand mosquitoes were released every 4 days into malaria sphere 1 for a month during which; swarming initiatiation time, duration and swarm sites observed were documented. Experiments were terminated when swarming ceased. One thousand mosquitoes were also released daily in malaria sphere 2 for 3 months and mating pairs caught using sweep nets. The pairs were maintained on sucrose, with females dissected three days post capture to confirm successful insemination by microscopy.

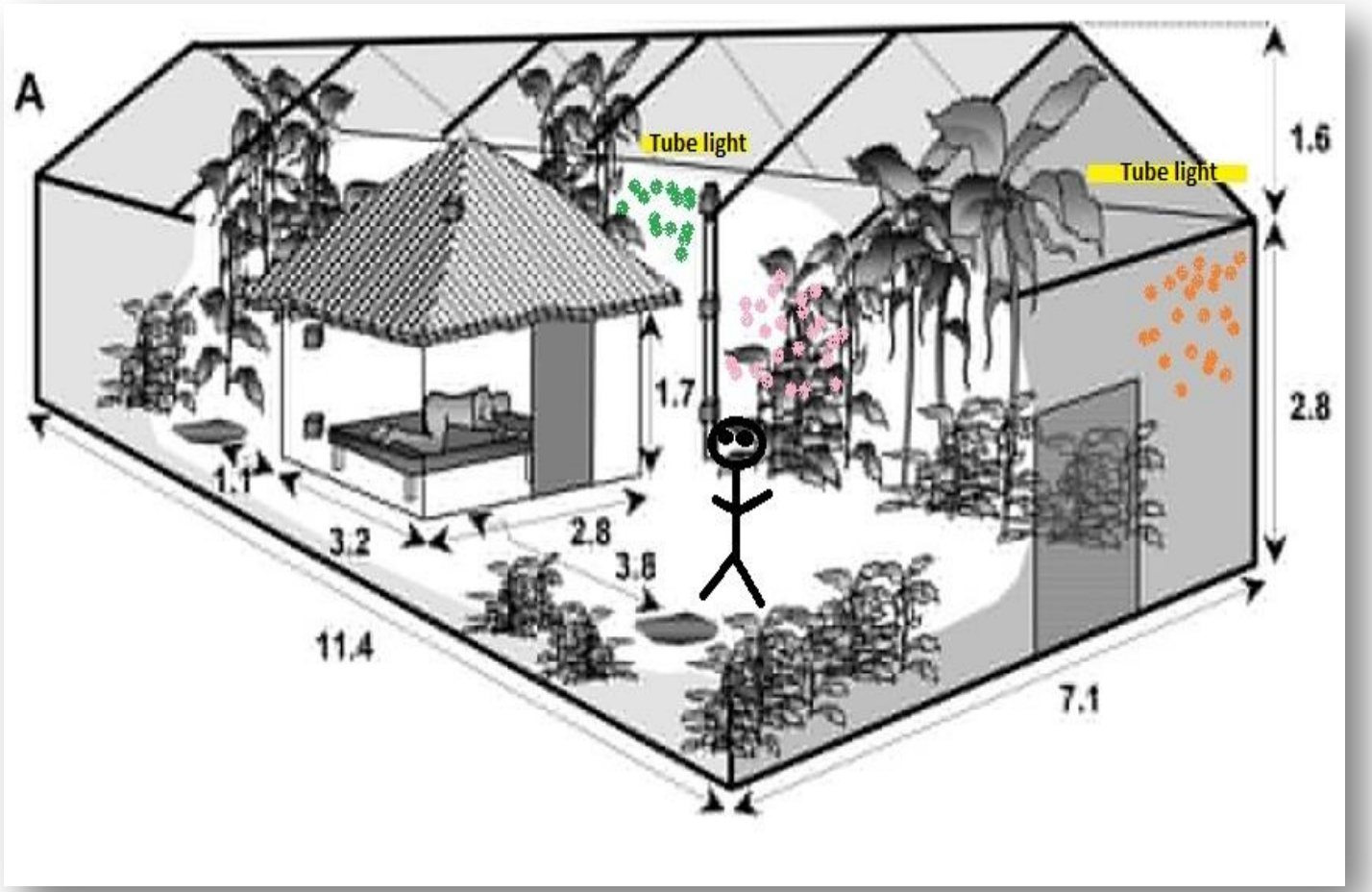


Figure 3: Structure of Mbita semi-field (Screen house) cage used to study swarm patterns in *Anopheles gambiae* Mbita strain; Color dots represent different swarm sites by mosquitoes. Tube lights were used to observe swarms after darkness.

3.5.2 Determination of propensity for mating in the female *Anopheles gambiae*

The propensity was tested in 30 x 30x 30cm cubical cages wherein females were sexed as pupae and kept to emerge as adults. Cages were labeled as male or female then two hours before dusk; the cages were linked by 10 cm aperture and kept overnight. The cages were separated at 6am in the morning of following day, and males and females within each cage documented. The females were dissected three days later and microscopy performed to determine presence/absence of sperm in the spermatheca for validation of mating. The same procedure as above was done for *An. gambiae* females injected with dsRNA against CATHEPSIN B (AGAP004533) as experimental and dsLacZ for controls.

3.6 Data analysis

For objectives 1 and 2; DNA sequences were translated *in silico* into their respective amino acid sequences. The sequences were subjected to multiple alignments using CLC workbench 6.6.1(CLC Bio, Aarhus, Denmark) and Genious version 5.5 created by Biomatters (Drummond *et al*, 2010). A phylogenetic tree was constructed by Neighbor Joining method by using *p*-distance estimates, and tested by using interior-branch test. Reliability of each node was assessed with 1,000 bootstrap replications. Amino acid pairwise identities and similarity scores were used for evaluation of amino acid sequences during protein alignments at both sites. Phylogenetic analysis was performed using CLC workbench and Geneious software, and visualized using TreeView software Version 1.6.6 (Roderic, 1996). Evolutionary relationships for the various haplotypes were calculated with the maximum likely hood ratio test (LRT) using PHYML to evaluate substitution models which were validated for significance using the Chi square test between various substitution models. Phylogenetic signals per sequence were validated using the K value on tree branches which represented transversion (change of purine to purine or pyrimidine to pyrimidine)/transition (change of purine to pyrimidine and vice versa). Orthologs were identified using STRING database with alignment score confirming identity. Artemis Comparison Tool was used with a >50% filter to confirm orthology from the STRING database. Gene ontology (GO) values were used to confirm functions of genes mapped on the interactome. Hyper geometric tests were performed in the Reactome database to map network proteins within pathways. Approximate likelihood ratio test (aLRT) was used to identify protein clusters of *An. gambiae* Plug proteins on a Whelan and Goldman (2001) maximum likelihood (ML) tree. For objectives 3, 4 and 5; the qRT-PCR data was analyzed using the comparative C_T, conducted for three independent experiments. Analysis of Variance (ANOVA) was done using

GraphPad Prism version 4.0 for Windows, (GraphPad Software, La Jolla California USA, www.graphpad.com) to; evaluate relative differences between expressions levels of various transcripts used, evaluate transcript level differences between experimental and LacZ controls groups after injections for knock down and also for analyzing differences between inseminated and non inseminated females. Statistical significance was considered as per the P-value provided by the software.

CHAPTER 4
RESULTS

4.1 Putative structural properties and selection pressures implicating transglutaminase and plugin as main proteins in mating plug formation in *Anopheles gambiae*

4.1.1 Identification and secondary structure analysis of male *Anopheles gambiae* plugin and transglutaminase proteins

Transglutaminase and Plugin sequences were obtained from VectorBase with both sequences constituting of 748 amino acids and 557 amino acids respectively. Protein analyses showed secondary structure predictions for transglutaminase having; 8.02% helices, 36.23% beta strands and 55.755% loops (Figure 4A) while that for plugin showed; 55.12% helices, 44.88% loops and no strands (Figure 4B). Frequency distribution of amino acids (Figure 4C) showed that plugin was dominantly Glutamine residues (24%) while transglutaminase was dominantly Valine residues (9%) compared to all other amino acids residues within their respective amino acid sequences. Plugin had more exposed residues of 89.59% and 10.41% buried while transglutaminase had 50.53% exposure and a 49.47% buried. Neither disulphide bonds nor protein-DNA interactions were predicted on both proteins. Possible sites for protein-protein interactions were predicted on both proteins with a higher proportion on transglutaminase than Plugin. Predicted PROSITE sequence motifs on both proteins with Asparagine glycolysation site (ASN-Glyco), Protein kinase C site (PKC-phospho) and Casein kinase II (CK2-phospho) was common. Three other sites specific to transglutaminase included; Tyrosine protein kinase (TYR-phospho site), N-myristoylation site (Myristyl) and Amidation site. The predicted Pfam domains showed Plugin as having GGDEF (PF00990) and HisKA (PF00512) whereas those for transglutaminase were Fer4 (PF00037), Metallophos (PF00149) and Pyr_redox (PF00070). These domains have their homologs in mammals.

The PSIBLAST alignments identified closely related proteins from UNIPROT database used for predicting secondary structure properties on *An. gambiae* transglutaminase and plugin. Plugin identified with; 1) B3RIJ8-1, a putative uncharacterised protein for *Trichoplax adherens* showed a sequence identity of 15% with molecular functions as chitin binding (GO: 0008061) and receptor activities (GO: 0004872) binding to an extracellular and intracellular messenger to initiate changes within the cell, 2) Q5EG54-1, a vitellogenin in *Dermacentor variabilis* showed sequence identity of 23% with known molecular functions: calcium ion binding (GO: 0005509), calcium-dependent phospholipid binding (GO:

0005544) and lipid transport activity (GO: 0005319). Transglutaminase amino acid predictions identified to protein-glutamine gamma-glutamyltransferase 2 with 30% of sequence identity. The glutamine gamma-glutamyltransferase 2 has similar molecular functions to acyltransferase activity (GO: 0008415) and transferase (GO: 0016740) involved in transfer of a chemical group from one compound (donor) to another (acceptor). The Hidden Markov model (HMM) predicted two conserved regions in the transglutaminase group of proteins found within the beta barrels (IWNYHVWS, YGQCWVY). Predictions for the plugin group of proteins showed several four sequence amino acids characteristic of helices within a sequence and were marked by the presence of several Gln residues. Low complexity regions (aa 1-260) were predicted for the plugin protein in *An. gambiae*.

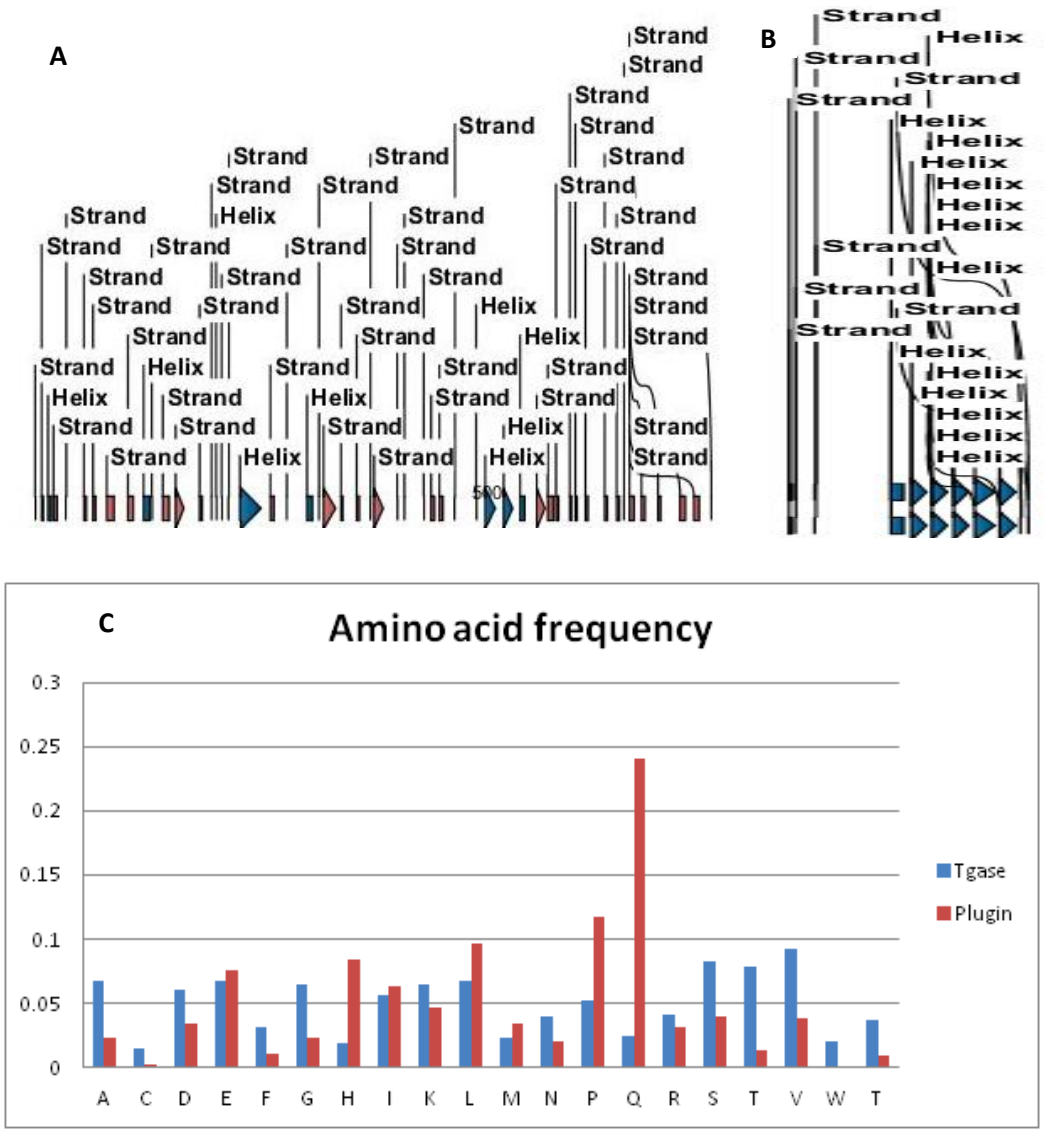


Figure 4: Amino acids distribution in tranglutaminase and plugin proteins of *Anopheles gambiae*; A) Secondary structure for transglutaminase with a high distribution of beta sheets and a few helices. B) Secondary structure for plugin with a higher preference for helices than beta sheets. C) Frequency distribution of amino acids within proteins and that of plugin characterized by the high presence of Gln (Q).

4.1.2 Contact map of male *Anopheles gambiae* plugin and transglutaminase proteins

Amino acid residue contact analysis performed on transglutaminase and plugin to understand their global structures with the 2D-Recursive Neural Network models predicted contacts at an 8Å and 12Å standard distance thresholds respectively for both transglutaminase and plugin proteins. The probability contact map (Figure 5) shows selected residue-residue contacts at an 8Å threshold with sequence separation 6 and predicted probability 0.1. The average contact order and contact number for plugin was 1.523 and 6.926 while that for transglutaminase was 5.484 and 16.687 respectively. These values predicted that transglutaminase was globular with more of its residues buried (49.47%) compared to plugin showing 10.41% buried otherwise 89.59% of it was exposed as seen in section 4.1.1 above.

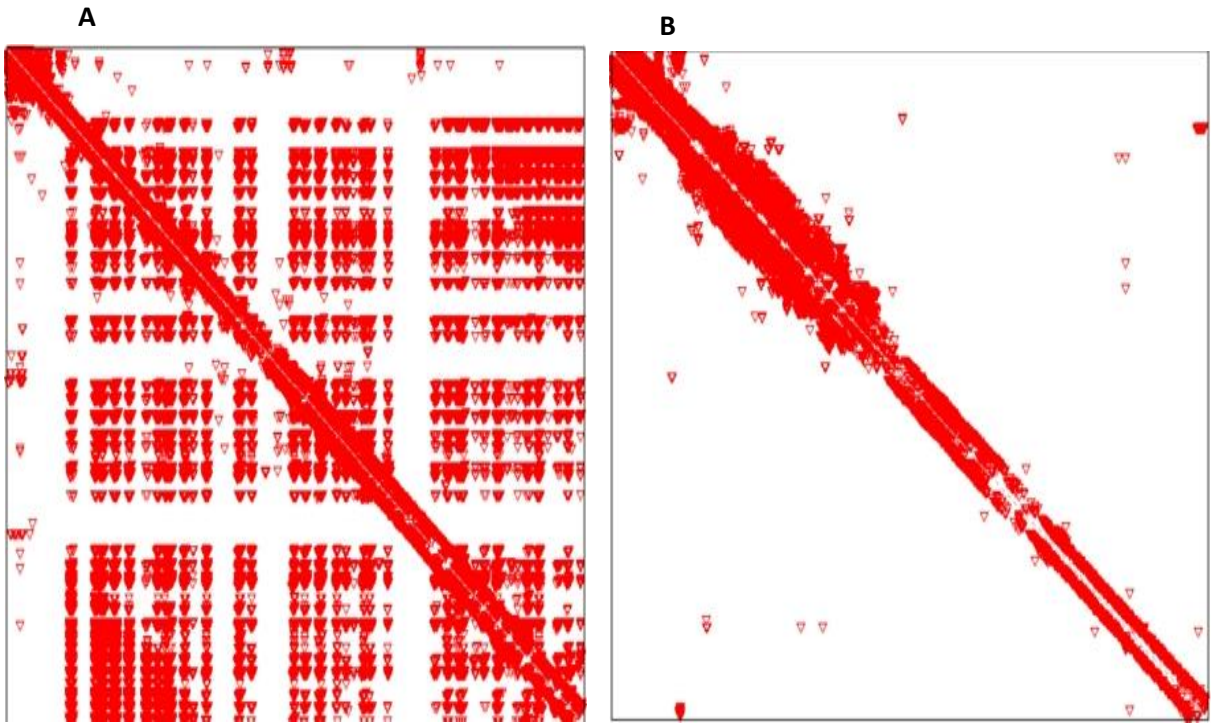


Figure 5: Contact map of transglutaminase and plugin in male *Anopheles gambiae*; Contact maps predicted by NNcon server. It was done with selected residue-residue contacts at an 8Å threshold with sequence separation ≥ 6 and predicted probability ≥ 0.1 A) Map for Tgase with average contact order and contact number of 5.484 and 16.687, respectively. B) Map for Plugin with average contact order and contact number of 1.523 and 6.926, respectively.

4.1.3 Superimposition (3D) of transglutaminase and plugin male *Anopheles gambiae* proteins

Identification of putative protein-protein interaction sites between transglutaminase and plugin was done using protein databank (pdb) codes for transglutaminase (A1kv3A) and plugin (3b8cA) which showed sequence similarity of 28.6%, a secondary structure similarity of 44.3% and an environmental score of 38.0 which characterises an exposition of 41.4%. The exposition value shows exposed regions which putatively interact (Figure 6). This interaction generated models consisting of triosephosphate isomerase (TIM) barrel consisting of eight alpha helices and eight beta strands alternating along the peptide backbone.

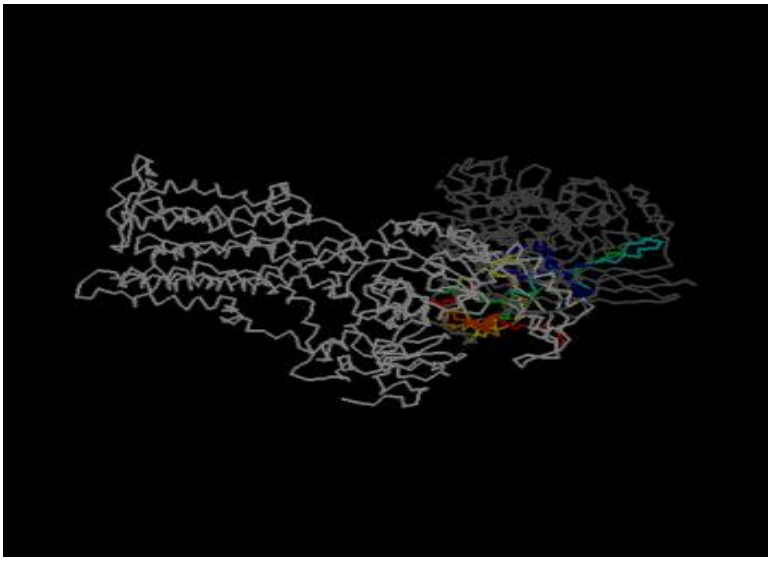


Figure 6: Superimposition (3D) of transglutaminase and plugin male *Anopheles gambiae proteins*; Super imposed 3D structures of related predictions were done in Matras using analog strands from PDB. It had for Plugin strand 3b8cA from ATPASE 2 (Light grey) and Tgase used strand 1kv3A from "PROTEIN-GLUTAMINE GAMMA-GLUTAMYLTRANSFERASE" (dark grey). Interacting residues (colored strands).

4.1.4 Ortholog identification of plugin and transglutaminase in closely related taxa

Transglutaminase (AGAP009099) and plugin (AGAP009368) from VectorBase and through BLAST Searches at NCBI identified; transglutaminase (accession # ACZ65013) and plugin (accession # ACZ65012) amino acid sequences in *An. gambiae*. The same BLAST searches at NCBI with transglutaminase (accession # ACZ65013) and plugin (accession # ACZ65012) identified homologous sequences for different taxa (section 3.1.1) as summarised in Table 1 A&B. In Table 1A, transglutaminase homolog sequences in other taxa were considered based on ≥95% identity for mammals and ≥90% identity in insects + others. In Table 1B, due to lack of high identity scores of plugin in related taxa, structural homologs with close similarity was considered for mammals and insects + others. The distances between the sequences could be identified on the phylogenetic trees (Figure A&B). Transglutaminase sequence in *An. gambiae* was closer in homology to mammalian sequences given that most variants were obtained in *Mus musculus* and *Homo sapiens* compared to a few in insects + others. Based on structural similarity of plugin protein sequence as shown in the previous section, it was observed that mammals equally possessed more proteins with similar features to that of the insect + other group. Their amino acid lengths and weights were quite similar. Homology was more conserved in the transglutaminase than plugin groups, indicated by varied protein lengths and sizes.

TABLE 1A: TRANSGLUTAMINASE RELATED GENES IN DIFFERENT TAXA

Organisnm	Code	Protein	Length	Weight
Mus musculus	gi|148674308	transglutaminase 2, C polypeptide, isoform CRA_a	693aa	77.644 kDa
Mus musculus	gi|148708973	coagulation factor XIII, A1 subunit, isoform CRA_c	763aa	86.9 kDa
Mus musculus	gi|120537330	Transglutaminase 3, E polypeptide	693aa	77.277kDa
Mus musculus	gi|124001551	protein-glutamine gammaglutamyltransferaseE		77.309 kDa
Homo sapiens	gi|339527	transglutaminase	817aa	89.754 kDa
Homo sapiens	gi|4507475	protein-glutamine gammaglutamyltransferase K	817aa	89.786 kDa
Homo sapiens	gi|339604	transglutaminase K enzyme	814aa	89.338 kDa
Homo sapiens	gi|119395709	coagulation factor XIII A chain precursor	732aa	83.267 kDa
Homo sapiens	gi|1353350	prostate-specific transglutaminase	684aa	77.143 kDa
Homo sapiens	gi|13937806	Transglutaminase 4(prostate)	684aa	77.134 kDa
Homo sapiens	gi|56202736	transglutaminase 6	625aa	70.515 kDa
Homo sapiens	gi|2895530|gb	transglutaminase X	720aa	80.767 kDa
Homo sapiens	gi|33331030	transglutaminase y	706aa	79.279 kDa
Homo sapiens	gi|114107591	Transglutaminase 5	720aa	80.804 kDa
Homo sapiens	gi|119613003	transglutaminase 7	683aa	76.792 kDa
Homo sapiens	gi|80478896	Transglutaminase 3 (E polypeptide, protein-glutaminegamma-glutamyltransferase)	693aa	76.757 kDa
Culex quinquefasciatus	gi|170034605	annulin	734aa	83.04 kDa
Culex quinquefasciatus	gi|170034603	glutamine gammaglutamyltransferase	917aa	104.221 kDa

Caenorhabditis elegans	gi|17508733	SPindle Defective family member (spd-5)	1198aa	135.147 kDa
Aedes aegypti	gi|157114825	protein-glutamine gammaglutamyltransferase	804aa	90.489 kDa
Drosophila melanogaster	gi|33589452	RE08173p	776aa	87.226 kDa
Drosophila pseudoobscura pseudoobscura	gi|125986229	GA20290	774aa	87.275 kDa
Drosophila persimilis	gi|194107225	GL19611	774aa	87.275 kDa
Drosophila ananassae	gi|190615868	GF14572	775aa	87.581 kDa
Anopheles gambiae	gi|270266161	transglutaminase	748aa	83.43 a

Table 1B: Plugin related genes in different taxa

Organisnm	Code	Protein	Length	Weight
Mus musculus	gi|244792359	traf2 and NCK-interacting protein kinase isoform 1	1360aa	154.943 kDa
Mus musculus	gi|124487407	protein bassoon	3942aa	418.838 kDa
Mus musculus	gi|61657921	kinesin-1 heavy chain	963aa	109.55 kDa
Mus musculus	gi|1245105	glutamine repeat protein-1	171aa	21.572 kDa
Mus musculus	gi|1079734	citron	1597aa	183.447 kDa
Mus musculus domesticus	gi|58803289	involucrin	610aa	71.182 kDa
Mus musculus domesticus	gi|13398531	SRY transcription factor	231aa	28.533 kDa
Mus musculus domesticus	gi|3228675	sex determining region of Y protein	231aa	28.533 kDa
Mus musculus domesticus	gi|2623365	sex determining protein	232aa	28.661 kDa
Mus musculus musculus	gi|2623367	sex determining protein	420aa	52.744 kDa
Mus musculus castaneus	gi|191889	apolipoprotein A-IV	435aa	49.253 kDa
Homo sapiens	gi|299890875	dynactin subunit 1 isoform 6	1271aa	140.885 kDa
Homo sapiens	gi|119579218	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2, isoform CRA_a	1590aa	181.277 kDa
Homo sapiens	gi|239735584	TRAF2 and NCK-interacting protein kinase isoform 4	1323aa	150.363 kDa
Homo sapiens	gi|196115147	cip1-interacting zinc finger protein	898aa	100.044

		isoform 1		kDa
Homo sapiens	gi|93141033	PAX-interacting protein 1	1069aa	121.34 kDa
Pongo abelii	gi|197100412	serine/threonine-protein kinase TAO3	898aa	105.36 kDa
Pongo abelii	gi|197098314	mediator of RNA polymerase II transcription subunit 15	566aa	62.003 kDa
Drosophila melanogaster	gi|54144924	DAXX	1620aa	179.546 kDa
Drosophila melanogaster	gi|7537266	defective chorion-1 eggshell protein	180aa	22.261 kDa
Drosophila melanogaster	gi|4755096	accessory gland protein Acp36DE	789aa	88.106 kDa
Drosophila melanogaster	gi|1336720	EN protein binding gene/engrailed nuclear homeoprotein-regulated gene	608aa	69.369 kDa
Drosophila melanogaster	gi|1203923	PsqB	646aa	70.426 kDa
Caenorhabditis elegans	gi|291201266	C. elegans protein D2045.1d, confirmed by transcript evidence	1026 aa	113.461 kDa
Caenorhabditis elegans	gi|44893809	C. elegans protein C07A9.3b, confirmed by transcript evidence	965 aa	109.273 kDa
Aedes aegypti	gi|108883429	Sec24B protein, putative	1188aa	131.915 kDa
Aedes aegypti	gi|108871039	circadian locomoter output cycles kaput protein (dclock) (dpas1)	900aa	101.826 kDa
Aedes aegypti	gi|108871862	hypothetical protein AaeL_AAEL011800	1154aa	129.14 kDa
Plasmodium falciparum	gi|33413782	normocyte binding protein 2a	3096aa	366.014 kDa
Plasmodium falciparum	gi|22086284	chimeric erythrocyte-binding protein MAEBL	2055aa	243.21 kDa
Apis mellifera	gi|110777193	PREDICTED: similar to Transcription factor Sp3 (SPR-2)	780aa	86.043 kDa

Apis mellifera	gi|110763985	PREDICTED: similar to Sex combs reduced CG1030-PA, isoform A	713aa	81.971 kDa
Anopheles gambiae	gi|270266159	plugin	557aa	65.534 kDa
Anopheles gambiae strain PEST	gi|58388669	AGAP006418-PA	266aa	29.279 kDa
Chagas	gi|33112398	Full=Male accessory gland secretory protein 57Dc; cAMPdependent phosphoprotein,	103aa	11.608 kDa

Legend: The organism names, protein IDs, protein names, protein lengths and protein weights are shown on the table.

4.1.5 Maximum likelihood phylogeny and evolutionary relationships in transglutaminase and plugin in *Anopheles gambiae* and their orthologs in other taxa

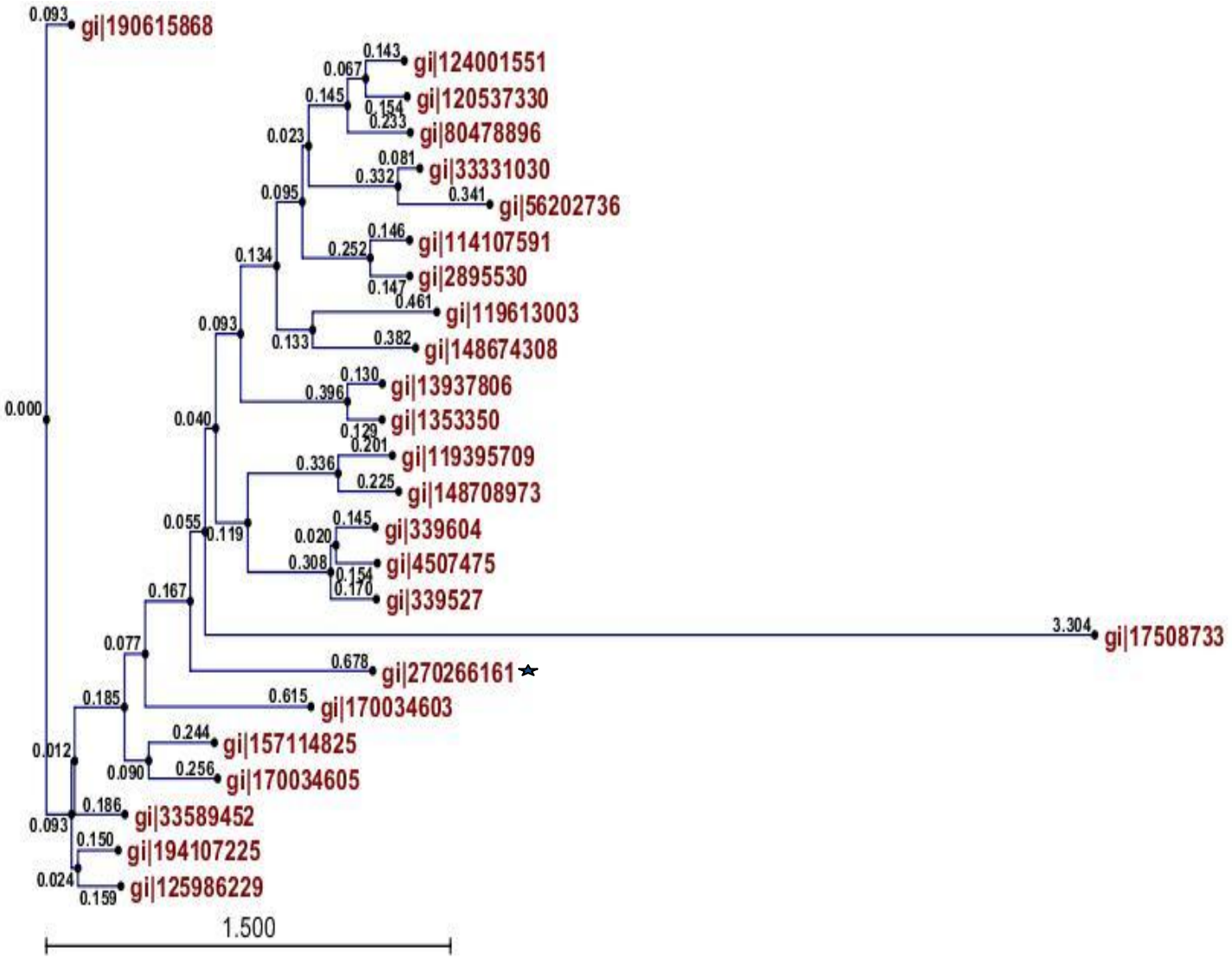
The K value used to determine the phylogenetic signal in these data showed a value of 2 for transglutaminase and plugin trees (Table 2). The value of 2 for the transglutaminase group could be due to the value of *C. elegans* (gi|17508733) which presents a K value of 3.304 (Figure 7A). The branch length K values for most proteins on the transglutaminase tree were less than one (<1). *An. gambiae* (gi|270266161) also presents a value of 0.678 which ranks second but shows transition rates higher than transversion rates hence lower substitution rates. The gamma value of 0.62 closer to 0.5 implies strong variation rate. In the plugin tree (Figure 7B), most genes had a substitution rate characterized by higher transversions but with constantant variation over sites given by a gamma value of 2.18. The ML ratio test of 1150.03 and 1061.17 for transglutaminase and Plugin group of proteins respectively were all statistically significant (p value < 0.001) and under selection as shown by the the K80 tree.

Table 2: Maximum likelihood analysis of transglutaminase and plugin sequence groups

	JC model values			K80 model values			X^2 (L_1-L_0) df=1
	L_0	K	α	L_1	K	α	
TGase	-57189.15	none	0.6	-56039.12	2.0	0.62	1150.03, p<0.001
Plugin	-133136.27	none	1.95	-132075.1	2.0	2.18	1061.17, p<0.001

Legend: L_0: likelihood value for null test (t=0), L_1: likelihood value for alternative test (t=1). K: transversion/transition, α: gamma value. JC: Jukes and Cantor 1969 substitution model, K80: Kimura 1980 substitution model. Df: degree of freedom

A


0.093
gi|190615868
0.143
gi|124001551
0.067
0.145
gi|120537330
0.154
0.233
gi|80478896
0.023
0.081
gi|33331030
0.332
0.341
gi|56202736
0.095
0.146
gi|114107591
0.134
0.252
gi|2895530
0.147
0.461
gi|119613003
0.093
0.133
0.382
gi|148674308
0.130
0.396
gi|13937806
0.000
0.040
gi|1353350
0.129
0.201
gi|119395709
0.336
0.225
gi|148708973
0.055
0.119
0.145
gi|339604
0.020
gi|4507475
0.308
0.154
0.170
gi|339527
0.167
3.304
gi|17508733
0.077
0.678
gi|270266161
0.185
0.615
gi|170034603
0.244
gi|157114825
0.012
0.090
0.256
gi|170034605
0.186
gi|33589452
0.093
0.150
gi|194107225
0.024
0.159
gi|125986229
1.500

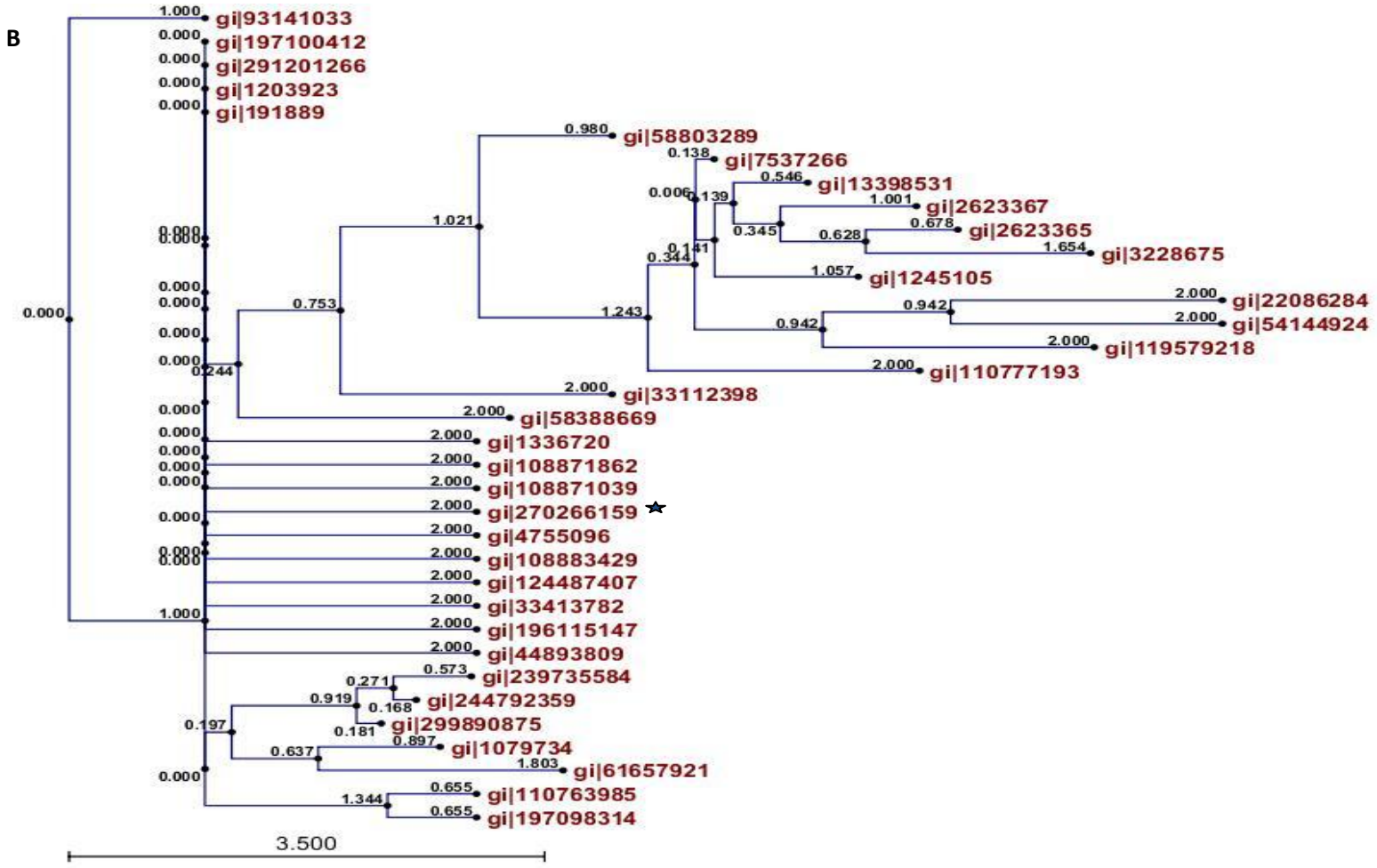


Figure 7: Maximum likelihood (ML) and evolutionary trees for plugin and transglutaminase.
This shows the ML evolutionary trees done with Neighbour joining (NJ) tree as a starting tree. The values on the branches represent K valuesA) Shows the tree for the transglutaminase group of proteins and the star represents *An. gambiae* transglutaminase sequence. B) Shows the tree for Plugin group of proteins and the star represents *An. gambiae* plugin sequence.

4.1.6 Population genetic analysis of transglutaminase and plugin sequence groups

For the transglutaminase group of proteins, 3596 number of sites were considered with number of polymorphisms = 1279 for nT1 (mammals) and 1333 for nT2 (Insects + others). The number of mutations (Mu) were nT1 = 2831 and nT2 = 2429. Nucleotide diversity (Pi) was nT1 = 0.48 and nT2 = 0.46. Mutations polymorphic in nT1 but monomorphic in nT2 were 950 meanwhile those polymorphic in nT2 but monomorphic in nT1 were 548. The average number of nucleotide substitutions per site (Dxy) between populations was 0.55. ω = Non-synonymous substitutions/synonymous substitutions were: nT1 = 20/35 = 0.57, nT2 = 33/33 = 1. None was found between populations. For the plugin group of proteins, total number of sites considered was 11826 with number of polymorphisms = 36 for nP1 (mammals) and nP2 (Insects + others). The Mu was 87 for nP1 and 103 for nP2. Pi was 0.62 for nP1 and 0.64 for nP2. Mutation polymorphic in nP1 but monomorphic in nP2 were 3 while those polymorphic in nP2, but monomorphic in nP1 were 19. This group had a Dxy of 0.64. No ω could be computed for this group of proteins because it considered only 36 sites without alignment gaps out of the total sites.

4.2 Putative protein-protein interactions mediating mating-plug formation and post-mating phenotype in *Anopheles gambiae*

4.2.1 Ortholog identification of mating plug proteins

Proteins identified in *An. gambiae* mating plug were 27 of which 16 were strings from STRING database (Table 3). About 34.88%, 30.23%, 20.93%, 11.63% and 2.33% of the 27 known plug proteins and 16 strings originated from Chromosomal arms 2L, 3R, 2R, 3L and X respectively. Among the unknown proteins on the mating plug; AGAP013150 and NOVEL ZCP7 amino acid sequences identified with AGAP004671 and AGAP008071 respectively, while NOVELACP1 had no similarity to proteins in the database. The prediction scores for the orthologs in the STRING database ranged from 40 to 1311. Of the 27 proteins identified 22.22% were from the female, 55.56% were from the male and the rest from both. Majority of these proteins were from MAGs, composed of Acps, enzymes and proteases. The majority of proteins identified by Rogers *et al,* (2009) originated from Chromosome 3R, but with added strings, majority originated from chromosome 3R and 2L (Table 3). Artemis Comparison Tool (ACT) confirmed orthology among the proteins. Comparisons done on Chromosomal arms with predictions from the STRING database showed 24.39% matches (M), 12.20% mismatches (MM) and 63.41% no matches (NM) (Table 3). Regions that matched for orthologs ranged from 79% to 97% identity. The STRING database also identified putative proteins (STRING column Table 3) which could function together in a protein-protein interaction manner with those already identified on the mating plug.

Table 3: Identified orthologs of *An. gambiae* in *D. melanogaster* using STRING database

An. gambiae ID (plug proteins)	STRING	Chromosome	Sex	Orthologs in *D. melanogaster*	Uniprot ID	Xsome	STRING Score	Xsome synteny
TRANSGLUTAMINASE (AGAP009099)		3R	male	CG7356	Q9VLU2	2L	108	MM
					Q8IPH0	2L	108	MM
AGAP009368		3R	male	CG15005	Q9VZG4	3L	40	NM
AGAP009370		3R	male					
AGAP012830		Unknown	male			unknown		
AGAP008276		2R	male	CG12350	Q7JPN9	3R	139	NM
AGAP008277		2R	male	CG12350	Q7JPN9		137	MM
AGAP013150 (AGAP004671)			male	CG4738	Q9VKJ3		362	NM
AGAP005791		2L	male	CG32834	Q9WIW6 D3DMG3	2R	74.7	NM
AGAP007041		2L	male	CG6676	Q95SM8	2R	172	NM
AGAP006418		2L	male	CG32679	Q8IRL3 D9PTU6	X	166	NM
AGAP009673		3R	male	CG5976	Q7KTY3 Q0GT94	3L	317	MM
AGAP003083		2R	male	CG6113	Q9VKT9	2L	180	NM
AGAP001649		2R	male	CG31414	Q8IMY3	3R	537	M
				CG3647	Q4V4A3 Q4V4J1			
	AGAP0012412	3L		CG6437	Q9W297	2R	583	NM
	AGAP002055	2R		CG3132	Q9VGE7	3R	642	NM
AGAP009584		3R	both	CG31884	Q9V429	2L	134	NM
	AGAP000565	X		CG2151	P91938	X	687	M
				CG11401	Q9VNT5	3L	687	
	AGAP011107	3L		CG6852	B7Z076	3L	144	NM
					Q9VVT6	3L	144	
	AGAP010517	3L		SOD1	B8YNX4	3L	302	NM
	AGAP007201	2L		TRX-2	Q6HI1	2L	169	NM
	AGAP007827	3R		CG17654	P15007	2L	736	M
	AGAP009623	3R		CG8893	P07487	X	544	M
	AGAP007120	2L		CG2210	P08879	3R	292	NM
	AGAP006818	2L		CG8975	P48592	2R	588	NM
				CG17797	O46197	2L		
	AGAP010198	3R		CG5371	P48591	2L	1311	M
	AGAP001325	2R		CG32920	Q960M4	3R	238	NM
				CG7217	Q6XHE3	3R		
AGAP012407		3L	both	CG6988	P54399	3L	642	M
	AGAP007393	2L		CG8983	Q3YMU0	2R	696	NM
	AGAP002816	2R		CG1333	Q9V3A6	3L	582	M
AGAP011630		3L	female	CG33998	Q6IG52	2R	66	NM
					B3DN29	2R	66	
CATHEPSIN B (AGAP004533)		2R	both	CG10992	Q9VY87	X	464	M
AGAP005194		2L	female	CG5255	Q9VEM5	3R	154	NM
TRYPSIN-LIKE SERINE PROTEASE (AGAP005195)		2L	female	CG4053	Q9VEM7	3R	136	NM
AGAP006904		2L	female	CG4859	Q9W122	2R	762	M
				CG4859	Q8MLN6	2R		
	AGAP003319	2R		CG6281	Q9VH14	3R	148	MM
AGAP007347		2L	female	CG7798	A1ZAB8	2R	131	NM
AGAP003139		2R	both	CG18525	Q9VFC2	3R	293	NM
				CG18525	Q9U1I4	3R		
AGAP006964		2L	female	CG32147	Q8SZB7	3L	155	NM
	AGAP009172	3R		CG5355	Q9VKW5	2L	980	M
AGAP006420		2L	both	CG32679	Q8IRL3	X	137	NM

AGAP009212	3R	both	CG7219	A4V9T5	2L	206	NM
NOVEL ACP1 (from female)	3R:b/w 9370 & 9371	male					
NOVEL ZCP7 (AGAP008071)	3R:b/w 5051000&5067900	male	CG8564	Q9VS63	3L	164	NM

Legend: The nucleotide sequences for chromosomal locations of proteins; NOVEL ACP1 and NOVELZCP were used to identify similar proteins and their orthologs. Xsome (chromosome).

4.2.2 Putative protein-protein interaction among *Anopheles gambiae* mating plug orthologs in *Drosophila melanogaster*

Alignments and scores of the orthologs predicted by STRING database predicted a network in Cytoscape of 555 nodes with 2344 edges is summarised in Figure 8. Degree (number of interactions per node) of distribution ranged from 1(150 nodes) to 52(25 nodes). The nodes clustered with coefficients 0.0(150 nodes) to 1.0(225 nodes). ClusterViz predicted 14 complexes with 238/400, 78/227, 53/1287, 38/160, 31/57, 30/72, 16/16, 15/21, 14/11, 10/41, 10/10, 10/14, 9/17 and 7/12 nodes/edges in descending order. These complexes originated from *D. melanogaster*, *Campylobacter jejuni*, *Arabdopsis thaliana* and *Homo sapiens*. *D. melanogaster* made up complexes 1, 8, 11 and 13. CytoHubba predicted the top 50 hubs (nodes) using EPC algorithm. The degree of interactions per node ranged from 3 to 30 (Figure 8). The largest complex was exhibited by *D. melanogaster*. Among the Cytohubber predicted top 50 hubs in the network, key orthologs to CATHEPSIN B (AGAP004533) were identified in mating plug, together with AGAP007120, AGAP009623, AGAP006818 and AGAP001325 (atrial protein) as strings to AGAP009584 on the plug (Table 3). The node protein ranked first (CG9083) in the hubbal interactions (Figure 9) had no close ortholog in *An. gambiae* but shared similar structural protein properties with plugin (AGAP009368) on the mating plug (Table 3). Accessory gland protein, Acp29AB (234 amino acid) in *D. melanogster* transferred by males during mating was attached to hubbal protein Ribonucleoside-diphosphate reductase (CG8975). The Acp29AB alongside its interactive partners (Acp33A, Acp26A, Acp62F, Acp53Ea, Acp26Ab, Acp36DE, Acp32CD, Acp63F, Acp98AB and Acp95EF) could not be identified interacting on the network despite their predicted interaction in the STRING database. There were no Acp29AB orthologs in *An. gambiae.* Several hubbal proteins had predicted interologs in *An. gambiae* but were not identified on mating plug. Only one female *An.gambiae* interolog (trypsin-like serine protease (AGAP005195) found on the mating plug was terminally identified on the network with one interaction.

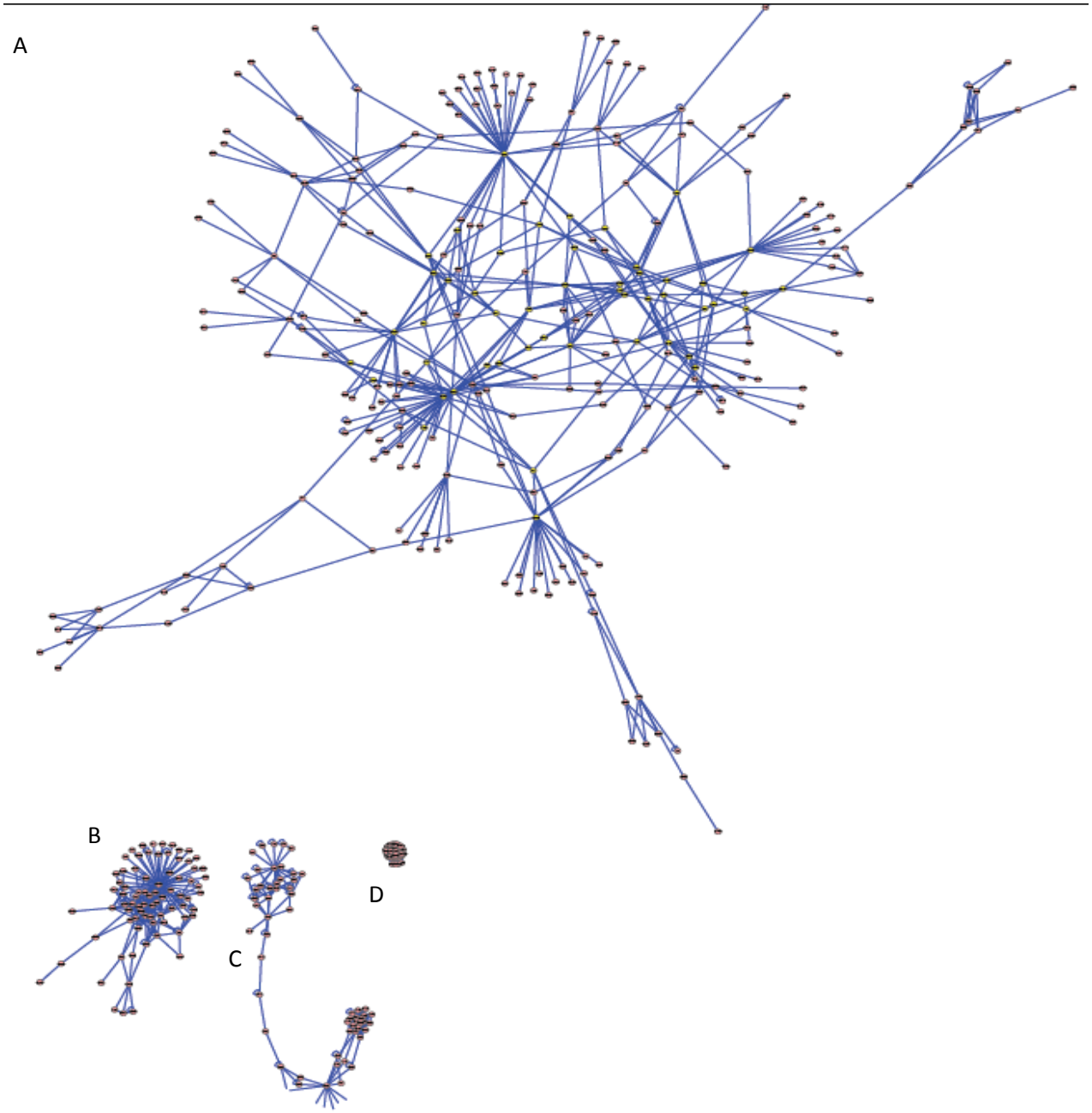


Figure 8: Network map for mating plug proteins in *Anopheles gambiae*; This network design is composed of 555 nodes and 2344 edges. It was done using Agile Protein Interaction DataAnalyzer (APID2NET) hosted in Cytoscape (V.2.8.0). Cluster A is composed mainly of *D. melanogaster* proteins. Clusters B, C and D are made up of *Campylobacter jejuni*, *Arabidopsis thaliana* and *Homo sapiens* respectively.

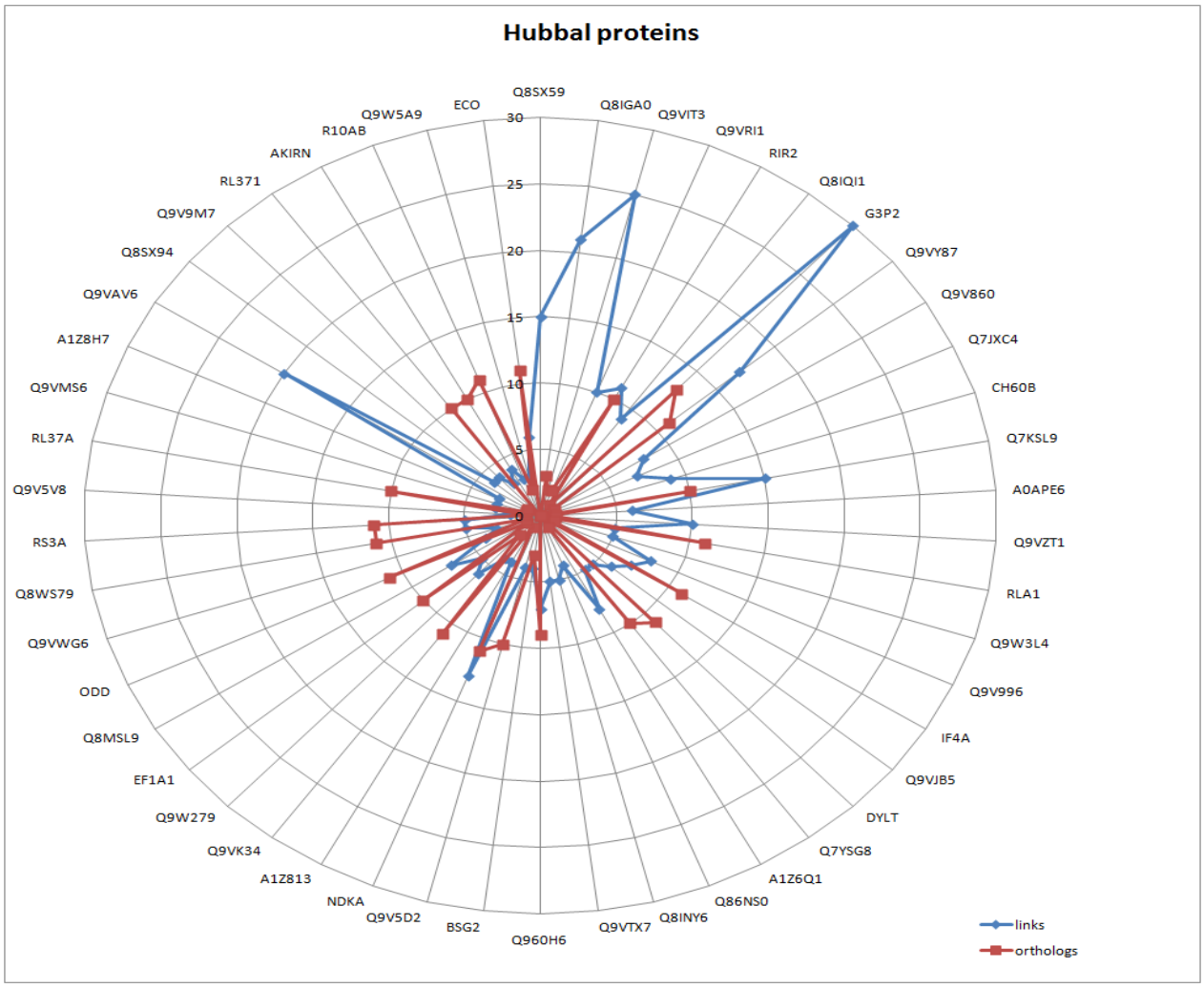


Figure 9: Hubbal nodes within protein-protein network in *Drosophila melanogaster*.
The proteins with their UniProt IDs were ranked in a clockwise direction with Q8SX9 as the top most protein as seen on the figure. The links in the blue graph showe various interactions per node. The red graph show number of orthologs these proteins are close to in *An. gambiae*.

4.2.3 Protein secondary structure prediction of *Drosophila melanogaster* orthologs (Q8SX59 and Q9VEM7) and their molecular and structural homology comparison in *Anopheles gambiae* mating plug proteins (plugin and trypsin-like serine protease (AGAP005195)

From the hubbal proteins; Q8SX59 (CG9083) in *D. melanogaster* was compared to plugin (AGAP009368) in *An. gambiae* using protein secondary structure properties given that the sequence identity between them was poor. The results are summarized in Table (4). The *D. melanogaster* protein had 317 residues, with 19.24% of the residues, vs 10.41% in plugin buried, 80.76% vs 89.59% for plugin exposed, with more than 16% of its surface. The protein contains 3.5% helices, 2.5% strands and 94.0% loops. A high proportion of the protein presents with low complexity regions (LCRs) ranging from aa 13-317 similar to plugin which ranges from aa 1-260. This protein is not globular and lacks DNA binding properties (Table 3). It is Protein-protein binding and its predicted localization in animals is nuclear. PROSITE predicted sequence motifs on this protein were Casein kinase II (CK2-phospho), and N-myristoylation site (Myristyl). The Gene ontology terms (Table 4) showed for molecular function; GO: 0016829 which catalyses the cleavage of C-C, C-O, C-N bonds and GO: 0005509 which interacts selectively and non covalently with Calcium ions. For biological process, it predicted GO: 0009987 responsible for cell growth and/or maintenance and GO: 0008152 which is involved in metabolic processes leading to cell growth. The protein shows no recognised domains across all databases. The protein property predictions for CG9083 were important because they looked very similar to those of plugin in *An. gambiae* (Table 4). In Q8SX59 protein large clusters of Glycine residues within the protein were identified, accounting for the LCRs. Both proteins share similar predicted molecular properties (Table 4), and also lack close orthologs in both species which can be seen by the spurious score for Plugin (Table 3) predicted by the STRING database. The protein Q9VEM7 in *D. melanogaster* having 243 amino acids was compared to the the only female *An. gambiae* interolog identified on the network, TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) having 243 amino acids (Figure 10) showed a 37.3% identity and also shared trypsine like and Gamma-aminobutyric acid (GABA) presented as PTHR similar domains.

Table 4: Protein and molecular properties of plugin in *An. gambiae* and Q8SX59 in *D. melanogaster*

Protein	Aa length	Secondary structure composition (%)			Solubility		S-S	P-DNA	LC region	Prosite	Biological process. GO term (score)
		H	E	L	e (>16%)	b					
PLUGIN (AGAP009368)	557	55.12	0	44.88	89.59	10.41	0	0	Aa 1–260	ASN glycosylation-site, PKC Phosphosite CK2 Phospho-site,	0005509 (0.085) 0008152 (0.207)
Q8SX59 (CG9083)	317	3.42	2.52	94.01	80.76	19.24	0	0	Aa 13-317	CK2 Phospho-site, MYRISTYL	0005509 (0.520) 0008152 (0.327)

Legend: Aa (amino acid), S-S (disulphide bond), P-DNA (protein-DNA binding), LC (low complexity region), GO (gene ontology), H (helix), E (strands), L (loops), e (exposed residues), b (buried residues), ASN (Asparagine), PKC (phosphokinase), CK2 (Casein kinase).

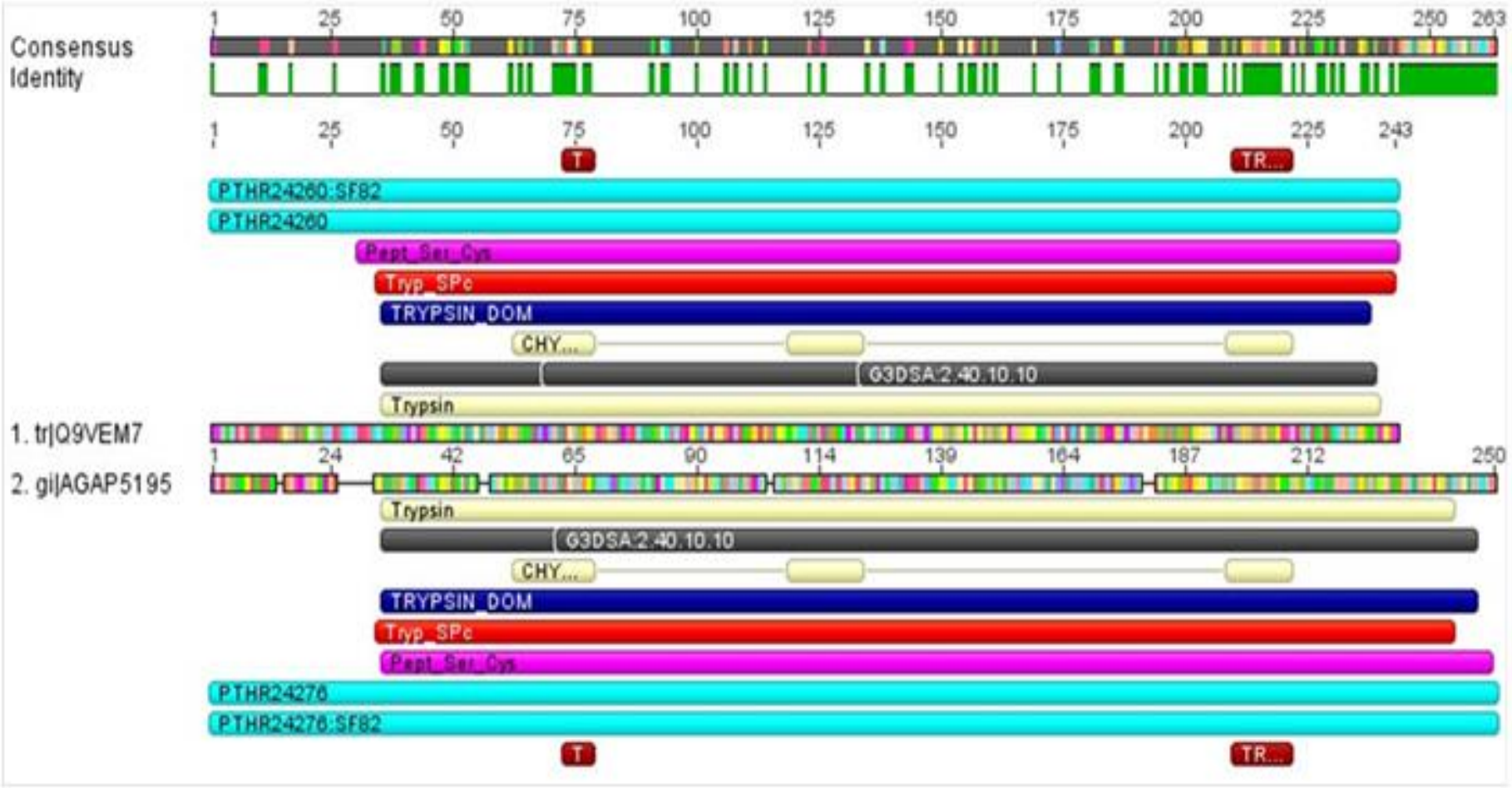


Figure 10: Genetic structural comparison between trypsin-like serine protease (AGAP005195) in *An. gambiae* females and Q9VEM7 in *D. melanogaster*; Dom: domain, CHY: chymotrpsin, Pept_Ser_Cyst: Peptidase_Serine_Cysteine, Tryp_SPc: Trysin_spliced variant, Q9VEM7 (*D. melanogaster* protein), AGAP5195 (AGAP005195, *An. gambiae* protein)

4.2.4 Pathway, Biological process and gene expression identification of *Anopheles gambiae* mating plug orthologs in *Drosophila melanogaster*

Complex 1 contained 113 processes out of 2021 known for *D. melanogaster* and 28 proteins in the network were involved in these processes (Appendix VI). These processes were grouped into 7 main pathways; membrane trafficking (1.1e-05), metabolism of proteins (3.0e-04), metabolism of amino acids and derivatives (4.5e-01), signalling by insulin receptor (2.1e-01), axon guidance (7.1e-01), regulatory RNA pathways (3.8e-02). Complex 8 identified 6 processes out of 2021 known processes with 8 genes involved. These 6 processes were grouped into hemostasis (2.5e-05) as the main process within which, two *D. melanogaster* orthologs (Q9U1I4 and Q9VFC2) of (AGAP003139) as an interolog of the mating plug proteins in *An. gambiae*. The orthologs were identified as serine protease inhibitors involved in platelet plug formation and degradation (Appendix VI). This protein is expressed in both males and females but more significantly in the latter. Complex 11 and 13 showed no identified processes for further analysis. The 50 hubbal proteins were also analysed in the Reactome database to see the main processes controlling this network and 49 processes out of the 2021 known processes were identified. Twelve proteins were involved in these processes which were: metabolism of proteins (8.8e-13), gene expression 2.0e-06, 3'-UTR-mediated translational regulation (7.7e-08), regulation of beta cell development (1.3e-06), diabetes pathways (6.8e-06), signal recognition (Preprolactin) (5.0e-07), and membrane trafficking (1.3e-03) (Appendix VI). The 50 hubs proteins put together included all the processes identified in the first complex with a wider coverage of pathways on the network. Protein metabolism and translational regulation which are the identified processes are aspects of the transglutaminase (AGAP009099) and Plugin (AGAP009369) main proteins involved in mating plug formation in *An. gambiae*. Signal recognition and membrane trafficking are processes required for secreted proteins from male and female reproductive tissues. Gene expression shows up-regulation and down-regulation of genes in the reproductive process. Hemostasis identified on the network is related to plug formation and degradation within the female. Q8SX59 (Table 4) identified as the top most protein in the network showed no expression in larval and adult stages but expressed in the embryo at 16-18hrs in FlyBASE.

The *D. melanogaster* proteins identified on a sub-network within the main network were putatively expressed in four analyzed reproductive tissues (ovary, mated spermatheca, male accessory glands and testes). A sub-network resulting from deriving interactions between Q8SX59 (plugin), Q9VEM7 (AGAP005195) and ACP29AB is summarised in Figure 11 below. Putative plugin protein (Q8SX59) was identified on the sub-network as a

major protein which interacts with Ribonucleoside-diphosphate reductase subunit M2 (RIR2), an interologous string to AGAP009584 found on the mating plug (Table 4). RIR2 showed interactions to two (Q9VCC9 and Q9U9A9) proteins identified in complex 1(Appendix VI) which are both involved in the target of rapamycin pathway (GO: 0032007) and involved in lipid transport of molecules for egg development. The possible interaction of Q8SX59 and RIR2 would be a phosphorylation reaction involved in *An. gambiae* because using CLC main work bench (V5.7.1) (Section 4.2.4), predictions for plugin showed a GGDEF domain (PF00990) which is a sensory receptor domain and a Histidine Kinase domain (PF00512) involved in a two component transduction system known in bacteria. Both domains need a response regulator in order to function properly. This domain (PF00072) was found on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) which was the only female protein identified on the plug network. This domain receives the signal from the sensor partner in a bacterial two-component system. The ortholog (Q9VEM7) of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) clusters on the right arm of Figure 11. This protein is linked to Q9VIT3 expressed both in the ovary and in the testes. This protein has protein binding (GO: 0005515) as the only known molecular function. The possible location of this interaction should be in the female atria because all the terminal proteins (Q7JY68, Q9V9M7, Q9VEM7, Q9W3L4, IF5A, and R10AB) carried by Q9VIT3 are expressed in the testes, male accessory gland and the mated spermatheca. This could be showing the liaison between male and female reproductive cells. Q7JY68 is involved in dephosphorylation (GO: 0016311) which is a close function to response regulator domains as observed on AGAP005195 interolog (Q9VEM7). IF5A is a translation elongation factor (GO: 0003746) and a translation initiation factor (GO: 0003743). It is the only eukaryotic protein to have a hypusine residue which is a post translational modification of a lysine by the addition of butyl amino (Park, 2006) group from spermidine found in semen. Q9V9M7 is a ribonucleoprotein (GO: 0030529) and involved in translation.

At the left arm of Figure 11, Acp29AB which is a MAG protein in *D. melanogaster,* was identified. This protein is involved in post mating behaviour in the female (GO: 0007617). Acp29AB is seen to interact in STRING datadase with Acp26Aa and Acp76A genes. The former (Acp26Aa) codes for a male accessory gland peptide that stimulates egg laying in mated *D. melanogaster* females during the first postmating day meanwhile the latter (Acp76A) which contains a serpin signature plays a role in the observed regulation of Acp proteolysis and/or in the coagulation of seminal fluid to form a mating plug. Interologs AGAP003139 and AGAP009212 on the *An. gambiae* mating plug showed similar serpin

signatures. Acp29AB interacts with Q8WS79 (ERR) expressed in the ovary and in the spermatheca. This is one of the hubbal proteins on the plug network and identified in the generic transcription pathway (Appendix VI). It is a steroid hormone receptor (GO: 0003707) and involved in transcription regulation (GO: 0045449). It combines with a steroid hormone to initiate a change in cell activity. We observe at the terminal (right arm of Figure 11), two proteins (Q9VTX7 and Q7YSG8) expressed in the MAGs, and involved in imaginal disc development (GO: 00007444) and multicellular organismal process (GO: 0032501). Acp29AB probably interacts with Protein similar (SIMA) expressed in the MAGs and in the ovary via Q8WS79 in the spermatheca. It has as function: signal transducer activity (GO: 0004871) and transcription regulation (GO: 0045449). The signal transducer role could aid Q8WS79 in changes in cell function. SIMA and Q8WS79 interact with the Q9VAV6 which is expressed in the testes. At this point Acp29AB could play the function of sperm competition. Q9VAV6 has as function histone acetyl transferase activity GO: 0004402. It is linked to two other proteins Q8IP72 and Q9VJY9 (identified in the Reactome database with complex 1(Appendix VI) involved in pre-microRNA processing (GO: 0031054) and female germline-sterm cell division (GO: 0048132). This protein interacts with IF4A which is a eukaryotic initiation factor involved in protein biosynthesis.

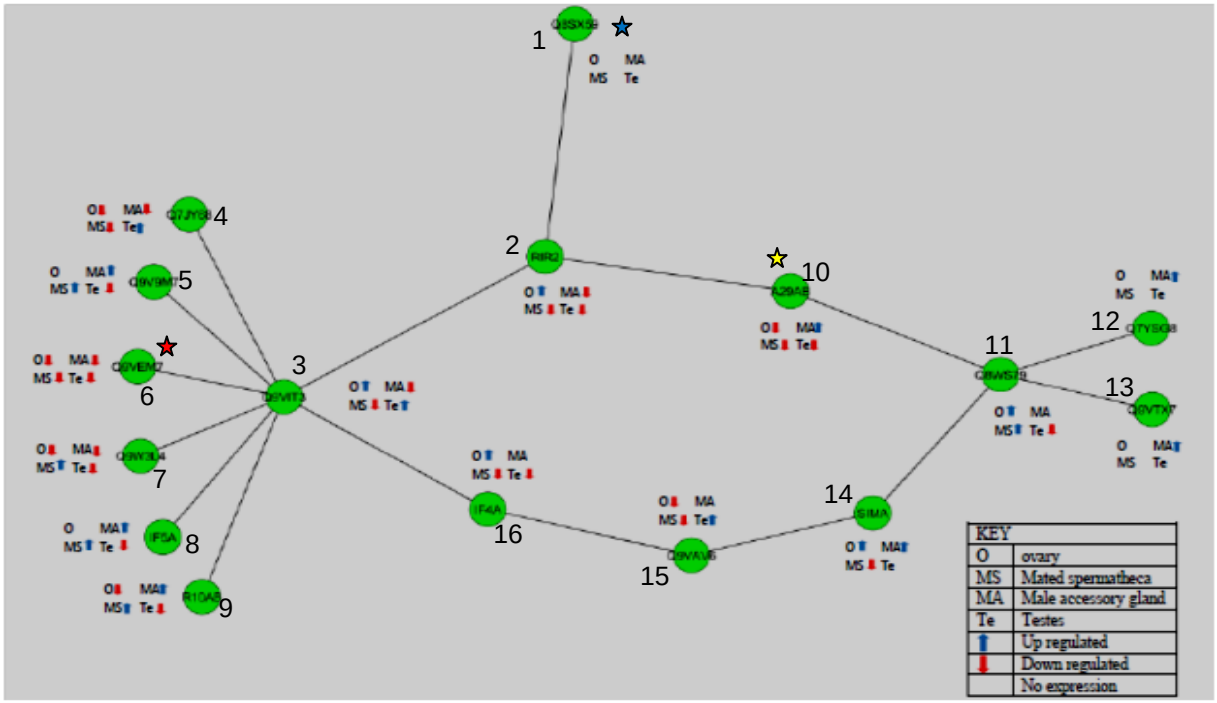


Figure 11: Sub-network predicting reproductive process in *Anopheles gambiae* through *Drosophila melanogaster* proteins; Q8SX59 (blue star) which is putative plugin protein (AGAP009368) is the main protein on the network which interacts with Ribonucleoside-diphosphate reductase subunit 2 (RIR2). Q9VEM7 is found on the right arm of the network alongside other proteins expressed in mated spermatheca. The left arm of the network involves Acp29AB (yellow star) produced mainly in the male accessory glands of *D. melanogaster* interacts with Q8WS79 an Estrogen Related Receptor. The expression values indicated gives a guide on the possible sites where the processes take place. 1 (Q8SX59), 2 (RIR2), 3 (Q9VIT3), 4 (Q7JY68), 5 (Q9V9M7), 6 (Q9VEM7), 7 (Q9W3L4), 8 (IF5A), 9 (R10AB), 10 (A29AB), 11 (Q8WS79), 12 (Q7YSG8), 13 (Q9VTX7), 14 (SIMA), 15 (Q9VAV6), 16 (IF4A).

4.2.5 Comparative analysis of maximum likelihood phylogenies of network identified *Anopheles gambiae* mating plug proteins and their orthologs in *Drosophila melanogaster*

Evolutionary relationships among orthologs of *An. gambiae* proteins in *D. melanogaster* were revealed on mating plug (Figure 12A). The mtREV showed a log-likelihood of -17103.06623, aLRT of 0.00, parsimony value of 3564, tree size of 57.26818, a gamma shape value of 4.280 and using 2423 seconds of computing time while WAG showed a log-likelihood of -16269.60100, aLRT of 0.00, parsimony value of 3560, tree size of 32.67632, a gamma shape parameter of 5.160 and using 850seconds of computing time. From these values, the WAG tree was the best tree due to a ML ratio test value for the two methods showed X^2 = 833.47 (P<0.001), its small tree size (Figure 12A) and evident least evolutionary change between the trees. Similar phenomana was revealed in its parsimony value and the gamma value. MAFFT alignment on the fourteen *An. gambiae* protein interologs was established, and when subjected to WAG maximum likelihood model revealed a likelihood of -16537.39231, aLRT of 0.001 (< 0.05), parsimony value of 3767, tree size of 33.86143, a gamma shape parameter of 5.140 and using 706seconds of computing time. These values were closely related to each other with similarity in both species (*D. melanogaster* and *An. gambiae*) and the proteins involved will have similar substitution rates and equally evolve the same hence directing similar processes in both organisms. The difference observed was the tree topology which rooted branches differently. This difference resulted from difference in percentage identity in multiple sequence alignments (0.8% for *gambiae* and 0.0% for *drosophila*) and percentage pairwise identities from alignments (5.6% for *gambiae* and 7.6% for *drosophila*). Orthologous proteins were identified with a similar colour scheme in both trees (Figure 12 A&B). Comparing the tree topologies in both species, most proteins were maintained in their branch groups though rooted differently. The most interesting difference was related to that of our putative plugin protein (Q8SX59) in the network which is grouped with transglutaminase (Q9VLU2 and Q8IPH0, though poorly supported) in the *drosophila* tree but both are apart in the *gambiae* tree. In the *gambiae* tree transglutaminase is closer to AGAP001649 (Q8IMY3 and A4V4A3) and AGAP009673 (Q7KTY3). The former is a glucosylceramidase (GO: 0008152) involved in glucosidase activity which trims mannosidases involved in N-linked glycoprotein biosynthesis probably activating the transglutaminase enzyme meanwhile the latter is an acyltransferase (GO: 0008415) which is responsible for acylating proteins. Both protein functions are linked to prosite motifs present in transglutaminase (section 4.2.1). The two closely related proteins above are MAGs specific. Plugin is seen closer to the female protein TRYPSIN-LIKE

SERINE PROTEASE (AGAP005195) (Q9VEM7). As seen above, both proteins are expected to be involved in a 2 response regulatory system of phosphorylation reactions. The proteins AGAP008276 and AGAP008277 were grouped together TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) in the *gambiae* tree and they share the same GO terms (GO: 0004252) as seen in section 4.2.4. TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) possesses both Trypsin-Histidine and Trypsin-Serine bindind sites confirming it as a Serine Protease and account for the specificity of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) on the plug. The placement of plugin protein and CATHEPSIN B (AGAP004533) (Q9VY87) [a thiol protease (GO: 0008234)] on the tree shows their independence to the related proteins given their vast interactions on the network [plugin (Q8SX59) as top node (15 interactions) and CATHEPSIN B (AGAP004533) as 8th (30 interactions)]. CATHEPSIN B (AGAP004533) is a cysteine type endopeptidase (GO: 0004197) site common to both transglutaminase and plugin proteins. The weak aLRT branch support (1.24x10-4) separating plugin (blue star), CATHEPSIN B (AGAP004533) , TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) (red star) (Figure 11B), AGAP008276 and AGAP008277 on the tree topology should imply a split of the tree in two parts, possibly those proteins involved in pre-mating grouped with Transglutaminase meanwhile Post mating proteins are grouped with Plugin. The identified maing plug proteins in *An. gambiae* identified on the network showed no direct interaction between them.

A

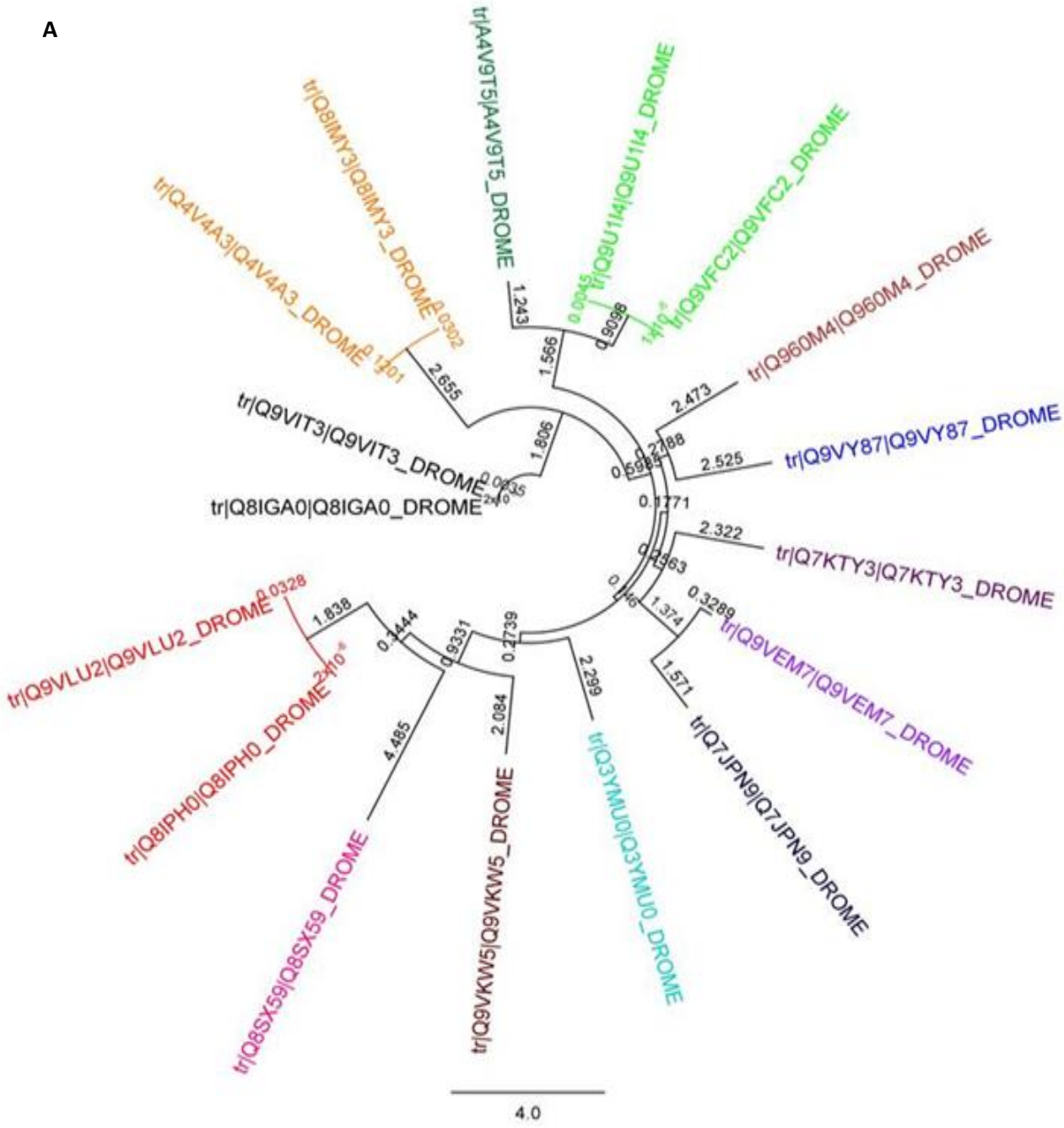


Figure 12 A: Molecular Phylogeny tree in *Drosophila melanogaster*; This figure presents a maximum likelihood tree using the Whelan and Goldman (WAG) amino acid substitution method. The tree was done using the geneious (v5.4) software. These proteins which are orthologs of *An. gambiae* proteins identified on the mating plug were 17 in number. A Multiple alignment program for amino acid or nucleotide sequences (MAFFT) was used with Blosum 62 matrix The colour codes represent the various identified orthologs on the network relative to the *An. gambiae* proteins on the plug. The low alignment identical sites are responsible for the very low aLRT branch lengths seen on the tree hence accounting for its topology.

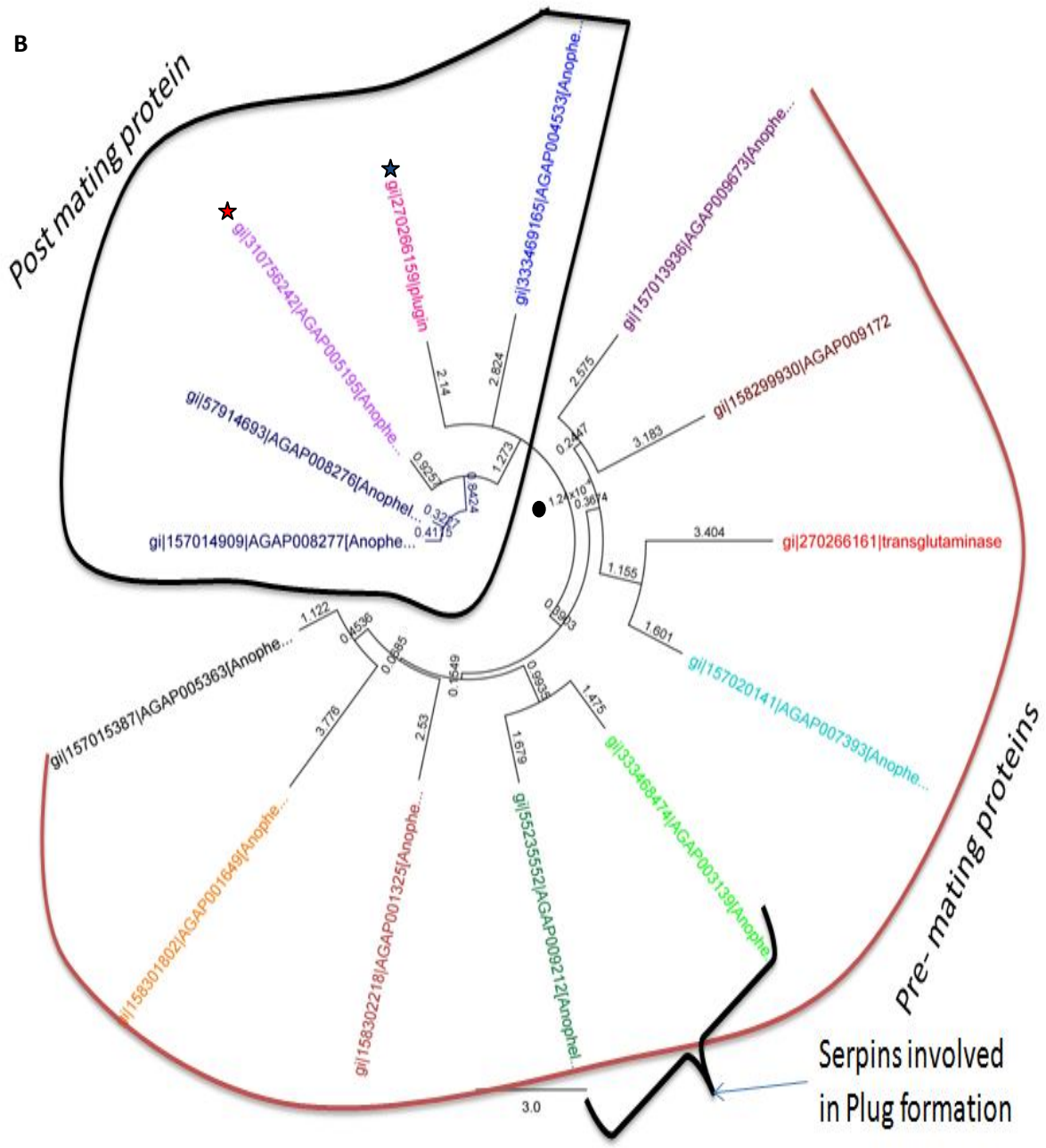


Figure 12 B: Molecular phylogeny tree and putative functional reproductive phenotypes in *Anopheles gambiae*; This figure presents a maximum likelihood tree using the Whelan and Goldman (WAG) amino acid substitution method. The tree was done using the geneious (v5.4) software. These proteins which are the *An. gambiae* proteins identified on the mating plug were 14 in number on the network. A Multiple alignment program for amino acid or nucleotide sequences (MAFFT) was used with Blosum 62 matrix. The colour codes represent the various identified proteins on the plug relative to its orthologs in *D. melanogaster*. The two proteins clusters on the tree predict possible reproductive events in males and females of *An. gambiae*. The low aLRT branch value (1.2×10^{-4}, black dot) separating the two clusters groups plugin (blue star) and CATHEPSIN B (AGAP004533) proteins with the female protein (TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) (red star)) on the tree leading most probably to possible post mating processes in the female. The Serpins which cluster together with transglutaminase should favour plug formation in the male.

4.3 Molecular factors among the interacting proteins mediating mating responses in *Anopheles gambiae*

4.3.1 Analysis based on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) in the female *Anopheles gambiae*

4.3.1.1 Expression analysis on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) in the female *Anopheles gambiae*

TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) is the only female protease identified to cluster with other male proteins on the network derived from mating plug proteins in *An. gambiae*. Its presence as the only female protein on the network resulted from the presence of the Trypsin domain which was absent in AGAP005194 a similar Trysin serine protease not identified on the network. TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) transcript is known to be highly expressed in virgins and then down-regulated more than 2 folds at 24hrs after mating, and only returns to virgin levels 5days post mating. qRTPCR experiments on pre-adult stages using female L4 larvae and pupae of *An. gambiae* (G3 strain) detected no expression of the gene in these stages and expression was limited to adult stages from 1 day old virgin females, where it was confined to the lower reproductive tract (LRT) (Figure 13A). TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) was confirmed to be down regulated 24hrs postmating, and transcript levels remained low up to 10 days post mating (Figure 13B). The same transcript regulation pattern was observed in the atria of the Mbita strain, which however showed higher expression ($P<0.001$, $P=0.0005$) than the G3 strain in virgins (Figure 13C) and is down-regulated in both strains 6 hours after mating. The difference between the two time points (virgins and 6hrs post-mating) was statistically significant in the Mbita strain ($P<0.001$, $t=12.43$) compared to the G3 strain whose difference was not significant ($P>0.05$, $t= 2.539$). The transcript expression in the spermatheca was lost completely 2 hours after mating.

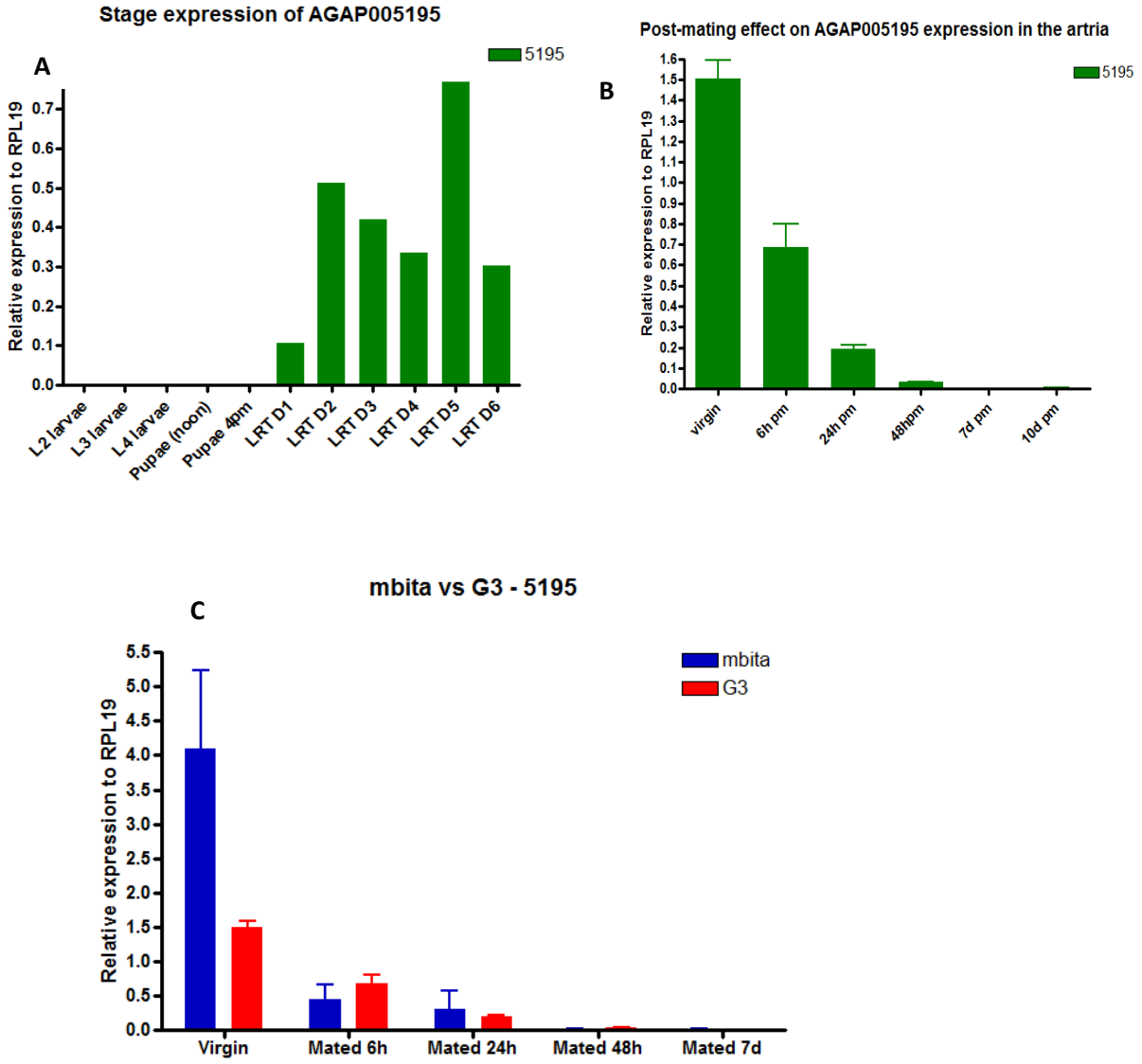


Figure 13: Expression analysis of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) in the atria of the female *Anopheles gambiae*; This figure shows the transcript variation of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) (5195) through qRTPCR analysis. A) This shows that TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) starts being expressed in the lower reproductive tract (LRT (atria and spermatheca)) of the female mosquito at adult day 1 then maintained throughout the virgin female. B) This shows the expression of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) in the 3 day old virgin female atria. After mating the transcript is down-regulated at 6hrs post mating (pm) to no expression at 10day post mating. This could imply transcript never returns to virgin levels. C) This shows the comparative expression of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) in the virgin females of the Mbita strain (blue) and G3 strain (red). The Mbita strain is 3 folds more expressed in the Mbita strain and it is more down-regulated at 6hrs pm than the G3 strain.

The primer pair (Appendix II) designed in Primer 3 (www.ncbi.nlm.nih.gov) was used to amplify an 801bp genomic region (spanning two introns) of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) (www.vectorbase.org) later used for RT-PCR analysis of cDNAs prepared from the spermatheca and atria of virgin and mated females. Different transcript band sizes were obtained. No bands were detected in 6hrs, 2days and 7days postmated spermathecae in females (Figure 14A), while bands were present in atria and virgin spermathecae. The bands showed stacks for the atria (Figure 14 B&C) with bands lesser than 801bp both in the G3 and the mbita strains which may represent different splice variants of the same gene.

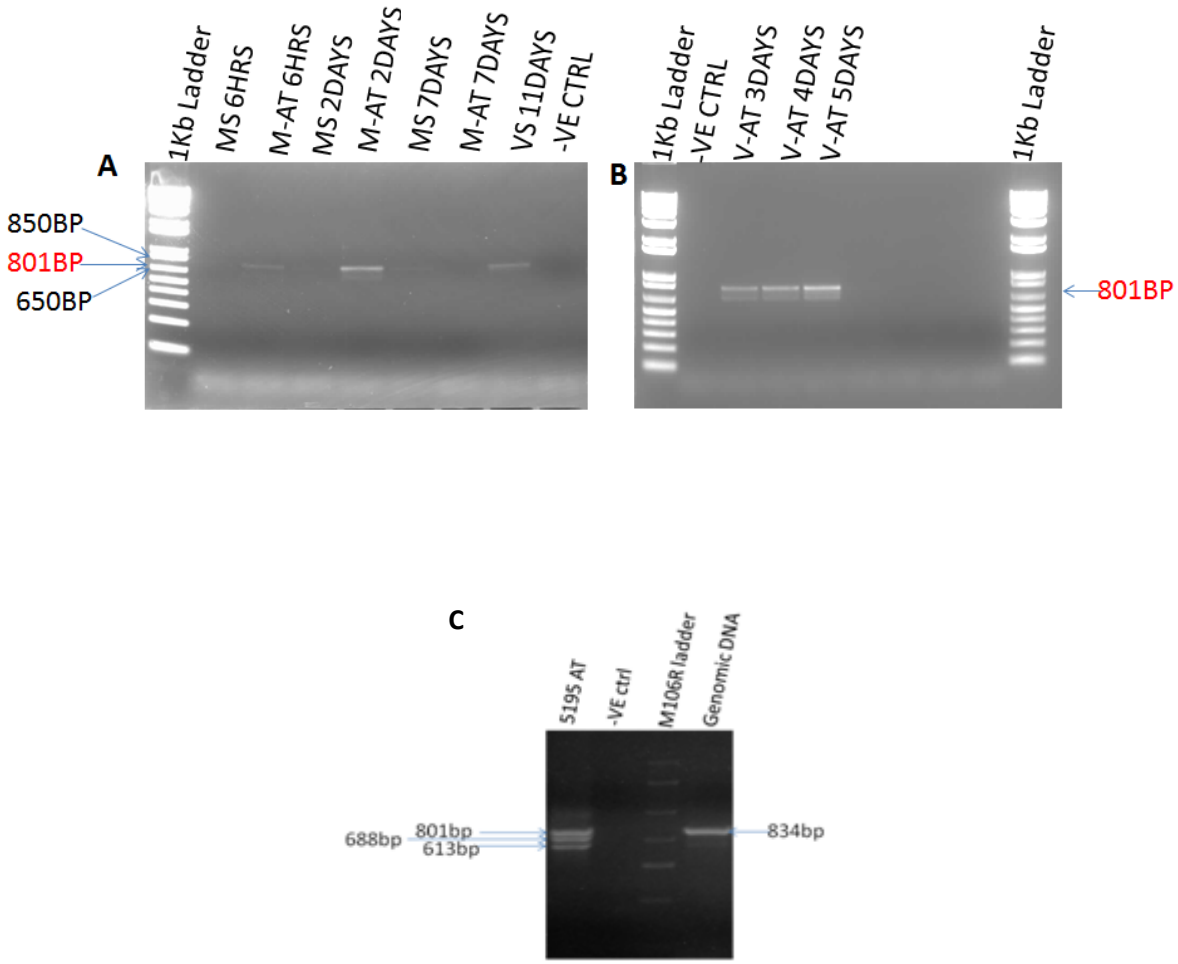


Figure 14: Gel analysis of AGAP005195 transcript in the atria of female *Anopheles gambiae*; This figure shows RT-PCR amplified transcripts for TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) in the female mosquito *An. gambiae*. A) The females were mated and later the atria and spermatheca were dissected at various time points to observe the transcript presence. It can be seen that the transcript is absent after 24hours. TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) is seen expressed in the female spermatheca up till 11days which could imply the transcript is expressed throughout adult life until mating takes place. B) The TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) transcript was amplified for days 3 to 5 in the female mosquito. The bands show stacks which could possibly be different variants. C) This shows the various bands as variants in TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) transcript of the mbita strain, 3 of them (613bp, 688bp and 801bp) in the artria and an 834bp in the genomic DNA. This shows that the 613bp and 688bp variant are the active bands since they are lost after mating as seen in figure A. MS - mated spermatheca, M-AT, mated atria, VS-virgin spermatheca, -VE, negative, CTRL, control

4.3.1.2 Cloning, sequencing and analysis of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) in the atria of *Anopheles gambiae* female

Cloning of the TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) PCR fragments amplified in the atrium and spermatheca revealed several white clones which after performing colony PCR identified; 4 (5195AT-2, 5195AT-3, 5195AT-4 and 5195AT-5) and 3 (5195VS-1, 5195VS-3 and 5195VS-4) possible alternative spliced variants for the virgin atria and spermatheca respectively (Figure 15). After purification the clones were sent for sequencing. Sequences were imported in Geneious software for analyses. The nucleotide and translated protein sequences all identified with TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) (ADP20253) after blast searches at the NCBI database using blosum 62 matrix. 5195VS-4 was not used for the alignment because it was translated on two protein frames with one frame different from other spermathecal clones so it did not align with the other sequences. The 3 sequences from the atria and 2 from the spermatheca (5195AT-2 and 5195AT-5) as described above were translated on two protein frames due to an insertion of an additional Arginine at amino acid (aa) 71 (Figure 18). This additional Arginine residue results from an insertion of AGG codon within the first intron causing a translational site before the second exon hence the break (Figure 17). The two other atrial sequences (5195AT-3 and 5195AT-4) precisely matched the sequence from Vectorbase but 5195AT-3 was higher because it contained sequences within the 2^{nd} intron which were not excised for mRNA formation and is likely to represent an immature form of the mRNA as only one protein band is observed on a western blot gel. The nucleotide alignments show a termination of our variants at 750bp (5195AT-2, 5195VS-2, 5195VS-3), 688bp (5195AT-3) and 613bp (5195AT-4) as opposed to 801bp expected because they were amplified from cDNA and all had a common termination codon (TAA) at this point (Figure 17). The presence of introns then makes the others smaller in size.

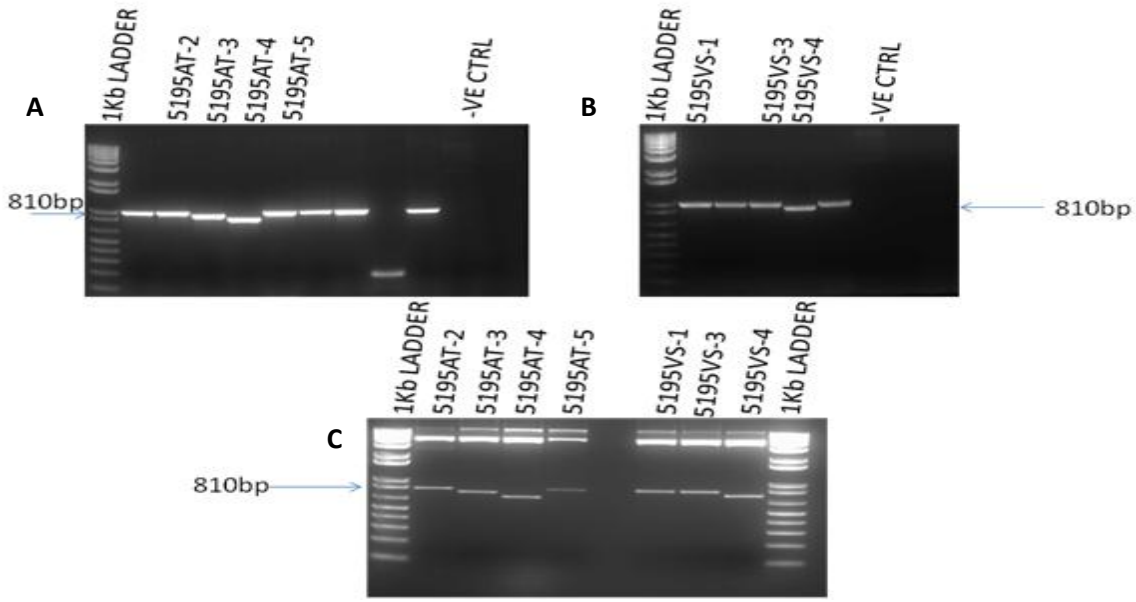


Figure 15: Colony PCR and amplicon identification of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) variants in the atria of female *Anopheles gambiae*; These gels show amplified bands for TRYPSIN-LIKE SERINE PROTEASE (AGAP005195). A,B) colony PCR with M13 primers on various clones. Three variants were identified in the artria and 2 in the spermatheca. Clones 2 to 5 for the artria and clones 1, 3 and 4 as seen above were sent for sequencing. C) Clones were grown in LB overnight and purified. They were digested with EcoR1.

TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) as seen in VectorBase was classified as a serine protease because of the nucleophilic Serine residue expected to be at the active site. This suggests that to consider TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) as a functional Serine protease, three residues (Asp-His-Ser) forming the active site domains are needed to form the active site in its 3D conformation. 5195AT-4 was aligned with AGAP005194, TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) (VectorBase) and ADP20253 (NCBI). Despite the fact that (CHYMOTRYSIN-LIKE SERINE PROTEASE (AGAP005194) was not detected in the network, it was used in the analysis because like TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) , it is specifically expressed in the female atria and also predicted as a serine protease and could be used to identify those molecular properties that differentiates it fromTRYPSIN-LIKE SERINE PROTEASE (AGAP005195).

The Interpro scan for protein domains identified an Isoleucine cleavage site at amino acid position 28 (UniProt P35035), Histidine 67 site (LTAAHC and UniProt ID PS00134) of the Trypsin-Histidine active site, the Aspartic acid site (D-112) and the Serine site (S-200) of the Trypsin-Serine active sites. They all form the Trypsin and Chymotrypsin domains (Figure 16). The serine site sequence (GAC**S**GDSGGPLM) with UniProt ID (PS00135) was identified only in AGAP005194 and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) (Vectorbase) meanwhile the other sequences had a nonsynonymous mutation which changed the Serine to a Glycine at aa position 200 alongside V201G, S202G and K203P. These mutations account for loss of the Serine active site. AGAP005194 had a nonsynonymous mutation which substituted a Glutamic acid for an Alanine at aa 73 at the Trypsin-Histidine motif hence losing the motif but all the proteins identified with the Chymotrypsin or structural and functional role (Figure 16). Our sequences were also compared to ADP20253 deposied at NCBI and several mutations were identified like; Valine for an Isoleucine at aa position 69, Leucine for Valine at aa 92, possibly Threonine for Isoleucine at aa198. However it cannot be ruled out that all these mutations could be due to sequencing artifacts, and more experiments will be needed to verify the accuracy of these data.

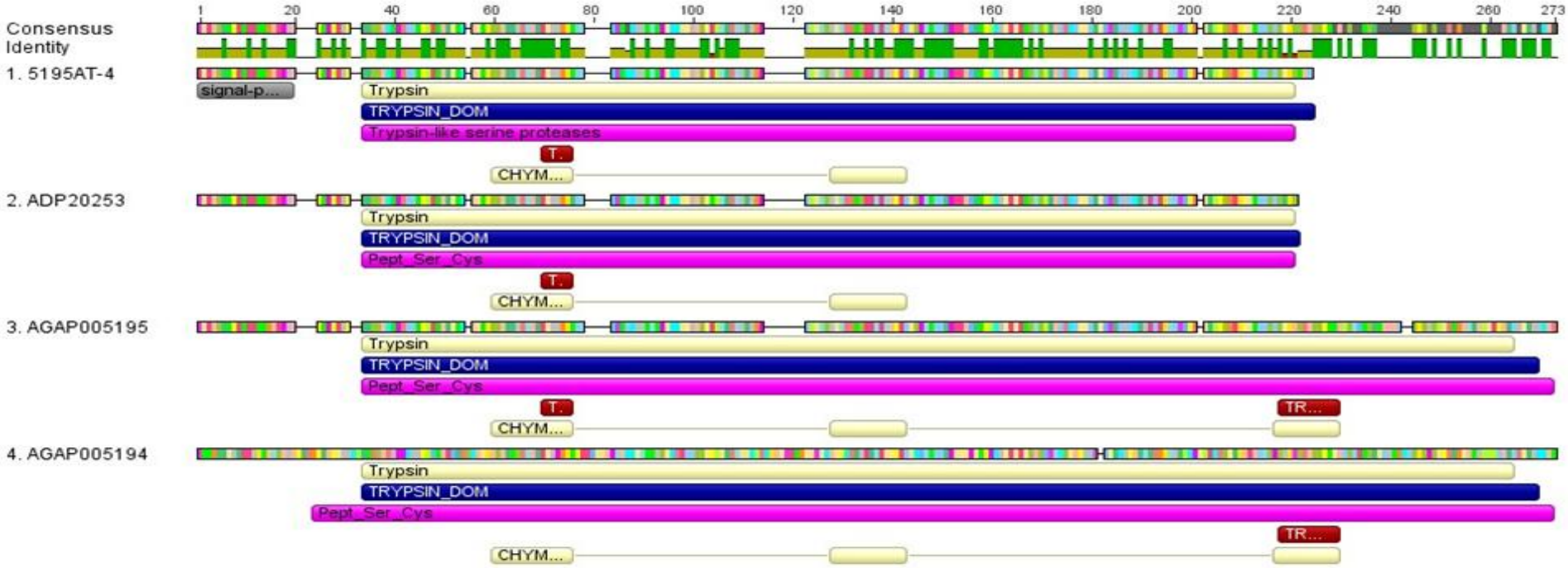


Figure 16: Gene structural variations of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) compared to its functional relatives in the female *Anopheles gambiae*; this shows an alignment of 5195AT-4, ADP20253 (NCBI), TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) and AGAP005194 from Vectorbase. The domains were obtained with InterPro scan. We can observe signal peptide I dentified in all the proteins. The chymotrypsin domain is seen discontinous for 5195AT-4 and ADP20253 with identification of only the Trypsin-Histidine sequence motif (LTAAHC). The loss of the Serine at position 200 deprives them from the Trypsin-Serine sequence motif (GACSGDSGGPLM). A complete domain is seen for AGAP005194 and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) . There is the loss of the Trypsin-Histidine motif on AGAP005194 (LTAGHC). The first two sequences could possibly not function as Serine proteases though posessing the domain.

4.3.1.3 Detection and analysis of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) alternative spliced variants in the female *Anopheles gambiae*

Alternative splicing which occurs through the attachment of the spliceosome complex at the 5'GT and 3'AG ends of the introns was seen as a prominent feature occurring within this sequence hence the reason for the observed variants. Variants 5195AT-5, 5195VS-1, 5195VS-3 and 5195VS-4 have all their introns retained and they show a mutation at the 5' end (GT to TA) of the first intron and an insertion of an additional Guanine at the 3' end (AGG) of this intron. The second intron has the 5'GT and 3'AG at both ends (Figure 17). The mutation observed at the intron affect the binding of the spliceosome complex at the 5' end hence a frame shift in the translated protein. The 4 mutant forms resulting from a frame shift mentioned above for the atria and the spermatheca separates the SERINE_Histidine site on one side of the protein and the Aspartic acid on the other side of the protein with mis-folding structural implications. This was the only molecular form observed in the spermatheca. The activation site identified on this protein is involved in activation of blood factors. Variant 5195AT-3 retains only the second intron having the 5'GT and 3'AG intact. This variant translates to a functional protein on one frame. Variant 5195AT-4 has all the introns excised and translates to a protein on one frame. Variants 5195AT-3 and 5195AT-4 differ by the presence of a non-synonymous mutation at position 75 on the protein sequence where Asparagine (N) replaces a Lysine (K) on 5195AT-4. This is the change of a positive aa to a polar uncharged aa (Figure 18)

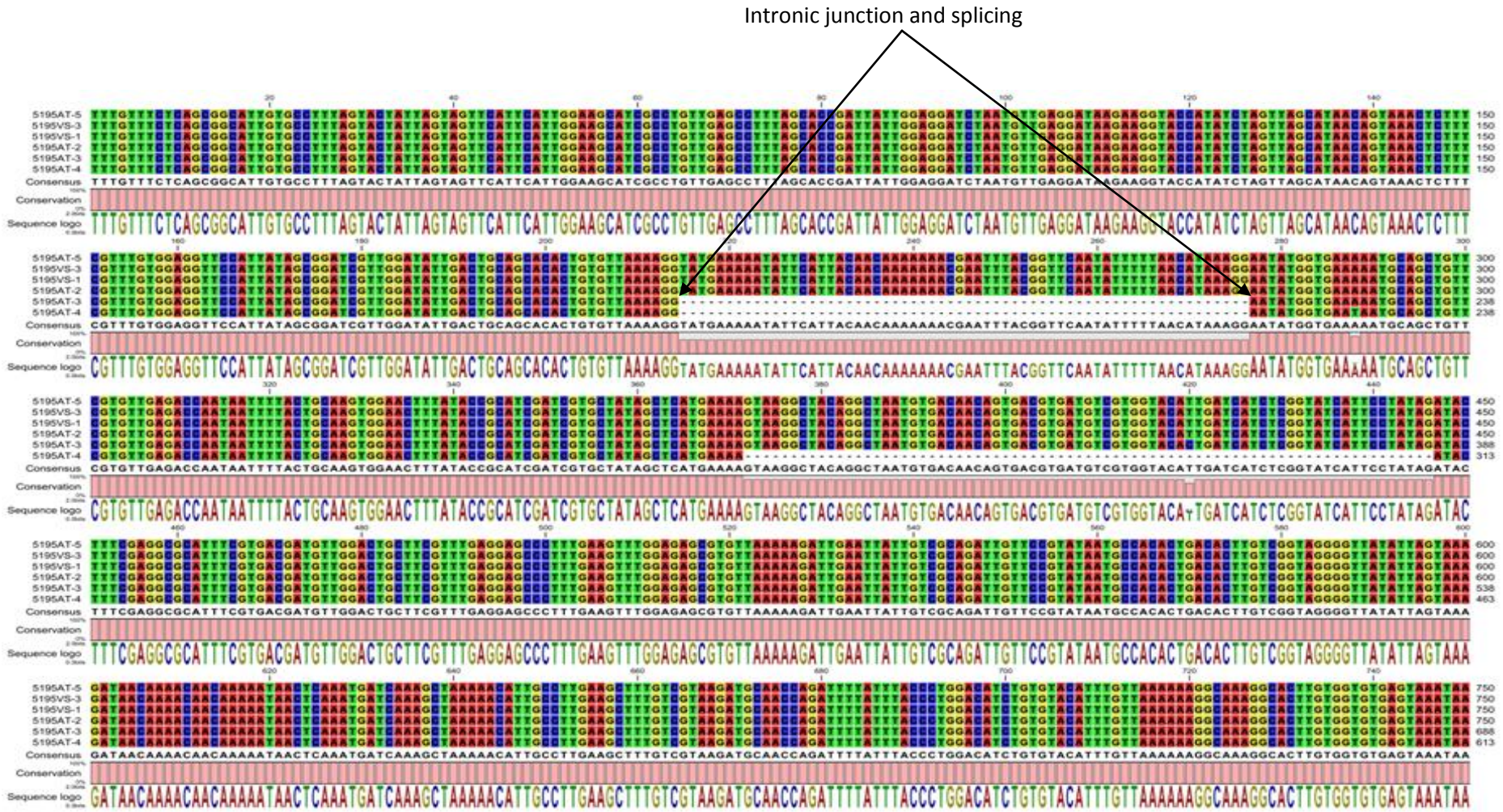


Figure 17: Nucleotide analysis of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) variants in the female *Anopheles gambiae*; This shows nucleotide alignment of the sequences with TTG as the start site. There is an insertion (AGG) at the beginning of the second codon for all the sequences except 5195AT-3 and 5195AT-4. 5195VS-4 like 5195AT-3 was a similar variant seen at the level of the spermatheca. This variant also has the AGG insertion which caused a frame shift in the sequence hence making it non functional. It is not included in the alignment because it was inserted in the reverse direction within the vector. 5195AT-3 and 5195AT-4 translated into functional proteins but with an SNP at nucleotide 226 changing the codon from Lysine into Asparagine

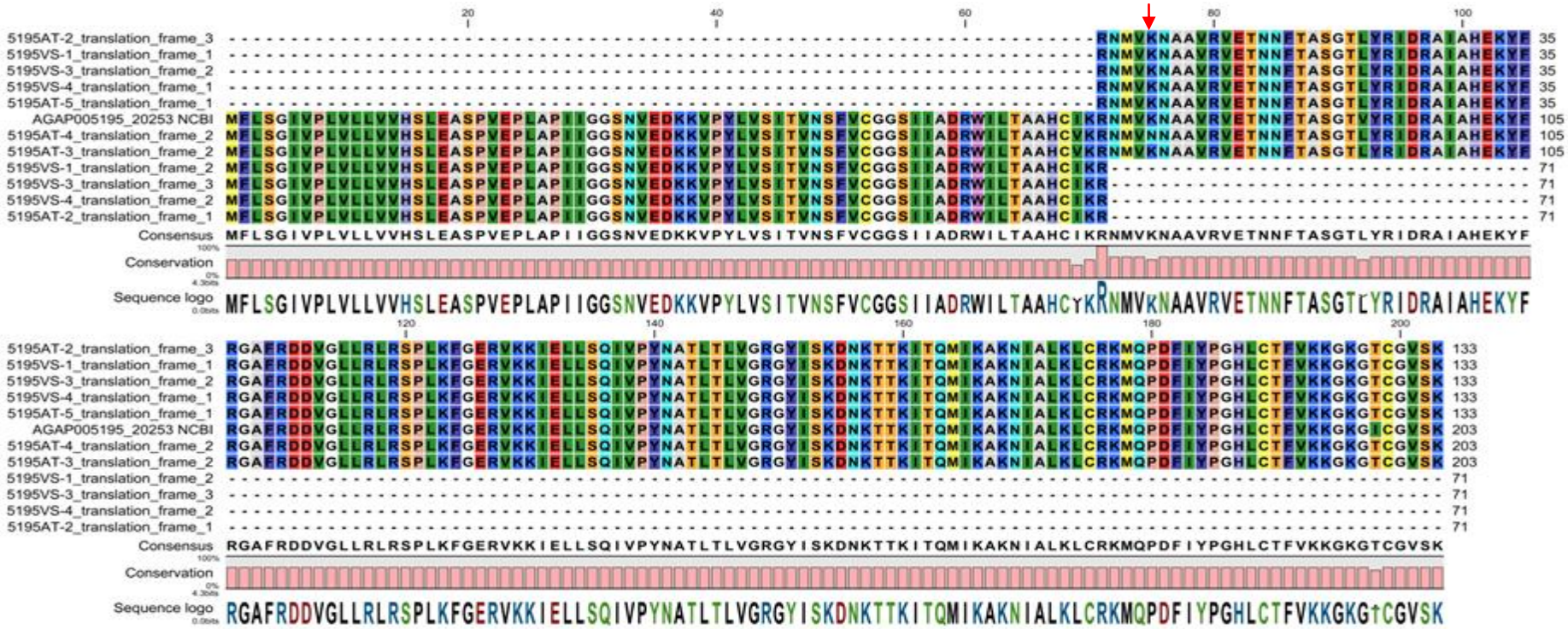


Figure 18: Protein analysis of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) variants in the female *Anopheles gambiae*; This shows the protein alignments for our sequences alongside ADP20253 from the NCBI database. Two of our variants 5195AT-3 and 5194AT-4 were translated into the complete protein on one frame but all the others had a frame shift in their sequences hence the introduction of an Arginine at position 71 on the protein sequence. 3 mutations relative ADP20253 were observed; V69I, L92V, T198I. N75K (red arrow) was observed only for 5195AT-4. 5195AT-3 and 5195AT-4 were similar for the 1st 203aa on the protein sequence deposited in the database with an SNP mutation seen at G200S. The loss of S200 should have a serious functional effect on the protein sequence since it is the serine active site.

4.3.1.4 Identity of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) protein in the atria of *Anopheles gambiae* females

The full length protein for TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) is 250 amino acids (aa). The protein contains a signal sequence of 19 aa (Figure 16). Cleavage of this sequence gives the protein a weight of 25.330 KDa. The Western blot performed on virgin females showed different bands corresponding to 40KDa, 160KDa and 260KDa in the atria (Figure 19), which are immunoreactive bands. Three N_Glycosylation sites involved in post translational protein modification were predicted on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) protein at aa; 85, 141 and 146 using PredictProtein server and could account for the number of bands observed. No band was observed in the spermatheca, despite the detection of transcripts in this tissue and less expressed than in the atria. Actin used as a loading control was seen in high amount in the atria but quite faint in the spermatheca (Figure 19).

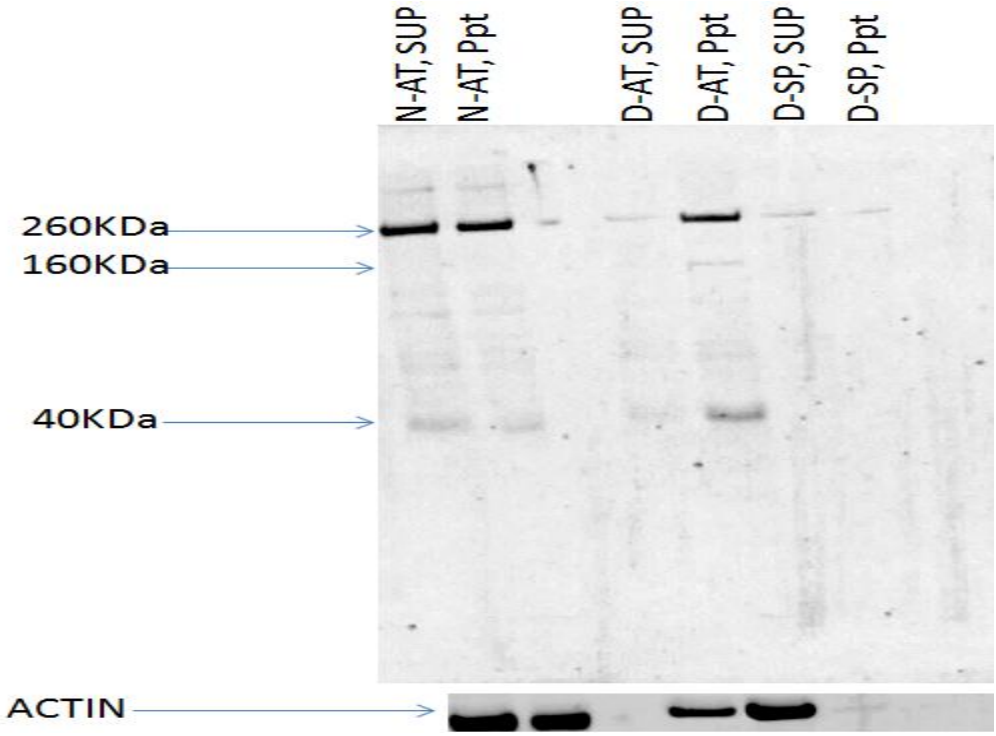


Figure 19: Western blot analysis on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) in the atria and spermatheca of *Anopheles gambiae*; This shows a western blot with specific antibodies detecting the protein TRYPSIN-LIKE SERINE PROTEASE (AGAP005195). A 40KDa band was identified for the protein with much higher bands at 160KDa and 260KDa which are possible immunoreactive reactive bands due to glycosylation sites on the protein. The protein size is 25KDa and two functional variants exist which could account for the protein size observed on the gel. No protein was observed in the spermatheca which could be due to the non functional variant forms of the protein. Actin was used as control and could be seen present at all the lanes though faint at the spermatheca. N (native), D (denatured), AT (atria), SP (spermatheca), SUP (supernatant), Ppt (precipitate).

4.3.1.5 Immunohistochemistry of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) and AGAP005194 as serine proteases on the mating plug, atria and spermatheca of female *Anopheles gambiae*

The location and specificity of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) protein compared to CHYMOTRYSIN-LIKE SERINE PROTEASE AGAP005194 (both serine proteases expressed in virgin females and located on the plug) in virgin and mated female mosquitoes using immunofluorescence staining in mated females 6 hours & 2 days post-mating (Figure 20) and in virgins 1 & 2 days post emergence (Figure 20) identified, TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) in red on the mating plug while CHYMOTRYSIN-LIKE SERINE PROTEASE (AGAP005194) in green was located both on the plug and atrial walls (Figure 20; A, B, C). TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) could also be detected in the secretory cells of the spermatheca while CHYMOTRYSIN-LIKE SERINE PROTEASE (AGAP005194) stained the epithelial cells of the spermatheca. For day one and two virgin females, TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) was identified separately in the atria and spermatheca but not in the duct linking both organs (Figure 20; G, H, I, J, K, L, M) while for the 6 hours mated females, TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) could be seen from the red fluorescence along the spermathecal duct while the green fluorescence staining AGAP005194 could not be located along the duct but localized on the epithelial cells of the spermathecal cuticle (Figure 20; D, E, F). CHYMOTRYSIN-LIKE SERINE PROTEASE (AGAP005194) and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) show differential localizations in the reproductive tissues.

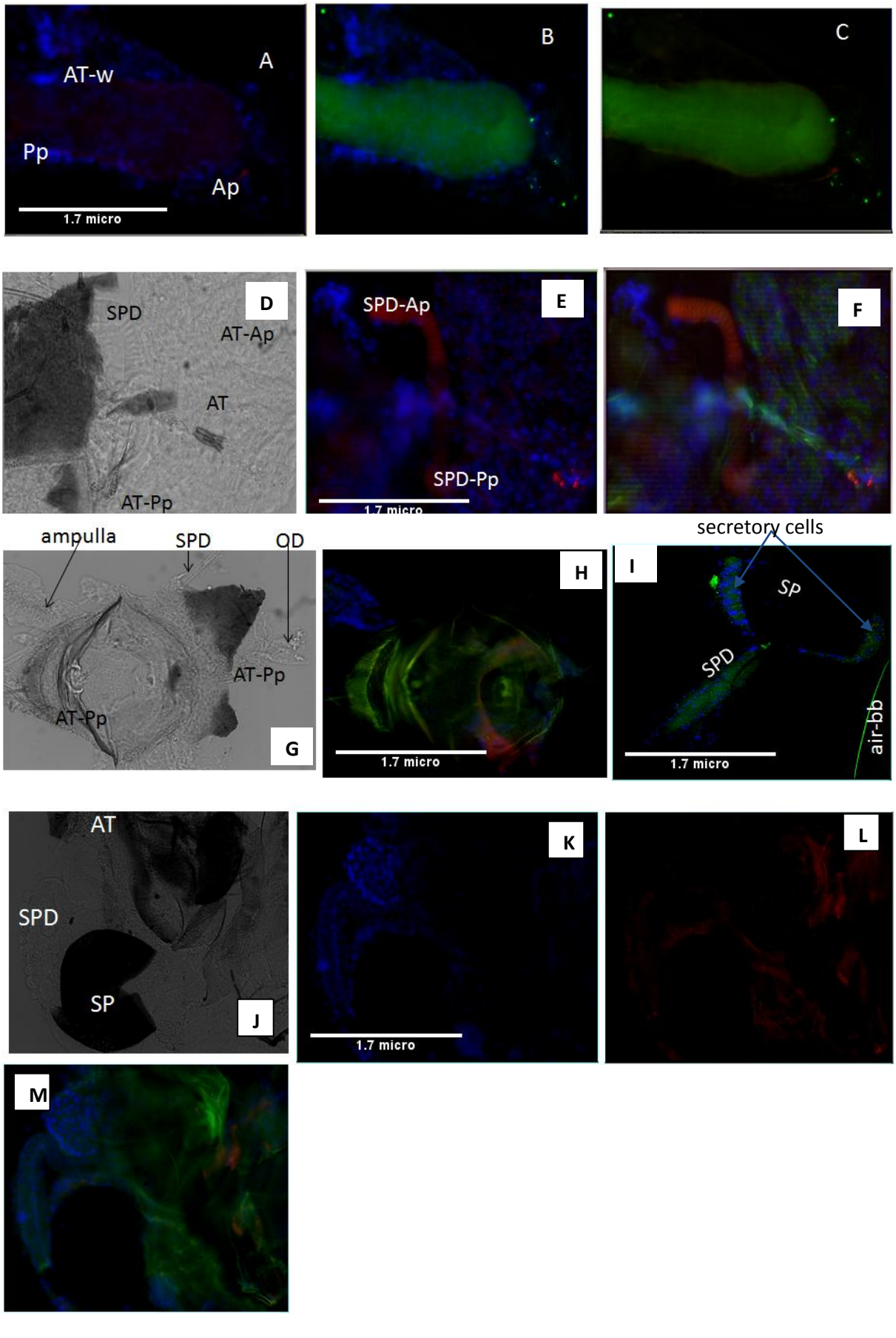

AT-w
A
Pp
Ap
1.7 micro
B
C
SPD
D
AT-Ap
AT
AT-Pp
SPD-Ap
E
SPD-Pp
1.7 micro
F
ampulla
SPD
OD
AT-Pp
AT-Pp
G
H
1.7 micro
secretory cells
I
SP
SPD
air-bb
1.7 micro
AT
SPD
SP
J
K
1.7 micro
L
M

Figure 20: Fluorescent detection of AGAP005194 and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) proteins in the atria and spermatheca of the female *Anopheles gambiae*; this shows the presence of the plug within the artrium. The dissections and staining were done 6hrs after mating. TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) stained with Alexa 546 (red) shows the protein specific to the plug meanwhile AGAP005194 stained with Alexa 488 (green) is also identified on the atrial walls. The blue is DAPI which is seen to stain the nuclei of atrial cells (AT-w). A) DAPI and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195), B) Dapi and AGAP005194, C) TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) and AGAP005194. This shows atrial stainings 6hours post mating. D) The spermathecal duct (SPD) can be seen exiting from the artria (AT). E) The SPD can be seen containing TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) possibly moving to the artria. F) AGAP005194 could be seen localized on the atrial wall structures but not the duct hence probably low involvement in reproduction. G) Bright field with ampulla identified for one day old virgin females. H) AGAP005194 super-imposed other stains with possible presence on cuticular structures. It is absent at the ampulla. TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) secretion appears to be very low. I) Presence of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) in the spermatheca and in the secretory cells though superimposed by AGAP005194. This suggests that the duct has few cells before the spermatheca. J) bright field showing spermatheca duct linking artria and spermatheca for two days old virgin females. K) DAPI stained structures. L) TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) stained structures with possible presence in the cells low expression but not in the duct. M) TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) overshadowed by AGAP005194 in all the structures. Ap-anterior pole, Pp-posterior pole, OD-ovarian ducts.

4.3.1.6 Secondary structure protein modelling of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) of the female *Anopheles gambiae*

Here, the two possible variants (5195AT-3 and 5195AT-4) which translated to the protein on one frame were compared to 5195VB. From the geneious software the full 5195VB protein shows 9 alpha helices, 20 beta strands, 17 coils and 20 turns but for the first 203 aa of 5195VB, it showed 8 alpha helices, 15 beta strands, 12 coils and 14 turns. 5195AT-3 showed for the first 203 aa; 8 helices, 15 beta strands, 12 coils and 13turns due to loss of a coil at the carboxylic acid end of the protein resulting from the mutations (Figure 21). 5195AT-4 showed for the first 203 aa; 8 alpha helices, 14 beta strands, 13 coils and 14 turns. This resulted from the N75K (Figure 18) mutation which introduced a turn and a coil at that position instead of a helix as seen in the former two protein structures. This secondary structure conformational change at this position is a point mutational effect with implications in its 3D structure formation. From the PredictProtein server, (Table 5) 5195AT-4 showed two predicted disulphide bonds between C52-C68 and C175-C189 forming two prominent halves of the protein in its 3D structure. 5195AT-3 showed predicted disulphide bonds at C68-C199 and C175-C189 but the former's prediction had a weak support compared to the latter. This makes a more loose 3D structure. 5195VB shows no disulphide bonds hence making the structure similar to 5195AT-3. 5195AT-3 and 5195AT-4 showed 12 protein binding sites which are all similar but differed for 5195VB at two points (aa 176-177 and 181-182) (Table 5). They all shared the same predicted PROSITES; ASN_Glycosylation (PS00001), PKC_Phospho_Site (PS00005), CK2_Phospho_Site (PS00006), TYR_Phospho_Site (PS00007), MYRISTYL (PS00008), TRYPSIN_HIS (PS00134) but 5195VB had one more site TRYPSIN_SER (PS00135). The last two sites for 5195VB confirm it as a Serine protease as identified in the Geneious software but not in our identified variants. The best protein hits for our variants showed for 5195AT-4 the pdb string 1AMH, an uncomplexed rat Trypsin mutant with ASP189 replaced with Serine (D189S). The pdb string for 5195AT-3 was 1BIT, the crystal structure of anioinic salmon trypsin in a second crystal form. They show that its structure differs from the standard trypsin with a more catalytic effect and with a deviation from substrate specifity by the wild type. The best pdb string predicted for 5195VB was tr/07Q9K2/07Q9K2_ANOGA in UniProt database as TRYPSIN-LIKE SERINE PROTEASE (AGAP005195), a serine endopeptidase. The three different predicted pdb strings showed structural and functional variations.

The Response Regulator receiver domain (RGYISKDNKTTKITQMIK) (PF00072) also known as signal transduction response regulator receiver domain (IPR001789) was

identified on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) at aa 149 to 166 (Aspartate at position 155) using CLC workbench. Arginine/Glycine (R/G) on the Response regulator sequence above is a site digested by Serine proteases and two of them (AGAP008276 and AGAP008277) from *An. gambiae* MAGs were identified on the mating plug. The protein binding site prediction was seen on aa 154–157 (KDNK) and a Protein Kinase C phosphorylation (PKC_Phospho) site was predicted at aa 158–160 (TTK). This PKC_Phosphosite is expected to interact with Plugin (AGAP009368) protein (also found on the mating plug) in a 2 Phosphorelay system. A Histidine Kinase A site was shown on the Plugin protein (section 4.1.1).

Figure 21: Secondary structure analysis of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) variants in the atria of female *Anopheles gambiae*; This figure shows the three variants for TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) (5195AT-3, 5195AT-4 and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) VB (VectorBase)) from top to bottom. TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) VB showed 9 alpha helices, 20 beta strands, 17 coils and 20 turns but for the first 203 aa it showed 8 alpha helices, 15 beta strands, 12 coils and 14 turns. 5195AT-3 showed for the first 203 aa; 8 helices, 15 beta strands, 12 coils and 13turns due to loss of a coil at the carboxylic acid end of the protein resulting from the mutations (Figure 10). 5195AT-4 showed for the first 203 aa; 8 alpha helices, 14 beta strands, 13 coils and 14 turns. This resulted from the N75K mutation (blue star) which introduced a turn and a coil at that position instead of a helix as seen in the former two protein structures.

4.3.2 Analysis based on CATHEPSIN B (AGAP004533) in *Anopheles gambiae* various life stages and reproductive tissues

4.3.2.1 Expression analysis on CATHEPSIN B (AGAP004533) in *Anopheles gambiae* various life stages and reproductive tissues

The CATHEPSIN B (AGAP004533) protein is a peptidase of 337 amino acids (aa) found on the mating plug of *An. gambiae*. According to VectorBase (www.vectorbase.org) it is expressed in both the male and the female. The protein was identified on the network with Q9VY87 as its ortholog in *D. melanogaster* (Figure 8). qRTPCR analysis of the transcript in the G3 strain showed that it was expressed in all developmental stages of the mosquito (Figure 22) (egg, larva, pupa and adult), with highest expression levels observed in pupae and young adults. The transcript was also found in all tissues of both male and female mosquitoes, with highest expression levels relative to the control gene RPL19, observed in female atria of the G3 strain (Figure 23 A). In the males of the G3 strain, CATHEPSIN B (AGAP004533) expression was higher in virgin testes than the MAGs (Figure 23B). The expression increased 6 hours to 10 days post mating in both tissues. In the female of the mbita strain (Figure 23C) the expression was highest in the atria than in the spermatheca and the ovary. There was a down-regulation of the gene in all the tissues 6 hours and 10 days post mating. Analysis done for CATHEPSIN B (AGAP004533) on the male *An. gambiae* mbita strain confirmed a higher expression of the gene in virgin male testes than MAGs (Figure 23D). This expression though not statistically significant ($P<0.05$) remains fairly stable even after mating meanwhile in the females the gene is highly expressed in the atrium than in the spermatheca and is down-regulated 6 hrs to 7 days postmating.

CATHEPSIN B (AGAP004533) is found in the same cluster of the network as TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) and Plugin (AGAP009368). PFam domain analysis done on the protein using CLC workbench identified two domains (Table 5); D-Amino Acid Oxidase (DAAO, PF01266) and RNA Recognition Motif (RRM, RRD, RNP domain, PF00076). When compared to transcripts levels of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195), the expression levels of CATHEPSIN B (AGAP004533) in the atria from 6 hours to 7 days postmating were down-regulated both in the G3 and mbita strains (Figure 23E). The expression levels in the spermatheca are also lower while in the ovary the expression levels are relatively very equal and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) is known to have an insignificant expression in the ovary (Figure 23C).

CHYMOTRYPSIN-LIKE SERINE PROTEASE AGAP005194 and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) from previous work are known to be down-regulated

more than two folds after mating and lost after 24 hours and two lipid related proteins AGAP004203 (vitellogenin) and AGAP002620were expressed after mating, 6 hours and 24 hours respectively therefore qRT-PCR analysis done on AGAP00533, AGAP005194 and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) transcripts in the mbita strain showed that these transcripts were down-regulated 24hours to 48hours postmating. The expression values of vitellogenin and AGAP2620 were seen to be upregulated 6hrs and 24hrs respectively and then lost up to 7 days post-mating (Figure 23F).

The expression of CATHEPSIN B (AGAP004533) in the males of the mbita strain was compared to the relative expression of TRANSGLUTAMINASE (AGAP009099), AGAP009370 and AGAP009368 (Plugin) which are transcripts expressed in virgin males and encode proteins present on the mating plug. In the MAGs, Plugin and AGAP009370 were seen to be up-regulated 6 to 24 hours after mating while the expression levels of Transglutaminase from 6 hours to 7 days was more stable. In the testes only Plugin and AGAP009370 were seen to be highly expressed at 48hours post mating more than 100folds (Figure 23G & H).

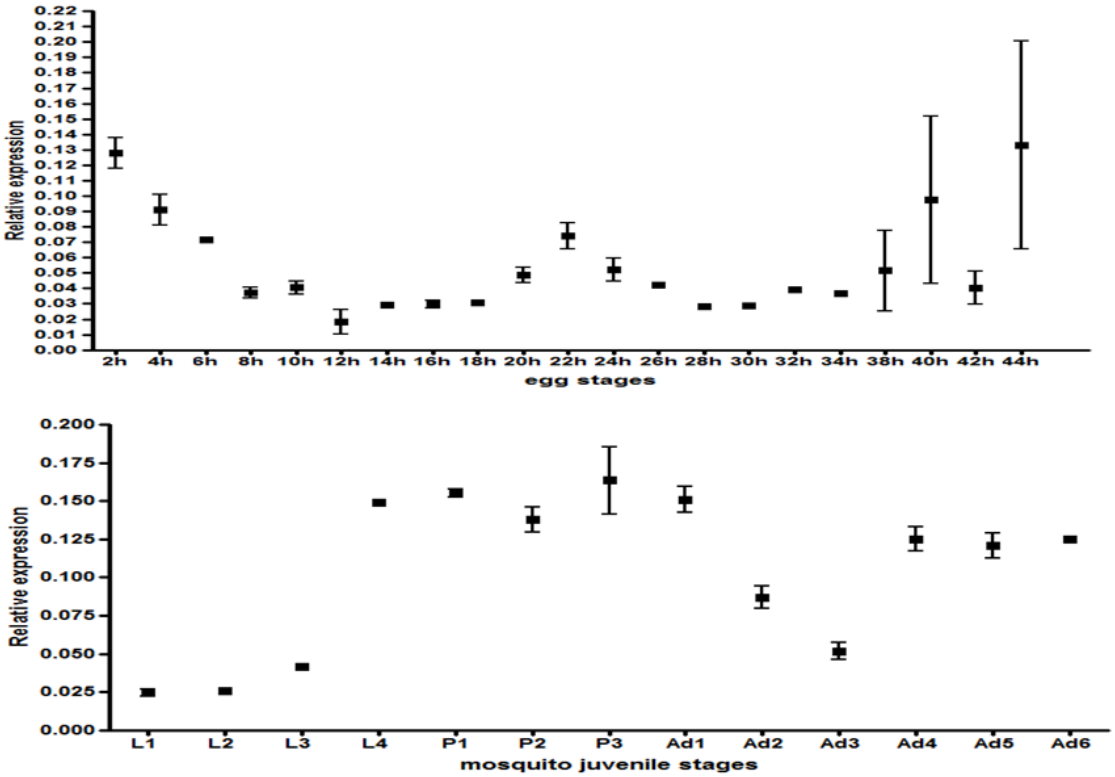
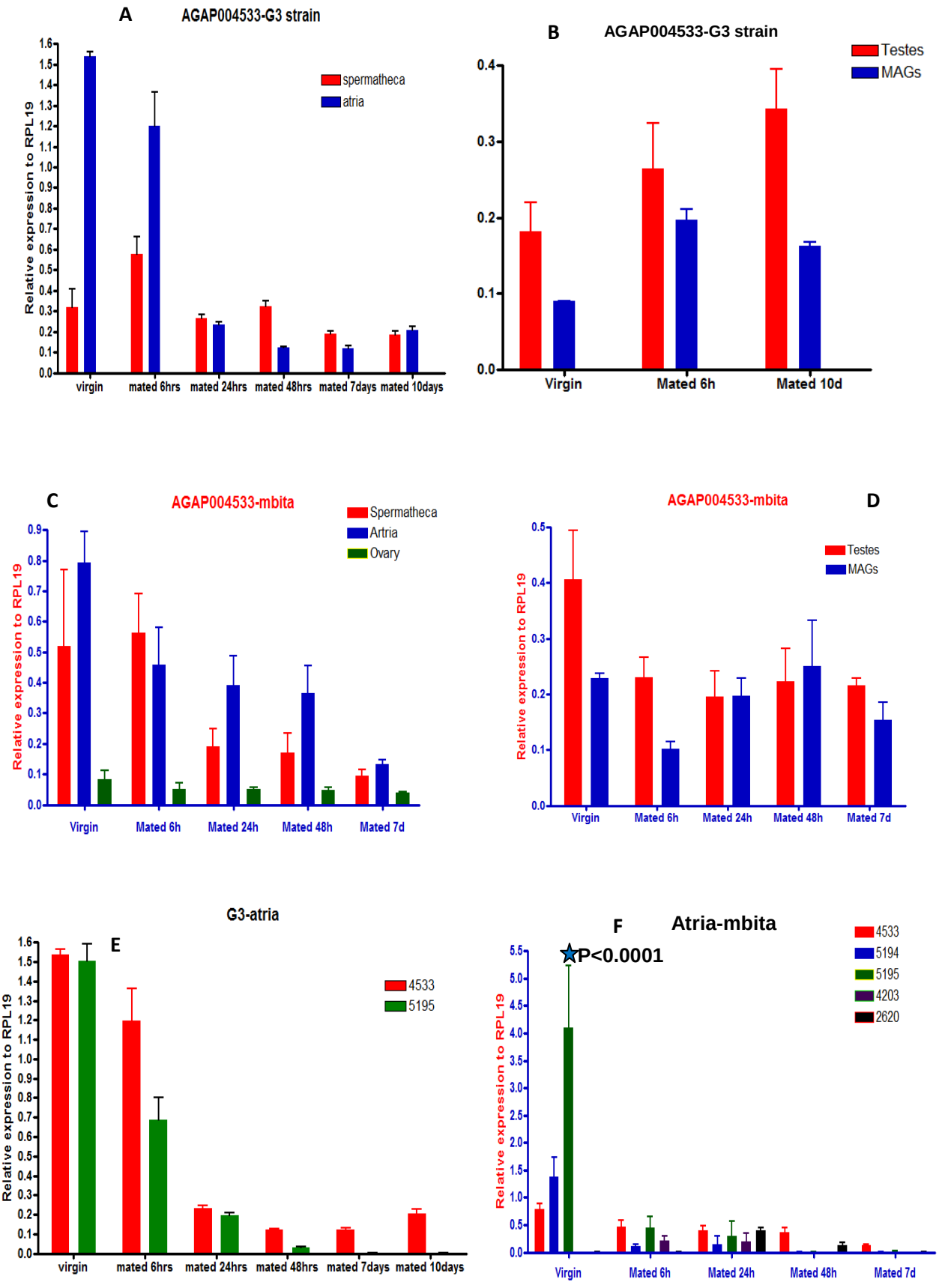


Figure 22: Life stage expressions of CATHEPSIN B (AGAP004533) in *Anopheles gambiae*; This shows the relative expressions of CATHEPSIN B (AGAP004533) at all life stages of the mosquito. Adults and pupae used were of the male. The transcript is expressed at all life stages with the highest seen in P3 (pupa before emergence). The expression of the gene from the eggs to the adult and with various expression levels could imply that it plays varying functions throughout the mosquito's life.


A
AGAP004533-G3 strain
Relative expression to RPL19
spermatheca
atria
virgin
mated 6hrs
mated 24hrs
mated 48hrs
mated 7days
mated 10days
B
AGAP004533-G3 strain
Testes
MAGs
Virgin
Mated 6h
Mated 10d
C
AGAP004533-mbita
Spermatheca
Artria
Ovary
Relative expression to RPL19
Virgin
Mated 6h
Mated 24h
Mated 48h
Mated 7d
D
AGAP004533-mbita
Testes
MAGs
Relative expression to RPL19
Virgin
Mated 6h
Mated 24h
Mated 48h
Mated 7d
E
G3-atria
4533
5195
Relative expression to RPL19
virgin
mated 6hrs
mated 24hrs
mated 48hrs
mated 7days
mated 10days
F
Atria-mbita
P<0.0001
4533
5194
5195
4203
2620
Relative expression to RPL19
Virgin
Mated 6h
Mated 24h
Mated 48h
Mated 7d

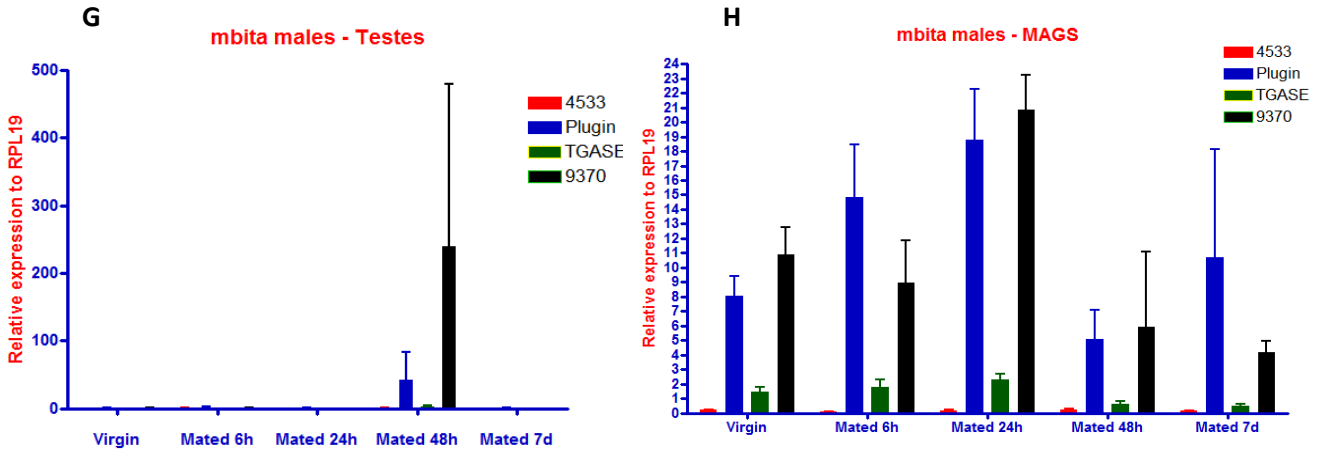
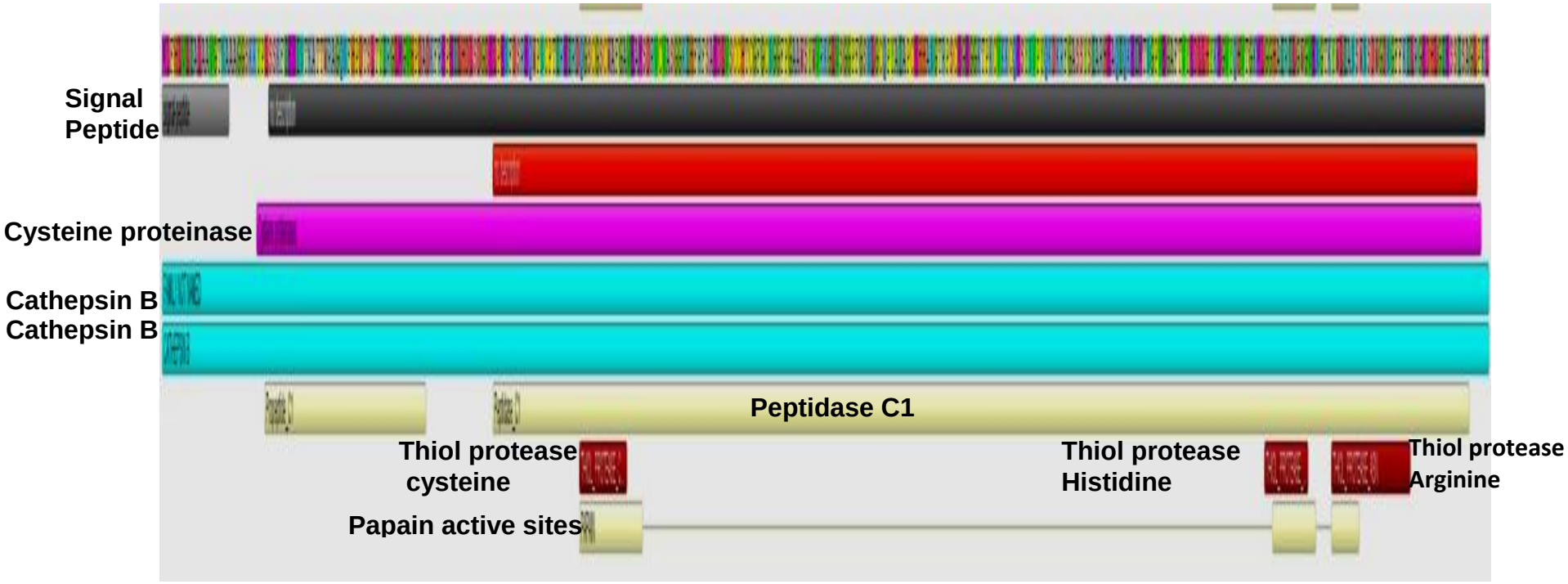


Figure 23: Expression of CATHEPSIN B (AGAP004533) transcript in various reproductive tissues of *Anopheles gambiae*; This figure shows relative expressions of CATHEPSIN B (AGAP004533) (4533) in reproductive tissues. A) In the G3 strain, the transcript is more expressed in the atria (blue) than in the spermatheca (red) and both are down-regulated after mating. B) The transcript is more expressed in the testes (red) than the MAGs (blue). The transcript increases 6hrs after mating and remains up-regulated till 10days. C) In the mbita strain, the transcript is more expressed in the atria (blue) than the spermatheca (red) and the ovary (green). There is a down-regulation of the transcript in the atria and spermatheca after mating but it remains constant in the ovary. D) The transcript is more expressed in the testes (red) than in the MAGs (blue). There is a down-regulation 6hrs after mating and both genes remain almost invariable after then. E) In the G3 strain, the transcript regulation varies like that for TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) (5195, green) with a down-regulation of both after mating. F) In the mbita atrain, the transcript is less expressed in the virgins relative to AGAP005194 (5194, blue) and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) (5195, green) but AGAP004203 (4203, vitellogenin, purple) and AGAP002620 (2620, black) showed no expression at this stage. 4533, 5194 and 5195 were all down-regulated after mating except 4203 and 2620 which were expressed 6hrs and 24hrs respectively after mating then down-regulated later. G) In the mbita males, testicular expressions of 4533 relative to Plugin (AGAP009368, blue), Transglutaminase (TGASE, TRANSGLUTAMINASE (AGAP009099), green) and AGAP009370 (9370) showed no expression in virgin and mated except at 48hrs post-mating where 50 fold and 300 fold expressions are observed for Plugin and 9370. H) The MAGs expression of the same genes seen in the testes of mbita males shows a high expression of Plugin, TGASE and 9370 compared to 4533. Plugin, TGASE and 9370 show increased expressions after mating then decrease from 48hrs after mating.

4.3.2.2 Cloning, sequencing and secondary structure analysis of CATHEPSIN B (AGAP004533) transcript in *Anopheles gambiae*

As described above, CATHEPSIN B (AGAP004533) was seen to be a member of the peptidase C1 family. The analysis performed on this gene with InterPro scan in the Geneious software showed a signal peptide from aa 1-17, a peptidase C1 (aa 85-332) domain (PF08127) with a propeptide_C1 sequence length from aa 27-67. This peptidase is made up of 3 active sites; THIOL_PROTEASE_CYSTEINE (aa 107-118), THIOL_PROTEASE_HISTIDINE (aa 281-291) and THIOL_PROTEASE_ASPARAGINE (aa 298-317) (Figure, 24A). The RT-PCR analysis done on this transcript amplified a 1145bp in all the reproductive tissues. This protein is expressed in all life stages of the mosquito and in all tissues of the adult. CATHEPSIN B (AGAP004533) was cloned and sequenced from larvae and pupae stages and from all reproductive tissues [spermatheca, atria and ovary, (female); testes and MAGs, (male)] of the G3 strain mosquito. Analysis with Geneious 5.5.5 and CLC workbench 6.6.1 softwares (Figure 24B) identified two singleton mutations at C584T in juvenile stages and at nucleotide A14T in the Ovary. The amino acid substitution in the ovary was nonsynonymous and translated to a protein sequence with loss of the signal peptide due to loss of 5 amino acids at the N-terminal region of the protein sequence while in the juvenile stages the amino acid substitution from Alanine to Valine (A195V) was synonymous and so no variation was observed in the sequence structure (Figure 24B). Our sequenced transcripts showed secondary structural variations at possible amino acid substitutional sites identified (Figure 24B) which included a nonsynonymous change of a Proline to a Leucine (P337L) in the atria and spermatheca hence extension of the termination codon. A common amino acid substitution (L6V) was identified in our sequences compared to that of NCBI (gi|347972085).

A


Signal Peptide
Cysteine proteinase
Cathepsin B
Cathepsin B
Peptidase C1
Thiol protease cysteine
Papain active sites
Thiol protease Histidine
Thiol protease Arginine

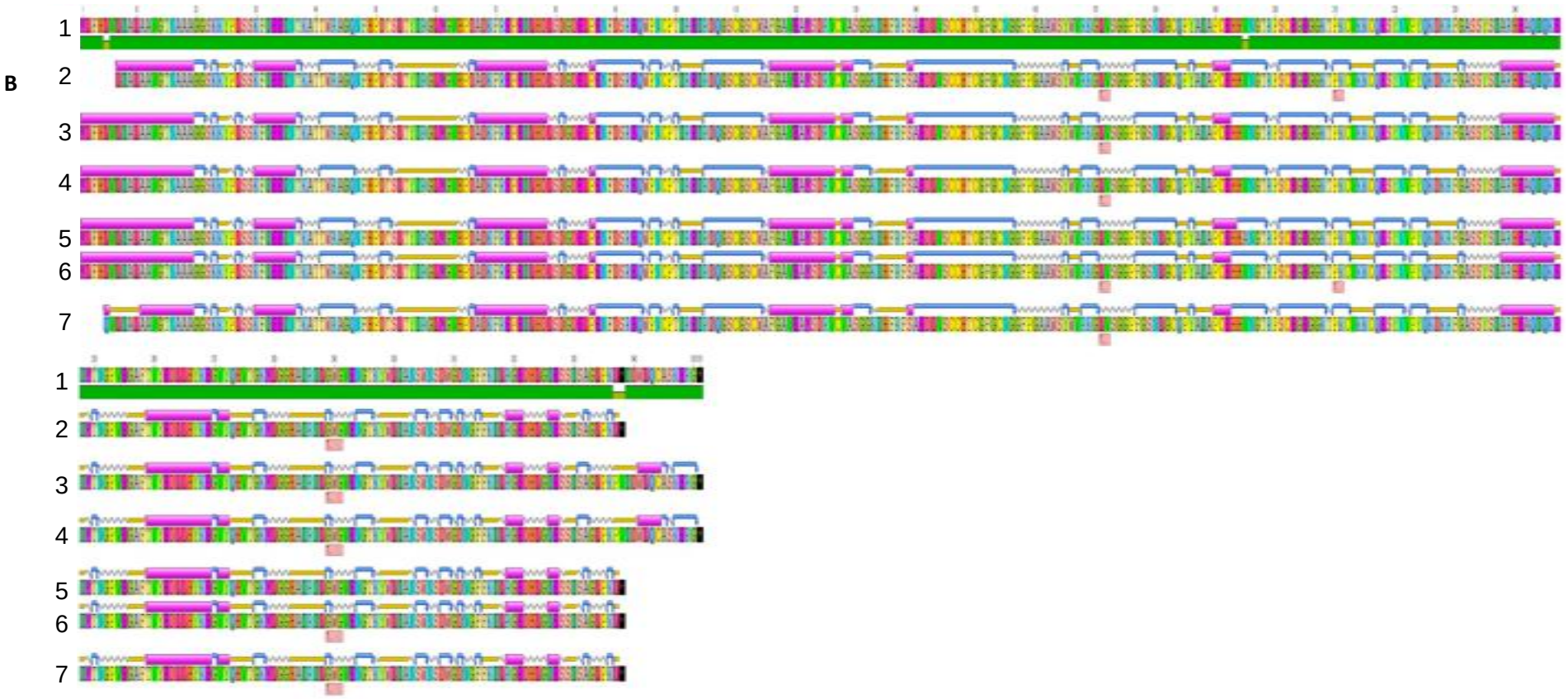


Figure 24: Genetic and secondary structural analysis of CATHEPSIN B (AGAP004533) transcript; these analyses done in Geneious software shows the gene structure of CATHEPSIN B (AGAP004533). A) The domain organization shows a signal peptide (light dark) for amino acid (aa) 1-17, Cysteine Proteinase for aa 25-335 (purple), Cathepsin B for aa 1-337 (blue), Propeptide_C1 for aa 27-67 (light brown) and Peptidase_C1 for aa 85-332 (light brown). The active site motifs were composed of; THIOL_PROTEASE_CYSTEINE (red) for aa 107-118, THIOL_PROTEASE_HISTIDINE (red) for aa 281-291 and THIOL_PROTEASE_ARGININE (red) for aa 298-317. The papain active sites (light brown) composed of aa 107-122, 283-293, 298-304. B) The secondary structures in the G3 strain were compared for testes (TE), atria (AT), spermatheca (SP), Pupae (Pup), MAGs (MA), Ovary (OV) from top to bottom. The testes and the ovary lacked the signal peptide due to mutations upstream of the sequence. The mutation at aa 337 (P337L) causes a loss of a stop codon in the atria and the spermatheca extending the sequence with 14aa. These variations could account for the differential expressions in CATHEPSIN B (AGAP004533) observed. 1 (consensus sequence), 2 (testes sequence), 3 (atria sequence), 4 (spermatheca sequence), 5 (larva and pupa sequence), 6 (MAGs sequence), 7 (ovary sequence)

4.3.3 Analysis based on plugin (AGAP009368) in the MAGS and testes of male *Anopheles gambiae* mosquitoes

4.3.3.1 Expression Analysis of Plugin in the MAGs and testes of the male *Anopheles gambiae*

Work done on Plugin was derived from the fact that it clustered together with the female protease TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) with possible molecular implications. Aligning both plugin sequences in **V**ectorbase and NCBI showed both transcripts were different relative to base pair identity therefore understanding their intervention in reproduction was of interest. qRT-PCR analysis done on the juvenile stages of the G3 strain showed that the transcript begins to be expressed in the pupae and increases in the adult (Figure 25).

RT-PCR analysis on the plug was carried out based on 4 primer pairs; Primer pair 1(FWD: GTCACCAGCAGCATACCCCGC, REV: AGCACAGGCACGAAAACCTGGG) and Size; 1492bp. Primer pair 2 (FWD: GTCACCAGCAGCATACCCCGC, REV: TGCGGGATTTGCTGATCGTGGG) and Size; 274bp, 343bp. Primer pair 3 (FWD: GCTCTGCCTGGTGGCATTTGG, REV: TGCGGGATTTGCTGATCGTGGG) and Size; 847bp, 781bp, 712bp. Primer pair 4 (FWD: GCTCTGCCTGGTGGCATTTGG, REV: AGCACAGGCACGAAAACCTGGG) and Size; none expected.

The idea behind the choice for the 4 pairs of primers was because two sequences for plugin were deposited at NCBI, one as XM559853 and the other as GU188739. Both gene bank accession numbers refer to AGAP009368. Primer pair (PP) 1 belonged to GU188739 and Primer pair 3 belonged to XM559853. The forward of each primer pair was used with the reverse of the other pair. The expected band size sfter PCR amplification with each pair of primers is stated above and it was observed that the reverse sequence of primer 2 was present in both plugin sequences. It was repeated two times in sequence 1 (GU188739) and three times in sequence 2 (XM559853). The forward in each primer were not identified in the other sequence indicating both sequences were different at different points along the sequence. RT-PCR done on cDNA from the MAGs and testes (Figure 26) showed that for PP 1, two different band sizes from the expected were observed in the MAGs (250bp and 1492bp) and testes (600bp and 1500bp), for PP 2 the bands identified were three each (200bp, 274bp and 343bp) for MAGs and testes as opposed to two expected bands, For PP 3, two bands were observed for MAGs (300bp and 343bp) and testes (343bp and 400bp) against 3 expected bands and the sizes were also smaller than expected, For PP 4, no band was observed in the MAGs meanwhile a 2Kb band was observed in the testes against no expected bands. Plugin protein and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) which are both sex-specific showed several bands on a gel with splicing related implications.

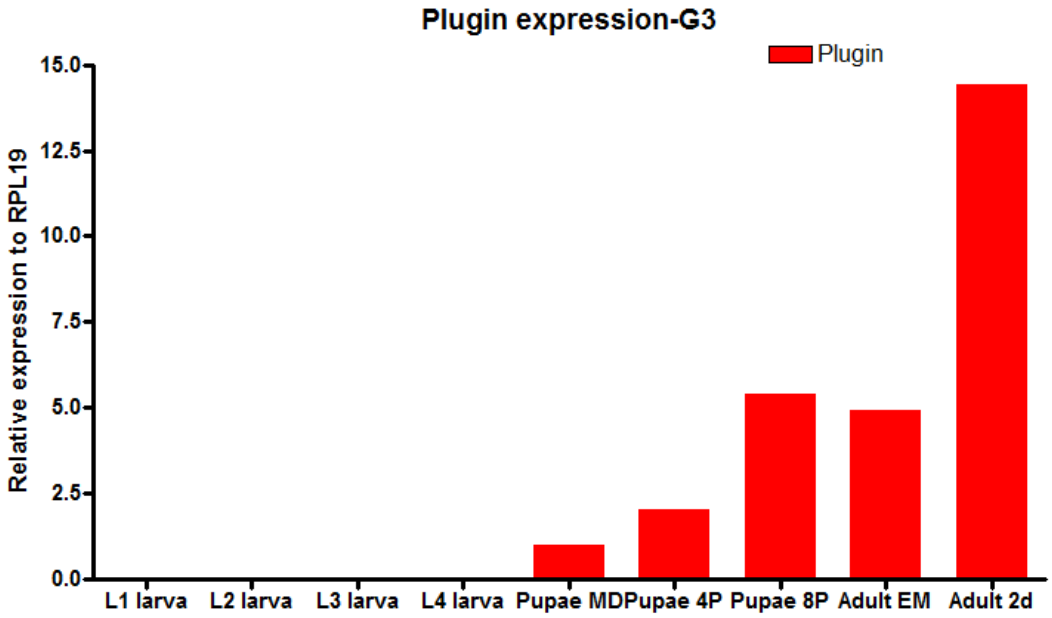


Figure 25: Stage expression analysis of Plugin (AGAP009368) in *Anopheles gambiae*; This shows Plugin expression in the various life stages. It starts being expressed only in the Pupae and continuous in the MAGs for the rest of the male's life time. This could imply it is essential only in the adult and specifically for mating. MD-midday, P-pm, EM-emerged, 2d-2 day.

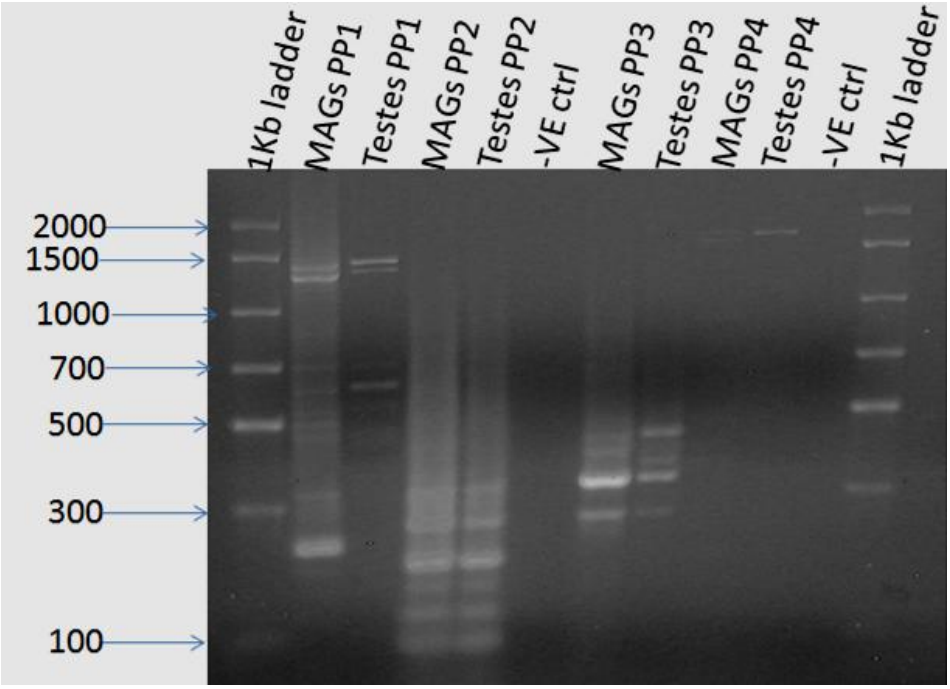


Figure 26: RT-PCR analysis on Plugin (AGAP009368) in *Anopheles gambiae* males; This gel shows Plugin amplification using different primer pairs (PP). Lane 2&3 shows bands amplified using PP 1 and with different sizes in the MAGs and testes. Lane 4&5 shows similar bands of the same size expressed in both MAGs and testes. Lane 7&8 shows bands of different sizes in the MAGs and testes. Lane 9&10 shows no band in the MAGs and a 2Kb band in the testes respectively. Lane 6&11 are the negative controls. Lane 1&12 are the 1Kb DNA ladder used. The variation in sizes from the expected sizes shows could imply different spliced sites within the gene in the various tissues hence resulting in different bands with the same primer.

4.3.3.2 Identity and 3D structural design of Plugin in the MAGs and testes of male *Anopheles gambiae*

Western blot analysis done on the testes and MAGs of three days old virgin males of the G3 strain and using virgin atria of the female mosquito *Anopheles gambiae* as a negative control showed that plugin protein was expressed solely in the MAGs (Figure 27) with a 40KDa molecular weight. The size observed here was different from the standard 80 KDa but it maintained the molecular dimer properties of Plugin though with a smaller band size. It was absent in the other tissues. The bands in the MAGs after incubation at 25ºC and 37ºC showed molecular dimers and trimers of the band at 80KDa and 160KDa.

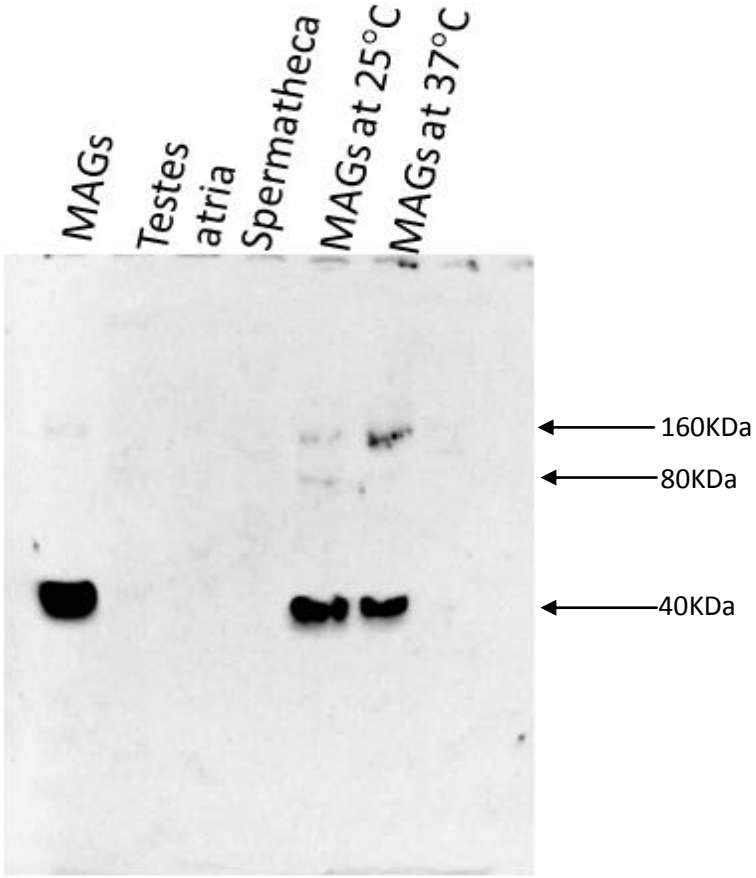


Figure 27; Identity of Plugin in the MAGs of *Anopheles gambiae*; The plugin size of 40KDa seen here showed no band in the testes with the atria and spermatheca as negative controls. The 37°C sample which was incubated for 30 minutes shows the molecular dimmer bands of Plugin at 80KDa and at 160KDa resulting from immunoreactions.

The Plugin protein sequence was used in the ITASSER software to predict its 3D structure. Homology modelling performed on plugin protein using known structures in the NCBI database (Table 2) based on sequence alignments done in CLC work bench showed no homologs in other taxa, hence no 3D structure was obtained. Based on this results, the I-Tasser server was used to conduct protein threading (fold modelling) analysis which considers the template in an alignment as a structure. The best 3D structure for Plugin (Figure 28A) had a C-score of -2.58; a TM-Score of 0.42±0.14 and a RMSD = 13.9±3.9Å. Plugin showed that the protein had no fixed 3D structure characterised by the presence of so many helices. Sites predicted on the protein was confirmed with a BS-score calculated by comparing the local structure and sequence similarity in the binding site region based on its TM-score, identity (IDEN) and coverage (Cov) of the structural alignment (Yang, 2008). The predicted binding site residues for Plugin (Figure 28B) with BS-score of 0.79, TM = 0.2952, IDEN = 0.04 and Cov = 0.47 on the strand 3b8cA (ATPASE 2 Plasma membrane type) were ASP 117, HIS 118, GLN 205, GLU 222, GLU 223, HIS 224, PRO 225, ASP 226, SER 227, GLN 228, PHE 229, PRO 230, LEU 231, PRO 232, ASP 233, ASN 234, MET 235, THR 236, GLU 237, ILE 252, GLN 266, LEU 267, ASN 268, GLN 274, ASN 275, HIS 276, GLN 277, MET 300. The molecular functions for each protein predicted by I-TASSER uses a GO-Score which is the average weight of the GO term based on the template on which the GO term was derived. Plugin showed the best GO-Score of 0.207 (GO: 0005488) which are ligands that bind to signal transducing receptors and alongside the PredictProtein server, it also predicted GO: 0005509 which is involved in Calcium ion binding in vitellogenin at a GO-Score = 0.085.

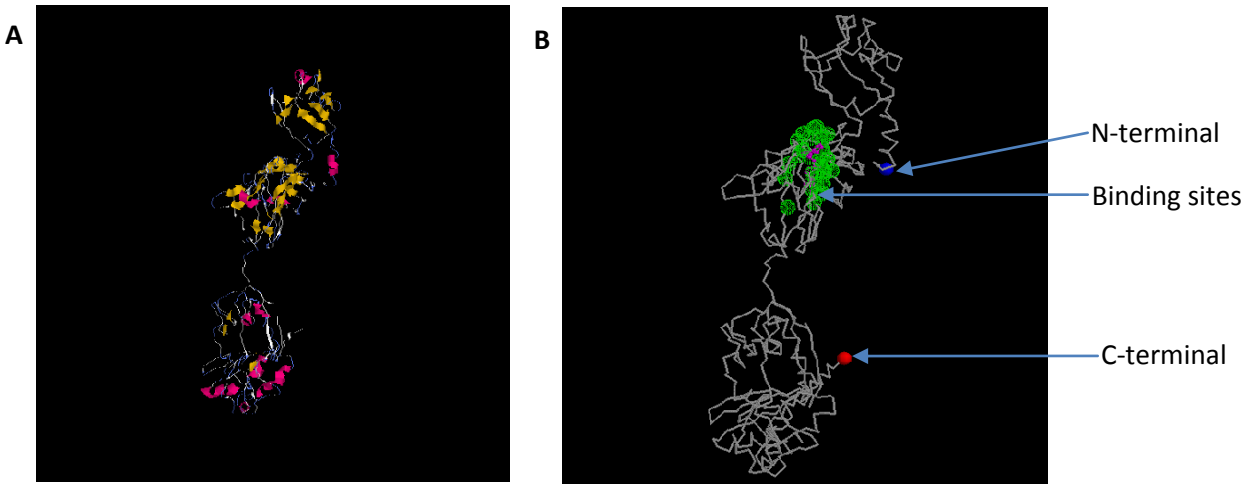


Figure 28; 3D structural predictions of Plugin protein in the male *Anopheles gambiae*; A) Plugin structure with the best C-score of -2.58 and a TM-Score of 0.42 ± 0.14 and a RMSD = 13.9 ± 3.9Å. It is constituted of more helices and loops. B) Plugin shows the ligand-protein complex with the best BS-score, 0.79 for Plugin. The ligand is depicted in magenta ball and stick. The predicted binding sites interacting with the ligand are shown as transparent green spheres while the N & C terminus in the model are marked by blue and red spheres respectively. The ligand(s) in the analog structures were then transferred onto the model and the fitness of the ligand-model complex (BS-score) was calculated by comparing the local structure and sequence similarity in the binding site region.

4.3.4 Predicted molecular properties of some male and female *Anopheles gambiae* mating responsive proteins based on protein predict software.

The Protein Predict software was used to analyze 7 proteins present on the plug [TRANSGLUTAMINASE (AGAP009099), AGAP009368, AGAP008276, AGAP008277, CATHEPSIN B (AGAP004533), CHYMOTRYPSIN-LIKE SERINE PROTEASE (AGAP005194) and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195)] after mating and 2 other proteins, AGAP004203 (Vitellogenin) and AGAP002620, that are up-regulated 6hours and 24hours postmating, respectively (Figure 23F). The various parameters analyzed by the software showed some variation relative to proteins present in the male only, those present in the female only and those present in both male and female reproductive tissues. All of these proteins were predicted to be secreted as expected except AGAP004203 (vitellogenin) which was seen with a Zinc finger domain which binds to DNA in the regulation of transcription within cells. The Metallo B Lactamase domain (PF00753) seen on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) involves it in odorant degrading enzyme (ODE) enzymatic properties which is known in *D. melanogaster* to terminate olfactory receptor neuron (ORN) responses. This domain in TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) could be linked to neural and odorant binding functions as seen in Table 5 below.

Table 5: Molecular properties of some *Anopheles gambiae* mating responsive proteins identified in the male and female reproductive tissues.

	Props	009099 (Tgase)	009368 (Plugin)	008276	008277	004533	005194	005195	005195A	005195B	004203	002620
Sec-str	Helix	8.02%	55.12%	13.97%	6.95%	20.47%	11.76%	12.32%	12.32%	11.33%	23.35%	11.18%
	Strands	36.23%	0	28.68%	41.70%	18.10%	33.82%	31.03%	31.03%	32.51%	21.89%	19.08%
	Loops	55.75%	44.88%	57.35%	51.35%	61.42%	54.41%	56.65%	56.65%	56.16%	54.75%	69.74%
Sur-acc	Exposed	50.53%	89.59%	52.57%	50.97%	54.60%	52.21%	56.65%	56.65%	56.16%	49.88%	73.03%
	Buried	49.47%	10.41%	47.43%	49.03%	45.40%	47.79%	43.35%	43.35%	43.84%	50.12%	26.97%
TMH		Absent	Absent	Present	absent	present	present	present	present	present	absent	absent
LC-region		Absent	present	absent	absent	present	absent	absent	absent	absent	present	present
Localisatn	Animals	Secretory extracell	Secretory extracell	Secretory extracell	Secretory extracell	Secretory extracell	Secretory extracell	Secretory extracell	Secretory extracell	Secretory extracell	Nuclear	Secretory extracell
S-S bond		Absent	absent	64 & 235	absent	absent	59 & 245	absent	68 & 199 175& 189	52 & 68, 175 &189	absent	absent
TM-BB		Absent	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
P-P int		Present	absent	present	present	present	present	present	present	present	present	present
P-DNA int		Absent	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
PROSITE	ASN_Gly PS00001	Present	Present	absent	present	present	Present	Present	Present	present	present	absent
	PKC_phos PS00005	Present	Present	absent	present	present	present	present	Present	present	present	present
	CK2_phos PS00006	Present	present	present	present	present	present	Present	Present	Present	present	absent
	TYR_phos PS00007	Present	Absent	present	absent	absent	Absent	Present	Present	Present	present	absent
	MYRISTYL PS00008	Present	Absent	present	present	present	present	present	present	Present	Present	present
Other PROSITEs		AMIDATION PS00009		TRY_SER PS00135	AMIDATION PS00009 TRY_SER PS00135	THIOL_CYS PS00139 THIOL_HIS PS00639 THIOL_ASN PS00640	TRY_SER PS00135	TRY_HIS PS00134 TRY_SER PS00135	TRY_HIS PS00134	TRY_HIS PS00134	CAMP_Phos PS00004 RGD PS00016	
Pfam domains		Fer4 PF00037 Metallophos PF00149 Pyr_redox PF00070	GGDEF PF00990 His_Kinase A PF00512 His_Kinase A PF00512	Trypsin PF00089 Trypsin PF00089	NADH dehydro PF00146 Trypsin PF00089 Trypsin PF00089	RRM, RBD, RNP PF00076 FAD Oxidoreduc PF01266	Trypsin PF00089 Trypsin PF00089 Trypsin PF00089	Metallo B lactamase PF00753 Response regulator PF00072 Trypsin PF00089	Metallo B lactamase PF00753 Response regulator PF00072 Trypsin PF00089	Metallo B lactamase PF00753 Response regulator PF00072 Trypsin PF00089	COX 2 PF00116 GTP_EFTU _D3 PF03143 Zf-C2H2 PF00096	

	Trypsin PF00089	Trypsin PF00089	Trypsin PF00089	TRP_2 PF07719

Legend: AGAP numbers; for example 009368 = AGAP009368, Tgase (transglutaminase), Sec-str (secondary structure), Sur-acc (surface accessibility), TMH (transmembrane helix), LC (low complexity), S-S (disulphide bond), TM-BB (transmembrane beta barrel), P-P int (protein-protein interaction), P-DNA int (protein-DNA interaction), Props (properties), extracell (extracellular), ASN_GLY (N-glycosylation site), PKC_Phospho (protein kinase c phosphorylation site), CK2_Phospho (Casein kinase II phosphorylation site), TYR_Phospho (tyrosine kinase phosphorylation site), TYR_SER (Trypsin-Serine active site), TRY_HIS (Trypsin-Histidine active site), THIOL_CYS (Thiol-Cysteine active site), THIOL_HIS (Thiol-Histidine active site), THIOL_ASN (Thiol-Asparaginne active site), CAMP_Phos (cAMP dependent protein kinase phosphorylation site),Fer4 (4Fe-4S), Pyr_redox (Pyridine nucleotide-disulphide oxidoreductase), HIS_Kinase (Histidine Kinase A), dehydro (dehydrogenase), RRM (RNA recognition motif), RBD (RNA binding domain), RNP (ribonucleoproteins), Oxidoreduc (oxidoreduction), COX2 (cytochrome c oxidase subunit II), GTP_EFTU_D3 (GTP-binding elongation factor), zf-C2H2 (zinc finger), TRP_2 (tetratricopeptide repeat).

4.4 Influence of the interactions among the interacting proteins on mating responses in *Anopheles gambiae*

4.4.1 Functional analysis through knockdown of CATHEPSIN B (AGAP004533) in *Anopheles gambiae* male mosquitoes

4.4.1.1 Expression analysis of CATHEPSIN B (AGAP004533) after knockdown in the male *Anopheles gambiae*

The RNA binding domain function of CATHEPSIN B (AGAP004533) was evaluated through knockdown of the transcript in the MAGs and testes in three days old adults injected with 6µg/µL of dsRNA as experimentals. The same concentration of dsLacZ was injected against the control gene LacZ. A total volume of 69nL was injected into each virgin male of the mbita strain and then the testes and MAGs were dissected from day one to day 4 post-injection. qRTPCR analysis showed a down regulation of CATHEPSIN B (AGAP004533) in all tested days in both tissues with statistical significance using two way ANOVA in the testes [P-value Testes day 1 (Pd1) < 0.01, Pd2 < 0.05, Pd3 < 0.001, Pd4 < 0.05] and only significant on day two (P-value MAGs < 0.01) for the MAGs (Figure 29 A&B). The knockdown effect was used to evaluate the relative expression of other genes (Plugin, Transglutaminase (AGAP009099) and AGAP009370) expressed in the virgin MAGs and identified on the plug. The knockdown showed a slight increase in plugin expression in the testes compared to the LacZ controls (Figure 29 C&D), however these values were not statistically significant. TRANSGLUTAMINASE (AGAP009099) (Transglutaminase) relative expression showed higher expression values in the testes for the experimental samples compared to the LacZ controls (P-value Transglutaminase day 4 < 0.05) (Figure 29 E&F) while in the MAGs the LacZ controls were slightly more expressed than in the experimental. AGAP009370 relative expression showed higher expression values in the experimental samples of the testes than in the LacZ controls with day 2 being statistically significant (P < 0.01) meanwhile in the MAGs the LacZ controls the transcript was more expressed than in the experimental group though not statistically significant (Figure 29 G&H).

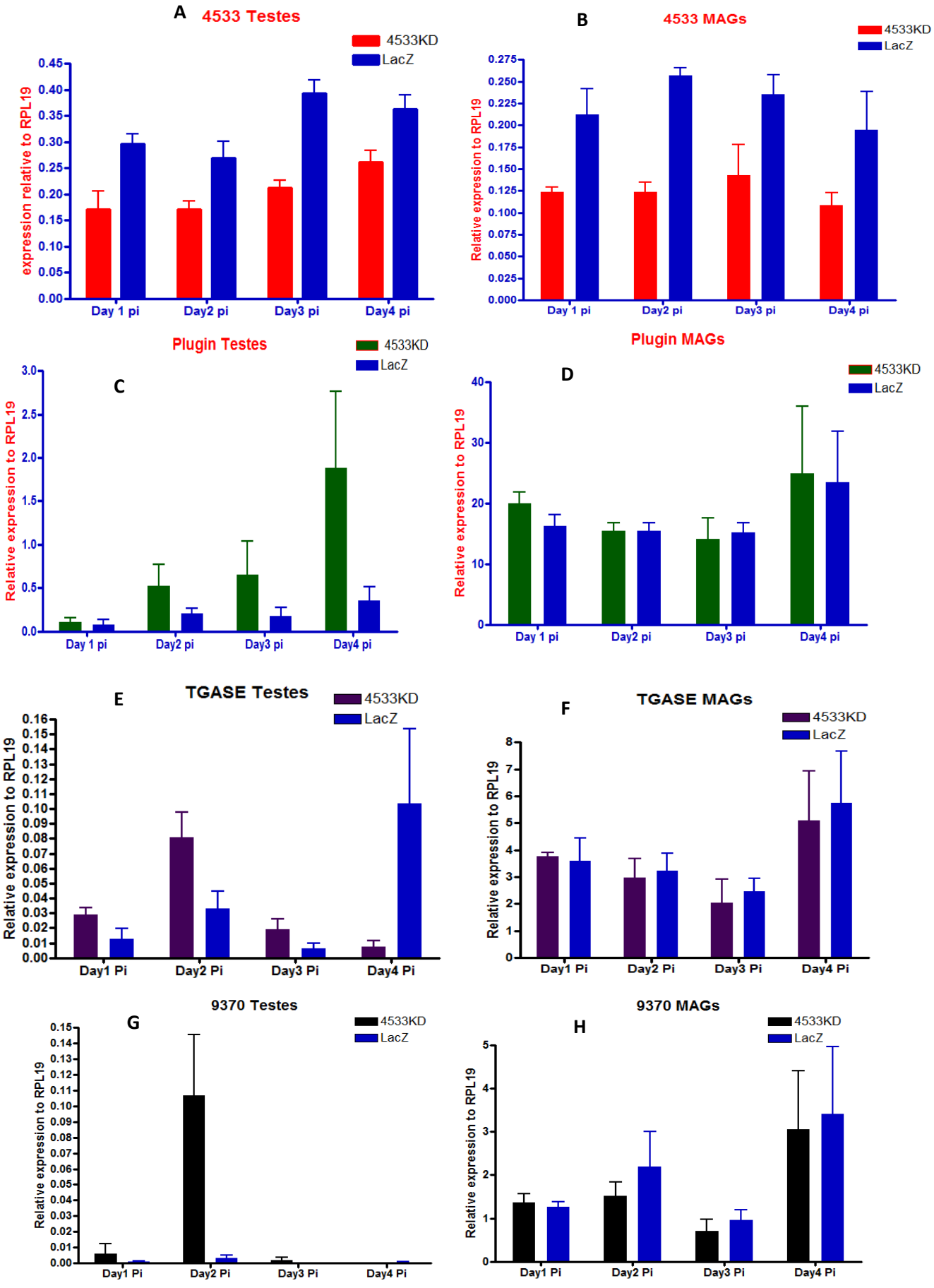

A
4533 Testes
4533KD
LacZ
expression relative to RPL19
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Day 1 pi
Day2 pi
Day3 pi
Day4 pi
B
4533 MAGs
4533KD
LacZ
Relative expression to RPL19
0.275
0.250
0.225
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
Day 1 pi
Day2 pi
Day3 pi
Day4 pi
Plugin Testes
4533KD
LacZ
C
Relative expression to RPL19
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Day 1 pi
Day2 pi
Day3 pi
Day4 pi
Plugin MAGs
D
4533KD
LacZ
Relative expression to RPL19
40
30
20
10
0
Day 1 pi
Day2 pi
Day3 pi
Day4 pi
E
TGASE Testes
4533KD
LacZ
Relative expression to RPL19
0.16
0.15
0.14
0.13
0.12
0.11
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
Day1 Pi
Day2 Pi
Day3 Pi
Day4 Pi
F
TGASE MAGs
4533KD
LacZ
Relative expression to RPL19
8
7
6
5
4
3
2
1
0
Day1 Pi
Day2 Pi
Day3 Pi
Day4 Pi
9370 Testes
G
4533KD
LacZ
Relative expression to RPL19
0.15
0.14
0.13
0.12
0.11
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
Day1 Pi
Day2 Pi
Day3 Pi
Day4 Pi
9370 MAGs
H
4533KD
LacZ
Relative expression to RPL19
5
4
3
2
1
0
Day1 Pi
Day2 Pi
Day3 Pi
Day4 Pi

Figure 29: Differential expression of CATHEPSIN B (AGAP004533) in *Anopheles gambiae* G3 males through RNAi; This shows relative expression of CATHEPSIN B (AGAP004533) to plugin (CATHEPSIN B (AGAP004533)), Transglutaminase (TGASE, TRANSGLUTAMINASE (AGAP009099)) and AGAP009370 in the testes and MAGs. 6μg/μL of double stranded RNA for CATHEPSIN B (AGAP004533) (experimental) and double stranded Lac Z for the control was injected in each 69μL per mosquito. A) Knockdown (KD) was obtained in the experimental relative to the control for all 4 days analyzed post injection (Pi). B) Knockdown was obtained in the experimental relative to the control for all 4 days analyzed Pi C) Plugin was more expressed in the testes for the experimental relative to the control and this was seen to increase with the days D) Plugin was more expressed in the experimental relative to the control E) TGASE was more expressed in the testes experimental for day 1-3 Pi and less on day 4 compared to the control F) TGASE was less expressed in the MAGs experimental relative to the control G) 9370 was highly expressed on day 2 Pi of the experimental compared to the control with statistical significance of $P<0.01$. H) 9370 was less expressed in the MAGs of the experimental relative to the control. 4533 (AGAP004533)

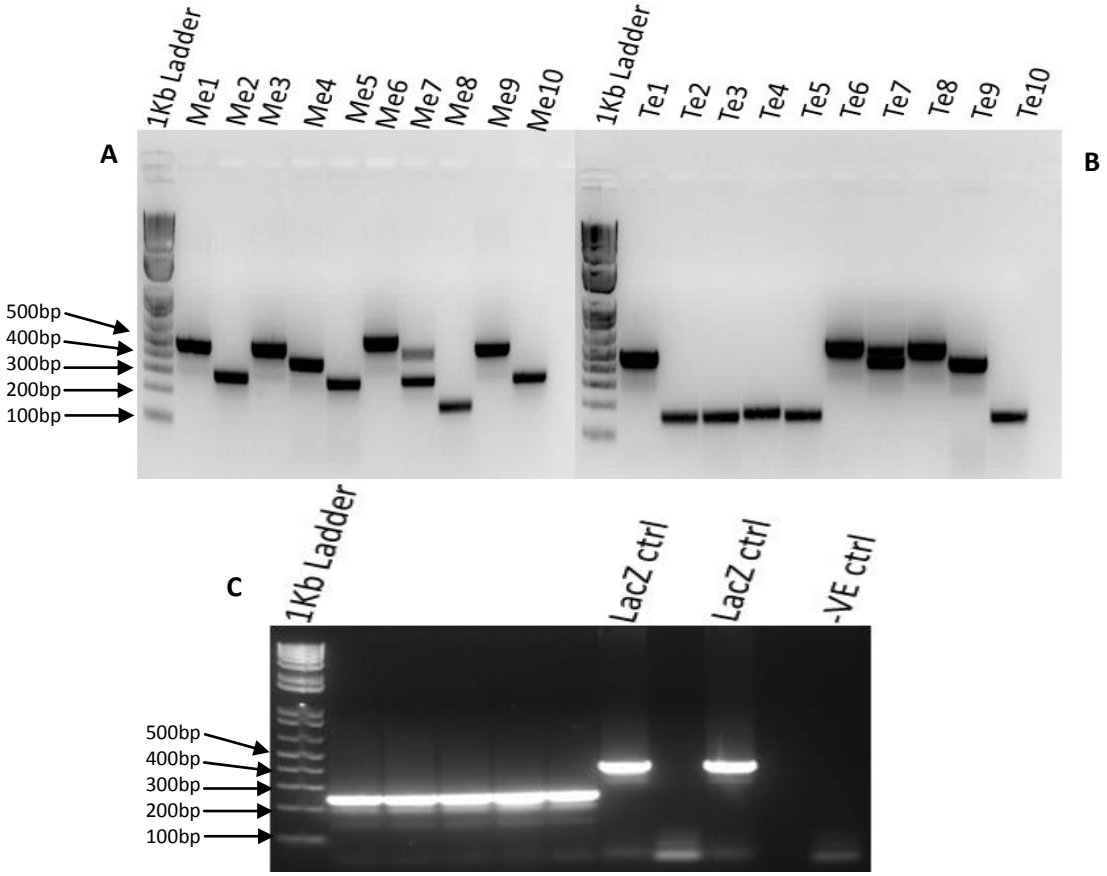
4.4.1.2 Cloning, sequencing and analysis of MAGs and testes transcripts in the male *Anopheles gambiae* after CATHEPSIN B (AGAP004533) knockdown.

The amplicons of Plugin using the various sets of primers (section 4.2.5) was cloned and sequenced. Several variants of plugin transcript were identified in the knockdown experimental sample as seen by various bands on the gel (Figure 30 A&B) meanwhile only one band was identified for the LacZ samples (Figure 30C). These variants were identified as spliced with different lengths and splicing junctions of 5'GT and 3'AG could be seen not excised in other variants giving those different lengths (Figure 30D). Indicating the down-regulation of CATHEPSIN B (AGAP004533) affects the spliceosome complex causing misplicing of the transcript leading to several variants. Seven and 10 variants were identified in the testes and the MAGs respectively (Figure 30 A&B) after sequencing 8 and 5 variants were obtained for the MAGs and testes, respectively (Figure 30D). The sequenced variants were used to do BLAST searches against the nucleotide database to confirm that the sequences matched the known Plugin sequence (GU188739). Based on similarity in variant length and splice junctions, they could be grouped as follows; 4, 5&7, 6, 9&10, 11, 12, 13, 14 and 15 for both the testes and MAGs (Figure 30D). Predicted transcription factors binding sites (TFs) were present in some variants meanwhile the TFs on the LacZ control sequences were the same for the bands amplified, indicating similarity between them (Table 6). The experimental variants had T00626 present at the start of the sequences with alternative splicing as coding sequence diversity. Meanwhile, LacZ controls had T00691present at the start of the sequences with alternative splicing as coding sequence diversity. The testes and MAGs variants of experimental sequences amidst misplicing also showed the insertion of a signal peptide and transmembrane regions (TMHMM) in the sequences while these sites were all absent in theLacZ controls (Table 6). The translation of the spliced variants on different frames produced sequences with signal peptides and transmembrane regions which were not identified in the LacZ controls with further implications in wrong splicing.

Table 6: Transcription binding factor (TF) predictions on Plugin spliced variants

TF binding site	TF name	Function	Coding sequence diversity	Kd MAGs	Kd Testes
GGGAGGG	T00490	Helicase	none	7	4
GAGGAGGG	T00105	Epimerase	none	4	5
GCTGCCATG	T00626	Transferase	Alternative splicing	6	5
TF binding site	**TF name**	**Function**	**Coding sequence diversity**	**LacZ MAGs**	**LacZ Testes**
TGTCCCGC	T00691	Transgenesis related protein	Alternative splicing	1	1
TTCAAG	T00999	Signal transduction	Methyl-accepting chemotaxis protein	1	1
GATGCAT	T01542	Ion channel	none	1	1

Legend: This shows transcription factor usage for the amplification of plugin gene after knockdown of CATHEPSIN B (AGAP004533). This amplification was based on primer pair 2 and from the table it can be observed that the transcription binding sites varied from one sequence to another for the experimental but the sites were uniform in the lacZ control. The functions of these TFs could be seen on the table



D

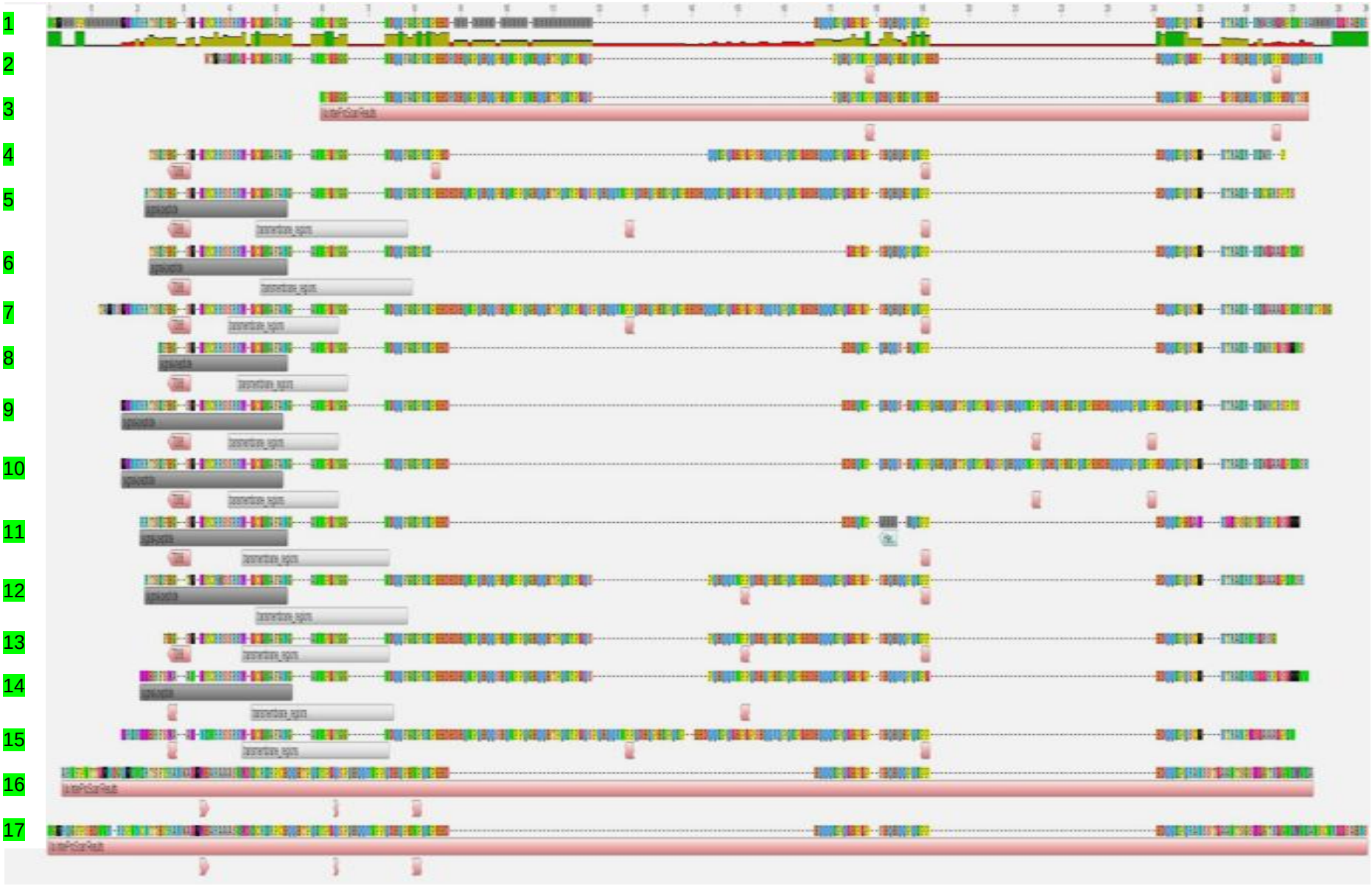


Figure 30: Splicing effects identified on Plugin after gene knockdown, RT-PCR and sequencing of CATHEPSIN B (AGAP004533) in the male *Anopheles gambiae*; Different variants of plugin amplified using the different primers. A,B) Colony PCR on day 3 Pi samples using primer pair 2 showed several bands (100-200bp) for Plugin in the MAGs (A) and testes (B) experimental. C) Colony PCR on day 3 Pi samples using primer pair 2 showed one variant for LacZ controls. D) Sequence analysis of Plugin variants amplified with primer pair 3, the first two sequences (2; experimental and 3; lacZ) showed the experimental longer than the lacZ sequence. The last 2 sequences (16&17; lacZ) amplified with primer pair 2 showed a sequence of the same length. The sequences in between testes (7, 9, 10, 14, 15) and MAGs (4, 5, 6, 8, 11, 12, 13) experimental samples amplified with primer pair 3 showed varying sequence lengths due to 5'GT and 3'AG junctions spliced in some variants and not in others. The presence of transcription factor (T00626) on these sequences responsible for alternative splicing (Table 5) may be responsible for the start codons and signal peptides (dark grey bands) created on these sequences and the epimerase (T00105) could be responsible for the transmembrane regions (light grey bands) seen on the sequences. These properties are not seen on the TTL sequences which possess a T00691 as an alternative splicing site which could account for proper splicing of the transcript in the testes of the controls. 1 (consensus sequence).

4.4.1.3 Identity of Plugin in MAGs of male *An. gambiae* mosquito

MAGs from both experimental samples and LacZ control in 1-3days post-injected mosquitoes extracted and analysed by Western blotting (Figure 31) showed that both groups had the 80KDa band size relative to the 1891bp nucleotide sequence deposited at NCBI. The protein was not observed in the testes as seen in section 4.3.5.2 was and was therefore not included in this analysis. The immunoreactive dimmer (160KDa) of the protein was higher in 3 days post injected experimental males than the LacZ controls. The actin control band was not identified on the gel after staining but the same number of MAGs tissues were crushed and the same volume loaded on the gel for consistency.

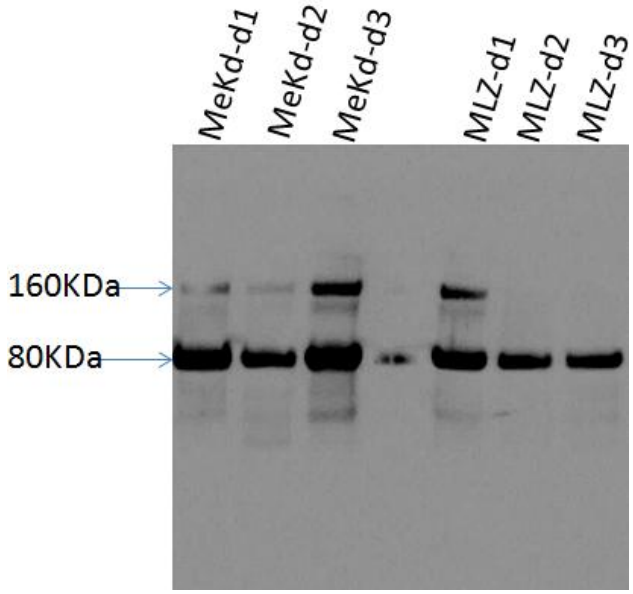


Figure 31: Identity of Plugin in the MAGs after CATHEPSIN B (AGAP004533) knockdown in the male *Anopheles gambiae*; This shows that the plugin size of 80kDa is maintained on the gel for both the experimental and the lacZ controls. The 160KDa band shows the molecular dimer of the protein in the cell due immunureaction of the protein. Despite misplicing, the band size in the experimental could only mean gene compensation by the cell to maintain the protein levels. Me (MAGs experimental), Kd (Knockdown), LZ (LacZ), d1, d2, d3 (days 1-3)

4.4.2 Functional analysis through knockdown of CATHEPSIN B (AGAP004533) in the female *Anopheles gambiae* mosquitoes

4.4.2.1 Expression analysis and mating propensity (mating drive) phenotype analysis after knockdown of CATHEPSIN B (AGAP004533) in the female *Anopheles gambiae* mosquitoes

The transcript levels of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) which was observed to be highly expressed in the atria where the plug is formed was seen to be down regulated up to 7 days post-mating as shown in section 4.2.1.1 above. This was the only female transcript identified on the network and it was the only transcript down regulated together with CATHEPSIN B (AGAP004533) which was our gene of interest for functional analyses. The transcript expression levels of CATHEPSIN B (AGAP004533) and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) were tested to find out if it was involved in an interaction that modulates the female mating propensity (mating drive) towards the male. In three independent trials (Figure 32A, A), virgin non injected females showed that 73.06% against 26.94% males were present in the male cage indicating that males will remain in the same cage for swarming which is equally an advantage and a strategy for them to mate with females.In the male cage; 84.3% against 15.7% females were present and were mated with successful insemination due to presence of sperm in the spermatheca indicating an attraction from males to the swarm. These values were statistically significant using 2 way ANOVA with the males having P-value<0.05 and females having P-value <0.05. Knockdown of CATHEPSIN B (AGAP004533) and its effect on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) was investigated for its co-effect on mating.

One day old virgin female *An. gambiae* of the Mbita strain were injectected with 5μg/μL of double stranded RNA against CATHEPSIN B (AGAP004533) transcript for the experimental and dsLacZ for the lacZ gene as controls. Three to five females were dissected two to three days post injection to obtain atria and spermatheca. Other females were mated 3 days post injection with same age males and couples were dissected for the atria and spermatheca 6hours and 24hours post mating. qRT-PCR analysis was done on cDNA from the atria and spermatheca to determine gene expression levels and evaluate knock down (Figure 32A, C&D). The spermathecal samples showed no significant regulation in the transcripts and were therefore not considered for these analyses. In the atrial samples, there was knockdown of CATHEPSIN B (AGAP004533) in the experimental group as observed on days two and three post injection and 6hrs post mating. The knockdown was statistically significant (6hrs post mating P<0.05) compared to the lacZ controls. TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) was down-regulated in the experimental samples

relative to the LacZ controls at two and three days post injection samples and at 6hrs post mating (Figure 32B, A & B). CHYMOTRYPSIN-LIKE SERINE PROTEASE (AGAP005194) also showed the same regulation pattern as TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) at the same time points. AGAP004203 (Vitellogenin) was more expressed in the experimental virgin samples than the LacZ controls though not statistically significant. A similar up-regulation pattern at 6hrs and 24hours for AGAP004203 and AGAP002620, respectively was observed both for the experimental samples and in the LacZ controls after mating. AGAP002620 was more expressed in the experimental than the LacZ controls 24hours post mating though not statistically significant. Signal transduction effects implicating TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) (Table 5) could be further confirmed by the loss of the two active variants in the virgins of both tested groups using RT-PCR (Figure 32B, A). The molecular RNA binding role of CATHEPSIN B (AGAP004533) identified in males through splicing in the plugin transcript was not observed in in TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) transcript (Figure 32B, A).

The experimental females and the LacZ controls were used in the mating drive experiments (Figure 32A, B). It is worth noting that some females died before dissections for qRT-PCR analysis and some died before dissections to check for insemination. Comparing the mean values for the mating events in both groups, 60.43% of males interacting with experimental females were in the male cage meanwhile 61.43% males interacting with LacZ females were in the male cage. The differences for each interacting group was not statistically significant but indicated that males remained in the male cage for swarming as seen earlier. For the experimental females; 69.05% in the male cage were non-inseminated compared to 30.95% inseminated while 55.17% of LacZ females in the male cage were non-inseminated compared to 44.83% inseminated, though the difference within each female group was not statistically significant but indicated that the mating drive was present and higher in the experimental females though insemination was higher for the LacZ controls. In the female cage, the experimental females showed that, 63.63% were non-inseminated compared to 36.36% inseminated while for the LacZ females, 68.42% were non-inseminated compared to 31.58% inseminated and though the difference within each group was not statistically significant, more LacZ control females stayed in the female cage and less were inseminated compared to the experimentals. Overall, insemination levels for both groups were low. The knockdown of CATHEPSIN B (AGAP004533) and its effect on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) shows that more experimental females are attracted to the swarm than LacZ females and few are successfully inseminated.

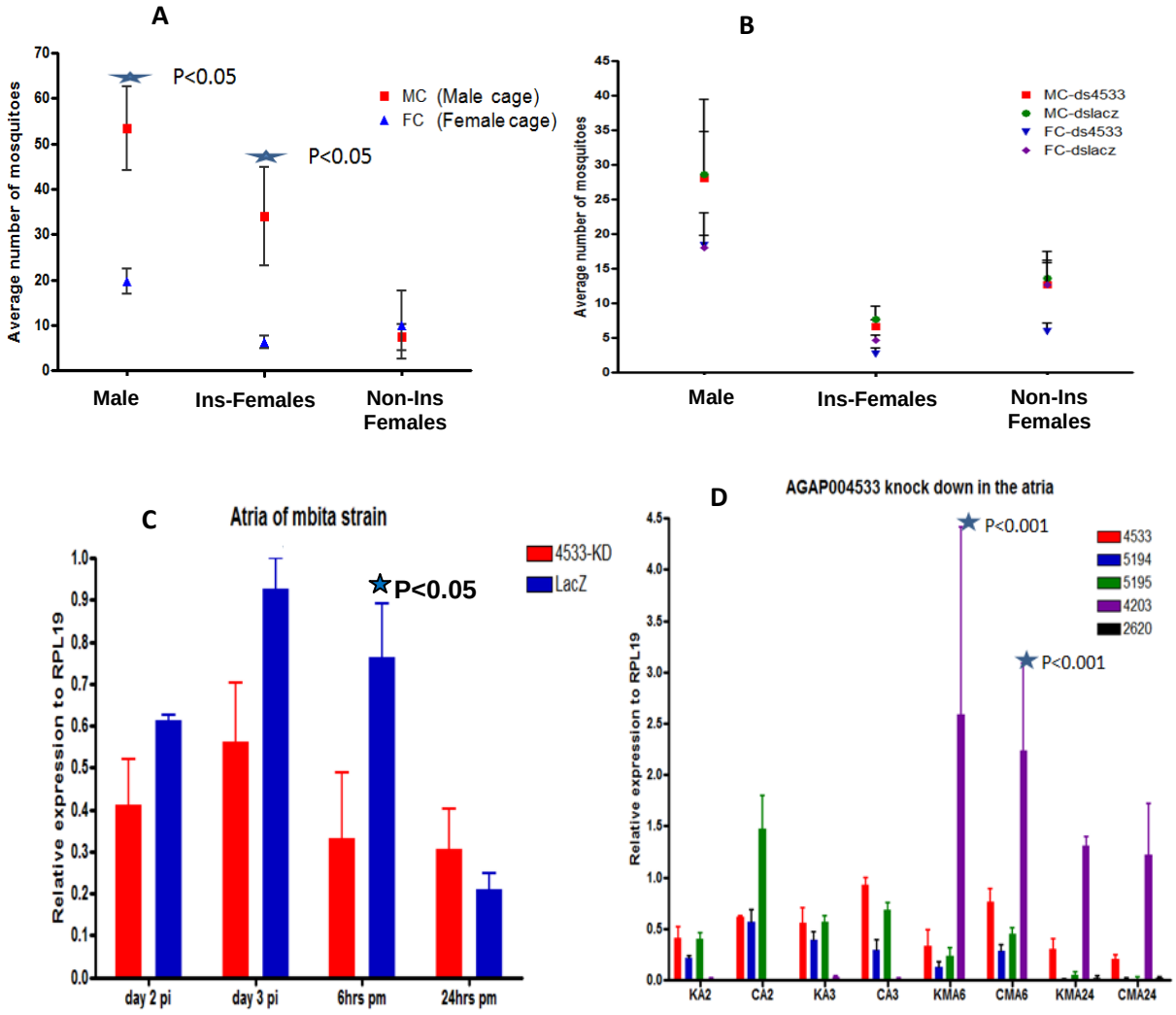


Figure 32A: Mating drive and gene expression analysis after CATHEPSIN B (AGAP004533) knockdown in the female *Anopheles gambiae*; This shows possible effects on mating response by the female *An. gambiae* relative to 4533 knockdown. A) It shows female response to mating as 73% of males remain in their cage to swarm and 84% of the females move into male cages and are mated with successful insemination. These were statistically significant with P<0.05. B) This shows mating response by females after knockdown (C). The response between experimental and control are very close. The difference was not statistically significant between compared parameters but it still confirms the fact that males will swarm together and the females are still attracted to the swarm but very few do mate with successful insemination. C) This shows knockdown in the experimental against the controls as seen on days 2&3 Pi and 6hrs Pm. The result was statistically significant at 6hrs pm (P<0.05). D) This shows other transcript responses relative to knockdown of 4533 (AGAP004533). 5194 (AGAP005194) and 5195 (AGAP005195) is more down regulated in the experimental group for days 2&3 Pi than the control but lost in both groups after mating. 4203 (AGAP004203) is still not expressed in the non mated group, meaning it is only activated by mating. It is more up-regulated in the experimental than control group at 6hrs and 24hrs pm. 2620 (AGAP002620) is only activated by mating and more up-regulated in the experimental than the controls 24hrs pm.

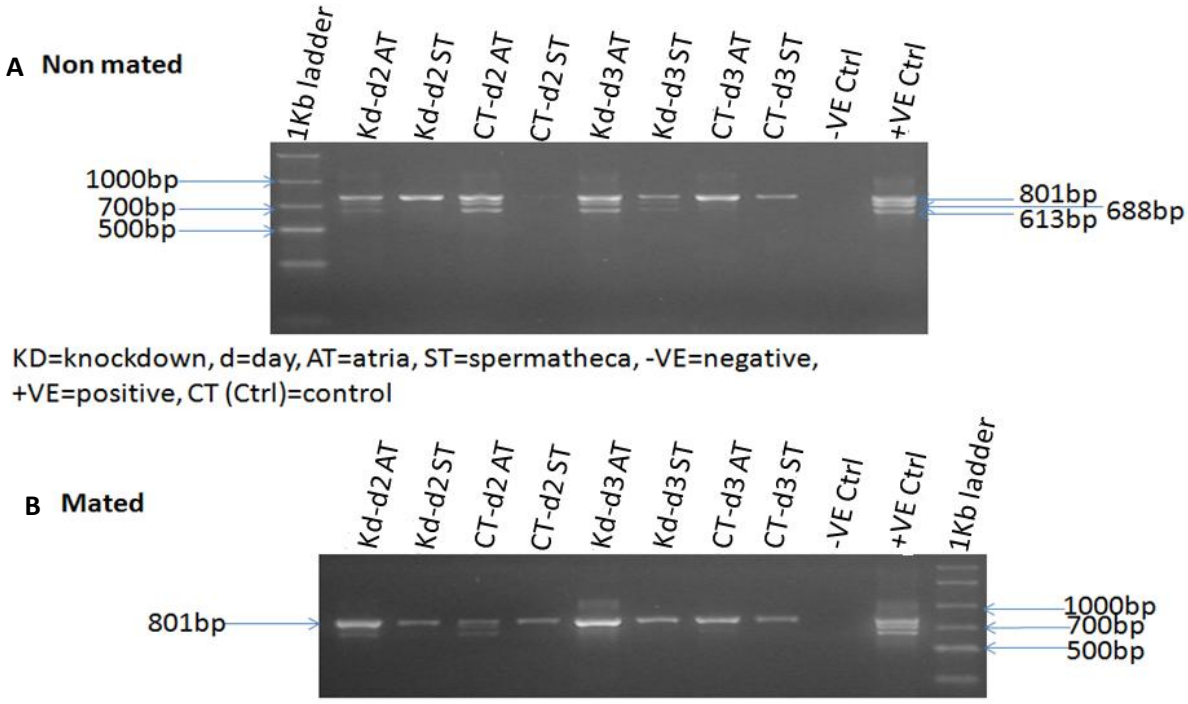


Figure 32B: PCR gel analysis on AGAP005195 after CATHEPSIN B (AGAP004533) knockdown in the female *Anopheles gambiae*; This shows possible signal transduction effects related to AGAP005195 A) This gel shows effects on the TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) active transcripts, showing that they are lost in the atria of both groups when injected and completely lost in the spermatheca. It is lost faster in the control group. B) This gel shows effects on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) active transcripts which are lost after mating. Only the 801bp band remains after mating and this band translates into a mutant protein after mating. Hence it should play no role in TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) function. KD (knockdown), Ins (inseminated).

4.5 Swarming patterns in male *Anopheles gambiae* mosquitoes and their role on mating responses in respective females

To date all the information obtained on the mating and post-mating biology in *An. gambiae* mosquitoes derives from studies performed in the laboratory. Although informative, these studies may not provide a fair representation of the processes happening in natural mosquito populations, and therefore need to be validated. As a first step towards validation of laboratory-based experiments, we tested the feasibility of setting up mating experiments in semi-field conditions provided by the malaria sphere at icipe, Thomas Odhiambo campus Mbita point, Western Kenya. Male mosquitoes sexed as pupae, 1000 in number, were released in malaria sphere 1 at 2 days post emergence. Swarm formation patterns and time were observed (Appendix VII). Swarming was categorized into 7 intervals of 5 minute each, from 18:45hrs when it was initiated till 19:20hrs when it was terminated (by which time no couples were seen) (Figure 33A). Swarming initiation was mainly observed from 18:45-18:54hrs in June, 18:55-19:04hrs in July and 18:50-18:59 in August. July and August saw a shift in swarm time. Swarms characterized as full due to the size and presence of many couples, were mainly observed from 19:00-19:04hrs in June, then 19:00-19:09hrs in July and 19:00-19:09hrs in August. Swarm termination was observed between 19:10-19:14hrs in June, and 19:10-19:19hrs in July and August. These swarming time periods observed for 3 months gave an insight on mating time frames possibly existing in actual field conditions. The males initiated the swarms and 2 - 3 swarms could be seen 3.5–4m above the hut, close to the roof top in malaria sphere 1 and at the corners in malaria sphere 2 (Appendix VII). Females later then found their way into the swarms and couples could be seen 5mins after swarm initiation, with a peak of mating occurring at 19:00hrs. Occurrence of copulation was followed up for 4 days, and then new mosquitoes were released on day 4. This was carried out for 1 month. Day 1 of release had the highest number of couples then the number of matings decreased gradually till day 3 (Figure 33B). At day 4, no couples of the released batch were actually observed. On day 1 of release the males swarmed mainly at the top of the hut but by day 2 and 3 swarms were seen at the corners of the screen house though couples were observed in them towards darkness which we think could result from the fact that most mated males were not so competitive hence the non mated males remained at the top of the hut though in small numbers. Swarming was affected by bad weather (wind and rainfall) and predators (spiders, geckos, ants, and praymantis). The mating window lasted for 30 minutes after which the females were gone. The males could still be observed swarming but no couples were seen.

During the day, the males could be seen resting under leaves while the females rested mainly in the hut simulating their host seeking behaviour.

One thousand mosquitoes in number were released at 2 days post emergence to observe how stable swarms are maintained in the field and this was carried out in the malaria sphere 2. Experiments with two days old mosquitoes were prefered as this is the age that females are postulated to mate in the field. Males swarmed on daily basis and couples were caught using sweepnets then transferred to cages for dissection at 3 days post-mating with aim to observe successful insemination of the females by the males. The age at which the female mosquitoes mated was not identified. A total of; 41.30% (n=74) of females were inseminated while the remaining 58.70% (n=107) were not inseminated as determined by the presence of sperm in the spermatheca, determined by microscopy (Figure 33C). The differences in both graphs were not statistically significant though in Figure 31B it showed that averagely 55 females mated on the day of release and averagely 30 mated on day two post release and averagely one ($P<0.01$) on day three post release indicating all females never mated on the same day after emerging as adults, meanwhile in Figure 37C it showed that fewer females were inseminated though caught as couples in the swarm.

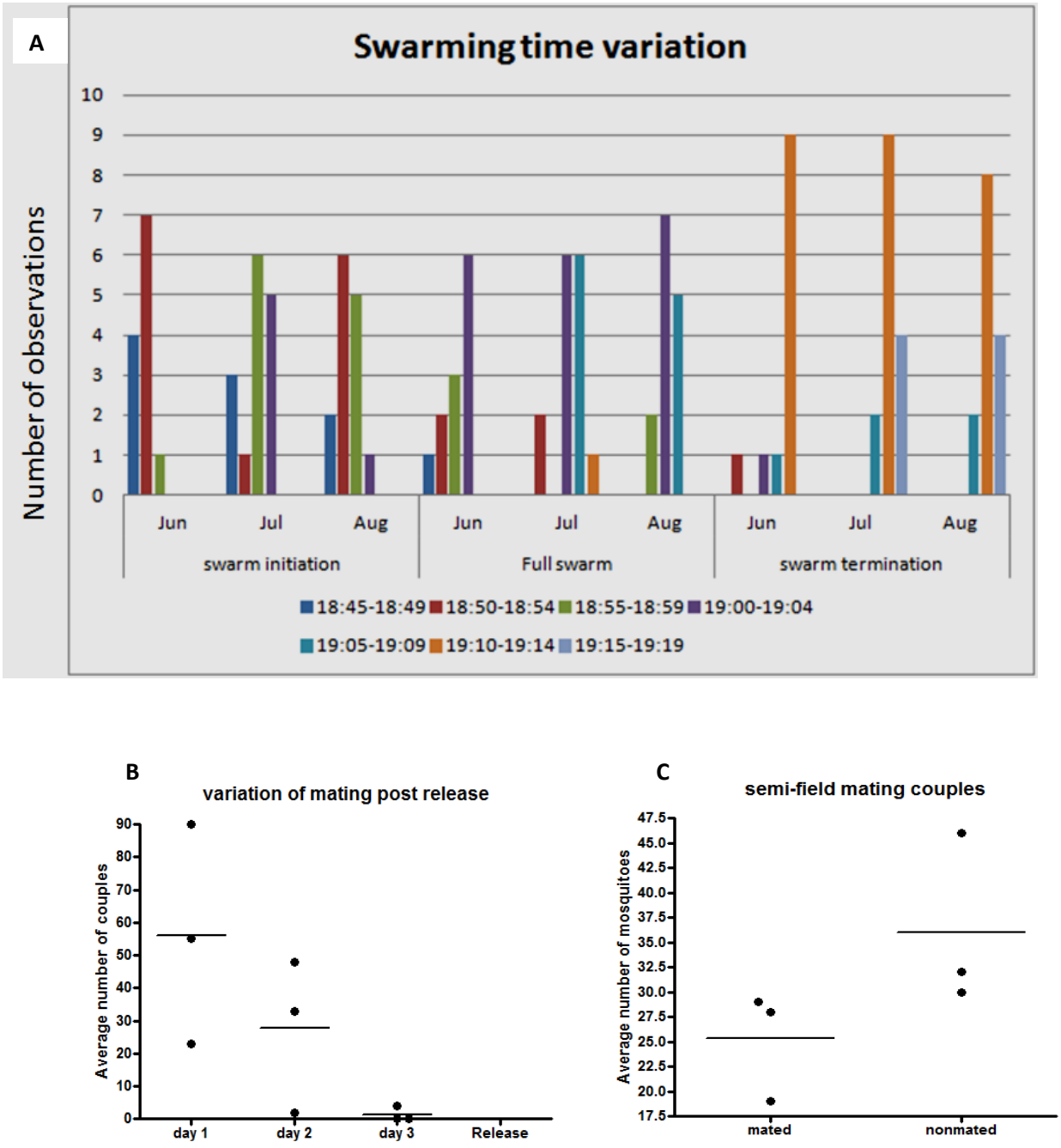


Figure 33: Swarming time detection, mating post release analysis and insemination post mating analysis in the mosquito *Anopheles gambiae*; This shows mating time variations and insemination success by males. A) Swarming variation with peak hours for swarm initiation from 18:45-18:54hrs in June with a shift observed in July and August from 18:50-18:59. Full swarms were seen more from 19:00-19:04hrs in June then from 19:00 -19:09 in July and August. Swarm termination was seen more from 19:10-19:14hrs in June then from 19:10-19:19hrs in June and July. There were time shifts observed in the months of July and August. B) The release of mosquitoes saw a decrease in swarming and mating from day 1 of release till day 4 when no couples and swarms where observed, then a new batch was released. C) Couples caught during swarming and dissected to check for successful insemination (sperm in spermatheca) after 3 days saw less females inseminated.

CHAPTER 5
DISCUSSION

These results showed that plugin protein constituted mainly helices which were confirmed by the presence of Gln residues within its structure and the various molecular states predicted by the HMM model. These residues are sites used by the transglutaminase enzyme in crosslinking action of the protein to form the mating plug (Rogers *et al,* 2009). The plugin protein contained low complexity regions known as regions with No Ordered Regular secondary structure (NORS) and are said to be dynamic helices (Sandhu *et al,* 2007). Such proteins do not respect the "protein-structure" paradigm but rather follow the "protein-trinity" paradigm which states that "the native proteins can exist in any of the three dynamic states Ordered, molten globule and random coils (Uversky, 2002). Most of these proteins are unfolded as seen for plugin. It is known that about 30% of the human proteome is predicted to be unfolded in the native state (Ward *et al,* 2004) and such proteins are involved in various functions such as; cell signalling, gene regulation, protein phosphorylation, storage of small molecules, oligomerization or self-assembly of large multiprotein complexes (Dyson *et al,* 2005). The predicted binding sites for plugin were concentrated at its unstructured region, which agrees to the fact that most of its functionally important sites are located on the loop regions that show large structural variability, although the scaffold of the protein structure is conserved as previously shown for unstructured proteins (Tress *et al,* 2005; Keedy *et al,* 2009). The presence of the active sites within this region behaves as a molecular switch in regulating certain biological processes. The plugin biological function (GO: 0005509) predicted by I-TASSER suggests a function similar to Vitellogenin linking it to post-mating responses within the female. Given that the *An. gambiae* plugin and transglutaminase are the two main proteins on the plug and expected to interact (Rogers *et al,* 2009), the absence of a fixed 3D structure for plugin may suggest an evolutionary modification in transglutaminase observed in the lost of its disulphide bonds in comparatively similar locations with human transglutaminases. This functional ability in plugin should permit it bind and move the transglutaminase protein around at the same time binding to other plug proteins. The protein interaction sites for Transglutaminase 2 have been shown to bind with ATP and GDP (Han *et al,* 2010) and some of these sites (LYS 544, GLN 548 and HIS 649) were predicted on the transglutaminase structure of *An. gambiae* hence suggesting that it could use the same sites to bind with plugin protein. The Phosphorylation PFam domains predicted on plugin could favour a possible interaction within the female involving plugin and other female proteases identified on the mating plug. The ML analysis carried out on the transglutaminase and

plugin group of proteins showed a closer relationship within the transglutaminase group of proteins as could be seen on the scale bar. The observed K values depicted on the branches show that there are more tranversions than transitions within this group which is an important aspect in conserving the protein structure within each taxa (Wolfner *et al,* 1997). This can be observed as K values get closer to zero. The ML ratio test confirms the K80 tree for transglutaminase and plugin group of proteins to be under selection meanwhile the K80 ML tree for plugin and its homologs suggests that these proteins are distantly related hence no perfect homolog for plugin (gi|270266159) was found in other taxa. This suggests that, the protein is unique to *An. gambiae* with a specific reproductive role which still needs to be identified. The ω value (indicative of selection) obtained in population genetic analysis (Felsenstein, 2004) revealed that the mammalian transglutaminase group of proteins was under purifying selection while that for the group of insect + others was under neutral selection. No ω value was predicted for the plugin group of proteins because of the low number of sites considered for the alignment but the K value could imply a phylogenetic signal favoring positive selection. Plugin was identified by Rogers *et al* (2009) as an Acp in *An. gambiae* and such proteins have been shown in *D. melanogaster* to be under positive selection (Wong *et al,* 2008). This selection could favor the specificity of plugin in reproduction. More work on the interacting surfaces on plugin protein will elucidate its possible roles in reproduction.

Plugin was projected on the network as the main hub protein given its structural similarity to Q8SX59 in *D. melanogaster*. Hubs are about three times more likely to be essential to *S. cerevisiae* compared to their non-hub counterparts (Jeong *et al,* 2001). Therefore plugin should be the main protein for male *An. gambiae* protein interaction network involved in plug formation within the male as previously shown by Rogers *et al* (2009). The domains predicted on plugin suggest that, it could interact in the female with TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) which is an ortholog to Q9VEM7 in *D. melanogaster*. Genomic comparisons between TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) and its ortholog Q9VEM7 confirmed their similarities in all the observed domains implying TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) could play a possible digestion and terminal cell signalling role in the mating process. The presence of Acp29AB known to be involved in sperm competition (Wong *et al,* 2008) and secreted by the MAGs in *D. melanogaster* goes to identify the network as a possible reproductory process given that Acps are only secreted by the males during mating and only take part in

reproduction (Kalb *et al,* 1993). The interaction of this Acp with sex peptide in *D. melanogaster* (Carvalho *et al,* 2006; Mueller *et al,* 2008), a hormone known to drive postmating behavioural patterns in the female suggests strong hormonal activity occurring within the reproductive process in both species. This hormonal function was further confirmed on the sub-network as Acp29AB was seen interacting with Q8WS79, an estrogen related receptor and a hormone involved in reproduction in *D. melanogaser*. The grouping of *An. gambiae* proteins on the ML tree identified two main clusters, considered as postmating and premating clusters. The premating cluster grouped most MAGs proteins with; transglutaminase known to be involved in the crosslinking action on plugin for mating plug formation (Rogers *et al,* 2009) and two Serine Protease inhibitors (AGAP003139 and AGAP009212) favoring successful plug formation. The postmating cluster grouped plugin, AGAP008276, AGAP008277, CATHEPSIN B (AGAP004533) and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) (only female protease on the mating plug). This interaction indicates a possible post mating molecular function in the female given that plugin had been predicted to interact with TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) given the identified PFam domains. TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) and CHYMOTRYPSIN-LIKE SERINE PROTEASE AGAP005194 were identified by Rogers *et al* (2008) as two female proteases to be downregulated by mating which was equally confirmed in our results. They were characterized as Serine proteases due to the presence of Trypsin_Histidine, Trypsine_Aspartate and Trysine_Serine active sites. These 3 sites are needed for the protein's proper function in its 3D conformation (Blow, 1997). Interestingly AGAP005194 was not identified on the network which could have resulted from a mutation at the Trypsin_Histidine active site with possible loss of its function as a Serine protease hence making TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) more specific to the plug as observed from the immunostainings. CATHEPSIN B (AGAP004533) is involved in the activation of Trypsin proteases in man through cleavage of their propeptide (Niels *et al,* 2002). They are also known in *D. melanogaster* to be involved in a controled cleavage of ovulin and sex peptide in post mated females hence favouring long term sex peptide effect in female postmating responses (Walker *et al,* 2006) therefore suggesting that, CATHEPSIN B (AGAP004533) could be involved in digesting the mating plug through the cleavage of the male Trypsins (AGAP008276 and AGAP008277) and other Acps. AGAP008277 a male specific protease (Rogers *et al,* 2009) presented an amidation site (PS00009) with hormonal related functions hence it could be involved in hormonal activation in the atria of the *An. gambiae* female mosquito after mating. In *D. melanogaster* it was

shown that Acp genes which code for small proteins originate from non coding DNA (Begun *et al,* 2006) making them species specific. These genes are known to be under positive selection (Aguadé, 1999) with specific sperm related functions and Acp29AB predicted on the network was shown to be involved in sperm competition and sperm storage through its lectin role (Wong *et al,* 2008). This suggests that plugin known as an Acp in *An. gambiae* could also have sperm related functions and be under positive selection as was observed on the ML K80 tree. Sperm cells considered as nonself due to antigenic surfaces acquired during spermatogenesis (Sharpe, 1994) could also be recognized by the female as nonself and result in the production of antibodies against them. It has been shown in man that sperm protection in the female is achieved by seminal plasma derived immunomodulatory factors that are carried with or coated on inseminated spermatozoa (Kelly, 1995). Semen also contains substances from male accessory glands which are geared towards proper sperm function. Most MAGs proteins in *An. gambiae* have been shown to be male specific (Rogers *et al*, 2009) suggesting that MAGs proteins and especially AGAP008276 and AGAP008277 which are male specific (Rogers *et al*, 2009) proteases could be more involved in digestion of the plug and drive post-mating behavioural phenotypes in the female since it is transferred alongside the plug to the female. The post-mating protein cluster on the *An. gambiae* phylogenetic tree was of great importance as it helped in the understanding of possible proteins in the male mosquito which could be involved in postmating events in the female mosquito. Further molecular analysis done on this cluster of proteins will help to shed more light on their possible roles in mating. Given that the identified *An. gambiae* interologs identified on the network showed no direct interactions with each other, this could suggest many MAGs proteins still need to be identified which will help in better understanding of the reproductive pathways in the male mosquito and possible targets for reproductive control.

These results showed that RT-PCR amplifications on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) identified 3 spliced variants; 5195AT-2, 5195AT-3 and 5195AT-4. All three variants confirmed alternative splicing as seen by the 5'GT and 3'AG splicing junctions for the introns. The identification of these variants was of great interest and could also imply different functions performed by the same gene when translated into the protein (Zahler, 2005) within the cell. 5195VS-1 and 5195VS-3 in the spermatheca and 5195AT-2 in the atria retained all the introns but translated into a truncated protein resulting from a frame shift caused by shifts in the 5'GT and 3'AG junctions within the first intron causing misplicing in these variants. This was also shown by Fay et al (1992) that the insertion in Plasminogen-activator inhibitors (PAI) of a TA at the 3' end of exon 4 possibly

due to DNA polymerase slippage causes a shift in the reading frame with the insertion of an abnormal stop cordon within the sequence hence breaking the protein structure. This could suggest that the mispliced proteins may have lost their functions and so no active protein will be expected in the spermatheca. This was further supported by Western blot analysis given that the TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) band could not be seen in the spermatheca but was observed in the atria. The absence of the band in the spermatheca could suggest the inability of the antibody to bind to the protein because of lack of a fixed 3D structure otherwise the transcript was targeted for destruction by the cell machinery. Ozvegy-LacZka *et al* (2005) showed that specific monoclonal antibody 5D3 could be used to differentiate between functional and non-functional ABCG2 proteins which had functional ATP- and drug-substrate-binding sites but differed in the formation of a catalytic intermediate. This technique could be used to differentiate our obtained variants given that TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) was seen to posess multi PFam domains suggesting multiple functions within the protein. The number of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) bands (40KDa, 80KDa and 160KDa) seen on the gel could possibly be related to the number of glycosylated sites on the protein (Beuckmann *et al*, 1996). Three N_Glycosylation sites were predicted on the protein at amino acid; 85, 141 and 146 suggesting glycosylation could be a possible post translational modification process taking place on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) protein sequence. N-glycosylation may also take part in proper folding of the protein. The two possible TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) functional variants (5195AT-3 and 5195AT-4) differed by one point mutation N75K resulting from T288A substitution on the latter variant. The functional effect of this point mutation was not tested but Ranson *et al* (2000) showed that a Leucine-Serine substitution (L1014S) within the voltage-gated sodium channel gene conferred knockdown resistance (*kdr*) to DDT and pyrethroid insecticides, in a colony derived from specimens from Kenya. This mutation could play a very vital role in differentiating these two functional variants and needs further analysis. All three variants showed an amino acid mutation at G200S located at the active serine site. The loss of this serine deprives it from the Trypsin_Serine active site suggesting that TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) may not perform the Trypsin protease function as originally proposed by Rogers *et al* (2008) given that the functional interaction of the three catalytic sites (Trypsin_Histidine, Trypsin_Aspartate and Trypsin_Serine) is vital for the protein's function. TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) still retained its Trypsin_Histidine (PS00134) active site which has been shown to be involved in electron

relay functions in the 3D structure of the protein and equally the response regulatory site identified by CLC workbench was still present in the variants. Therefore the two response regulatory function with possible signal transduction related function may be the possible functional role effected by this protein within the female atria of *An. gambiae*. Immunostainings done on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) suggests its specificity on the plug and in mating and this difference in localization compared to AGAP005194, a female specific Serine protease identified on the plug, could result from loss of the Trypsin_Histidine active site which goes further to confirm the absence of the AGAP005194 on the network and also suggesting a nonspecific role in mating. The presence of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) within the spermatheca duct from atria to spermatheca in 6hrs postmated females could imply a possible role of the protein in sperm related function which was inferred by the $GABA_A$ domain predicted on this protein. $GABA_A$ receptors have been found in the spermatozoon plasma membrane and GABA stimulates hyperactivation and promotes the acrosome reaction (Calogero *et al,* 1996). The immunostainings on day one and day two virgin females showed a low expression of this protein in the atria and spermatheca suggesting possible non secretion because the ducts were all void of this protein implying it could be maintained in the cells and not secreted until mating occurs. The secretion of this protein could possibly be triggered by the presence of the plug in the atria as seen with the secretion of Trypsin in the stomach only after the presence of food (Khorasani *et al,* 1990). The secondary structural analysis of the variants 5195AT-3 and 5195AT-4 showed the insertion of a turn at the N75K mutation on the latter hence facilitating the formation of cysteinic disulphide bonds in the latter variant which could give it a different 3D conformation from the former hence different functional proteins. The response regulator site with the phosphor_Tyrosine kinase (GO: 0004721) receptor site confirms its possible involvement in signal transduction related functions. Indepth molecular analysis on this domain will help to clarify its role in signal transduction.

CATHEPSIN B (AGAP004533) was shown by Rogers *et al* (2009) to be expressed in the virgin female atria and it was also identified on the plug. qRT-PCR analysis performed showed that CATHEPSIN B (AGAP004533) was expressed in all life stages of the mosquito contrary to TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) which was expressed only in adult day one females and plugin which was expressed from late male pupae. It has been shown in *An. gambiae* that the detoxification P450 enzyme varies in the larvae, pupae and adults due to different environmental exposures (Strode *et al,* 2006). This could mean

that the protein has specific functions at different life stages of the mosquito and this fact is also backed by multiple domains identified on this protein.The transcript expression of CATHEPSIN B was also identified in all tissues of the male and female mosquitoes and expression variations were also observed. Expression variations of the same gene in various tissues known as differential expression resulting from different molecular properties of the gene in these tissues could be a possible mechanism involving CATHEPSIN B (AGAP004533). Regions of DNA that affect gene expression are highly variable, containing 0.6% polymorphic sites (Stephens *et al,* 2001). These naturally occurring polymorphic nucleotides can alter *in vivo* transcription rates (Segal *et al,* 1999; Crawford *et al,* 1999). This suggests that this gene could play different roles in the various reproductive tissues for reproductive success in *An. gambiae*. Sequencing transcripts from different reproductive tissues identified SNPs suggesting that, they could play important roles in the expression variation levels observed for the transcript. The secondary structure predictions of CATHEPSIN B (AGAP004533) variants in the various reproductive tissues both in males and females of *An. gambiae* showed structural variations which could affect the 3D conformation and function of the protein. From the network analysis, the ortholog of CATHEPSIN B (AGAP004533) in *D. melanogaster* (Q9VY87) was not directly interacting with any of the MAGs proteins implying it could be expressed earlier in the plug formation pathway in the male given that the plug is an end product of the reproductive process in the male (Rogers *et al,* 2009). This interactive property of CATHEPSIN B (AGAP004533) could help to explain its absence from observed virgin MAGs transcripts (Rogers *et al*., 2009) though seen to present the most favoured interactions for plug formation as predicted on the network. The qRT-PCR analysis performed in the various reproductive tissues showed a higher expression of the gene in the female atria and spermatheca compared to the male MAGs and testes implying the gene could play a terminal role in the female reproductive process. In mated females the transcript regulation of CATHEPSIN B (AGAP004533) was relative to TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) as both were down regulated after mating suggesting these two genes could be co-expressed in the female with one affecting the expression of the other. Grigoriev (2001) analyzed physical interactions in yeast and observed that proteins encoded by co-expressed genes interact with each other more frequently than with random pairs. In the virgin atria of the G3 strain, there was no statistical difference observed in the expreesion levels of CATHEPSIN B (AGAP004533) and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) but this difference was significant ($P<0.0001$) in the atria of the Mbita strain with the expression of TRYPSIN-LIKE

SERINE PROTEASE (AGAP005195) very high. The mbita strain is more of a field strain and so the difference could be strain specific based on their sexual competitive role and the way both mosquitoes are reared. For example femlale mosquitoes in the wild will have to identify swarms before being mated unlike full lab colonized strains where both couples are reared together and mating is easier and faster. The expression levels of both genes in the mbita strain could imply different roles played by these two proteases during mating. Relative expressions of the proteins CATHEPSIN B (AGAP004533), AGAP005194, TRYPSIN-LIKE SERINE PROTEASE (AGAP005195), AGAP004203 (vitellogenin) and AGAP002620 to RPL19 (house keeping gene) showed varied expression levels in virgins and postmated females. The first 3 proteins were expressed in the virgins and down-regulated after mating meanwhile the last two were only expressed at 6hrs and 24hrs postmating respectively. All proteins were down-regulated after 24hrs. The same transcript expression pattern of these proteins in the G3 strain of the *An. gambiae* mosquito had been shown by Rogers *et al* (2008). The similar response by these female specific proteins, 24hrs after mating could imply that they are co-regulated and possibly involved in a common interacting reproductive pathway. In the males of the Mbita strain, transcript levels of CATHEPSIN B (AGAP004533), Plugin, transglutaminase and AGAP009370 were compared and transcripts of plugin, transglutaminase and AGAP009370 were seen to be more expressed in the testes and MAGs of virgins and mated suggesting further the eclipsed role of CATHEPSIN B (AGAP004533) within the reproductive pathway in the males as predicted on the network. At this point it was not too clear if any of the tested male genes could have any effect on the other and so further analysis through knockdown of CATHEPSIN B (AGAP004533) in both males and females was used to evaluate relative male and female reproductive transcript expressions.

These results showed multiple bands for plugin on the gel with unexpected number of bands and different band sizes. The absence of the forward primers of primer pair 1 and primer pair 3 sequences in both the genebank and VectorBase mRNA sequences shows possibility of two different plugin sequences which was further confirmed by the absence of the reverse primer of primer pair 1 sequence in XM559853 (VectorBase). This could suggest; gene duplication with loss of several sequence fragments in both variants, or site specific expression of the same gene and these sequence variations could be further supported by the fact that, plugin at VectorBase was identified in the salivary glands (Arca *et al*, 2005) meanwhile that at NCBI was deposited by Rogers *et al* (2009) identified in the MAGs. The repeat regions seen in the plugin protein could play significant roles in the generation of

several variants in this gene as seen by the various bands on the gel for a given primer pair. Plugin was identified as an Acp in the male *An. gambiae* and Acps in *D. melanogaster* are known to be small in protein size (Begun *et al,* 2006) and from the smaller then expected band size variants obtained with Primer Pair 3, could account for such events taking place within this gene locus. This could suggest that, this protein could be involved in the generation of Acps which are transferred to the female with the plug during mating with aim of performing several functions after plug digestion. The small sizes could also suggests the presence of different splice sites flanking introns on the transcript leading to different variants derived from this gene during translation. This splicing events in plugin transcript as observed in the male MAGs and testes injected with double stranded RNA for CATHEPSIN B (AGAP004533) (section 4.4.1.2) showed novel start sites within the variants and different 5'GT and 3'AG splice sites within the variants as well meanwhile only one variant was seen in the LacZ injected controls. Further analysis on the possible translational events on plugin transcript will probably help in identifying the various variants produced by this gene and their functional analysis will help to explain the identification of only a few Acps known so far in the MAGs of *An. gambiae*. The lack of a fixed 3D structure in Plugin due to low complexity regions in the protein could be an advantage of the protein in holding the plug in place as seen by its putative properties of a top-most hubbal protein in the network. The gradual breakdown of this protein in the female atria should release proteins of different sizes and various functional roles with objective to drive postmating responses within the female. So far nothing is known about the domain structure of this protein and it has no close homologs in other taxa therefore more work on the molecular analysis of Plugin is still required to circumvent its role in mating.

These results showed that, two main domains were predicted on CATHEPSIN B (AGAP004533) which were; RNA binding domain and the D-Amino Acid Oxidase (DAAO). The former domain was of particular interest because of the spliced variants identified in TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) and also because they have similar expression levels. The RNA binding domain is known to interact with the spliceosome responsible for proper splicing of mRNA at the 5'GT and 3'AG junctions hence excising the introns for the right splice variants (Zahler, 2005). The molecular function of this domain was evaluated through knockdown of CATHEPSIN B (AGAP004533) in the G3 males. From these results plugin expression in the experimental sample showed an increase in expression levels in the MAGs and testes relative to the LacZ controls though the difference was not statistically significany ($P<0.05$). The same expression pattern was observed in the testes for

transglutaminase and AGAP009370. This expression profiles showed possible co-expression of CATHEPSIN B (AGAP004533) and plugin given that the knockdown of CATHEPSIN B (AGAP004533) affects plugin expression levels relative to the control hence CATHEPSIN B (AGAP004533) and plugin could be involved in the same functional pathway. Rogers *et al* (2009) showed that 15% of females who mated to males knockdown for the transglutaminase enzyme lacked a mating plug and transglutaminase is involved in the cross-linking action on plugin to form the plug therefore, CATHEPSIN B (AGAP004533), plugin and transglutaminase could be involved in the pathway leading to the formation of the mating plug. The same expression profiles for TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) and AGAP004533 were seen in the female probably confirming their presence on the protein-protein network as seen above. This goes further to confirm the clustering of plugin, CATHEPSIN B (AGAP004533) and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) on the ML Phylogenetic tree hence CATHEPSIN B (AGAP004533) could be involved in the male reproductive machinery at a given point in time hence with an important role in the mating plug formation pathway (Rogers *et al*, 2009). Multiple spliced variants of plugin transcript were observed in the testes and in the MAGs while only one variant was seen in theMAGs and testes of the control. These variants resulting from the knockdown of CATHEPSIN B (AGAP004533) confirmed the role of the RNA binding domain identified in this protein. The spliced variants showed a varied use of predicted transcription binding factors (TFs) on the experimental sequences meanwhile those of the controls showed similar TFs along the sequences confirming the fact that they were the same and could therefore be the actual fragment expressed in the testes and MAGs of *An. gambiae* males. Transcription factors known as proteins that bind to specific DNA sequences thereby controlling the genetic flow of information from DNA to RNA (Latchman, 1997) could be wrongly inserted within the transcripts of the experimantals probably by polymerase slippage hence inserting wrongly spliced sites, for example T00626 (Diebold *et al*, 2002) which will be wrongly used by the RNA binding domain of CATHEPSIN B (AGAP004533) and the spliceosome complex hence misplicing the experimental transcripts while T00691 (Devoto *et al*, 2003) splice site predicted on the LacZ controls favors proper splicing of the transcript as observed hence possible justification for a common variant observed in the control transcripts. The difference in spliced variants as observed in the alignments, justified the presence of 5'GT and 3'AG spliced junctions which are alternatively used by the spliceosome complex to produce several transcript lengths. This gene could form a complex locus wherein the DNA binding sequences by the TFs are located further away from the start

sites and also from the known protein region to facilitate successful splicing functions. This is a similar case seen in the haemoglobin complex locus wherein the complex program of transcriptional regulation leading to the differentiation and developmental stage specific expression in the globin locus is mediated by DNA-regulatory sequences located both proximal and distal to the gene-coding regions. The most prominent distal regulatory element in the human b-globin locus is the locus control region (LCR), located from about 6 to 22 kb upstream of the ε-globin gene (Tuan *et al,* 1985; Forrester *et al,* 1987). Identifying these TF binding sites can help in gene modification at these sites hence haulting expression of the gene and possible post mating effects in the female.

These results showed that, knockdown of CATHEPSIN B (AGAP004533) in the female mosquitoes of the mbita strain observed a down-regulation of the transcript in the experimental group compared to the control group. The difference was statiscally significant at 6days postmating. TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) was down-regulated in the experimental group at days two, three post-injection and 6hrs postmating which was not observed in the control group. This goes further to confirm its possible co-expression with CATHEPSIN B (AGAP004533). AGAP004203 (vitellogenin) and AGAP002620 showed no expression before mating in both groups but were only expressed after mating. This further confirms that both transcripts are only induced by mating as showm by Rogers *et al* (2008). AGAP004203 was significantly expressed ($P<0.001$) in both the experimental group and the lacZ control group compared to the other transcrips though the transcript expression difference compared in both the experimental and control groups was not statistically significant suggesting that, this could result from the loss of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) which was absent in non mated and mated females after RT-PCR analysis. Experiments to evaluate the mating drive phenotype in *An. gambiae* females showed a high number of females in the male cage during swarming for both non-injected and injected females, indicating that mating drive in the female mosquito could be accounted for by two factors, a possible pheromone released by the males to attract females and the maturity of the female reproductive system ready for mating. Based on the results, more experimental females went to the swarm though few were inseminated compared to less lacZ control females who went to the swarm though more were inseminated. This suggests that the interaction between CATHEPSIN B (AGAP004533) and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) could be involved in female maturity and influencing her decision to go to the swarm to be mated. The functional effects of these proteins could be in correlation with swarm response genes located on the antennae therefore down-regulation of

both transcripts in the experimental favors a swarm drive decision but compromising its maturity and willingness to mate while the opposite effect is seen in the controls. This could be further ascertained by the resistance in experimental females towards mating as shown by a higher proportion of inseminated females in the female cage thereby suggesting a struggle by the males to mate the females, hence ending up in female cages. The hormonal component of this interaction could be another possible factor since the presence of transcripts like; AGAP0013533 (Juvenile hormone binding protein) known to be linked to circadian clock and feeding behavior in *D. melanogaster* and also modulate aggregation behavior in *Locusta migratoria* (Sarov-Blat *et al,* 2000), and AGAP000278 an odorant binding protein both identified to be present in virgin female atria (Rogers *et al*., 2009). The ability of virgin females to identify swarms (section 4.4.2.1) and their active nature only at dusk shows a spatio-temporal specificity in their reproductive behavior. Juvenile hormone could be involved in circardian clock patterns in the female *An. gambiae* and it has also been shown to stimulate the transcription of vitellogenin in *D. melanogaster* (Sarov-Blat *et al,* 2000), so the maintained expression of vitellogenin could suggest that AGAP004203 (vitellogin), CATHEPSIN B (AGAP004533), TRYPSIN-LIKE SERINE PROTEASE (AGAP005195), AGAP000278 (Obp) and juvenile hormone should interact in a common pathway. CATHEPSIN B (AGAP004533) could be involved in activation of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) because CATHEPSIN B (AGAP004533) is known to activate pancreatic trypsinogen which acts on trypsin (Niels *et al,* 2002). TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) posesses the trypsin PFam domain hence absence of CATHEPSIN B (AGAP004533) could affect its expression. Both active transcripts of TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) were lost as seen by RTP-CR, implying CATHEPSIN B (AGAP004533) could play a role different in the male than in the female mosquito of *An. gambiae*. It was also observed from inmmunostainings done on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) that the protease was only released into the spermatheca duct after mating confirming the fact that TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) needs to be activated somehow before functioning. The PFam Metallo beta lactamase domain identified on TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) is known to belong to the esterase family. This family is known in *D.melanogaster* (Est 6) to degrade cis-Vaccenyl acetate (cVA), a pheromone (Chertemps *et al,* 2012) transferred by males to females during mating and through its odorant binding receptor (LUSH, GO: 0019236) induces post mating behaviors in the female. This implies that TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) could be degrading Acps with

odorant binding related properties released by the male on the plug during mating therefore loss of the transcript in injected virgin experimental and LacZ control females could account for the increased expression in vitellogenin after mating compared to non injected females. In the LacZ controls we think that the cell machinery manages to balance the female reproductory system after mating since CATHEPSIN B (AGAP004533) is still actively expressed, which will not be the case in the experimental because TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) is also down-regulated due to knockdown of CATHEPSIN B (AGAP004533). RNA binding PFam domain in CATHEPSIN B (AGAP004533) was not seen to play the same role at the molecular level as in the males after knockdown. This was confirmed through RT-PCR analysis on plugin transcript, which showed alternative spliced variants in the male but we only observed a loss of the TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) transcript after knockdown in the female. The amino acid substitutions on CATHEPSIN B (AGAP004533) and its possible secondary structural modifications suggests that, CATHEPSIN B (AGAP004533) could play different roles in different reproductive tissues at the molecular level thereby affecting the mosquito's reproductive process at a phenotypic level. Identifying interacting sites between CATHEPSIN B (AGAP004533) and Plugin in the male and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) in the female will help in understanding their interactions for vector control interventions.

Understanding mosquito behavioural patterns in the field has been of great interest in vector biology especially for vector reproductive control interventions. Understanding swarm patterns for better clarity of mosquito reproductive process in the wild, it was observed that; the swarming and mating time was consistent with an interval of 20–25 minutes. The swarm peaks (full swarm) saw a lot of mating couples. Their swarming in open space could be due to predator avoidance, which was a serious issue observed for swarms at the corners of thes ame screen house. Those couples, which rested on cages, could have been eaten up by the predators available. Males were seen going back to the swarms after mating without the females as shown by (Charlwood *et al,* 1979). Since mated females never went back to the swarm, couples seen and caught decreased towards swarm termination hours. Males could still be seen swarming at the corners without couples and then finally rested on the cages or leaves. It was also interesting to note that females went to the swarm after male initiation (Diabate *et al,* 2003) and despite the human odour present there were no bites experienced until when couples couldn't be observed anymore. This aspect of the female showed an odour preference distinguished by the female at various physiological aspects in their life and also

possibly implying the release of a pheromone by the males during swarming. The mating depletion experiments suggested that for every batch of mosquitoes emerging from the field, there is a time frame for the females to mate after which males of that batch probably stop swarming and another group takes over. There could be possible overlap between batches but with possible out competition of the previous batch by the most recent batch. The aspect of catching mating couples on different days could imply a decision by females to go to the swarm and be mated and this could be based on their sexual maturity. The mating window was also quite small and given that the process involved a lot of energy (Yuval *et al*, 1993), all females could possibly not be mated on the same day. Day one of release saw little or no bites, which increased with day two and three. This could possibly be due to the fact that the mated females then seek a host for blood meal after mating. It was also important to note that some mosquitoes got preyed upon which could also account for the loss in numbers. The reduction in couples post release could confirm the idea that females mate just once in their lives (Holt *et al*, 2002). These results also ties with that of Rogers *et al* (2008) showing that mating drives postmating responses in female *An. gambiae* mosquitoes given the fact that the females don't go back to the swarm but rather respond more to human odours after mating than as virgins. The decision taken by the female to go to the swarm and be mated is very important. It has previously been shown that; genetically it is more advantageous for the female to mate first before looking for the blood meal since she will need to host seek for the rest of her life though virgin females are also known to seek blood meals (Lyimo and Takken, 1993) but their specificity in terms of odour selection is very narrow. Therefore there could be a relationship between the odour profiles and mating expressed genes between mated and non-mated females. The fact that most of the females are not inseminated could mean the females need some age for proper insemination given that mosquitoes were released at two days old and according to Charlwood *et al* (2003); two days old mosquitoes are not properly inseminated. This suggests that, the majority of non-inseminated females were the two days old. This could also imply that in a real field situation very few female mosquitoes are properly inseminated after mating and so very few will lay eggs per batch of mosquitoes post emergence. This also shows that the genetic program for swarm drives could be linked to the mating machinery but the female cannot distinguish when she is actually ready for mating. Therefore targeting what takes a female to the swarm will be of greater importance for reproductive control than controlling the mating process itself. Mating events in the mosquito are still poorly understood and needs a lot of work to be used for vector control interventions. From this work, it shows that the events leading to mating in the female are quite different

from the reproductive events that occur after mating in the female. This could be seen by the fact that CATHEPSIN B (AGAP004533) expressed in both males and females could perform different regulatory roles in both sexes of the mosquito relative to mating. To better clarify and understand mating behaviour in these species, a lot of field work application is needed over long periods of time and across different geographical zones. Mating behavioural patterns linked with genes expressed in the reproductive organs of both males and females will identify targets for reproductive control, and identifying genes linked to sex drive in the female *An. gambiae* will help in targeting virgin females to prevent them from going to be mated by the males during swarming.

CHAPTER 6
CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

This study attempted to analyze the roles of MAGs and female reproductive proteins and the possible roles they play in reproduction. The study showed that the protein properties and interacting residues identified on plugin and transglutaminase could contribute to the successful formation of the mating plug. The positive selection process on plugin makes it special in the reproductive process in *An. gambiae*.

Plugin could be the main protein in mating plug formation in the male *An. gambiae* due to its unstructural advantage from low complexity regions. Its putative interaction with the female TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) as seen on the network could drive the primary reactions that occur within the female after mating. Plugin, CATHEPSIN B (AGAP004533) and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) could co-function together after mating as identified on the post-mating cluster on the phylogenetic tree.

TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) could be mating specific as seen by its presence on the mating plug within the female atria and in in the spermathecal duct after mating. CATHEPSIN B (AGAP004533) is differentially expressed in the various mosquito life stages and the various male and female reproductive tissues therefore showing functional variability in the male and female *An. gambiae* mosquito. Plugin transcript in the male could be alternatively spliced to produce different functional variants in the testes and the MAGs of the male mosquito.

CATHEPSIN B (AGAP004533) is possibly involved in splicing events on plugin through its RNA binding domain hence affecting the method in which it is delivered to the female during mating. CATHEPSIN B (AGAP004533) is involved in a common pathway with TRYPSIN-LIKE SERINE PROTEASE (AGAP005195), which defines a possible mating drive phenotype in the female *An. gambiae.*

Swarming is the main method for *An. gambiae* males to identify females during mating. Females are attracted to the swarm but not all are properly inseminated hence this could affect their reproductive efficiency in the wild.

6.2 Recommendations

1. X-ray analysis based on the protein-protein interacting surface residues identified on plugin and transglutaminase is required to better understand their 3D conformation for vector control interventions through amino acid residue modifications on the protein.
2. The identification of more reproductive proteins involved in the mating process in the male and female *An. gambiae* is neccessary for a better interactive network, which will identify more putative targets for vector control interventions.
3. Intense molecular analysis through yeast-2-hybridization and liquid chromatography mass spectrometry will elucidate the interacting sites and active fragments of Plugin, CATHEPSIN B (AGAP004533) and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195). These sites can be chemically rendered inactive in the mosquito hence preventing reproduction in *An. gambiae*.
4. The knockdown of more proteins alongside CATHEPSIN B (AGAP004533) in the male and the female *An. gambiae* is neccessary to understand the reproductive pathways implicating plugin in the male and TRYPSIN-LIKE SERINE PROTEASE (AGAP005195) in the female. This will help in identifying other proteins used by the male and female mosquito for reproductive success.
5. Understanding mosquito reproductive biology through swarming patterns and temporal distribution in the wild is necessary to elucidate swarm effects on female behaviour after being mated by the males. This will help in understanding post-mating effects (oviposition, ovulation and lifetime refractoriness to mating) in the female.

CHAPTER 7

REFERENCES

Abraham, E. G., Donnelly-Doman, M., Fujioka, H., Ghosh, A., Moreira, L., & Jacobs-Lorena, M. (2005). Driving midgut-specific expression and secretion of a foreign protein in transgenic mosquitoes with AgAper1 regulatory elements. *Insect Mol Biol*, 14: 271-9.

Adachi, J., and Hasegawa, M. (1996). Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J Mol Evol*, 42: 459-468.

Aguadé, M. (1999). Positive selection drives the evolution of the *Acp29AB* accessory gland protein in Drosophila. *Genetics,* 152: 543-551.

Aikawa, M., & Seed, T. M. (1980). Morphology of Plasmodia in: malaria. *Academic Press, New York*, 1: 285-344.

Aksoy, S. (2008). Transgenesis and the management of vector-borne disease New York. *Springer Science and Business Media, LLC Landes Bioscience*.

Alphey, L. (2002). Re-engineering the Sterile Insect Technique. *Insect Biochemistry and Molecular Biology*, 32: 1243-7.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol,* **215**: 403-410.

Altschul, S. F., Madden, T. L., Schaeffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25: 3389–3402.

Ambrish, R., Alper, K., Yang, Z. (2010). I-TASSER: a unified platform for automated protein structure and function prediction, *Nat Protoc*, 5: 725–738.

Arca, B., Lombardo, F., Valenzuela, J. G., Francischetti, I. M., Marinotti, O., Coluzzi, M., Ribeiro, J. M. (2005). An updated catalogue of salivary gland transcripts in the adult female mosquito, *Anopheles gambiae, J Exp Biol*. 208(Pt 20)**:** 3971-3986.

Armstrong, S. J. R., Abdullah, S & Minja H. *et al.* (1999). KINET: a social marketing programme of treated nets and net treatment for malaria control in Tanzania, with evaluation of child health and long-term survival. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 93: 225-231.

Begun, D. J., Lindfors, H. A., Thompson, M. E., Holloway, A. K. (2006). Recently evolved genes identified from Drosophila yakuba and D. erecta accessory gland expressed sequence tags. *Genetics*, 172: 1675–1681.

Beier, J. C., Oster, C. N., Onyango, F. K., Bales, J. D., Sharwood, J. A., Perkins, P. V., Chumo, D. K., Koech, D. K., Whitmire, R. E., Diggs, C. L., & Hoffman, S. L. (1994). *Plasmodium falciparum* incidences relative to entomologic inoculation rates at a site proposed for testing malaria vaccines in Western Kenya. *American Journal of Tropical Medecine and Hygiene*, 50: 529-36.

Beuckmann, C. T., *et al*. (1996). "Lipocalin-type prostaglandin D synthase (beta-trace) is located in pigment epithelial cells of rat retina and accumulates within interphotoreceptor matrix." *J Neurosci*, 16(19): 6119-6124.

Birkhead, T. R. & Møller, A. P. (1998). Sperm competition and sexual selection. *London: Academic Press,* 826: 619-628.

Blandin, S., Moita, L. F., Kocher, T., Wilm, M., F.C. Kafatos, F. C., Levashina, E. A. (2002). Reverse genetics in the mosquito *Anopheles gambiae*: targeted disruption of the Defensin gene. *EMBO Rep*, 3: 852–856.

Blow, D. M. (1997). The tortuous story of Asp...His...Ser: structural analysis of α-chymotrypsin. *Trends Biochem, Sci*, 22 (1997): 405–408.

Boggs, R. T., Gregor, P., Idriss, S., Belote, J. M., and McKeown, M. (1987). Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the *transformer* gene. *Cell,* 50: 739-747.

Brengues, J., and Coz, J. (1973). Quelques aspects fondamentaux de la biologie *d'Anopheles gambiae Giles* (Sp. A.) et *d'Anopheles funestus Giles*, en zone de savane humide d'Afrique de l'Ouest. Cah. O.R.S.T.O.M., Sér. *Ent. méd. et Parasitol,* 11: 107-126.

Bryan, J. H. (1972). (Further studies on consecutive matings in the *Anopheles gambiae* complex. *Nature,* 239: 519–520.

Calogero, A., Hall, J., Fishel, S. *et al.* (1996). Effects of g-aminobutyric acid on human sperm motility and hyperactivation. *Mol. Human Reprod,* 2: 733–738.

Carvalho, G. B., Kapahi, P., Anderson, D. J., Benzer, S. (2006). Allocrine modulation of feeding behavior by the Sex Peptide of *Drosophila*. *Curr. Biol.* 16:692–96.

Carver, T. J., *et al*. (2005). ACT: the Artemis Comparison Tool. *Bioinformatics,* 21: 3422-3423.

Catteruccia, F., Crisanti, A., & Wimmer, E. A. (2009). Transgenic techniques to induce sterility. *Malaria Journal*, 8(Suppl 2):S7.

Catteruccia, F., Nolan, T., Loukeris, T., Blass, C., Savakis, C., Kafatos, F., & Crisanti. A. (2000). Stable germline transformation of the malaria mosquito *Anopheles stephensi*. *Nature*, 405: 959-962.

Chambers, G. M., & Klowden, M. J. (2001). Age of *Anopheles gambiae* Giles male mosquitoes at the time of mating influences female oviposition. *J Vector Ecol,* 26: 196–201.

Chapman, T., & Davies, S.J., (2004) Functions and analysis of the seminal fluid proteins of *Drosophila melanogaster* fruitflies. *Peptides* 25: 1477–1490.

Charlwood, J. D., & Jones, M. D. R. (1979). Mating behaviour in the mosquito, *Anopheles gambiae s.l.* I. Close range and contact behaviour. *Physiol Entomol,* 4: 111–120.

Charlwood, J. D., & Jones, M. D. R. (1980). Mating in the mosquito, *Anopheles gambiae*. s.l. II. Swarming behaviour. *Physiol Entomol*, 5: 315–320.

Charlwood, J. D., Pinto, J., Sousa, C. A., Ferreira, C., & Do Rosario, V. E. (2002). Male size does not affect mating success (of *Anopheles gambiae* in Sao Tome) *Med Vet Entomol,* 16: 109–111.

Charlwood, J. D., Thompson, R., & Madsen, H. (2003). Observations on the swarming and mating behaviour of*Anopheles funestus* from southern Mozambique. *Malar J,* 2: 2.

Chen, B. E., Kondo, M., Garnier, A., Watson, F. L., Puettmann-Holgado, R., Lamar, D. R., & Schmucker, D. (2006). The molecular diversity of Dscam is functionally required for neuronal wiring specificity in Drosophila. *Cell,* 125(3): 607--620 .

Cheng, J., Baldi, P. (2005). Three-stage prediction of protein beta-sheets by neural networks, alignments and graph algorithms. *Bioinformatics,* 21(Suppl. 1): i75-i84.

Chertemps, T., Adrien, F., Durand, N., Rosell, G., Dekker, T., Philippe L., Maibeche-Coisne, M. (2012). A carboxylesterase, Esterase-6, modulates sensory physiological and behavioural response dynamics to pheromone in Drosophila. *BMC Biology*. 10: 56.

Chevone, B. I., Richards, A. G. (1976). Ultrastructure of the atypic muscles associated with terminalial inversion in male *Aedes aegypti* (L) *Biol Bul,* 151: 283–296.

Chin, C-S., and Manoj, P. S. (2003). Global snapshot of a protein interaction network—a percolation based approach. *Bioinformatics,*19: 2413-2419.

Chintapalli, V. R., Wang, J., Dow, J. A. T. (2007). Using FlyAtlas to identify better Drosophila models of human disease. *Nat. Genet,* 39: 715–720.

Chizema-Kawesha, E., Miller, J. M., Steketee, R. W., Mukonka, V. M., Mukuka, C., Mohamed, A. D., Miti, S. K., Campbell, C. C. (2010). Scaling up malaria control in Zambia: progress and impact 2005-2008. *Am J Trop Med Hyg*, 83: 480-488.

Christie, M. (*1958)*. Predation on larvae of *Anopheles gambiae*. Giles. *Journal of Tropical Medicine and Hygene,* 61: 168-176.

Clements, A. N. (1999). The Biology of mosquitoes Sensory Reception and Behavior. *Vol. 2. London: CABI.*

Cline, T.W. (1993). The *Drosophila* sex determination signal: how do flies count to two?. *Trends Genet*, 9: 385–390.

Coetzee, M. (2000). Craig, M., le Sueur D. Distribution of African malaria mosquitoes belonging to the *Anopheles gambiae* complex. *Parasitol Today*. 16:74–77.

Coetzee, M. (2004). Distribution of the African malaria vectors of the *Anopheles gambiae* complex. *Am J Trop Med Hyg*. 70:103–104.

Collins, F.H., & *Paskewitz,* S.M. (1995). Malaria: Current and future prospects for control. *Annual Review of Entomology*, 40: 195-219.

Collins, F.H., Petrarca, V., Mpofu, S., Brandling-Benneth, A.D., Were, J.B., Rasmusssen, M.O., & Finerty, V. (1988). Comparison of DNA-Probe and cytogenetic methods for identifying field collected *Anopheles gambiae* complex mosquitoes. *American Journal of Tropical Medecine and Hygiene,* 39: 545-550.

Coluzzi, M., Petrarca, V., & Di Deco, M.A. (1985). Chromosomal inversion integration and incipient speciation in *Anopheles gambiae. Bulletin of Zoology,* 52: 45-63.

Coluzzi, M., Sabatini, A., Petrarca V., & Di Deco, M. A. (1979). Chromosomal differentiation and adaptation to human environment in the *Anopheles gambiae* complex. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 37: 483-497.

Crawford, D. L., Segal, J. A., and Barnett, J. L. (1999). Evolutionary analysis of TATA-less proximal promoter function. *Mol. Biol. Evol,* 16: 194-207.

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., Hermjakob, H., D'Eustachio, P., *and* Stein, L. *(*2011). Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Res,* 39: D691-D697.

Curtis, C. F., Miller, J. E., Hodjati, M. H., KolacZinski, J. H., & Kasumba, I. (1998) Can anything be done to maintain the effectiveness of pyrethroid-impregnated bednets against malaria vectors? *Philosophical Transactions of the Royal Society, Series B: Biological Sciences.* 353: 1769-1775.

Dao, A., Adamou, A., Yaro, A. S., Alpha, S.Y., Hamidou, M.M., Kassogue, Y., Traore, S. F., & Lehmann, T. (2008). Assessment of alternative mating strategies in *Anopheles gambiae*: Does mating occur indoors? *J Med Entomol,* 45: 643–652.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood for incomplete data via the EM algorithm (with Discussion). *Journal of the Royal Statistical Society*, Series B39, 1-38.

Detinova, T. S., & Gillies, M.T. (1964). Observations on the determination of the age composition and epidemiological importance of populations of *Anopheles gambiae Giles* and *Anopheles funestus Giles* in Tanganyika. *Bull. WHO,* 30: 23-28.

Devoto, A., Hartmann, H. A., Piffanelli, P., Elliott, C., Simmons, C., Taramino, G., Goh, C. S., Cohen, F. E., Emerson, B. C., Schulze-Lefert, P., Panstruga, R. (2003). Molecular phylogeny and evolution of the plant-specific seven-transmembrane MLO family. *J Mol Evol,* 56(1): 77-88.

Diabate, A., Baldet, T., Brengues, C., Kengne, P., Dabire, K. R., Simard, F., Chandre, F., Hougard, J.M., Hemingway, J., Ouedraogo, J. B., & Fontenille, D. (2003). Natural swarming behaviour of the molecular M form of *Anopheles gambiae*. *Trans R Soc Trop Med Hyg*, 97: 713–716.

Diabate, A., Dabire, R. K., Kengne, P., Brengues, C., Baldet, T., Ouari, A., Simard, F., & Lehmann, T. (2006). Mixed swarms of the molecular M and S forms of *Anopheles gambiae* (Diptera: Culcidae) in sympatric area from Burkina Faso. *J Med Entomol*, 43: 480–483.

Diebold, R., Schuster, J., Däschner, K., and Binder S. (2002). The Branched Chain Amino Acid Transaminase Gene Family in Arabidopsis Encodes Plastid and Mitochondrial Proteins. *Plant Physiology*, 129 (2): 540-550.

Dottorini, T., Nicolaides, L., Ranson, H., Rogers, D. W., Crisanti, A., & Catteruccia, F. (2007). A genome-wide analysis in *Anopheles gambiae* mosquitoes reveals 46 male accessory gland genes, possible modulators of female behavior. *Proc Natl Acad Sci USA*, 104: 16215–16220.

Downes, J. A. (1968). The swarming and mating flight of Diptera. *Annu Rev Entomol*, 14: 271–298.

Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., *et al*. (2010) Geneious v5.5, Available: http://www.geneious.com. Accessed 2010.

Dyson, M. H., Thomson. S., Inagaki, M., Goto, H., Arthur, S. J., Nightingale, K., Iborra, F. J., Mahadevan, L. C. (2005). MAP kinase-mediated phosphorylation of distinct pools of histone H3 at S10 or S28 via mitogen- and stress-activated kinase 1/2. *J Cell Sci* 118(10): 2247-2259.

Eggert, A. K., & Sakaluk, S. K. (1994). Sexual cannibalism and its relation to male mating success in sagebrush crickets, *Cyphoderris strepitans* (Haglidae: Orthoptera). *Animal Behaviour*, 47: 1171-1177.

Fay, W. P., Shapiro, A. D., Shih, J. L., Schleef, R. R., and Ginsburg, D. (1992). Complete deficiency of plasminogen-activator inhibitor type 1 due to a frame-shift mutation. *New Eng. J. Med*, 327: 1729-1733.

Felsenstein, J. (2004). Inferring Phylogenies. *Sinauer Associates*: Sunderland, MA.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S.E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391: 806–811.

Foley, D.H., Bryan, J.H., Yeates, D., & Saul, A. (1998). Evolution and systematics of *Anopheles*: Insights from a molecular phylogeny of Australasian mosquitoes. *Molecular Phylogeny and Evolution*, 9: 262-275.

Forrester, W. C., Takegawa, S., Papayannopoulou, T., and Groudine, M. (1987). Evidence for a locus activation region: The formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res*, 15: 10159-10177.

Garver, L., and Dimopoulos, G. (2007) Protocol for RNAi assays in adult mosquitoes (*A. gambiaeAn. gambiae*). *J Vis Exp*, 5: 230.

Gerber, G. H. (1970). Evolution of the methods of spermatophore formation in pterygotan insects. *Canadian Entomologist journal*,102: 358–362.

Giglioli, M. E. C., Mason, G. F. (1966). The mating plug of anopheline mosquitoes. *Proc R Ent Soc Lond*, 41: 123–129.

Gillies, M. T. (1955). The pre-gravid phase of ovarian development in *Anopheles funestus*. *Annual reviews of Tropical Medecine and Parasitology*, 49: 320-325.

Gillies, M. T. (1956). A new character for the recognition of nulliparous females of *Anopheles gambiae*. *B World Health Organ*, 15: 451–459.

Gillies, M. T. (1972). Some aspects of mosquito behaviour in relation to transmission of parasites. *Zoological Journal of Linnean Society*, 51: 69-81.

Gillies, M.T. (1954). The recognition of age-groups within populations of Anopheles gambiae by the pre-gravid rate and the sporozoite rate. *Ann. Trop. Med. Parasitol*, 48: 58-74.

Gilliesm, M. T., & Coetzee, M. (1987). A supplement to anophelinae of Africa south of Sahara (Afro-tropical region). *Publication of the South Africa Institute of Medical Research*, 55: 1-143.

Githeko, A. K., Service, M. W., Atieli, F. K., Hill, S. M., Crampton, J. M. (1993). Field Testing an Enzyme-Linked Synthetic Oligonucleotide Probe for Identification of *Anopheles gambiae* ss and *An. arabiensis*. *Annals of Tropical Medicine and Parasitology*, 87: 595-601.

Githeko, A.K., Service, M.W., Mbogo, C.M., Atieli, F.K., & Juma, O. (1994). Origin of blood meals in door and out-door resting malaria vectors in Western Kenya. *Acta Tropica*, 58: 307-316.

Grigoriev, A. (2001). A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast Saccharomyces cerevisiae. *Nucleic Acids Res*. 29: 3513–3519.

Grossman, G., Rafferty, C., Clayton, J., Stevens, T., Mukabayire, O., & Benedict, M. (2001). Germline transformation of the malaria vector, *Anopheles gambiae*, with the *piggyBac* transposable element. *Insect Molecular Biology.* 10: 597-604.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*,52: 696-704.

Haddow, A. J. (1942). The mosquito fauna and climate of native huts at Kisumu, Kenya. *Bulletin of Entomological Research*, 33: 91-142.

Han, B. G., Cho, J. W., Cho, Y. D., Jeong, K. C., Kim, S. Y., & Lee, B. I. (2010). Crystal structure of human transglutaminase 2 in complex with adenosine triphosphate. *Int J Biol Macromol*, 47: 190–195.

Hernandez-Toro, J., Prieto, C., De las Rivas, J. (2007). APID2NET: unified interactome graphic analyzer. *Bioinformatics* 23: 2495–2497.

Highton, R.B., Bryan, H., Boreham, P.F.L., & Chandler, J.A. (1979). Studies on the sibling species *Anopheles gambiae* Giles and *An. arabiensis* Patton (Diptera : Culicidae) in the Kisumu area, Kenya. *Bulletin of Entomological Resolutions*, 69: 43-53.

Holt, B. F., 3rd, Boyes, D. C., Ellerstrom, M., Siefers, N., Wiig, A., Kauffman, S., Grant, M. R., & Dangl, J. L. (2002). An evolutionarily conserved mediator of plant disease resistance gene function is required for normal Arabidopsis development. *Dev. Cell* 2: 807-817.

Huang, X., Suyama, K., Buchanan, J.A., Zhu, A.J., & Scott, M.P. (2005). A Drosophila model of the Niemann-Pick type C lysosome storage disease: dnpc1a is required for molting and sterol homeostasis. *Development*, 132(22): 5115--5124.

Huho, B. J., Ng'habi, K. R., Killeen, G. F., Nkwengulila, G., Knols, B. G., Ferguson, H. M. (2006). A reliable morphological method to assess the age of male *Anopheles gambiae***.** *Malar J*, 5: 62.

Jaenisch, T., Sullivan, D. J., Dutta, A., Deb, S., Ramsan, M., Othman, M. K., Gaczkowski, R., Tielsch, J., Sazawal, S. (2010). Malaria incidence and prevalence on Pemba Island before the onset of the successful control intervention on the Zanzibar archipelago. *Malar J*, 9: 32.

Jeong, H., Mason, S. P., Barabasi, A. L., and Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, 411: 41–42.

Joshi, G. P., Service, M. W., & Pradhan, G. D. (1975). A survey of Species A and B of the *Anopheles gambiae* Giles complex in the Kisumu area of Kenya prior to insecticidal spraying with OMS-43 (fenitrothion). *Annual reviews of Tropical Medecine and Parasitology,* 69: 91-104.

Jukes, T.H., and Cantor, C. R. (1969). Evolution of protein molecules. In Munro.H.N. (ed.) *Mammalian Protein Metabolism,* Academic Press, New York, 21-123.

Kabiru, E.W., Mbogo, C.M., Muiruri, S.K., Ouma, J.H., Githure, J.I., and Beier, J.C. (1997). Sporozoite loads of naturally infected *Anopheles* in Kilifi District, Kenya. *Journal of American Mosquito Control Association,* 13: 259-262.

Kalb, J. M., Dibenedetto, A. J., and Wolfner, M. F. (1993). Probing the function of *Drosophila melanogaster* accessory glands by directed cell ablation. *Proc. Natl. Acad. Sci*. USA 90**:** 8093–8097.

Kawabata, T., and Nishikawa, K. (2000). Protein structure comparison using the Markov transition model of evolution. *Proteins,* 41:108-122.

Kazutaka, K., and Hiroyuki, T. (2010). Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics*, 26 (15): 1899-1900.

Keedy, D. A., Williams, C. J., Headd, J. J., Arendall, W. B., 3rd, Chen, V. B., Kapral, G. J., Gillespie, R. A., Block, J. N., Zemla, A., Richardson, D. C., and Richardson, J. S. (2009). The other 90% of the protein: assessment beyond the Calphas for CASP8 template-based and high accuracy models. *Proteins* 77 (Suppl 9), 29–49.

Kelly, R. W. (1995) Immunosuppressive mechanisms in semen: implications for contraception. *Hum Reprod* 10: 1686-1693.

Kettle, D. S. (1992). *Medical and Veterinary Entomology*. Centre for Agriculture and Biosciences International. *Wallingford, UK*, pp 99-136.

Khorasani, G. R., Sauer, G. R., Ozimek, L., and Kennelly, J. J. (1990). Digestion of soybean meal and canola meal protein and amino acids in the digestive tract of young ruminants. *J. Anim. Sci.* 68: 3241–3248.

Kim, W., Koo, H., Richman, A. M., Seeley, D., Vizioli, J., Klocko, A. D., & O'Brochta, D. A. (2004). Ectopic expression of a cecropin transgene in the human malaria vector mosquito Anopheles gambiae (Diptera: Culicidae): effects on susceptibility to Plasmodium. *J Med Entomol,* 41**:** 447-55.

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences.*J. Mol. Evol.,* 16, 111-120.

Klowden, M. J. (2001). Sexual receptivity in *Anopheles gambiae* mosquitoes: absence of control by male accessory gland substances. *J Insect Physiol*, 47: 661–666 .

Knipling, E. F. (1955). Possibilities of insect control or eradication through the use of sexually sterile males. *Journal of Economic Entomology,* 48: 459-467.

Knipling, E. F., Laven, H., Craig, G. B., Pal, R., Kitzmiller, J. B., Smith, C. N., & C. N. Brown, C. N. (1968). Genetic control of insects of public health importance. *Bull. World Health Organ. 38,* 421-438.

Knipling, E. F. (1959). Screwworm eradication: concepts and research leading to the sterile-male method. *In: Smithsonian Report for 1958*. Smithsonian Institution, *Washington.* 409-418.

Krzywinski, J., & Besansky, N. J. (2003). Molecular systematics of *Anopheles*: From Subgenera to Subpopulations. *Annual Reviews of Entomology,* 48: 111-139.

Kubli, E. (2003). Sex-peptides: seminal peptides of the Drosophila male. Cell. Molec. Life Sci, 60(8): 1689-1704.

Latchman, D. S. (1997). Transcription factors: an overview. *Int J Biochem Cell Biol. 29,* 1305-12.

Lawson, D., Arensburger, P., Atkinson, P., Besansky, N. J., Bruggner, R.V., Butler, R., *et al.* (2009). VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Res* 37: D583–D587.

Levashina, E. A., Moita, L. F., Blandin, S., Vriend, G., Lagueux, M., and Kafatos, F. C. (2001). Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. *Cell* 104: 709-718.

Librado, P., & Rozas, J. **(**2009**).** *DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics*, 25*:* 1451–1452.

Lin, C.Y., Chin, C.H., Wu, H.H., Chen, S.H., Ho, C.W., and Ko, M.T. (2008) Hubba: hub objects analyzer—a framework of interactome hubs identification for network biology. *Nucleic Acids Res*, 36: W438-W443.

Lindh, J. M., Borg-Karlson, A. K., & Faye, I. (2008). Transstadial and horizontal transfer of bacteria within a colony of *Anopheles gambiae* (Diptera: Culicidae) and oviposition response to bacteria-containing water. *Acta Trop*, 242-250.

Lindsay, S., Ansell, J., Selman, C., Cox, V., Hamilton, K., & Walraven, G. (2000). Effect of pregnancy on exposure to malaria mosquitoes. *The Lancet*, 335: 1972.

Lyimo, E. O., & Takken, W. 1993. Effects of adult body size on fecundity and the pregravid rate of *Anopheles gambiae* females in Tanzania. *Med. Vet. Entomol.* 7: 328-332.

Macdonald, G. (1957). The epidemiology and control of malaria. London, Oxford University press.

Magrane, M., and UniProt Consortium. (2011). UniProt Knowledgebase: a hub of integrated protein data. The Journal of Biological Databases and Curation.

Mahmood, F. (1997). Age-related changes in development of the accessory glands of male *Anopheles albimanus*. *J Am Mosq Control Assoc,13*, 35–39.

Mahmood, F., & Reisen, W. K. (1982). *Anopheles stephensi* (Diptera: Culicidae): changes in male mating competence and reproductive system morphology associated with aging and mating. *J Med Entomol*, 19: 573–588.

Mahmood, F., & Reisen, W. K. (1982). *Anopheles stephensi* (Diptera: Culicidae): changes in male mating competence and reproductive system morphology associated with aging and mating. *J Med Entomol*, 19: 573–588.

Mahmood, F., & Reisen, W. K. (1994). *Anopheles culicifacies*: effects of age on the male reproductive system and mating ability of virgin adult mosquitoes. *Med Vet Entomol*, 8: 31–37.

Marchand, R. P. (1984). Field observations on swarming and mating in *Anopheles gambiae* mosquitoes in Tanzania. *Neth J Zool, 34*, 367–387.

Marshall, J. F. (1938). The British mosquitoes. London: British Museum (Natural History).

Mbogo, C. M., Snow, R.W., Kabiru, E., Ouma, J. H., Githure, J. I., Marsh, K. & Beier, J. (1993). Low-level *Plasmodium falciparum* transmission and the incidence of severe malaria infection on the Kenyan Coast. *American Journal of Tropical Medicine,* 49: 245-253.

Mbogo, C. N. M., Snow, R.W., Khamala, C. P. M., Kabiru, E. W., Ouma, J. H., Githure, J. I., Marsh, K., & Beier, J. C. (1995). Relationships between *Plasmodium falciparum* transmission by vector populations and the incidence of severe disease at nine sites on the Kenyan coast. *American Journal of Tropical Hygiene*, 52: 201-206.

McCrae, A. W. R. (1983). Oviposition by African malaria vector mosquitoes I. Temporal activity pattern of caged, wild-caught, freshwater *Anopheles gambiae* Giles sensu lato. *Ann Trop Med Parasitol*, 77: 615–625.

McCrae, A. W. R. (1984). Oviposition by African malaria vector mosquitoes II. Effects of site tone, water type and conspecific immatures on target selection by freshwater *Anopheles gambiae* Giles sensu lato. *Ann Trop Med Parasitol,* 78: 307–318.

McQuilton, P., St. Pierre, S. E., Thurmond, J and the FlyBase Consortium. (2011). FlyBase 101 – the basics of navigating FlyBase. *Bioinformatics*, 40, Issue D1: Pp. D706-D714.

Meyrowitsch, W. D., Pedersen, M. E., Alifrangis, M., Scheike, H. T., Malecela, N. M., Magesa, M. S., Derua, A. Y., Rwegoshora, T. R., Michael, E., and Simonsen, E. P. (2011). Is the current decline in malaria burden in sub-Saharan Africa due to a decrease in vector population? *Malaria Journal*, **10**:188.

Molyneux, D., Floyd, K., Barnish, G., & Fevre, E. M. (1999). Transmission control and drug resistance in malaria: a crucial interaction. *Parasitology Today,* 15: 238–240.

Moreno, M., Cano, J., Nzambo, S., Bobuakasi, L., Buatiche, J. N., Ondo, M., Micha, F., Benito, A. (2004). Malaria Panel Assay versus PCR: detection of naturally infected *Anopheles melas* in a coastal village of Equatorial Guinea. *Malar J.* 4: 3.

Mosha, F.W., & Petraca, V. (1983). Ecological studies on *Anopheles gambiae* complex on the Kenya coast. *Transactions of the Royal Society of Tropical Medecine and Hygiene,* 77: 344-345.

Mueller, J. L., Linklater, J. R., RaviRam, K., Chapman, T., Wolfner, M. F. (2008). Targeted gene deletion and phenotypic analysis of the *Drosophila melanogaster* seminal fluid protease inhibitor Acp62F. *Genetics*, 178: 1605–14.

Muirhead-Thomson, R. C. (1951). The distribution of Anophelinae mosquito bites among different age groups- a new factor in malaria epidemiology. *British Medical Journal*, 1: 1114-1123.

Munga, S., Minakawa, N., Zhou, G., Barrack, O. J., Githeko, A. K., & Yan G. (2005). Oviposition site preference and egg hatchability of Anopheles gambiae: Effects of land cover types. *J Med Entomol*, 42: 993–997.

Mutero, A., Pralavorio, M., Bride, J.M., *et al*. (1984). Resistance associated point mutations in insecticide-insensitive acetylcholinesterase. *Proceedings of the National Academy of Sciences of the United States of America*, 91: 5922-5926.

Nguyen, D. M., Tran, D. H., Harbach, R. E., Elphick, J., & Linton, Y. M. (2000). A new species of the Hyrcanus Group of *Anopheles*, subgenus *Anopheles*, a secondary vector of malaria in coastal areas of southern Vietnam. *Journal of the American Mosquito Control Association.* 16: 189–198.

Niels, T., Hans, B., and Volker, K. (2002). Cathepsin B cleavage of the trypsinogen activation peptide. *BMC Gastroenterology*. 2, 16.

Nijhout, H. F., & Sheffield, H. G. (1979). Antennal hair erection in male mosquitoes: a new mechanical effector in insects. *Science,* 206, 595–596.

Omar, A. H., & Shaurub, E. H. (1987). Mating behaviour and insemination in *Anopheles multicolour* Cambouliu. *Journal of the Egyptian Society of Parasitology*, 17: 381–387.

O'Meara, W. P., Bejon, P., Mwangi, T. W., Okiro, E. A., Peshu, N., Snow, R.W., Marsh, K. (2008). Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya. *Lancet*, 372: 1555-1562.

O'Meara, W. P., Mangeni, J. N., Steketee, R., Greenwood, B. (2010). Changes in the burden of malaria in sub-Saharan Africa. *Lancet Infect Dis,* 10**:** 545-555.

Otienoburu, P. E., Bayoh, N., Gimnig, J., Huang, J., Walker, E. D., Otieno, M. F., Vulule J., & Miller, J. R. (2007). Anopheles gambiae (Diptera: Culicidae) oviposition as influenced by type of water infused into black and red soils of western Kenya. *International Journal of Tropical Insect Science,* 27: 2-5.

Ozvegy-LacZka, C., Va´rady, G., Koblos, G., Ujhelly, O., Cervenak, J., Schuetz, J. D., Sorrentino, B. P., Koomen, G. J., Varadi, A., Nemet, K., Sarkadi, B. (2005). Function-dependent conformational changes of the ABCG2 multidrug transporter modify its interaction with a monoclonal antibody on the cell surface. *J Biol Chem*, 280: 4219–4227.

Papathanos, P.A., Bossin, H.C., Benedict, M.Q., Catteruccia, F., Malcolm, C.A., Alphey, L., & Crisanti, A. (2009). Sex separation strategies: past experience and new approaches. *Malaria Journal*, 8(Suppl 2):S5.

Parham, P. E., Michael, E. (2010). Modeling the effects of weather and climate change on malaria transmission. *Environ Health Perspect*, 118**:** 620-626.

Parker, G. A. (1970). Sperm competition and its evolutionary consequences in the insects. *Biological Reviews, 45,* 525–567.

Pavlopoulos, G. A., Secrier, M., Moschopoulos, C. N., Soldatos, T.G., Kossida, S., Aerts, J., Schneider, R., Bagos, P. G. (2011). Using graph theory to analyze biological networks. *BioData Mining,* 4:10.

Perera, O. P., Harrell R. A. I., & Handler A. M. (2002). Germ-line transformation of the South American malaria vector, *Anopheles albimanus*, with a *piggyBac*/EGFP transposon vector is routine and highly efficient. *Insect Mol. Biol 11,* 291–297.

Petrarca, V., & Beier, J.C. (1992). Intraspecific chromosomal polymorphism in the *Anopheles gambiae* complex as a factor affecting malaria transmission in the Kisumu area of Kenya. *American Journal of Tropical Medecine and Hygiene, 46,* 229-237.

Petrarca, V., Beier, J. C., Onyango F., Asiago, C., Koech, D. K., & Roberts, C. R. (1991). Species composition of the *Anopheles gambiae* complex (Diptera : Culicidae) at two sites in western Kenya. *Journal of Medical Entomology, 28,* 307-313.

Ranson, H., Jensen, B., Vulule, J. M., Wang, X., Hemingway, J., Collins, F. H. (2000). Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan *Anopheles gambiae* associated with resistance to DDT and pyrethroids. *Insect Mol Biol*, 9: 491-497.

Ravi, R. K., & Wolfner, M. F. (2007). Seminal influences: Drosophila Acps and the molecular interplay between males and females during reproduction. *Integrative and Comparative Biology*, 47: 427–445.

Rees, D. M., Onishi, K. (1951). Morphology of the terminalia and internal reproductive organs, and copulation in the mosquito, *Culiseta inornata* (Williston) (Diptera, Culicidae) *Proceedings of the Entomological Society of Washington. 53,* 233–246.

Reisen, W. K., & Aslamkhan, M. (1976). Observations on the swarming and mating behaviour of *Anopheles culicifacies* Giles in nature. *Bull World Health Organ, 54,* 155–158.

Roderic, D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, 12: 357-358

Rogers, D. J., Randolph, S. E. (2006). Climate change and vector-borne diseases. *Adv Parasitol*, 62**:** 345-381.

Rogers, D. W., Baldini, F., Battaglia, F., Panico, M., Dell, A., Morris, H. R., Catteruccia, F. (2009). Transglutaminase-mediated semen coagulation controls sperm storage in the malaria mosquito. *PLoS Biol* **7**: e1000272.

Rogers, W. D., Whitten, M. M. A., Thailayil, J., Soichot, J., Levashina, E., & Catteruccia F. (2008). Molecular and cellular components of the mating machinery in *Anopheles gambiae* females. *Proc Natl Acad Sci USA, 105,* 19389–19394.

Ross, H. H. (1952). Conflict with *Culex*. *Mosquito News,* 11,128–32.

Rost, B., Yachdav, G., and Liu, J. (2004). The PredictProtein server. Nucleic Acids Research, 32, Issue suppl 2: Pp. W321-W326.

Roth, L. M. (1948). A study of mosquito behavior. An experimental laboratory study of the sexual behavior of *Aedes aegypti* (Linnaeus) *Am Midl Nat. 40*, 265–352.

Sakaluk, S. K., & Eggert, A.K. (1996). Female control of sperm transfer and intraspecific variation in sperm precedence: antecedents to the evolution of a courtship food gift. *Evolution, 50*, 694-703.

Sandhu, K. S., and Dash, D. (2007). Dynamic a-Helices: Conformations that Do Not Conform. *PROTEINS: Structure, Function, and Bioinformatics* 68:109–122.

Sarov-Blat, L., So, W.V., Liu, L., Rosbash, M. (2000). The Drosophila takeout gene is a novel molecular link between circadian rhythms and feeding behavior. *Cell* 101(6): 647-656.

Scali, C., Catteruccia, F., Li, Q., Crisanti, A. (2005). Identification of sexspecific transcripts of the *Anopheles gambiae* doublesex gene. *J. Exp. Biol. 208*, 3701–3709

Schuster-Bockler, B., Bateman, A. (2007). Reuse of structural domain-domain interactions in protein networks. *BMC Bioinformatics. 8*:259.

Segal, J. A., Barnett, J. L., and Crawford, D. L. (1999). Functional analysis of natural variation in Sp1 binding sites of a TATA-less promoter. *J. Mol. Evol*, 49: 736-749.

Service, M. W. (1973). Identification of predators of Anopheles gambiae resting in huts, by the precipitin test. *Trans R Soc Trop Med Hyg, 67*, 33–34.

Service, M.W. (1980). A Guide to Medical Entomology. Macmillan, New York 226 *pp*. 204-210

Sharpe, R. M. (1994), Regulation of spermatogenesis. In. Knobil, E.; Neil, J. D. (Eds.). *The physiology of reproduction*. 2. ed. *New York: Raven Press. pp.* 1363-1434.

Sigrist, C. J. A., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L., Xenarios, I. (2012). New and continuing developments at PROSITE. *Nucleic Acids Res*, 41(D1), D344–D347.

Steele, R.H. (1986). Courtship feeding in *Drosophila subobscura*. II. Courtship feeding by males influences female mate choice. *Animal Behaviour, 34*, 1099-1108.

Stephens, M., *et al*. (2001). A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet*. 68: 978-989.

Steve Rozen., and Helen J. Skaletsky. (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.

Stresman, G. H. (2010). Beyond temperature and precipitation: Ecological risk factors that modify malaria transmission**.** *Acta Trop*, 116**:** 167-172.

Strode, C., Steen, K., Ortelli, F. and Ranson, H. (2006). Differential expression of the detoxification genes in the different life stages of the malaria vector*Anopheles gambiae. Insect Molecular Biology*. 15(4): 523–530.

Szklarczyk, D., *et al*. (2010). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*. 39: D561-D568.

Thailayil, J. (2011). Analysing the molecular basis of postmating behaviours in *Anopheles* mosquitoes. PhD Thesis Imperial College London.Natural Sciences.

Thomas, D. D., Donnelly, C. A., Wood, R. J., and Alphey, L. S. (2000). Insect population control using a dominant, repressible, lethal genetic system. *Science 287*, 2474–2476.

Thornhill, R. (1976). Sexual selection and nuptial feeding behavior in *Bittacus apicalis* (Insecta: Mecoptera). *American Naturalist*, 110: 529-548.

Touré, Y. T., Oduola, A. M., Morel, C. M. (2004). The *Anopheles gambiae* genome: next steps for malaria vector control. *Trends Parasitol*, 20: 142-149.

Toure´, Y. T., Petrarca, V., Coulibaly, A., Maı¨ga H. M. *et al.* (1994). Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae s. str.* in Mali, West Africa. *Genetica*, 94: 213–223.

Tress, M., Tai, C. H., Wang, G., Ezkurdia, I., Lopez, G., Valencia, A., Lee, B., and Dunbrack, R. L., Jr. (2005). Domain definition and target classification for CASP6. *Proteins* 61 (Suppl 7), 8–18.

Tripet, F., Touré, Y. T., Dolo, G. & Lanzaro, G. C. (2003). Frequency of multiple inseminations in field-collected *Anopheles gambiae* females revealed by DNA analysis of transferred sperm. *Am. J. Trop. Med. Hyg. 68*1-5.

Tsy, J., Duchemin, J. B., Marrama, L., Rabarison, P., Le Goff, G., Rajaonarivelo, V., Robert, V. (2003). Distribution of the species of the *Anopheles gambiae* complex and first evidence of *Anopheles merus* as a malaria vector in Madagascar. *Malar J.* 2.

Tuan, D., Solomon, W., Li, Q., and London, I. M. (1985). The "Blike globin" gene domain in human erythroid cells. *Proc. Natl. Acad. Sci. 82,* 6384-6388.

Uversky, V. N. (2002). What does it mean to be natively unfolded? *Eur J Biochem,* 269: 2–12.

Van Bortel, W., Harbach, R. E., Trung, H. D., Roelants, P., Backeljau, T., & Cooseman, M. (2001) Confirmation of *Anopheles varuna* in Vietnam, previously misidentified and mistargeted as the malaria vector *Anopheles minimus*. *American Journal of Tropical Medicine and Hygiene, 65,* 729-32.

Verhoek, B. A., & Takken, W. (1994). Age effects on the insemination rate of *Anopheles gambiae s.l.* in the laboratory. *Entomol Exp Appl, 72,*167–172.

Wagbatsoma, V. A., & Ogbeide, O. (1995). Towards malaria control in Nigeria: A qualitative study on the population of mosquitoes. *J R Soc Heal, 115,* 363–365.

Walker, M. J., Rylett, C. M., Keen, J. N., Audsley, N., Sajid, M., Shirras, A. D. and Isaac, R. E. (2006). Proteomic identification of *Drosophila melanogaster* male accessory gland proteins, including a pro-cathepsin and a soluble γ-glutamyl transpeptidase. *Proteome Science*, 4: 9.

Ward, J. J., Sodhi, J. S., McGuffin, L. J, Buxton, B. F., Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol*, 337: 635–645.

Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol*, 18: 691–699.

White, G. B. (1974). *Anopheles gambiae* complex and disease transmission in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 68: 278-301.

White, G. B. (1985). *Anopheles bwambae* sp. n., a malaria vector in Semliki Valley, Uganda, and its relationships with other sibling species of the *An. gambiae* complex. *Syst Entomol.*10:501–522.

Wolfner, M. F., H. A. Harada, M. J., Bertram, T. J., Stelick, K. W., Kraus *et al.* (1997). New genes for male accessory gland proteins in *Drosophila melanogaster. Insect Biochem.Mol. Biol.* 27: 825–834.

Wong, A., Shannon, N., Albright, J. D., Giebel, K., Ravi, R., Shuqing Ji, Fiumera, A. C., and Wolfner, M. F. (2008). A Role for Acp29AB, a Predicted Seminal Fluid Lectin, in Female Sperm Storage in *Drosophila melanogaster. Genetics*, 180: 921-931.

World Health Organization. (2009). *WHO World Malaria Report*. Geneva, Switzerland.

World Health Organization. (2010). *Roll Back Malaria: Key malaria facts*. Geneva, Switzerland.

YANG, Z. (1994*a*). Estimating the pattern of nucleotide substitution. *J. Mol. Evol*. 39:105–111.

YANG, Z. (1994*b*). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol*. 39: 306–314.

Yuval, B. (2006). Mating systems of blood-feeding flies. *Annu Rev Entomol, 51,* 413–440.

Yuval, B., Wekesa, J. W., & Washino, R. K. (1993). Effect of body size on swarming behaviour and mating success of male *Anopheles freeborni* (Diptera: Culicidae) *J Insect Behav,* 6:155–158.

Zahler, A. M. (2005). Alternative splicing in C. elegans. WormBook, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.31.1.

CHAPTER 8

APPENDICES

I) Proteinase K method for DNA extraction protocol

1. In 2 ml Eppendorf Safe-Lock tube with mechanically disrupted animal or plant tissues (seeds, leaves or herbarium) add fresh 1 ml of grinding buffer (see below) with 200 µg of proteinase K, vortex very well and incubate the samples at 50°-60°C for several hours or better overnight at 37°-55°C (the longer the better, until dissolve tissue) with occasional vortexing.
2. Add 600 µl of chloroform, vortex very well for 2 minute creating an emulsion (in the MM300 Mixer Mill at 30 Hz) and incubate the samples at 65°C for one hour.
3. Spin at maximum speed in a microcentrifuge for 5 minutes.
4. Transfer the supernatant into a new 2 ml tube containing 800 µl of 2-propanol and 100 µl 3M Na-acetate, vortex very well, and centrifuge the tubes at maximum speed in a microcentrifuge for 4 minutes.
5. Discard the supernatant and add 1.8 ml of 70% ethanol into tube and vortex well; centrifuge the tube for 5 minutes at 14000 rpm and again discard the supernatant.
6. Do not dry DNA pellet and dissolved immediately in 300 µl of 1xTE, pH 8.0 (with RNAse A) at 55°C for 10-20 minutes.

DNA extraction (Collins *et al,* 1987)

Grinding buffer

0.08M NaCl

0.16M Sucrose

0.06M EDTA

0.5% SDS

0.1M Tris-HCl

TE Buffer (Tris EDTA) (Sambrook *et al,* 1989)

100ml 1M tris (pH 8)

20ml 0.5M EDTA

Make up volume to 1 L

II) Reverse Transcriptase and Quantitative Real Time PCR primers used

Primer Name	Sequence (5'->3')	Name/Function
AGAP005194F_qRT	GCATGTATGGGAGATTCTGGTG	Female protease
AGAP005194R_qR	GTGTACTGCTTTGAATATACCGACC	Female protease
AGAP005195F_qRT	CGCATCGATCGTGCTATAGC	Female protease
AGAP005195R_qRT	AAGTAGTCCAACATCGTCACGAAA	Female protease
AGAP002620F_qRT	CGGTCCGGTAGAGAGAGATG	
AGAP002620R_qRT	CCACCGTTACCACCGAAA	
AGAP004203_FqRT	GGAAACTACCATCCATCTTGCG	Vitellogenin
AGAP004203_RqRT	CCAAGCTCATAGTCGTAGGTCTTT	Vitellogenin
9370_F-qRT	GAAAAGGTTTCCAACCCACA	Fusillo
9370_R-qRT	TTATGATCGTCCATTCCCATC	Fusillo
Plugin_F_qRT	TGATTCAACCGTAGACATGAAGG	Acp
Plugin_R_qRT	CCACCATACAACGGAACGAC	Acp
TGASE_F_qRT	CAACCATGTCTACCACCTCTACC	Male protease
TGASE_R_qRT	ATCCTTTATCGCACTGTTCGTC	Male protease
AGAP004533_F_qRT	CCTGGAGCTATTGGGTCCGG	Male/female protease
AGAP004533_R_qRT	CCAAGTTGGAACCGAACGGG	
T7LacZF	TAATACGACTCACTATAGGGCTCGAGGTCGACGGTATCG	
T7LacZR	TAATACGACTCACTATAGGGCGGCCGCTCTAGAACTAG	
S7 Forward	GGCGATCATCATCTACGTGC	
S7 Reverse	GTAGCTGCTGCAAACTTCGG	
AGAP004533 Forward	AACAGCAATAGAAAATTACAAGGCA	
AGAP004533 Reverse	CCTAATACGCTGTTTTGCGCT	
AGAP005195 Forward	TGTTTCTCAGCGGCATTGTGCCT	
AGAP005195 Reverse	CAAGCAGAGAAGCGCATGCTGG	
ds4533 Forward	TAATACGACTCACTATAGGGACAACATGGCGTGCTGGTCA	
ds4533 Reverse	TAATACGACTCACTATAGGGCACCTTCGCCCTCACACGAA	

III) Solutions and reagents for western blot used

Extraction Bufer

Tris HCl PH 7.4 (1M)	-	25µL
NaCl (5M)	-	30µL
SDS 20%	-	5µL
NP4O lysis buffer (10%)	-	100µL
Triton X-100 (10%)	-	100µL
EDTA PH 8.0 (0.5M)	-	20µL
Protease inhibitors 7x	-	143µL
Double distilled water	-	577µL

10X PBS

Dissolve the following in 800ml distilled H_2O.

- 80g of NaCl
- 2.0g of KCl
- 14.4g of Na_2HPO_4
- 2.4g of KH_2PO_4

Adjust pH to 7.4.
Adjust volume to 1L with additional distilled H_2O.
Sterilize by autoclaving.
(Dilute to 1X to use)

Tween PBS

TPBS (0.05% Tween 20), pH 7.4

Add 250 µl Tween 20 to 500 ml TBS.

Blocking Solution (either solution may be used)

Starting Block T20 Blocking Buffer*, Pierce #37543

(*A proprietary protein formulation in Tris buffered saline at pH 7.5 with

0.05% Tween)

5% Milk in TTBS, made from dried skimmed powdered milk

IV) Quantification of protein using the Bradford assay

This method is according to the manufacturer's instructions for the Bio-Rad Protein microassay.

Reagent Company Catalogue number

Bio-Rad Protein Assay Dye

Reagent Concentrate Bio-Rad 500-0006

BSA

(Stock concentration = 1 µg/µL) Sigma A7030

Procedure

1. Samples and standards are assayed in duplicate.

2. Add 800 µL MilliQ water to microfuge tube.

3. Remove 1 µL of water using a P2 Gilson pipette. NOTE: this is done as you can't pipette 799 µL using a P1000 Gilson pipette.

4. Remove x µL of water for the other BSA standards. The amount is dependent on the amount of standard that will be added at a ratio of 1:1, e.g. If you require a 7µg standard, remove 7µL water and then add 7 µL 1 µg/µL BSA standard.

5. Add 1 µL of sample to a microfuge tube.

6. Vortex and pulse spin.

7. Add 200 µL Bio-Rad Protein Assay Dye Reagent Concentrate.

8. Vortex briefly.

9. Incubate at room temperature for at least 5 min. During this time, turn on the spectrophotometer and set wavelength to 595 nm.

10. Place sample into disposable cuvette.

11. Measure and record the absorbance at 595 nm.

12. Enter values from standard curve and samples into Excel.

V) Permealization and Blocking Buffer

Bovine Serum Albumin (1%) - 0.5g

Triton x-100 (0.1%) - 50µL

PBS (1%) - 450µL

Procedure

Weigh 0.5g of Bovine serum albumin

Dissolve it in 450µL of PBS

Filter the solution to remove any impurities

Add 50µL of Triton x-100

Place on a shaker for 30minutes

Store at 4°C

VI) Identified complexes in the protein-protein Network

BIOLOGICAL PROCESS IDENTIFICATION OF PROTEINS IN COMPLEX 1 (*Drosophila melanogaster*)

Each Event is coloured according to the **un-adjusted, i.e. not corrected for multiple testing**, probability (from hypergeometric test) of seeing given number or more genes in this Event by chance. Please note that only those "child" events are shown which have a p-value lower than the "parent" event. The top-level (root) Events are ordered according to the lowest p-value of their components.

Colour key for probabilities:

1e+00 3e-01 1e-01 3e-02 1e-02 3e-03 1e-03 3e-04 1e-04 3e-05 1e-05 3e-06 1e-06 3e-07 1e-07 3e-08 1e-08 3e-09 1e-09 3e-10 >

- Membrane Trafficking 1.1e-05, 7/67
- Metabolism of proteins 3.0e-04, 9/230
- Metabolism of amino acids and derivatives 4.5e-01, 2/187
- Signaling by Insulin receptor 2.1e-01, 2/107
- Axon guidance 7.1e-01, 2/303
- Gene Expression 3.2e-01, 5/471
- Regulatory RNA pathways 3.8e-02, 2/39

Total number of events assessed: 2021

Number of matching events (i.e. individual hypergeometric tests performed): 113

Number of genes matching submitted identifiers: 28

Results in a tab-delimited text file.

False discovery rate	Un-adjusted probability of seeing N or more genes in this Event by chance	Number of genes in your query which map to this Event	Total number of genes involved in this Event	Identifier of this Event	Name of this Event	Submitted identifiers mapping to this Event
-	2.6e-09	6	18	1463150	Vesicle Budding	A1Z7J6, Q9V5D2, A1Z813, Q9V4Z0, Q8MKS2, Q5U0U9, Q86NS0
-	2.6e-09	6	18	1463170	Vesicle Uncoating	A1Z7J6, Q9V5D2, A1Z813, Q9V4Z0, Q8MKS2, Q5U0U9, Q86NS0
-	2.6e-09	6	18	1463171	Cargo, Sec31p:Sec13p, and v-SNARE recruitment	A1Z7J6, Q9V5D2, A1Z813, Q9V4Z0, Q8MKS2, Q5U0U9, Q86NS0
-	5.3e-09	6	20	1471331	COPII (Coat Protein 2) Mediated Vesicle Transport	A1Z7J6, Q9V5D2, A1Z813,

						Q9V4Z0, Q8MKS2, Q5U0U9, Q86NS0
-	5.3e-09	6	20	1471332	ER to Golgi Transport	A1Z7J6, Q9V5D2, A1Z813, Q9V4Z0, Q8MKS2, Q5U0U9, Q86NS0
-	5.3e-09	6	20	1471333	Transport to the Golgi and subsequent modification	A1Z7J6, Q9V5D2, A1Z813, Q9V4Z0, Q8MKS2, Q5U0U9, Q86NS0
-	7.7e-08	6	30	1471334	Asparagine N-linked glycosylation	A1Z7J6, Q9V5D2, A1Z813, Q9V4Z0, Q8MKS2, Q5U0U9, Q86NS0
-	3.3e-06	6	55	1471181	Post-translational protein modification	A1Z7J6, Q9V5D2, A1Z813, Q9V4Z0, Q8MKS2, Q5U0U9, Q86NS0
-	1.1e-05	6	67	1471261	Membrane Trafficking	A1Z7J6, Q9V5D2, A1Z813, Q9V4Z0, Q8MKS2, Q5U0U9, Q86NS0
-	9.6e-05	3	12	1463102	Loss of Sar1b GTPase	Q9V5D2, A1Z813, Q86NS0
-	9.6e-05	3	12	1463160	Coat Assembly	Q9V5D2, A1Z813, Q86NS0
-	3.0e-04	8	230	1470931	Metabolism of proteins	Q9V9M7, Q9V5D2, A1Z813, Q8IQG4, Q86NS0,

						A1Z7J6, Q5U0U9, Q8MKS2, Q9V4Z0
-	3.7e-04	2	4	1452720	trimethyllysine + alpha-ketoglutarate + O2 => beta-hydroxy-trimethyllysine + succinate + CO2	Q4V6C2, Q9VDM7
-	9.1e-04	2	6	1458197	Phosphorylation of complexed TSC2 by PKB	Q9U9A9, Q9VCC9
-	9.1e-04	2	6	1465991	TSC2 activates intrinsic GTPase activity of Rheb	Q9U9A9, Q9VCC9
-	9.1e-04	2	6	1471198	Inhibition of TSC complex formation by PKB	Q9U9A9, Q9VCC9
-	2.1e-03	2	9	1470908	Carnitine synthesis	Q4V6C2, Q9VDM7
-	4.6e-03	2	13	1471460	SEMA3A-Plexin repulsion signaling by inhibiting Integrin adhesion	RAC1, Q9VSL8
-	9.7e-03	2	19	1465562	Formation of NR-MED1 Coactivator Complex	Q9VSE9, Q8WS79
-	9.7e-03	2	19	1471411	Nuclear Receptor transcription pathway	Q9VSE9, Q8WS79
-	9.7e-03	2	19	1471412	Generic Transcription Pathway	Q9VSE9, Q8WS79
-	1.1e-02	2	20	1462993	Dicer cleaves pre-miRNA to yield mature miRNA	Q8IP72, Q9VJY9
-	1.5e-02	2	24	1465973	Phosphorylated AMPK phosphorylates TSC2	Q9U9A9, Q9VCC9
-	1.7e-02	2	25	1471418	Regulation of Rheb GTPase activity by AMPK	Q9U9A9, Q9VCC9
-	2.0e-02	2	28	1471461	Semaphorin interactions	RAC1, Q9VSL8
-	2.6e-02	2	32	1471320	Energy dependent regulation of mTOR by LKB1-AMPK	Q9U9A9, Q9VCC9
-	3.8e-02	2	39	1471329	MicroRNA (miRNA) Biogenesis	Q8IP72, Q9VJY9
-	3.8e-02	2	39	1471330	Regulatory RNA	Q8IP72, Q9VJY9

pathways

BIOLOGICAL PROCESS IDENTIFICATION OF PROTEINS IN COMPLEX 8 (*Drosophila melanogaster*)

Each Event is coloured according to the **un-adjusted, i.e. not corrected for multiple testing**, probability (from hypergeometric test) of seeing given number or more genes in this Event by chance. Please note that only those "child" events are shown which have a p-value lower than the "parent" event. The top-level (root) Events are ordered according to the lowest p-value of their components.
Colour key for probabilities:
1e+00 3e-01 1e-01 3e-02 1e-02 3e-03 1e-03 3e-04 1e-04 3e-05 1e-05 3e-06 1e-06 3e-07 1e-07 3e-08 1e-08 3e-09 1e-09 3e-10 >

Hemostasis2.5e-05, 5/195
Total number of events assessed: 2021
Number of matching events (i.e. individual hypergeometric tests performed): 6
Number of genes matching submitted identifiers: 8

Results in a tab-delimited text file.

False discovery rate	Un-adjusted probability of seeing N or more genes in this Event by chance	Number of genes in your query which map to this Event	Total number of genes involved in this Event	Identifier of this Event	Name of this Event	Submitted identifiers mapping to this Event
-	1.5e-07	5	70	1470548	Exocytosis of platelet alpha granule contents	Q9U1I4, Q8T0M5, Q7K8Y5, Q9VFC2, Q9I7G6
-	6.2e-07	5	93	1471380	Platelet degranulation	Q9U1I4, Q8T0M5, Q7K8Y5, Q9VFC2, Q9I7G6
-	6.9e-07	5	95	1471135	Response to elevated platelet cytosolic Ca2+	Q9U1I4, Q8T0M5, Q7K8Y5, Q9VFC2, Q9I7G6
-	1.1e-05	5	164	1471136	Platelet Activation	Q9U1I4, Q8T0M5, Q7K8Y5, Q9VFC2, Q9I7G6
-	1.4e-05	5	173	1471137	Formation of Platelet plug	Q9U1I4, Q8T0M5, Q7K8Y5, Q9VFC2, Q9I7G6
-	2.5e-05	5	195	1471138	Hemostasis	Q9U1I4, Q8T0M5, Q7K8Y5, Q9VFC2,

BIOLOGICAL PROCESS IDENTIFICATION OF THE HUBBAL PROTEINS (*Drosophila melanogaster*)

Each Event is coloured according to the **un-adjusted, i.e. not corrected for multiple testing**, probability (from hypergeometric test) of seeing given number or more genes in this Event by chance. Please note that only those "child" events are shown which have a p-value lower than the "parent" event. The top-level (root) Events are ordered according to the lowest p-value of their components.
Colour key for probabilities:
1e+00 3e-01 1e-01 3e-02 1e-02 3e-03 1e-03 3e-04 1e-04 3e-05 1e-05 3e-06 1e-06 3e-07 1e-07 3e-08 1e-08 3e-09 1e-09 3e-10 >

Metabolism of proteins 8.8e-13, 11/230

Gene Expression 2.0e-06, 9/471

3' -UTR-mediated translational regulation 7.7e-08, 7/136

Regulation of beta-cell development 1.3e-06, 6/123

Diabetes pathways 6.8e-06, 6/163

Signal Recognition (Preprolactin) 5.0e-07, 6/105

Membrane Trafficking 1.3e-03, 3/67

Total number of events assessed: 2021
Number of matching events (i.e. individual hypergeometric tests performed): 49
Number of genes matching submitted identifiers: 12

Results in a tab-delimited text file.

False discovery rate	Un-adjusted probability of seeing N or more genes in this Event by chance	Number of genes in your query which map to this Event	Total number of genes involved in this Event	Identifier of this Event	Name of this Event	Submitted identifiers mapping to this Event
-	8.8e-13	11	230	1470931	Metabolism of proteins	Q9VXX8, Q9VMU4, Q9V9M7, Q9V5D2, P08570, A1Z813, Q9VTP4, Q86NS0, P08736, P55830, Q02748

-	3.2e-09	8	146	1470930	Translation	Q9VXX8, P55830, P08736, Q9VMU4, Q9V9M7, Q02748, P08570, Q9VTP4
-	8.2e-09	7	99	1457288	Aminoacyl-tRNA binds to the ribosome at the A-site	Q9VXX8, P55830, P08736, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	8.2e-09	7	99	1457304	Hydrolysis of eEF1A:GTP	Q9VXX8, P55830, P08736, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	8.8e-09	7	100	1457302	Translocation of ribosome by 3 bases in the 3' direction	Q9VXX8, P55830, P08736, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	8.8e-09	7	100	1471160	Peptide chain elongation	Q9VXX8, P55830, P08736, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	8.8e-09	7	100	1471161	Eukaryotic Translation Elongation	Q9VXX8, P55830, P08736, Q9VMU4,

						Q9V9M7, P08570, Q9VTP4
-	7.7e-08	7	136	1471158	L13a-mediated translational silencing of Ceruloplasmin expression	Q9VXX8, P55830, Q9VMU4, Q9V9M7, Q02748, P08570, Q9VTP4
-	7.7e-08	7	136	1471159	3' -UTR-mediated translational regulation	Q9VXX8, P55830, Q9VMU4, Q9V9M7, Q02748, P08570, Q9VTP4
-	9.4e-08	7	140	1470928	Cap-dependent Translation Initiation	Q9VXX8, P55830, Q9VMU4, Q9V9M7, Q02748, P08570, Q9VTP4
-	9.4e-08	7	140	1470929	Eukaryotic Translation Initiation	Q9VXX8, P55830, Q9VMU4, Q9V9M7, Q02748, P08570, Q9VTP4
-	3.1e-07	6	97	1457294	Peptide transfer from P-site tRNA to the A-site tRNA	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	3.1e-07	6	97	1463500	MAFA-, NKX2-2-, PAX6-, and PDX1-dependent synthesis of	Q9VXX8, P55830,

					insulin precursor protein	Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	3.5e-07	6	99	1454192	Release of 40S and 60S subunits from the 80S ribosome	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	3.7e-07	6	100	1454190	eIF5B:GTP is hydrolyzed and released	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	3.7e-07	6	100	1454191	The 60S subunit joins the translation initiation complex	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	3.7e-07	6	100	1457206	Polypeptide release from the eRF3-GDP:eRF1:mRNA:80S Ribosome complex	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	3.7e-07	6	100	1457209	GTP Hydrolysis by eRF3 bound to the eRF1:mRNA:polypeptide:80S Ribosome complex	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	3.7e-07	6	100	1457210	GTP bound eRF3:eRF1 complex binds the peptidyl tRNA:mRNA:80S Ribosome complex	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570,

						Q9VTP4
-	3.7e-07	6	100	1470933	GTP hydrolysis and joining of the 60S ribosomal subunit	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	3.7e-07	6	100	1471152	Eukaryotic Translation Termination	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	5.0e-07	6	105	1463959	Signal Recognition (Preproinsulin)	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	5.0e-07	6	105	1469883	Signal Recognition (Preprolactin)	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	5.9e-07	6	108	1463974	Interaction between SRP and SRP Receptor	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	7.4e-07	6	112	1470934	Formation of a pool of free 40S subunits	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	8.3e-07	5	59	1457275	Dissociation of L13a from the 60s	Q9VXX8, Q9VMU4,

					ribosomal subunit	Q9V9M7, P08570, Q9VTP4
-	1.0e-06	6	118	1471363	Regulation of gene expression in beta cells	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	1.2e-06	6	122	1464040	Translocation of Preproinsulin to Endoplasmic Reticulum	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	1.3e-06	6	123	1471356	Regulation of beta-cell development	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	2.0e-06	6	132	1471364	Insulin Synthesis and Processing	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	2.0e-06	9	471	1470914	Gene Expression	Q9VXX8, P55830, P08736, Q9VMU4, Q9V9M7, Q8WS79, Q02748, P08570, Q9VTP4
-	6.6e-06	3	12	1463102	Loss of Sar1b GTPase	Q9V5D2, A1Z813, Q86NS0

-	6.6e-06	3	12	1463160	Coat Assembly	Q9V5D2, A1Z813, Q86NS0
-	6.8e-06	6	163	1471187	Diabetes pathways	Q9VXX8, P55830, Q9VMU4, Q9V9M7, P08570, Q9VTP4
-	2.4e-05	3	18	1463150	Vesicle Budding	Q9V5D2, A1Z813, Q86NS0
-	2.4e-05	3	18	1463170	Vesicle Uncoating	Q9V5D2, A1Z813, Q86NS0
-	2.4e-05	3	18	1463171	Cargo, Sec31p:Sec13p, and v-SNARE recruitment	Q9V5D2, A1Z813, Q86NS0
-	3.4e-05	3	20	1471331	COPII (Coat Protein 2) Mediated Vesicle Transport	Q9V5D2, A1Z813, Q86NS0
-	3.4e-05	3	20	1471332	ER to Golgi Transport	Q9V5D2, A1Z813, Q86NS0
-	3.4e-05	3	20	1471333	Transport to the Golgi and subsequent modification	Q9V5D2, A1Z813, Q86NS0
-	1.2e-04	3	30	1471334	Asparagine N-linked glycosylation	Q9V5D2, A1Z813, Q86NS0
-	7.3e-04	3	55	1471181	Post-translational protein modification	Q9V5D2, A1Z813, Q86NS0
-	1.3e-03	3	67	1471261	Membrane Trafficking	Q9V5D2, A1Z813,

						Q86NS0
-	2.7e-02	2	77	1453936	Ribosomal scanning	P55830, Q02748
-	2.7e-02	2	77	1457395	Formation of translation initiation complexes containing mRNA that does not circularize	P55830, Q02748
-	2.7e-02	2	77	1470927	Ribosomal scanning and start codon recognition	P55830, Q02748
-	2.7e-02	2	77	1471169	Translation initiation complex formation	P55830, Q02748
-	2.8e-02	2	78	1457273	Association of phospho-L13a with GAIT element of Ceruloplasmin mRNA	P55830, Q02748
-	2.8e-02	2	78	1470932	Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S	P55830, Q02748

VII) Screen house datasheet

Date of Realease	Follow up	Number released	Time released	Age	Screen house	Weather conditions	Resting sites post-release	Swarm initiation time	Full swarm time	swarm termination time	swarm sites	Number of swarms	Average couples seen	couples caught	End of experiment	predators	coupulating sites	weather challenges	Other challenges
9-Jun		500	18:06	2days	malaria sphere	bright sky	hut	18:51	19:00	19:10	corners 3, 4	one	seven	sixteen	19:17	spiders	Thached roof	early darkness	
						not windy	lower surface plantain leaves				ontop of hut	three				Geckos	Net		
							napier grass leaves									Mantis	air		
	6/10/2012				malaria sphere	windy, dark sky	hut	18:45	18:49	18:52	ontop of hut	two	none	none	18:53	Geckos		windy	
							roof top											early darkness	
																		threatening to rain	
	6/11/2012				malaria sphere	bright sky	hut	18:45	18:55	19:10	ontop of hut	three	thirty three	none	19:12	Geckos		early darkness	
						not windy										mantis			screen house cleared
	6/12/2012				malaria sphere	bright sky	hut	18:50	19:00	19:10	ontop oh nut	three	four	none	19:12		roof top	early darkness	
						not windy													
6/13/2012		1000	18:10	2days	malaria sphere	not windy	leaves	18:50	19:00	19:10	ontop of hut	one							
						bright sky	hut				corners1,2,3,4	one	eighty	ten	19:20	Geckos	roof top	early darkness	
																Mantis	net		
																spiders			
	6/14/2012				malaria sphere	bright sky	hut	18:48	19:00	19:10	corners 1,3,4	one	fourty two	six		Geckos	roof top	early darkness	
						not windy					ontop of hut	three					Net		
	6/15/2012				malaria sphere	bright sky	hut	18:50	19:00	19:10	ontop of hut	two	none	none	19:20			early darkness	
						not windy													
6/16/2012		1000	18:06	2days	malaria sphere	bright sky	leaves	18:55	19:01	19:11	ontop of hut	three	thirty five	tewenty	19:15	Geckos	roof top	early darkness	
						not windy	hut				corner 3	one					Net		
	6/17/2012				malaria sphere	bright sky	hut	18:50	18:59	19:03	ontop of hut		two	none	19:20	Geckos		early darkness	
						not windy	roof top				corners 2,3	few				spiders			
							pots												
6/18/2012																			experiments by owners of green house
6/28/2012																			
	6/28/2012	*	18:17	2days	Screen house	not windy	corners next to the door,pots,plants	18:50	18:53	19:10	corners 1,2,3 and behind the door	four	three	six	19:13	playing mantis and Geckos	net	early darkness	
	6/29/2012	*	18:10	2days	Screen house	not windy	corners next to the door,pots,plants	18:50	18:58	19:10	corners 1,2,3 and behind the door	three	twenty seven	fourteen	19:13				
	6/30/2012	*	18:20	2days	Screen house	bright sky	corners next to the door,pots,plants and behind the door	18:47	18:53	19:08	corners 1,2,3 and behind the door	three	six	nine					
7/1/2012		*	18:14	2days	screen house	cloudy	in the basins,pots,nets,top of the roof and plants	18:49	19:00	19:17	Non	Non	Non	Non	19:17				mosquitoes and swarms almost on the ground
	7/2/2012	*	18:30	2days	screen house	calm	on-top of the net and behind the door	18:47	18:53	19:15	corners 1,2,3 and behind the door	four	twenty six	twelve	19:13			early darkness	
	7/3/2012	*	0:00	*	screen house	calm	on-top of the net and behind the door	18:55	19:00	19:17	corners 1,2,3 and behind the door	three	seven	fifteen	19:17	gecko and spiders			
	7/4/2012	*	18:30	2days	screen house	windy	pots,roof-top,plants and on the net	18:50	19:10	19:17	corners 1,2,3 and behind the door	five	fifteen	twelve	19:17	spiders and geckos		early darkness	swarm starts when its already dark
	7/5/2012	*	18:30	2days	screen house	windy	pots,roof-top,plants,in the hut and on the net	18:47	18:53	19:10	corners 1,2,3 and behind the door	three	six	fourteen	19:14			early darkness	
7/8/2012																			
		*	18:25	2days	screenhouse	cool evening	pots and roof-top	19:00	19:05	19:13	corners 1,2 and 3	three	eight	1	19:15		root-top	early darkness	late swarming with darkness
	7/9/2012	*	not released	*	screenhouse	windy	pots and roof-top	19:00	19:05	19:11	corners 1,2 and 3	two	eight	1	19:13	gecko and spider	net	early darkness	
	7/10/2012	*	not released	*	screenhouse	windy	pots and roof-top	19:02	19:05	19:09	corner 1	one	five	0	19;13		net	early darkness	
	7/11/2012	*	not released	2days	screen house	windy	pots and roof-top	18:57	19:03	19:10	corner 3	one	*	eight	19:13			early darkness	

7/12/2012																					
		*		18:30	2days	screen house	cool evening	pots,roof-top and on the net		18:58	19:05		19:13	corner 1,2,3 and behind the door	two	eleven	seven	19:17		air		
	7/13/2012	*		not released	2days	screen house	windy and cloudy	roof-top and on the net		18:58	19:03		19:09	corners 1 and 3	two	six	six	19:13	gecko and skink		early darkness	
7/25/2012																					
		*		18:15	2days	screen house	calm and cold	behind the door,roof-top,pots on the wals		19:00	19:05		19:10	corner 3	n/a	one	one	19:13			darkness	no swarming but could see the mosquitoes
	7/26/2012	*		not realesed	*	screen house	calm and cold			19:00	19:05		19:10	behind the door and corner 1	one	seven	n/a	19:12	spiders		cloudy	swarm scartered
	7/27/2012	*		not realesed	*	screen house	cool evening	pots on the walls of the hut	n/a	n/a	n/a		n/a		n/a	n/a	n/a	19:13			cloudy and dark	scartered swarm
7/29/2012																					
		*		18:30	2days	screen house	cool evening	on-top of the net,behind the door and on the pots		18:57	19:03		19:10	corner 1,2,3,behind the door	four	twelve	one	19:13	gecko and skink	air	early darkness	
	7/30/2012	*		18:25	2days	screen house	calm weather	in the hut,pots and behind the hut		18:57	19:03		19:10	corner 1,2,3,behind the door	four	twenty	seven	19:14		behind the door	early darkness	early darkness
	8/4/2012	*		18:35	2days	screen house	calm weather	on the net,pots,in the hut and on the net		19:00	19:04		19:13	corners 1,2,3,behind the door and behindhut	five	twenty eight	four	19:13		corner 1	early darkness	early darkness
8/7/2012																					
		*		18:30	2days	screen house	drizling	on the net		18:51	19:04		19:15	corners 1,2,3 and bahind the door&hut	four	thirty five	eleven	19:13			early darkness	"
	8/8/2012		700	18:28	2days	screen house	cloudy,drizling&windy	on the net,pots,in the hut		18:49	18:59		19:05	corners 1,2,3 and bahind the door&hut	three	seventeen	fourteen	19:10			early darkness	"
	8/9/2012		500	18:30	2days	screen house	calm	on the net,pots,in the hut		18:50	18:57		19:09	corners 1,2,3 and bahind the door&hut	five	thirty three	twenty two	19:13	gecko	net	early darkness	"
	8/10/2012		700	18:45	2days	screen house	calm	pots and on the net		18:55	19:02		19:16	corners 1,2,3 and bahind the door&hut	six	seventeen	fifteen	19:18	gecko,mantis	net	early darkness	"
	8/11/2012		800	18:29	2days	screen house	calm	pots and on the net		18:56	19:06		19:13	corners 1,2,3 and bahind the door&hut	six	seventeen	fifteen	19:14		net	early darkness	"
	8/12/2012		80	18:30	2days	screen house	windy	net and in the hut		18:54	19:02		19:13	corners 1,2,3 and bahind the door&hut	six	six	twenty	19:16			early darkness	"
	8/13/2012		700	18:33	2days	screen house	calm	pots,inside the hut and on plants		18:57	19:03		19:14	corners 1,2,3 and bahind the door&hut	six	fourtyeight	twenty	19:16			early darkness	"
	8/14/2012		1200	18:30	2days	screen house	calmand cloudy	pots,inside the hut and on plants		18:49	19:05		19:13	corners 1,2,3 and bahind the door&hut	six	thiry seven	fifteen	19:19	gecko and mantis		early darkness	"
	8/15/2012		600	18:36	2days	screen house	calm	pots,inside the hut and on plants		18:54	19:05		19:13	corners 1,2,3 and bahind the door&hut	six	eleven	eleven	19:15	gecko		early darkness	"
	8/16/2012		500	18:41	2days	screen house	calm and bright sky	pots,inside the hut and on plants		18:55	19:03		19:16	corners 1,2,3 and bahind the door&hut	five	ten	four	19:18			early darkness	"
	8/17/2012	*	*		*	screen house	calm	pots,inside the hut and on plants		18:54	19:05		19:13	corners 1,2,3 and bahind the door&hut	five	twenty one	three	19:15			early darkness	"
	8/18/2012		600	18:30	2days	screen house	cloudy	pots,inside the hut and on plants		18:54	19:05		19:14	corners 1,2,3 and bahind the door&hut	six	eleven	seven	19:16			early darkness	"
	8/19/2012		500	18:30	2days	screen house	calm and cloudy	pots,inside the hut and on plants		18:57	19:04		19:15	corners 1,2,3 and bahind the door&hut	six	twelve	four	19:17			early darkness	"

VIII) Ethical clearance

KENYA MEDICAL RESEARCH INSTITUTE

P.O. Box 54840-00200, NAIROBI, Kenya
Tel (254) (020) 2722541, 2713349, 0722-205901, 0733-400003; Fax: (254) (020) 2720030
E-mail: director@kemri.org info@kemri.org Website:www.kemri.org

KEMRI/RES/7/3/1 **September 28, 2012**

TO: **PROF. WOLFGANG RICHARD MUKABANA,**
HEAD, HUMAN HEALTH DIVISION,
ICIPE,
P. O. BOX 30772 – 00100,
NAIROBI

Dear Sir,

RE: NON-SSC PROTOCOL No. 349 – REVISION 2 (*RE-SUBMISSION 2*): GENETICS OF MOSQUITO INTERATION WITH MALARIA, FUNGUS AND INSECTICIDES

Reference is made to your letter dated September 27, 2012. The ERC Secretariat acknowledges receipt of the revised proposal.

This is to inform you that the Committee determines that the issues raised at the 205th ERC meeting of 21st July 2012 are adequately addressed. Consequently, the study is granted approval for implementation effective this **28th day of September 2012** for a period of one year. Please note that authorization to conduct this study will automatically expire on **September 27, 2013**. If you plan to continue data collection or analysis beyond this date, please submit an application for continuation approval to the ERC Secretariat by **August 16, 2013**. The regulations require continuing review even though the research activity may not have begun until sometime after the ERC approval.

You are required to submit any proposed changes to this study to the SSC and ERC for review and the changes should not be initiated until written approval from the ERC is received.

Please note that any unanticipated problems resulting from the implementation of this study should be brought to the attention of the ERC and you should advise the ERC when the study is completed or discontinued.

Work on this project may begin.

Sincerely,

DR. CHRISTINE WASUNNA,
ACTING SECRETARY,
KEMRI ETHICS REVIEW COMMITTEE

In Search of Better Health

IX) Conference Abstracts

Deciphering the reproductive protein-protein interaction network in *Anopheles gambiae* with *Drosophila melanogaster* as a framework

Daniel Achinko[1*], **Paul Mireji**[2], **Flaminia Catteruccia**[3] and **Dan Masig**[1]

- * Corresponding author: Daniel Achinko

Author Affiliations

[1] Molecular Biology and Biotechnology Department, International Center of Insect Physiology and Ecology (icipe), Nairobi 30772-00100, Kenya

[2] Biochemistry and Molecular Biology Department, Egerton University, Njoro, Kenya

[3] Division of Cell and Molecular Biology, Imperial College London, London SW7 2AZ, UK

For all author emails, please log on.

Genome Biology 2011, **12**(Suppl 1):P32 doi:10.1186/gb-2011-12-s1-p32

The electronic version of this article is the complete one and can be found online at: http://genomebiology.com/2011/12/S1/P32

Published:19 September 2011

Background

Protein-protein interactions (PPIs) are the most fundamental biological processes at the molecular level. The experimental methods for testing PPIs are time-consuming and are limited by analogs for many reactions. As a result, a computational model is necessary to predict PPIs and to explore the consequences of signal alterations in biological pathways. Reproductive control of the vector *Anopheles gambiae* using transgenic techniques poses a serious challenge. To meet this challenge, it would help to define the biological network involving the male accessory gland (MAG) proteins responsible for successful formation of the mating plug [1]. This plug forms in the male and is transferred to the female during mating, hence initiating the PPIs in both sexes. As is the case in *Drosophila melanogaster*, a close relative of *A. gambiae*, some MAG proteins responsible for the formation of the mating plug have been shown to alter the post-mating behavior of females.

Methods and results

The STRING database for known PPIs was used to identify orthologs of *A. gambiae* proteins in *Drosophila* (Table 1). Twenty-seven proteins are known to form the mating plug in *A. gambiae*, and 16 others were obtained as strings in the STRING database. Chromosome synteny comparisons for proteins with more than 50% identity between species were carried out using the Artemis Comparison Tool (ACT version 9.0), and this showed 24.39% matches (M), 12.20% mismatches (MM) and 63.41% unmatched (NM). The network built in Cytoscape (version 2.8.0) with the UniProt IDs for these *Drosophila* orthologs showed 14 complexes, with 4 of them being for *Drosophila*. The network showed 555 nodes and 2,344 edges. The top 50 identified hubs in the network showed a range of 3 to 30 interactions. The expression values for these proteins in FlyAtlas showed that they are upregulated in the reproductive tissues of both sexes. To understand the processes involved in plug formation, the Reactome database was used, and the hub proteins were identified in 49 of the 2,021 known processes in *Drosophila*. Twelve proteins were involved in the following processes: metabolism of proteins (8.8e–13), gene expression (2.0e–06), 3'-UTR-mediated translational regulation (7.7e–08), regulation of β-cell development (1.3e–06), diabetes pathways (6.8e–06), signal recognition (preprolactin) (5.0e–07) and membrane trafficking (1.3e–03). Of the top 50 proteins, 92% had orthologs in *A. gambiae*, with one identified in the mating plug and four others identified as strings to AGAP009584, which is found in the mating plug. Acp29AB was identified in the network and is known to induce post-mating responses in *Drosophila*, confirming that the network is reproductive and giving an insight into the possible pathways involved. The CG9083 (Q8SX59) protein was ranked first among the hub proteins but has no ortholog in *A. gambiae*. Interestingly, it has the same protein properties as the Plugin protein (AGAP009368) in *A. gambiae*, suggesting that Plugin may be the main protein in the PPI reproductive network in *A. gambiae.* The Whelan and Goldman (WAG) maximum likelihood tree evaluations of the plug proteins in *A. gambiae* and their orthologs in *Drosophila* showed that these proteins are involved in similar biological processes in both species, but the *A. gambiae* protein evaluation provided a better explanation for the expected process as it clustered in both pre-mated and post-mated PPIs.

Table 1. Orthologs of *Anopheles gambiae* proteins in *Drosophila* identified using the STRING database

This table shows the 27 proteins known to be in the mating plug of *A. gambiae*[1], derived predominantly from the male. The 16 strings predicted as orthologs in *Drosophila*, using the STRING database, have varying scores. Scores above 60 can be trusted following their alignments. Plugin, which has the lowest score, has no good ortholog in *Drosophila*. Most of the proteins are encoded on chromosome arms 2L and 3R in both species. The chromosome synteny comparisons using ACT showed 24.39% matches (M), 12.20% mismatches (MM) and 63.41% unmatched (NM). The presence of gaps between the alignments resulted in the observed MM and NM. The nucleotide sequences at the chromosomal locations where the proteins NOVEL ACP1 and NOVEL ZCP7 are encoded were used to identify similar proteins and their orthologs.

Conclusions

The identification of *A. gambiae* proteins in this network creates more targets for functional analysis and reproductive control of the malaria vector.

References

1. Rogers DW, Baldini F, Battaglia F, Panico M, Dell A, Morris HR, Catteruccia F: **Transglutaminase-mediated semen coagulation controls sperm storage in the malaria mosquito.**

   PLoS Biol 2009, **7:**e1000272. PubMed Abstract | Publisher Full Text | PubMed Central Full Text

Achinko et al. BMC Proceedings 2012, 6(Suppl 6):P1
http://www.biomedcentral.com/1753-6561/6/S6/P1

BMC Proceedings

POSTER PRESENTATION Open Access

Life stage and tissue speciation of cathepsin B (AGAP004533) derives different functional properties in the G3 strain of the mosquito *Anopheles gambiae*

Daniel Achinko[1*], Dan Masiga[1], Paul Mireji[1,2], Flaminia Catteruccia[3]

From Beyond the Genome 2012
Boston, MA, USA. 27-29 September 2012

Cathepsin B is a lysosomal papain-like cysteine peptidase that is expressed in all tissues and functions primarily as an exopeptidase through its carboxydipeptidyl activity. Together with other cathepsins, it is involved in the degradation of proteins, proenzyme activation, antigen processing, metabolism and apoptosis. AGAP004533 is a cathepsin B peptidase of 337 amino acids known to be found on the mating plug. This plug is known to be produced in the male *Anopheles gambiae* mosquito and transferred to the female during mating [1]. The female digests this plug in 24 h. The protein is expressed in all life stages of the mosquito and in all tissues of the adult. We cloned and sequenced the protein in the larvae and pupae stages and all reproductive tissues (spermatheca, atria and ovary of the female; testes and male accessory glands (MAGs) of the male) of the G3 mosquito strain. These sequences were analysed with Geneious 5.5.5 and cLc workbench 6.6.1 software. Within the coding sequence, two single mutations at C584T (juvenile stages) and nucleotide A14T (ovary) were identified. The latter translates into a glutamine for leucine (Q6L), which causes the loss of the signal peptide due to loss of five amino acids at the N-terminal region of the protein sequence, meanwhile the former translates into an alanine for valine (A195V). Both mutations cause structural modifications within the secondary structure of the protein that eventually affect its 3D conformation. The sequences in the artria and spermatheca showed insertion of a cytosine at nucleotide 1010, which translates to a proline for a leucine (P337L) substitution, and hence loss of the stop codon at amino acid 338. This loss causes an extension of 14 amino acids at the carboxylic end of the protein, resulting in secondary structure modification. The sequence for the testes appeared transposed, and hence was not considered in the analysis. All the sequences translated on the same frame except for that of the ovary, which translated on a different frame. Protein BLAST of these sequences at NCBI using the blosum62 matrix, identified with AGAP004533 of *Anopheles gambiae* alongside other mosquito species, although that of the artria and spermatheca also identified with species of distant taxa such as *Manduca sexta* (FM957999.1) and *Gallus gallus* (NM205371.1). This relation was due to the amino acid extensions at the carboxylic end relating to parasite killing in the former, and embryonic apoptosis in the latter. Transcription factor predictions on all sequences identified equal binding sites (T00821, T00752), and that of the male accessory glands identified an extra binding site (T00360) known in humans as a bifunctional protein nuclear cytoplasmic *O*-*N*-acetylglucosaminidase and acetyltransferase. This site also has alternative splicing functions, which could be important for the variations observed in this gene. Sequence variations of this protein in the different stages and tissues of the mosquitoes may also be highly related to their functions and relative positions in the same or different biological processes within the various tissues. In-depth analysis of the reproductive role of AGAP004533 will help in reproductive control of the vector.

[1]Molecular Biology and Biotechnology Department, International Center of Insect Physiology and Ecology, Nairobi, Kenya, PO Box 30772-00100
Full list of author information is available at the end of the article

Author details
[1]Molecular Biology and Biotechnology Department, International Center of Insect Physiology and Ecology, Nairobi, Kenya, PO Box 30772-00100. [2]Biochemistry Department, Egerton University, Njoro, Kenya, PO Box 20115-536. [3]Dipartimento di Medicina Sperimentale e Scienze Biochimiche, Università degli Studi di Perugia, 05100 Terni, Italy.

Published: 1 October 2012

Reference

1. Rogers DW, Baldini F, Battaglia F, Panico M, Dell A, Morris HR, Catteruccia F: Transglutaminase-mediated semen coagulation controls sperm storage in the malaria mosquito. *PLoS Biol* 2009, **7**:e1000272.

doi:10.1186/1753-6561-6-S6-P1
Cite this article as: Achinko *et al.*: Life stage and tissue speciation of cathepsin B (AGAP004533) derives different functional properties in the G3 strain of the mosquito *Anopheles gambiae*. *BMC Proceedings* 2012 **6**(Suppl 6):P1.