This work is protected by copyright and other intellectual property rights and duplication or sale of all or part is not permitted, except that material may be duplicated by you for research, private study, criticism/review or educational purposes. Electronic or print copies are for your own personal, non-commercial use and shall not be passed to any other individual. No quotation may be published without proper acknowledgement. For any other use, or to quote extensively from the work, permission must be obtained from the copyright holder/s.

Genetic and behavioural mechanisms of reproductive isolation and speciation in recently-diverged sibling species of the malaria vector *Anopheles gambiae* s.l.

Nahla Alhafez

In consideration for the degree of Doctor of Philosophy

Research centre for applied Entomology and Parasitology

October 2016

Keele University

Keele University

Declaration Part 1. To be bound in the thesis

SUBMISSION OF THESIS FOR A RESEARCH DEGREE

Part I. DECLARATION by the candidate for a research degree. To be bound in the thesis

Degree for which thesis being submitted

Title of thesis

This thesis contains confidential information and is subject to the protocol set down for the submission and examination of such a thesis.

YES/NO [please delete as appropriate; if YES the box in Part II should be completed]

Date of submission 22/June/2016 Original registration date 6/June/2011
(Date of submission must comply with Regulation 2D)

Name of candidate Nahla Alhafez

Research Institute Life science Name of Lead Supervisor Frederic Tripet

I certify that:

(a) The thesis being submitted for examination is my own account of my own research
(b) My research has been conducted ethically. Where relevant a letter from the approving body confirming that ethical approval has been given has been bound in the thesis as an Annex
(c) The data and results presented are the genuine data and results actually obtained by me during the conduct of the research
(d) Where I have drawn on the work, ideas and results of others this has been appropriately acknowledged in the thesis
(e) Where any collaboration has taken place with one or more other researchers, I have included within an 'Acknowledgments' section in the thesis a clear statement of their contributions, in line with the relevant statement in the Code of Practice (see Note overleaf).
(f) The greater portion of the work described in the thesis has been undertaken subsequent to my registration for the higher degree for which I am submitting for examination
(g) Where part of the work described in the thesis has previously been incorporated in another thesis submitted by me for a higher degree (if any), this has been identified and acknowledged in the thesis
(h) The thesis submitted is within the required word limit as specified in the Regulations

Total words in submitted thesis (including text and footnotes, but excluding references and appendices) : 47,637

Signature of candidate ...Nahla Alhafez... Date 22/06/2016

Note

Extract from Code of Practice: If the research degree is set within a broader programme of work involving a group of investigators – particularly if this programme of work predates the candidate's

registration – the candidate should provide an explicit statement (in an ‘Acknowledgments’ section) of the respective roles of the candidate and these other individuals in relevant aspects of the work reported in the thesis. For example, it should make clear, where relevant, the candidate’s role in designing the study, developing data collection instruments, collecting primary data, analysing such data, and formulating conclusions from the analysis. Others involved in these aspects of the research should be named, and their contributions relative to that of the candidate should be specified (*this does not apply to the ordinary supervision, only if the supervisor or supervisory team has had greater than usual involvement*).

Keele University

Declaration Part 2. To be bound in the thesis

SUBMISSION OF THESIS FOR A RESEARCH DEGREE

Part II. DECLARATIONS by the candidate for a research degree, the academic and (where appropriate) non-academic supervisor. Not to be bound in the thesis nor to be made available to the examiners.

Degree for which thesis being submitted: In consideration for the degree of Doctor of

Philosophy

Title of thesis: Genetic and behavioral mechanisms of reproductive isolation and speciation in

recently-diverged sibling species of the malaria vector *Anopheles gambiae* s.l.

This thesis contains confidential information and is subject to the protocol set down for the submission and examination of such a thesis.

YES [please delete as appropriate; if YES the box below should be completed]

Date of submission 22/June/2016 Original registration date 6/June/2011

(Date of submission must comply with Regulation 2D)

Name of candidate Nahla Alhafez

Research Institute Life science Name of Lead Supervisor Frederic Tripet

<u>1. DECLARATION by the candidate for a research degree</u>

I certify that: (delete as appropriate)

the decision to submit this thesis has been taken following consultation with, and having received advice from, my academic and (where appropriate) non-academic (e.g. industrial) supervisor(s)

Signature of candidate Nahla Alhafez… Date 22/06/2016

2. DECLARATION by the Lead Supervisor

I have read the thesis*/I have read part of the thesis*/I have not read the thesis*, and I am aware that the student intends to submit.

Signature of lead supervisor ……………………………… Date …………………
Please print name ……………………………………

3. DECLARATION where the thesis contains confidential information
Delete the following box if it does not apply

I have read the thesis*/I have read part of the thesis*/I have not read the thesis*, and I am aware that the student intends to submit. Signature of non-academic supervisor …………………………………………………… Please print name ……………………………………………… On behalf of ………………………………………………… (non-academic sponsor) The copy of the thesis lodged in the University Library shall be subject to an embargo on access for a period of ………..years. Signature of lead supervisor ……………………………………………… Signature of non-academic supervisor ………………………………………………… Signature of research student …………………………… Date ………………… *Delete as appropriate

Abstract

Anopheles gambiae s.l. species are reported as the most important vectors due to their ability to spread and exploit both temporary and man-made breeding sites. These factors combined with their capacity to transmit malaria, make them wide-open subject for researchers to find alternative different methods to control malaria disease. Historically, there were many attempts to limit or eradicate malaria disease by controlling the mosquitoes. The control methods varied between using insecticide-treated bed nets and indoor residual spraying, in addition to other methods that based on sterile insect technique (SIT) or genetic modification techniques. However, the success of any attempts to eliminate malaria depends on our understanding of the genetic diversity of the vectors in field. In this regard, two distinct sibling species *Anopheles coluzzii* and *Anopheles gambiae s.s.* were defined in *An. gambiae* complex based on single nucleotide polymorphisms intergenic spacer (IGS) near the centromere of the X chromosome. These species were found to mate assortatively even within mixed swarms. Moreover, genetic differences between them were limit to three regions that were 2L, 3L and X chromosome near the centromere regions. Several studies tried to explain the sympatric speciation process between the *An. gambiae* s.l. sibling species based on these islands of speciation and their role in assortative mating behaviour, but no full explanation was concluded so far. Therefore, two objectives were taken into consideration in this thesis. The first was genetic objective that aims to identify the islands of speciation that more likely responsible for assortative mating. The second was behavioural objective that aims to identify the mechanisms of assortative mating using recombinant and parental strains. These studies were important step for trying to explain the speciation process within *An. gambiae* s.l. species that in turn important for malaria control strategies.

Table of contents

Declaration II

Abstract VI

Table of contents VII

Acknowledgements XII

CHAPTER 1. Introduction to Malaria mosquitoes 13

1.1. The mysterious history of human's small enemy 13

1.2. Malaria infection (plasmodium life cycle) 19

1.3. *Anopheles gambiae* life cycle 22

1.4. Malaria control history 26

1.5. Reproductive system in females and males of *An. gambiae* 28

1.6. Mating behaviour 30

1.7. The *Anopheles gambiae* complex 33

1.8. Chromosomal forms 35

1.9. Molecular forms 39

1.10. Reproductive isolation and ecological adaptation in *An. gambiae s.s.* and *An. coluzzii* 40

1.11. Mechanisms of assortative mating in *An. gambiae s.s.* and *An. coluzzii*, ... 41

1.11.1. Intrinsic and extrinsic post-mating reproductive barriers 41

1.11.2. Pre-mating reproductive barriers. 42

1.12. Genetic structure and models of speciation. 45

1.13. Research objectives 51

CHAPTER 2 . General material and methods 52

2.1. Insectary conditions 52

2.2. *An. gambiae* strains breeding 53

2.3. Sexing the pupae 55

2.4. Females dissection and obtaining the sperm bundle..57

2.5. Standard amplification for differentiation between S and M genotype in X-divergence-island..57

2.6. DNA restriction reaction for differentiation between S and M genotype in X-divergence-island and gel separation..58

2.7. Wings measurements..58

2.8. ChargeSwitch method for DNA extraction..60

2.9. DNAzol method for DNA extraction..60

2.10. Creating recombinants..61

CHAPTER 3. Genomic characterization of X, 2L, 3L divergence islands and assortative mating in females and males of X-island recombinant strains, and parental strains 64

3.1. Introduction..64

3.2. Materials and methods..69

3.2.1. Genotyping X, 2L and 3L..69

DNAzol extraction 69

DNA restriction and gel separation 70

Inversions characterization 71

3.2.2. Recombinant and parental strains mating preference..71

Dissection of mated females and genetic analysis of sperm..72

3.3. Results..72

3.3.1. Genotyping X, 2L, 3L islands of recombinant and parental strains..72

3.3.2. Analysing assortative mating behavior..73

Preliminary assortative mating experiments to test the recombinant females preference 73

Assortative mating analyses within parental strains..76

Assortative mating analyses within recombinant strains 78

3.4. Discussion..80

CHAPTER 4. Genome sequencing of recombinant and parental strains, and comparative genomic with field sympatric populations 82

4.1. Introduction..82

4.2. Materials and Methods 87

4.2.1. DNA preparation of recombinant and parental strains 87

DNA extraction 87

Genome amplification 87

Genome purification 87

Picogreen assay (DNA Concentration) 88

4.2.2. Genome sequencing of recombinant and parental strains 90

4.2.3. Deep targeted sequencing of field sympatric populations.........................92

4.3. Results 93

4.3.1. Sequencing the full-genome of recombinants and parental strains 93

4.3.2. Comparison with X-island from sympatric field populations 95

4.4. Discussion..99

CHAPTER 5 Analysing the stages of mating behavior and sperm quantification in *An. coluzzii* and *An. gambiae s.s.* 102

5.1. Introduction..102

5.2. Materials and methods...109

5.2.1. Sexing the pupae..109

5.2.2. Set the experiment cages 109

5.2.3. Videos recording analyses...111

5.2.4. Wings measurements..112

5.2.5. Females dissection...113

5.2.6. Creating gDNA standard curve for QPCR 113

5.2.7. Sperm quantification 117

5.2.8. Data analyses 117

5.3. Results 118

5.3.1. Phenotypic quality of individuals for mating experiment 118

5.3.2. Matched islands combinations and not-matched islands combinations analyses..119

5.3.2.1. Mating behavior..119

5.3.2.2. Proportion of swarming and non-swarming individuals 133

5.3.3. Matched and not-matched X-island analyses 149

5.3.3.1. Mating behavior..149

5.3.4. Sperm quantification 158

5.3.4.1. Phenotypic quality of individuals for the sperm quantification samples 158

5.3.4.2. The sperm quantification in different mating combinations 159

5.3.4.3. The sperm quantification in different females types 161

5.4. Discussion...162

CHAPTER 6 A role of wing morphology in the wing-tone hypothesis of assortative mating? ……………………………………………………………………169

6.1. Introduction..169

6.2. Materials and Methods 175

6.2.1. Preparing the wings for studies 175

6.2.2. Comparisons wing length and width 176

6.2.3. Detailed wing morphometry study 176

6.3. Results 183

6.3.1. Wing length and width ANOVAs 183

6.3.1.1. Dimorphism between females and males 183

6.3.1.2. Variation between strains 183

6.3.2.Morphometrics analyses of wings 186

6.3.2.1. Dimorphism between females and males 186

6.3.2.2. Variation between strains 194

6.4. Discussion..205

CHAPTER 7. General discussion about behavioural mechanisms of reproductive isolation and sympatric speciation in *An. gambiae* s.l species, and their effects on malaria control strategies ……………………………………………………………………208

7.1. Introduction...208

7.2. The mechanisms of assortative mating in *An. gambiae s.s.* and *An. coluzzii*....210

7.3. Explaining sympatric speciation models based on the divergence islands........210

7.4. Malaria control strategies 218

APPENDICES 218

REFERENCES..223

Acknowledgements

I am thankful to ALLAH (the God) who helped me and guided me in every step through my research, in addition to my parents Mohammad Fateh Alhafez and Fathiah Alagha who stood beside me in my ambition for completing my study.

First of all, I would like to thank Dr. Frederic Tripet for his successful supervision, aiding and being patient through the last 4 years as well as co-supervisor Dr Srabasti J.Chakravorty for her help in giving me comments, suggestions and guide. Thanks to the Panel team for the very critical discussion for my work: Lauren Cator, Paul Eggleston and the chair Peter Thomas.

Many thank for my colleagues in Tripet group: Fred Antwi, Esther Ekechukwu and Douglas Graeme Paton and staff working in the laboratory both Ann Underhill and to Chris Bain were always available.

Special thanks for the Ministry of the Higher Education of Syrian and Keele.University for funding this work.

Newcastle-under-Lyme,

The United Kingdom

31st January 2016

CHAPTER ONE

Introduction to Malaria mosquitoes

1.1. The mysterious history of human's small enemy

Studies confirmed that mosquitoes originated in the Cretaceous era when Poinar (2000) and his colleagues found some studies refer to fossils of mosquitoes belonging to the Culicidae family in Canadian Cretaceous amber (Table 1) (Poinar et al. 2000; Lukashevich & Mostovski 2003). On basis of Wilson studies the fossils age date back to around the time of continental drift (Wilson 1963 Cited in (Rai 1999). This in turn gave information about the origin of early mosquitoes geographical subdivision and isolation which led to the first events of speciation (Rai 1999). In general, the order Diptera includes the Culicidae family that includes three subfamilies, Anophelinae, Toxorhynchitinae and Culicinae. This family includes more than 3300 species of mosquitoes (Russell & Rozeboom 1943; Service 2004). Based on morphological characters, they are divided in the subfamily Anophelinae which comprises three genera: *Anopheles* Meigen is distributed globally and consists of the majority of anopheline species, *Bironella* Theobald genus exists in Australia, and *Chagasia* Cruz genus spans over tropical areas (Russell & Rozeboom 1943; Harbach 2004).

Several species have annoying traits for mammals, when they suck blood from man and other animals causing irritant bites, and sometimes lead to dangerous diseases due to transmitting pathogens of human malaria, dengue, viral infections, and other fatal illnesses (Rueda 2008). The most pathogenic species belong to the subgenera *Culex* (causes Wuchereria bancrofti and arboviruses), *Aedes* (causes yellow fever, dengue and transmits encephalitis viruses), *Ochlerotatus* (transmits filariasis and encephalitis viruses), *Haemagogus* and *Sabethes* both cause yellow fever. On the other hand, some species do

not transmit any pathogens but they could cause allergic reactions as a result of their bites (Service 2004). Despite the serious infections transmitted by the later genera, *Anopheles* species are considered as the most important vectors because they transmit malaria , and also their species distribute all over the world (Service 2004). According to records, Anopheline mosquitoes were found in Europe, North Africa, Middle East, India subcontinent, South-East Asia, Mexico, Central America, South America, Australia and Sub-Saharan Africa (Fig. 1) (Service 2004).

Table 1.1: Fossils described as mosquitoes from different eras and areas (Poinar et al. 2000)

Species	Age and location	Reference , notes
Recent		
Aedes ciliaris (LINNAEUS)	Copal, Sweden	Linnaeus, 1767; Bloch, 1776; *nomen dubium* (Natvig, 1948); possibly same as *Aedes cinereus* MEIGEN (Edwards, 1932; Natvig, 1948).
Culex flavus GISTL	Copal, Brazil	Gistl, 1831
Culex loewi GIEBEL	Copal, Africa	Giebel, 1862; this is *Toxorhynchites brevipalpis* THEBALD 1901 (an extant species); *loewi* has been suppressed by the International Commission on Zoological Nomenclature (Opinion 1213) (Melville, 1982).
Culex tanzaniae	Copal, Tanzania	Capasso, 1991; *nomen nudum*
Tertiary		
Aedes petrifactellus (COCKERELL)	Oligocene, Isle of Wight, England	Cockerell, 1915; synonym of *A. protolepis* (Cock.) (Edwards, 1923).
Aedes protolepis (COCKERELL)	Oligocene, Isle of Wight, England	Cockerell, 1915
Anopheles? rottensis STATZ	Oligocene, Rott, Germany	Statz, 1944
Culex damnatorum SCUDDER	Eocene, Green River, Wyoming	Scudder, 1890
Culex erikae SZAD. AND SZAD.	Eocene, Poland	Szadziewski and Szadziewska, 1985
Culex pipiens LINNAEUS	Eocene, Poland	Keilbach, 1982; *Culex pipiens* is a recent species and further studies of the fossil may show it to be *C. erikae* or an undescribed species
Culex proavitus SCUDDER	Eocene, Fossil Canyon, Utah	Scudder, 1877; a psychodid (Edwards, 1923)
Culex protorhinus COCKERELL	Oligocene, Isle of Wight, England	Cockerell, 1915; generic position doubtful (Edwards, 1932)
Culex vectensis EDWARDS	Oligocene, Isle of Wight, England	Edwards, 1923
Culex winchesteri COCKERELL	Eocene, Cathedral Bluffs, Colorado	Cockerell, 1919
Culicites tertiarius HEYDEN	Upper Oligocene, Germany	Von Heyden, 1862; a chaoborid (Edwards, 1923)
Mansonia cockerelli (EDWARDS)	Oligocene, Isle of Wight, England	Edwards, 1923
Mansonia martinii STATZ	Oligocene, Rott, Germany	Statz, 1944
Mansonia varivestita STATZ	Oligocene, Rott, Germany	Statz, 1944
Neoculicites arvenensis (PITON)	Oligocene, Lac Chambon, France	Piton, 1936; Evenhuis, 1994
Neoculicites ceyx (HEYDEN)	Upper Oligocene, Germany	Von Heyden, 1870; possibly an *Aedes* (Edwards, 1923); Evenhuis, 1994
Neoculicites depereti (MEUNIER)	Upper Oligocene, Aix, France	Meunier, 1917; Evenhuis, 1994

Mesozoic		
Amblylexis gibberata BODE	Jurassic, Germany	Bode, 1953; none of the genera by Bode listed here are considered assignable to family by Carpenter (1992), and none were considered to be Culicidae by Knight and Stone (1977)
Amianta eurycephala BODE	Jurassic, Germany	Bode, 1953
Amphipromeca acuta BODE	Jurassic, Germany	Bode, 1953
Apistogrypotes inflexa BODE	Jurassic, Germany	Bode, 1953
Asioculicus damiaoensis HONG	Jurassic-Cretaceous, China	Anonymous,1976; not a mosquito (present study)
Asioculicus longipodus HONG AND WANG	Cretaceous, China	Anonymous, 1976; not a mosquito (present study)
Chironomaptera gregaria (GRAGAU)	Cretaceous, China	Kalugina, 1980; family assignment doubtful (Carpenter, 1992); listed as a Chaoboridae (Evenhuis, 1994)
Culex fossilis BRODIE	Jurassic, England	Brodie, 1845; a chironomid (Edwards, 1923)
Cormophora arucaeformis BODE	Jurassic, Germany	Bode, 1953
Culiciscolex gibberatus BODE	Jurassic, Germany	Bode, 1953
Cyrtomides maculatus BODE	Jurassic, Germany	Bode, 1953
Ellipes laesa BODE	Jurassic, Germany	Bode, 1953
Empidocampe retrocrassata BODE	Jurassic, Germany	Bode, 1953
Propexis incerta BODE	Jurassic, Germany	Bode, 1953
Rhopaloscolex brevis BODE	Jurassic, Germany	Bode, 1953
*R. longu*s BODE	Jurassic, Germany	Bode, 1953
Sphallonymphites decuratus BODE	Jurassic, Germany	Bode, 1953

There are about 70 species of Anopheline that can transmit malaria and only 40 of them are considered as important species (Service 2004). Despite some species being primary vectors in some areas, they can also be secondary vectors in other places (Service 2004). Most mosquitoes are active in specific times when they can be active outdoors and their mouth parts allows them to bite through light clothing (Service 2004). Studies estimate that around 300 million clinical cases of malaria occur around the world, and about one million people die each year due to the infection (Knell 1991; WHO 1999). Almost 90% of these deaths are found in Sub-Saharan Africa in tropical areas that are rainy for most of the year. Under such conditions malaria transmission is continuing throughout the year, affecting mostly young children with a ratio one in five children deaths/year (Knell 1991; WHO 1999). In the drier tropical areas where rainfall is seasonal, malaria is transmitted only in the wet seasons (Knell 1991; WHO 1999). On the other hand, there are no reports of

malaria in Europe except for some cases of imported malaria that are usually reported when travellers to the tropical areas carry the infection back to their countries. For instance, in Britain the number of imported malaria cases has significantly increased from 101 in 1970 to 2212 in 1985 (Knell 1991; WHO 1999).

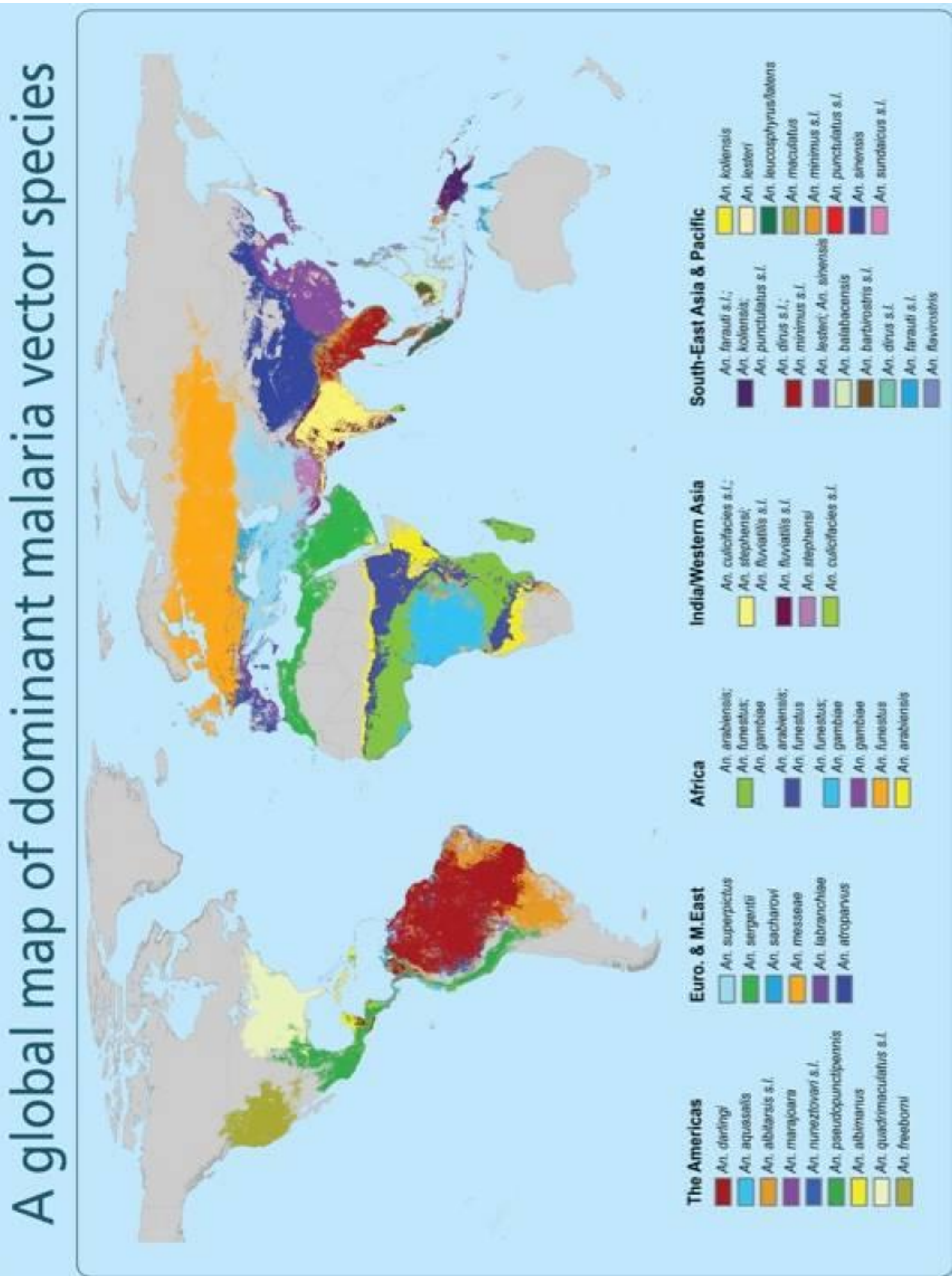


Figure 1.1. Malaria vector species and their distribution around the world - many varieties within the Anopheles genus that carry malaria infection around the world (Sinka et al. 2012).

1.2. Malaria infection (plasmodium life cycle)

Malaria disease is caused by intracellular parasites that are called *Plasmodium* species (Vlachou et al. 2006). They parasitize generally vertebrate animals as their intermediate host, whereas the mosquitoes are their final host (Vlachou et al. 2006). About 200 species of *Plasmodium* were classified that can infect birds, reptiles and mammals (Nishimoto et al. 2008; Al-Mekhlafi et al. 2010). Five species of them could cause malaria to humans: *P. falciparum, P. vivax, P. malariae, P. ovale, and P. knowlesi* (Nishimoto et al. 2008; Al-Mekhlafi et al. 2010). *P. falciparum* has the most virulent progression and it stands behind the majority of deaths in humans (Nishimoto et al. 2008; Al-Mekhlafi et al. 2010). Understanding the parasite life cycle is important in order to support our knowledge about the disease and to know how it develops, so that we can find suitable methods to control it. The *Plasmodium* life cycle includes four stages: fertilization is a sexual stage that occurs in the stomach of a mosquito, sporogony is the first asexual stage and takes place in the stomach wall and mosquitoes body, hepatic schizogony is the second asexual stage found in the liver of humans, erythrocyte schizogony is the third asexual stage and occurs in blood cells of humans (Knell 1991).

The life cycle starts when a female-mosquito takes a blood meal including *Plasmodium* male and female gametocytes from an infected vertebrate host (Knell 1991). Following ingestion, the sexual stages of the *Plasmodium* parasites start when the male-gametes produce eight flagella within minutes in a process called exflagellation and release them into the plasma of blood (Knell 1991). Thereafter, fertilization occurs between the male and female macrogametes to produce zygotes, which turn within a few hours into mobile ookinetes (Knell 1991). These ookinetes go through mid-gut epithelium and arrive to the basal side and penetrate the basal membrane but cannot pass within the basal lamina, and here they round up to start the next phase (Knell 1991). At that time, the ookinetes

grow quickly and start multiplying and transforming into oocysts (Knell 1991). The oocysts in turn pass through the cytoplasm of mid-gut epithelium cells. This process induces the apoptosis and extruded of damaged cells (Knell 1991). After about one week, the oocysts develop into thousands of worm-shaped sporozoites in the intercellular space between the basal lamina and mid-gut epithelium (Knell 1991). After two weeks, the oocysts burst releasing thousands of sporozoites that move through the mosquito's haemocoel to reach its salivary glands and stay there ready to be transferred into a new host through a mosquito bite. (Fig.1.2).

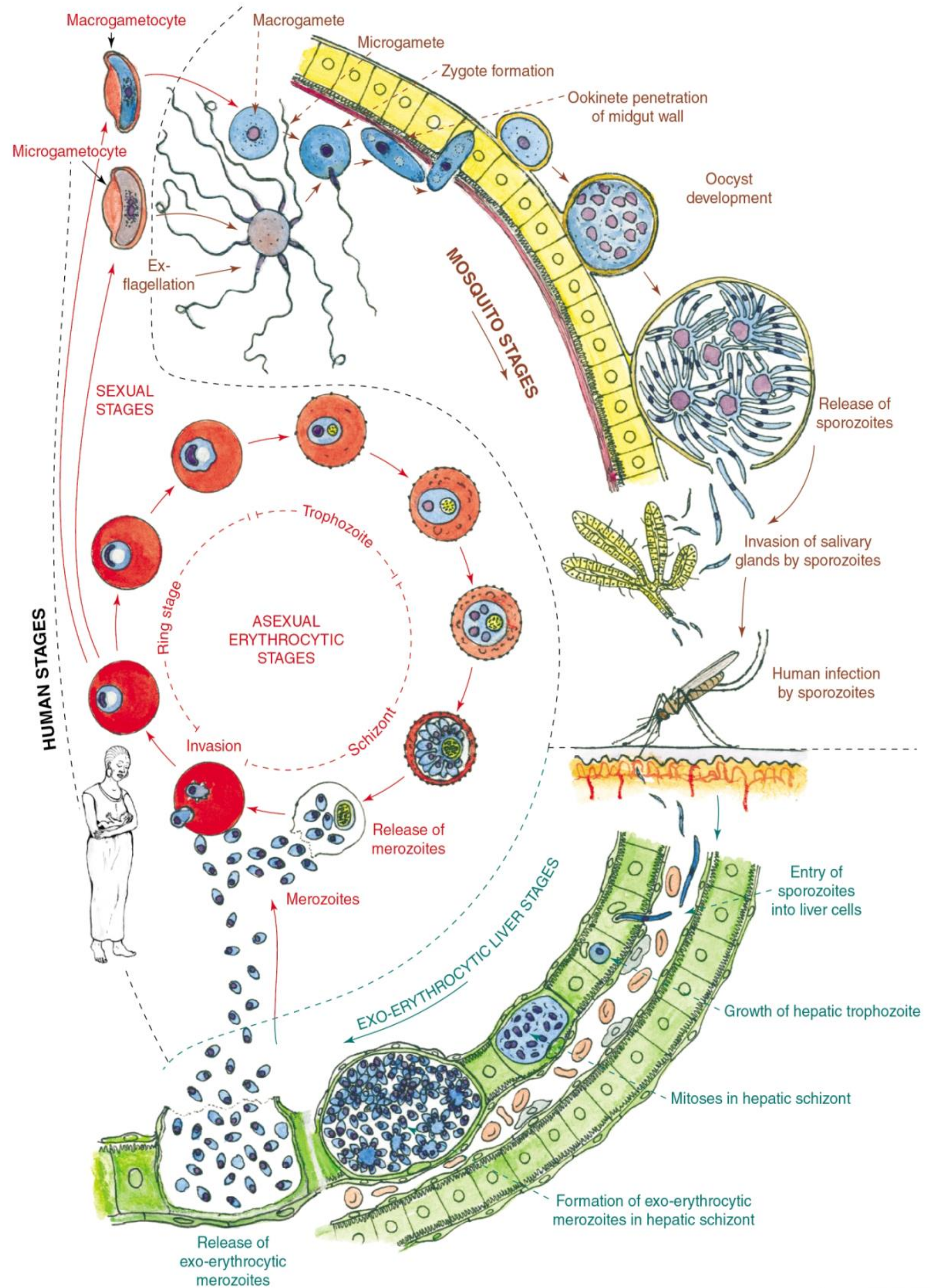


Figure 1.2. Plasmodium falciparum life cycle in human and mosquito - *Plasmodium* life cycle passes through four phases: fertilization is a sexual phase in the stomach mosquito, sporogony is the first asexual phase in the stomach wall and mosquitoes body, hepatic schizogony is the second asexual phase in the liver of humans, erythrocytic schizogony is the third asexual phase in blood cells of humans (Bannister & Mitchell 2003).

As soon as the mosquito take a blood meal again, sporozoites are transferred into the blood stream of the a host and they start a new stage of their life in human liver cells, where they become hepatic trophozoites and develop quickly into hepatic schizonts within two days (Knell 1991). At this time, they start producing numerous invasive merozoites (Knell 1991). After one week from the infection, the mature merozoites move from hepatic vessel to the blood stream, where they initiate another asexual stage (Knell 1991). So far, the infected person does not have any symptoms, therefore this period is called the incubation period which covers the whole of hepatic schizogony and the first few cycles of erythrocyte schizogony (Knell 1991). The third asexual phase starts when the merozoites go through erythrocytes and form ring-shapes that grow into trophozoites (Knell 1991). Within two to three days these trophozoites divide into rosette-shaped shizonts that in turn release merozoites from the exploded red blood cells and re-inter red blood cells to continue the asexual life cycle (Knell 1991). At this time, the multiplication in the blood cells causes periodic fever and as the number of the merozoites increases the person becomes more ill (Knell 1991). After a few blood cycles, a number of merozoites mature to male and female gametocytes within four days, and they stay ready to be taken by mosquitoes to start a new life cycle (Knell 1991).

1.3. *Anopheles gambiae* life cycle

Mosquitoes can be described as small insects with body size that varies between 3mm to 19mm in different species. In all mosquitoes species, the body generally is divided into the: head, thorax and abdomen (Fig.1.3) (Service 2004). Mosquitoes have a pair of fore wings and two small halteres (Service 2004). However, the mosquitoes use only the fore wings for flying (Service 2004). They are different from other flies by holding prominent proboscis and many scales on their: bodies, wing veins and on the posterior edge of their wings (Fig.1.3) (Service 2004). The males of mosquitoes can be distinguished from the

females by having feathery antennae (Service 2004). In addition, the last abdominal segment has a pair of claspers in males, and pair of small finger-like cerci in females (Service 2004).

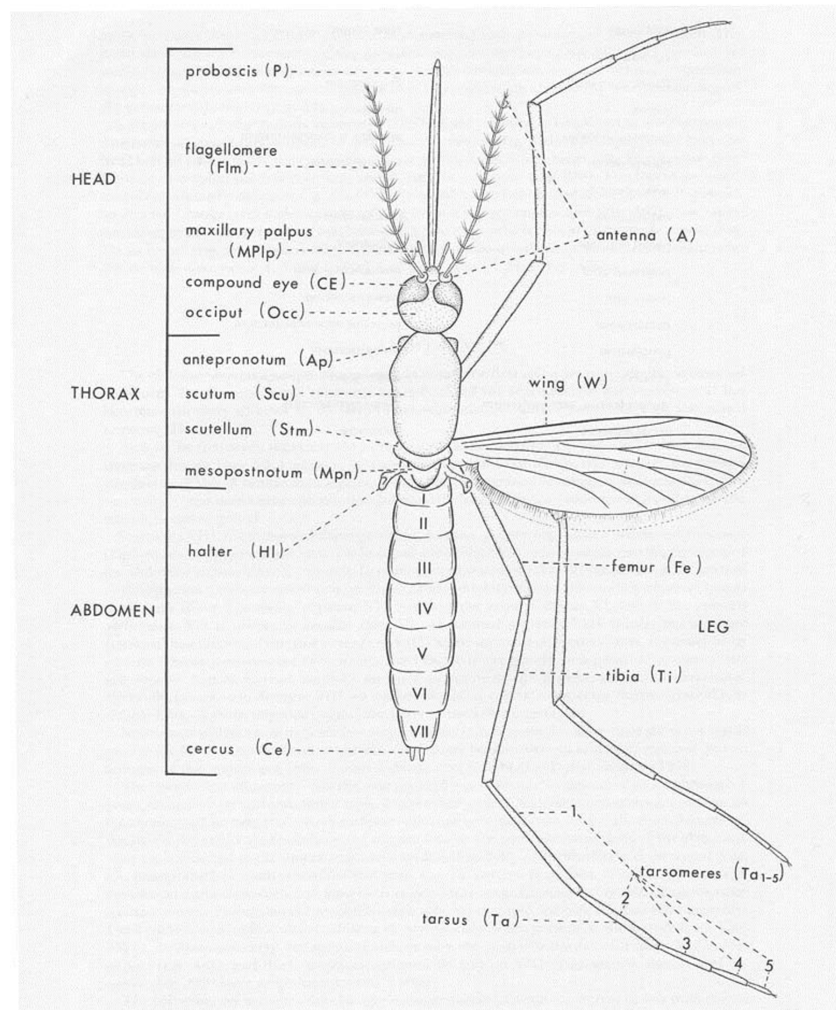


Figure 1.3. The female adult mosquito body - consists of a head, thorax and abdomen (Service 2004).

The genus *Anopheles* is characterized by the distinctive resting position where they put the head down, the body at an angle, and they raise the hind legs (Knell 1991). In addition, they have blackish spots and pale scales on the wing veins, and blackish maxillary palps that are as long as the proboscis (Fig. 1.4) (Knell 1991).

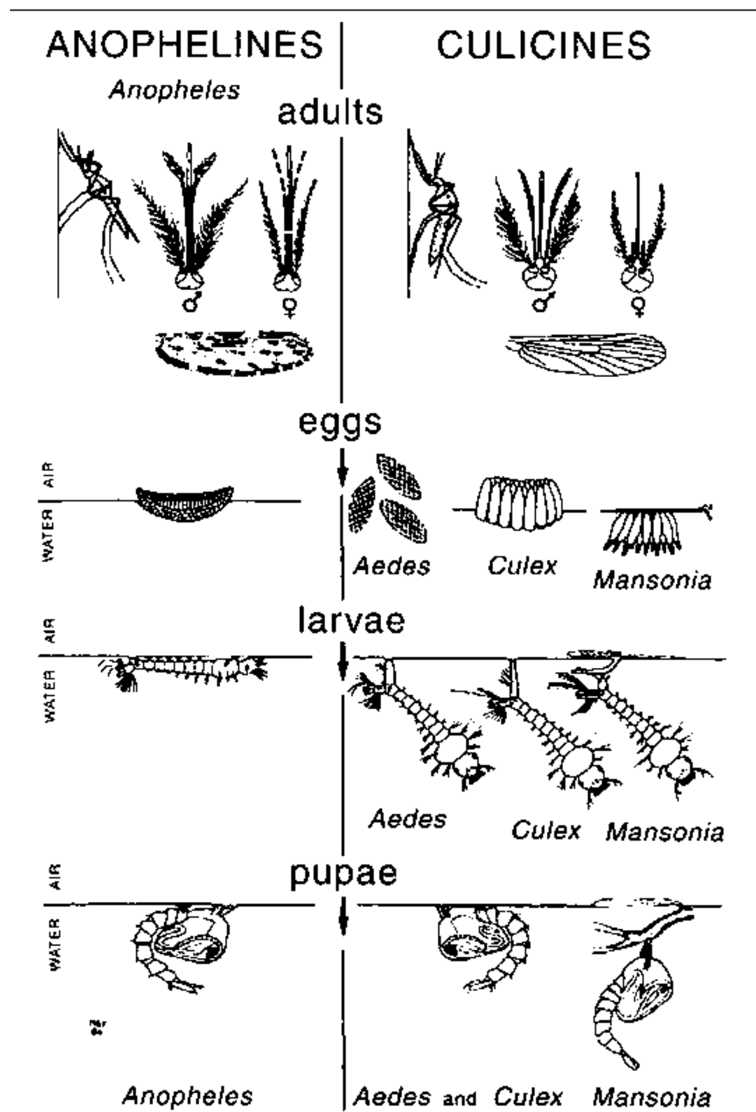


Figure 1.4. The differences between Anopheles and other Mosquitoes in different life stages (Service 2004).

The life cycle of *Anopheles* mosquitoes consists of four stages that take about 7-21 days depending on the temperature (Knell 1991). In *Anopheles* mosquitoes, both males and females can feed on nectar or fruit juices, but the females often need blood meal in order to produce eggs (Service 2004), therefore only the females can transmit malaria (Knell 1991; Service 2004). After a blood meal, the gravid females lay blackish eggs individually at dusk after resting for two to three days (Cook & Zumla 2003). Generally, the oviposition

sites are determined by females, and vary between species (Cook & Zumla 2003). Preferred oviposition sites are usually uncontaminated with plants or animal pollutants (Cook & Zumla 2003; Service 2004). They can lay eggs in damp soil or vegetation, moist holes or directly in water (Cook & Zumla 2003; Service 2004). The eggs of mosquitos look like boats, and they have air floats to stay on water surface (Cook & Zumla 2003; Service 2004). Their hatching period varies between 2-20 days depending on the temperature (Cook & Zumla 2003; Service 2004). After hatching, the larvae stay parallel to the water surface, and can be disturbed easily and move in water (Service 2004). They feed on small organisms in water and breathe through spiracles (Service 2004). Depending on the surrounding conditions, they can grow through three larval stages within 1-4 weeks to become coma-shaped pupae (Cook & Zumla 2003; Service 2004). The pupae do not feed and only breathe using two air-trumpets staying on the water surface unless they are disturbed, in which case they move randomly (Cook & Zumla 2003; Service 2004). After about two days in warm weather and about two weeks in colder areas, the adults emerge and stay with an angle on the water surface, and keep their proboscis and abdomen in a straight line to become harder and drier before they fly (Fig.1.5) (Knell 1991; Cook & Zumla 2003; Service 2004). After two to three days, the adults start seeking mates that they can fertilize to start a new life cycle (Knell 1991; Cook & Zumla 2003; Service 2004).

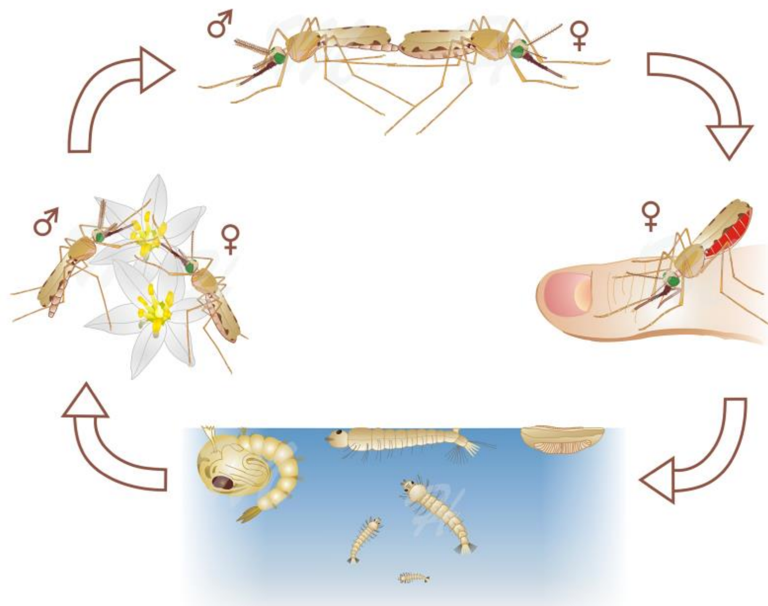


Figure 1.5. *Anopheles gambiae* life cycle - *Anopheles gambiae* life cycle passes through four phases: larvae, pupae and adults (Anitram 2012).

1.4. Malaria control history

Although there have been many difficulties, malaria eradication is not impossible and it is achievable. These beliefs increased after 1945, when researchers found that starting new global projects to eradicate malaria is feasible (Knell 1991). In this context, they specified the host that is the human and in some cases primates, as well as they determined the vectors that are the *Anopheles* females that need to rest after every blood meal (Knell 1991). They also determined the sporogony duration of *Plasmodium*, which is about 10 – 30 days depending on the temperature (Knell 1991). During this time, the carrier female needs 4 blood meals within the rest period, and it is possible to be destroyed before transmitting *Plasmodium* to another human (Knell 1991). This information helped the Eighth World Health Assembly to start their plan for global malaria eradication in 1955, and the World Health Organisation supported this by starting a basic plan for every national eradication campaign in all the affected countries (Knell 1991). During that time, DDT was an effective option that is non-toxic to humans, cheap and persistent with no

need to apply frequently (Knell 1991). Unfortunately the mosquitos resistance started to appear in the beginning of the campaign and it took about 5 years to develop (Knell 1991). By 1970, malaria was destroyed in whole the Europe, South America, Caribbean, North America, Australia, several Middle Eastern countries, Japan, Taiwan, and Singapore (Knell 1991). Nevertheless, the infections continued in many tropical countries due to some difficulties that faced the campaign in 1970s, e.g.: the mosquitoes and plasmodium resistance increased rapidly against different treatment, in addition to inflation and political problems (Knell 1991). Therefore, the number of infections increased again sharply to 2.5 times between 1973 and 1977 according to WHO. Today, the attempts are continued to decrease the number of malaria infections in the main affected countries in Africa (Knell 1991).

In general, to be able to control any disease we need to know the strengths and weaknesses. The strengths of malaria disease lay in the strong productivity for both the mosquitoes and parasites, whereas the weakness is the long time that the parasite takes inside the mosquito to become infectious (Knell 1991; Service 2004). Therefore, the control process focuses on three directions: the human that can be educated to avoid infection by using bed nets, the parasite that requires treatment with several drugs within several days but they are no longer effective due to the parasite resistance, and the mosquitoes control (Knell 1991; Service 2004). However, the mosquitoes control gave better results in a shorter time by attacking the larvae in breeding places (Knell 1991; Service 2004). This method was useful in small towns where the water collections are few relatively, but can fail if any breeding places are missed (Knell 1991; Service 2004). Mosquitoes control can be achieved also by attacking the females using insecticides, but this method is still weak due to the mosquito's resistance (Knell 1991; Service 2004).

Other methods have been used like environmental control by reducing the number of mosquitoes breeding places (Service 2004). In this regard, various types of biological control methods were successful for example: breeding a type of fish that eats mosquitos' larvae, or use several types of mosquito pathogens and predators (viruses, bacteria, protozoa, fungi and nematode worms) (Knell 1991). Additionally, mosquitoes' males can be sterilized using chemicals or X ray, thereafter release them in the field. These males are able to mate and give eggs that include dead embryos. These methods were successful in Aedes mosquitoes (Bellini, Balestrino, et al. 2013; Bellini, Medici, et al. 2013) but did not meet the expected results in *An. gambia* due to the reduction in their ability to compete with field mosquitoes because of the effect of radiation dosages (Helinski & Knols 2008; Helinski et al. 2009). However, some alternative control technologies including entomopathogenic fungi and viral paratransgenesis showed good results in laboratory but still need to be assessed in the field (Catteruccia 2007; Takken 2009). In this regard, using Wolbachia bacteria is promising method by transmitting endosymbionts that destroy the embryos of the eggs that are produced when the infected laboratory males mated with the field females (Werren et al. 2008). These new approaches to succeed require more information about the target mosquitos' biology, and still need to be assessed in the field.

1.5. Reproductive system in females and males of *An. gambiae*

There is not a lot of information for the sexual behaviour of Anopheline mosquitoes (Howell & Knols 2009). After the males emerge, their sexual organs and antennal fibrillae must mature before they can seek females. Therefore, during the first 12-24h the male's terminalia turns upside down about 180° in order to be in an accurate position for mating (Cook & Zumla 2003; Howell & Knols 2009). Thereafter, male becomes sexually active about 48h after emergence and fully active within 3-7 days after emergence (Mahmood & Reisen 1982).

The male reproductive system includes two pear-shaped testes that are placed in the 5th and 6th abdominal segments (Clements 1992). Each testis consists of a follicle surrounded by a sheath, a pair of vasa efferentia that are derived from testis rudiments, a pair of seminal vesicles that have slightly thicker wall than the vasa efferentia, ejaculatory duct and two accessory glands that are derived from imaginal disks of the ninth abdominal segment (Clements 1992). The male produces immature spermatozoa that are stored in spermatocytes in the testes (Huho et al. 2006). After the spermatocytes mature, they break down to release the mature spermatozoa that in turn are stored again in a sperm reservoir (Huho et al. 2006). When the male starts mating, these spermatocytes leave the reservoir and pass the vas vesicle to the seminal vesicle, and then the ejaculatory duct before they are transmitted to a female (Huho et al. 2006) (Fig.1.6).

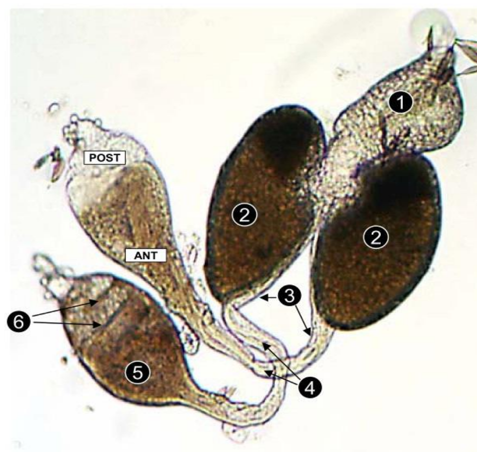


Figure 1.6. Male reproductive system in *An. gambiae* – 1) Ejaculatory ducts; 2) Accessory glands; 3) Seminal vesicle; 4) Vas efferentia; 5) Sperm reservoir 6) Spermatocysts (Huho et al. 2006).

Generally, females are able to mate at their night of emergence (Mahmood & Reisen 1982). The female reproductive system includes pair of ovaries that are situated in the posterior section of the 4th and 5th segment of the abdomen. Each one composed of many egg-follicles that pass the follicular tubes through their two development stages. Usually

their second stage is commenced when the female gets the blood meal (Clements 1992). The ovaries are connected to lateral oviducts that continue to common oviduct, then form atrium (Clements 1992). The atrium connects to the sperm duct that related to spermatheca (Clements 1992) (Fig.1.7).

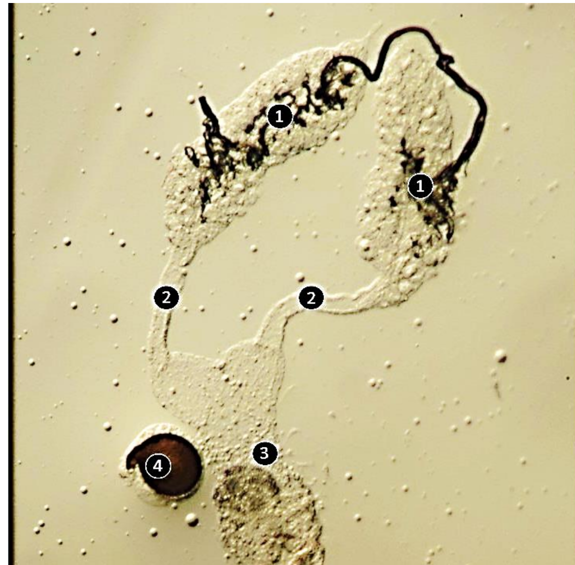


Figure 1.7. Female reproductive system in *An. gambiae* – 1) Ovaries, 2) Lateral oviduct, 3) Common oviduct, 4) Spermatheca.

1.6. Mating behaviour

Generally, *An. gambiae* s.l. individuals prefer to mate at dusk (Cook & Zumla 2003) which can be associated with the inherent circadian rhythm in Anophelines (Charlwood & Jones 1979). Charlwood (1979) found a reduction in the males inseminations when he tried to make changes in the circadian rhythm (Charlwood & Jones 1979). This interaction between swarming stimulation, circadian rhythm and light level can be a sign of the start of mating and also it might play a main role as a barrier between different species (Howell & Knols 2009). In addition, *Anopheles* mosquitoes are believed to swarm at various heights in different places. Several studies showed that the different species of mosquitoes

tend to swarm during different times over different markers (Charlwood et al. 2003; Howell & Knols 2009). Therefore, mosquitoes situate themselves in specific places at specific times to ensure that they are available for mating (Howell & Knols 2009). It seems impossible for species swarms to mix even in areas where sympatric species occur, for instance: although M- and S-forms of *An. gambiae* s.l. swarm at similar times, they are rarely found in the same place (Diabate et al. 2003; Diabaté et al. 2006).

When dusk comes, the mosquitoes males antennal fibrillae erect at specific times in different species, one hour later the males leave their resting areas (Charlwood & Jones 1979). Swarms are initially started by one or two males who move with zigzag-like flight patterns (Diabate et al. 2003; Dao et al. 2008). Then more males join them to make a loose cloud (Diabate et al. 2003; Dao et al. 2008). Thereafter, the males begin to move in a way that makes a denser cloud, which start moving as a single unit (Reisen & Aslamkhan 1975; Howell & Knols 2009). The males start swarming and females join the swarms later (Reisen & Aslamkhan 1975; Diabate et al. 2003). These swarms usually include between 5-5000 males (Manoukis et al. 2009). The males usually prefer to fly in the centre to increase their chances to mate by either staying in the space that females frequently pass through or being in a known place by females (Howell & Knols 2009). It was proved that the males become active before the females due to a different phototactic response which gives the males time to optimize their chances for mating (Howell & Knols 2009). The copulations start about 5 to 20min after the swarm begins (Howell & Knols 2009). During this time, females and males try to harmonize their wing beats with the nearby mosquitoes to check if they are the same or opposite sex (Pennetier et al. 2010). During the copulation process, the male continues approaching the female, then he uses the tarsal claws on his legs to catch the female legs (Clements 1992). Thereafter, he swings himself under her, bents his abdomen and his genitalia to interlock them with female's ones (Clements 1992).

Later the male starts releasing the female's legs to take an end-to-end position, they stay so for about 13-14sec, then the male fluctuates himself for a further 2-3sec until he opens his claspers and leaves the female to re-join the swarm again (Clements 1992; Howell & Knols 2009) (Fig.1.8). In contrast, the female commonly does not mate again due to the mating plug that is inserted by the male during the mating process (Shutt et al. 2010).

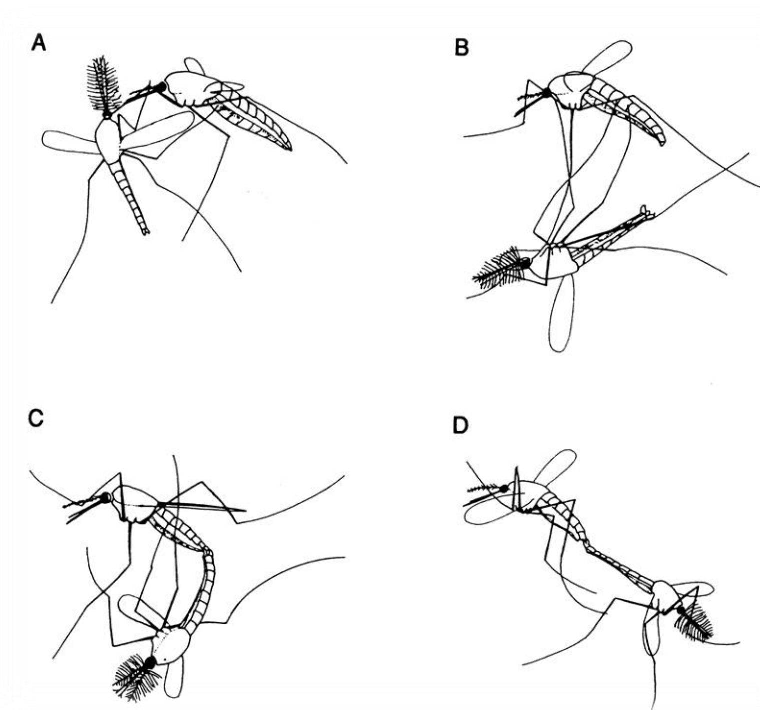


Figure 1.8. Mating process in *Anopheles gambiae* – A) A male continues approaching a female, B) The male uses tarsal claws on his legs to catch the female legs, C) The male bents his abdomen and his genitalia to interlock them with female's ones, D) The male starts releasing the female's legs to take an end-to-end position (Clements 1992).

Finally, the number of mating decrease in the swarm, when the males leave it to rest and replenish energy via nectar feeding for flight and daily survival (Charlwood & Jones 1979; Howell & Knols 2009). This is because the males use about 50% of their energy reserves during 25min of swarming (Maïga, Niang, Simon P Sawadogo, et al. 2014). The females leave the swarm to chase their hosts for a blood meal which enables them to produce eggs

(Cook & Zumla 2003). Usually, some anopheline mosquitoes rest indoors on the walls or ceilings before and after feeding, therefore this type of mosquitoes are called endophilic (Knell 1991). Some of them rest outdoors in a variety of shelters for example: trees, caves and rock fissures, and are called exophilic (Service 2004).

1.7. The *Anopheles gambiae* complex

Mosquitoes control methods are important aspect of malaria control programs. These control methods are influenced by mosquitos adaptation to different environments including the ones adapted to human activities changes (Coluzzi et al. 1978). In this context, *An. gambiae* s.l. species are considered as the most important species to be studied (Coluzzi et al. 1978).

A few decades ago as a result of previous research, it was reported that *Anopheles* mosquitoes included several distinct species that were isolated by pre-mating and post-mating mechanisms (Coluzzi et al. 2002; Slotman et al. 2006; Slotman et al. 2007) . These species were known as *Anopheles gambiae* complex, or *An. gambiae* s.l. (Slotman and Tripet et al, 2007). Based on the distribution of the inversions on the right and left arm of the second chromosome, they found that this complex consisted of seven morphologically indistinguishable species, that have genetic and eco-ethological differences affecting their ability to transmit malaria (della Torre et al. 2002; Coluzzi et al. 2002). These species were: *An. gambiae sensu stricto*, *An. arabiensis*, *An. quadriannulatus* A and B, *An. merus*, *An. melas* and *An. bwambae* (Coluzzi,. et al 2002). Later, another taxon called *Anopheles comorensis* was discovered by Brunhes (1997) in the Indian Ocean islands of the Comoros in the Mozambican channel (Brunhes et al. 1997). However, the relation between *An. comorensis* and *An. gambiae* complex still unclear and more studies are in progress to reveal it (Coetzee et al. 2013). In addition, Hunt (1998) found differences between *An.*

quadriannulatus species B from Ethiopia and *An. quadriannulatus* species A from southern Africa. He found that the hybrid males between these two species are sterile (Hunt et al. 1998). Therefore they assigned the Ethiopian member of the *An. gambiae* complex as *An. Amharicus* (Coetzee et al. 2013). Subsequently, due to the genetic and ecological differences within *An. gambiae senso strecto* population, it has been split into *An. coluzzii* and *An. gambiae s.s.* resulting in 8 sister species within the complex (Coetzee et al. 2013). In general, The crosses between the field collected samples showed that the F1 hybrid males were sterile whereas the females were viable (Weetman et al. 2014). Therefore, the sibling species of *An. gambiae* s.l. were thought to mate assortatively but they are incompletely reproductively isolated, which indicates that 'a narrow channel for genetic communication between these species is still open and was presumably wider in the past' (Coluzzi et al., 2002).

Generally, *An. gambiae* s.l. species vary in their adaptation to different environments and this reduces competition between them for limiting resources (della Torre et al. 2005). There are four fresh water species: *An. coluzzii*, *An. gambiae s.s.*, *An. arabiensis*, and *An. quadriannulatus*, alongside with three brackish water that are: *An. bwambae*, *An. melas*, and *An. merus* (Lehmann & Diabate 2008). This adaptation to different climates is due to their chromosomal structure that consists of two pairs of autosomes and one pair of sex chromosome. The right (2R) and left (2L) arm of the second chromosome plays the main role through chromosomal inversions in ecological adaptation within the mosquitoes population (Coluzzi et al., 2002). Moreover, most scientists agree that the *Anopheles gambiae* complex species arose from the same ancestor. The ancestral taxon could have the chromosome arrangement similar to *An. quadriannulatus* due to its chromosomal structure that have central position amongst other species in the complex, beside its relict distribution and its tolerance to moderate conditions (Coluzzi et al., 2002).

An. gambiae.s.s., An. coluzzii and *An. arabiensis* are considered as the species most adapted to human changes in the environment, and they spread widely in the Afrotropical areas (Coluzzi et al., 1978). Additionally, these species have the highest number of inversion polymorphisms, followed by *An. melas* (Coluzzi et al. 1978; Coluzzi et al. 2002), whereas, a few inversion polymorphisms have been found in *An. bwambae* and *An. quadriannulatus* species A, and no inversion polymorphisms have been recorded in *An. quadriannulatus* species B and *An. merus* (Coluzzi et al. 2002). It is believed that chromosomal differentiation may contribute to the speciation processes in these species. For example: despite their similar ecological and morphological traits, *An. melas* exist in West Africa whereas *An. merus* has been found in East Africa. On the other hand, *An. gambiae* and *An. arabiensis* are the most similar species ecologically in their adaptation to domestic environments (Coluzzi et al. 2002). More studies are in process to reveal the role of chromosomal inversions in adaptation differences and speciation process.

1.8. Chromosomal forms

Anopheles gambiae (that includes the sibling species currently known as *An. coluzzii* and *gambiae s.s.*) is reported as the most important vector due to its ability to spread and exploit both temporary and man-made breeding sites as well as its ability to distribute in different climatic regions (Coluzzi et al., 1978; della Torre et al., 2002; della Torre, Tu, & Petrarca, 2005). These factors combined with its capacity to transmit malaria, which make it an extremely efficient vector of malaria and wide-open subject for researchers to find alternative different methods to control malaria disease.

Della Torre (2005) believes that human influence in the environment through the last thousand years has created changes in the nature of west African regions, and has found new ecological habitats for mosquitoes to avoid intraspecific competition (della Torre et al. 2005). This proves that speciation process started in West Africa by establishing co-

adapted chromosomal inversions that led to create new species within *An. gambiae* s.l., and this enables it to distribute widely throughout Africa (della Torre et al. 2005; Boulesteix et al. 2007). Coluzzi and collaborators revealed in their studies on *An. coluzzii* and *An. gambiae* s.s. in East and West Africa that populations in forest areas often have standard chromosomal arrangements, whereas savannah samples were different and have more inversions that also increase in the frequency towards drier savannah (Coluzzi et al. 1985). Cytogenetic analyses of chromosomal polymorphisms within *An. coluzzii* and *An. gambiae* s.s. showed that the inversions are non-randomly distributed and they mostly exist on right (2R) and left (2L) arms of the second chromosome (Coluzzi et al. 1978; Coluzzi et al. 1985; Coluzzi et al. 2002). Consequently, these studies led to the description of five different chromosomal forms: Mopti, Bamako, Bissau, Forest and Savanna based on the geographical regions that they collected from (della Torre et al., 2005; Slotman et al, 2007; Lee et al., 2009). Therefore they are considered as indicators to different ecological habitats and some of these forms may also be partially reproductively isolated (della Torre et al., 2002).

Coluzzi and colleagues (1985, 2002) suggested a scenario of sympatric speciation between chromosomal forms in which chromosomal inversions played a central role by protecting gene associations arising within populations that were temporarily isolated geographically or/and in ecologically marginal zones that became subsequently stable. This led to assortative mating behaviour and reproductive isolation between the forms (Coluzzi et al. 1985; Coluzzi et al. 2002). This process started with Forest chromosomal form that is thought to be the ancestral form of *An. gambiae* due to its standard karyotype structure (Coluzzi et al. 1985) (Fig.1.9).

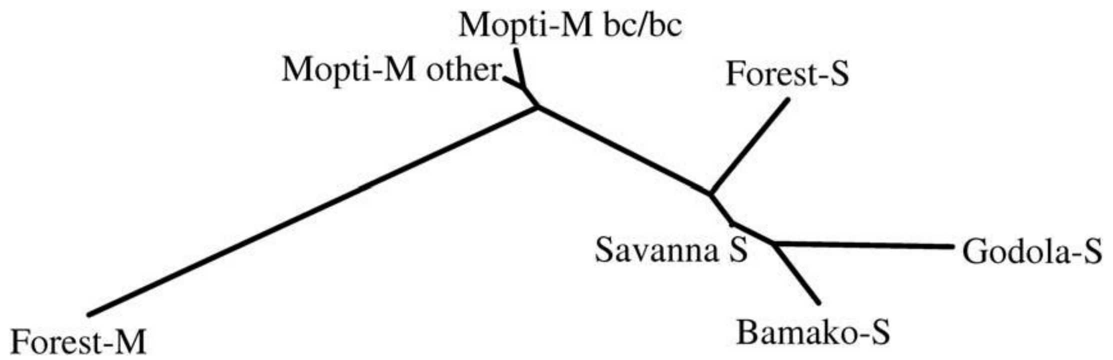


Figure 1.9. Unrooted phylogenetic tree (neighbour joining) of the seven groups of *An. gambiae* identified by Bayesian analysis (Lee et al. 2009).

Coluzzi (1985) also described the chromosomal forms based on the distribution of the most common chromosomal arrangements (Fig.1.10):

(1) Forest form has typical arrangement without inversions 2R+/+, 2L+/+, or by a single inversion polymorphism 2Rb, 2Rd or 2La, and this form is associated with humid tropical forest environments and the transitional environments between forest and humid savannah areas.

(2) Bissau is characterized by high frequencies of the 2Rd inversion and standard 2L+ arrangement, and it is found in coastal areas of West Africa and frequently associated with agriculture areas.

(3) Savanna shows high frequencies of 2Rb and 2La inversions as well as polymorphism involving the 2Rcu arrangements and polymorphism in the j, d and the rare k inversion. This form spreads widely across Africa and its frequency highly depends on season due to its association with the wet weather.

(4) Bamako is known by the fixed 2Rjcu arrangement and polymorphism in the 2Rb inversion, and is found basically in Southern Mali.

(5) Mopti usually shows high frequencies of 2Rbc, 2Ru and nearly fixed for 2La. This form is typically associated with arid savannah and Sahel savannah such as areas of Central Mali and also with man-made breeding sites and thus can maintain a relatively high population frequency even over the tropical dry season (Coluzzi et al., 1985; della Torre et al., 2002; Lanzaro & Tripet, 2003; della Torre et al., 2005). However, these different chromosomal forms overlap in different regions, and the hybrids are still expected in the mutual areas (Coluzzi et al. 1985).

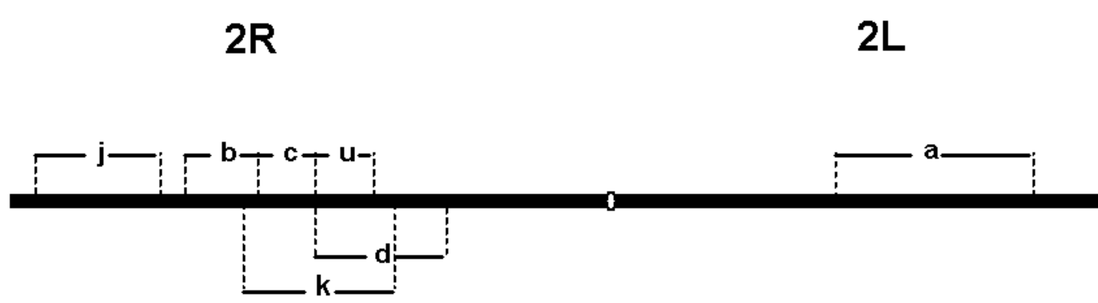


Figure 1.10. Position of the inversions on the second chromosome of *An. gambiae* s.l.- Five chromosomal forms were described based on the arrangements of such inversions described from the banding patterns of polytene chromosome (Lanzaro & Tripet 2003).

Karyotype frequencies studies using Hardy–Weinberg equilibrium in the areas where Bamako, Mopti and Savanna forms occur, showed that the frequencies significantly departure from the equilibrium (Coluzzi et al. 1978; Bryan et al. 1982; Taylor et al. 2001). Moreover, the hetero-karyotypes presence suggests hybridisation occurrence between Savanna/Mopti and Savanna/Bamako but there were no hybrids found between Bamako/Mopti (Coluzzi et al. 1978; Bryan et al. 1982; Taylor et al. 2001). However, there was no departure from the (H-W) equilibrium between these forms in Mali, indicating that these forms have a single gene pool in this area (Lanzaro & Lee 2013). Moreover, there

were about 26% of 2.459 individuals in Mali, and about 39% of 632 in Cameroon could not be defined as any of the described chromosomal forms (Lee et al. 2012). Recent comparative genomic studies found a lack of differences amongst the inversion regions of the different chromosomal forms (White et al. 2007; White et al. 2009). These results suggest that there is no evidence for the inversions' role in evolution of reproductive isolation between the forms (Lanzaro & Lee 2013).

1.9. Molecular forms

In 1997, a new PCR method was developed to attempt to differentiate the chromosomal forms within *Anopheles gambiae* complex based on single nucleotide polymorphisms in the intergenic spacer (IGS) near the centromere of the X chromosome. After extensive sequencing studies a nucleotide difference in the 3 end of the 28S coding region and part of the adjacent intergenic spacer (IGS) on the X chromosome was identified and characterized as C/C genotype in Mopti, and T/T genotype in Bamako, Savanna and Bissau (Favia et al. 1997). Based on these findings, they defined two morphologically undistinguishable M and S molecular forms (Favia et al. 1997; della Torre et al. 2005a). Subsequently, they considered both M and S molecular forms as distinct sibling species within the *Anopheles gambiae* complex due to the genetic and ecological differences, and they named the 'M form' as *Anopheles coluzzii*, whereas the 'S form' was named as *Anopheles gambiae s.s.* (Coetzee et al. 2013). Interestingly, this association between chromosomal and molecular forms was obvious in Burkina Faso and Mali, but it breaks down outside these areas due to existence of populations that have Savanna chromosomal arrangements but are characterised as M molecular form (della Torre et al., 2002).

1.10. Reproductive isolation and ecological adaptation in *An. gambiae s.s.* and *An. coluzzii*

The population structure of *An. gambiae* has been the subject for extensive studies that revealed complex patterns of gene flow and reproductive isolation. Despite *An. gambiae s.s.* and *An. coluzzii* species overlapping in many areas, *An. coluzzii* was mostly associated with more permanent man-made habitats and some regions of Central Africa, whereas *An. gambiae s.s.* can be found in rainy locations widely throughout Africa. Therefore, spatial swarm segregation has been found in different places (Lanzaro & Lee 2013). However, in Cameroon populations the Forest chromosomal form could be either M or S molecular form (della Torre et al. 2005; Lee et al. 2009). In this context, they accepted that there are Forest M and Forest S molecular forms that often live in sympatry (della Torre et al. 2005a; Lee et al. 2009). Therefore, in regard to the speciation process scenario, the Forest chromosomal form was considered as the ancestral form (Coluzzi et al. 2002), that split into the M and S molecular forms that later accumulated inversions that helped them to adapt to and conquer different ecological zones (Slotman et al. 2006). Gentile in 2001 reported that these molecular forms are partially reproductively isolated (Gentile et al. 2001). Mixed swarms that included both sibling species have been reported in different areas of Africa: e.g. Soumousso (a typical savannah village of Burkina Faso) and Vallée du Kou (VK7) (Diabaté et al. 2006; Diabaté et al. 2007; Dabire et al. 2013). The genetic analysis of sperm from mated females collected in the field suggested that females strictly mated with their own species (Tripet et al. 2001), suggesting extreme assortative mating between species (Tripet et al. 2001; Dabiré et al. 2014). Hybrids frequencies varied in different areas of Africa. In Mali, the hybrid frequency was approximately 1% (Tripet et al. 2001), whereas no hybrids were found in Cameroon (Diabaté et al. 2006; Diabaté et al. 2007). However, in the Western Coastal countries of Guinea Bissau and Senegal, hybrids can occur locally at much higher frequencies (Caputo et al. 2011). Hybrids in laboratory were found to be fully viable and fertile (Diabaté et al. 2006; Diabaté et al. 2007)

suggesting that there are no intrinsic post-mating barriers, but there are incomplete pre-mating barriers responsible for the reproductive isolation between *An. gambiae s.s.* and *An. coluzzii* species in field (Lanzaro & Lee 2013).

1.11. Mechanisms of assortative mating between *An. gambiae s.s.* and *An. coluzzii*

1.11.1. Intrinsic and extrinsic post-mating reproductive barriers

The male accessory gland secretions (MAG) also known as Accessory gland proteins (Acps) has been reported to inhibit any further mating in mosquitoes females (Shutt et al. 2010). These secretions are found in male seminal fluid and they cause post-mating changes in female behaviour and physiology (Gillott 2003). Tripet study (2005) showed that cross-mated females may mate a second time and this behaviour might constitute a post-mating reproductive barrier between the sibling species (Tripet et al. 2005). Tripet explained, that the proteins associated with MAG, which are responsible for preventing females from further mating, are likely to not prevent re-mating in reciprocal crosses in these forms (Tripet et al. 2005). This hypothesis is consistent with Goma conception in 1963, when he reported that the double-mated females used the sperm preferentially in *An. gambiae* forms. In this context, Tripet originally proposed that there could be interactions between male seminal products and the female reproductive system causing females to re-mate with their own males when first mated accidentally to males of an alternative cryptic species (Tripet et al. 2005). This process could decrease the chance of hybrids progeny by diluting the sperm of the opposite type (Lanzaro & Tripet 2003; Tripet et al. 2005). Therefore, this hypothesis would be an important step in our knowledge of reproductive isolation barriers in *An. gambiae s.s.* and *An. coluzzii* species (Lanzaro & Tripet 2003; Tripet et al. 2005). However in controlled laboratory experiments, neither Tripet et al (2005) nor Shutt et al (2010) found that cross-inseminations or injection of MAG extracts

between *An. gambiae s.s.* and *An. coluzzii*, were less likely to cause monogamy in reciprocal combinations (Tripet et al. 2005; Shutt et al. 2010) suggesting that there are no intrinsic post-mating barriers.

The idea that extrinsic barriers to reproduction isolation occur in the form of fitness costs to hybrids is supported by observations of divergent ecological adaptation between *An. coluzzii* and *An. gambiae s.s.*. Several studies have reported notable differences between the sibling species in larval and adult stages. The larval stage of the two species are adapted to different water body-type (Diabaté et al. 2007). In this regard, field studies showed that *An. coluzzii* larvae are able to survive better than those of *An. gambiae s.s.* in habitats that include aquatic predators (Diabaté et al. 2007; Gimonneau et al. 2012; Roux et al. 2013). In addition, newly emerged *An. coluzzii* females were heavier and had higher proteins and lipids reserves than those in *An. gambiae s.s.* of similar wing length (Mouline et al. 2012). Therefore, *An. coluzzii* females had higher ability to develop and mature eggs after a single blood meal than those of *An. gambiae s.s.* (Mouline et al. 2012). Furthermore, *An. coluzzii* is thought to remain in Sahelian regions and undergo aestivation in the dry season, which helps its population to build-up faster when the next rainy season starts, whereas *An. gambiae s.s.* is thought to possibly migrate away from Sahelian areas during the same dry seasons (Dao et al 2014).

1.11.2. Pre-mating reproductive barriers

Different studies have confirmed that *An. gambiae s.s.* and *An .coluzzii* mate assortatively (Tripet et al. 2001; Dabire et al. 2013), and the existence of hybrids is rare (Nwakanma et al. 2013). However, the hybrids were fully viable and fertile in the lab (Diabaté et al. 2006; Diabaté et al. 2007). This in turn confirms that there is pre-mating barriers that prevent the gene flow between these species in absence of the post-mating barriers (Diabaté et al.

2006; Diabaté et al. 2007). This raises a number of questions as to how do these species recognize their own form during mating, and what mechanisms are used in this process?

Spatial segregation of swarms is considered as one of the pre-mating isolation mechanisms in *An. gambiae s.s*. and *An. coluzzii* in Burkina Faso (Diabaté et al. 2006) and Mali (Diabaté et al. 2009a; Diabaté et al. 2011). However, hybrids have been reported in Western Coastal regions of Africa (Caputo et al. 2008), which confirm that the spacial segregation is not the main barrier between the sibling species. In addition, mixed swarms that included both sibling species have been reported in different areas of Africa: e.g. Soumousso (a typical savannah village of Burkina Faso) and Vallée du Kou (VK7) (Diabaté et al. 2006; Diabaté et al. 2007; Dabire et al. 2013), and the species were found to mate extremely assortatively even in mixed swarms (Tripet et al. 2001; Dabiré et al. 2014).

Other studies argued the importance of flight tones as pre-mating barrier between *An. gambiae s.s.* and *An. coluzzii* species, in this area of research Gibson found in 2010 that the female mosquitoes flight tone usually attract males that use their sensitive antennae and Johnston's organ at the base of each antenna to identify females place of their own species (Gibson et al. 2010). Therefore, the differences in flight tone could be considered as a reproductive isolating mechanism (Gibson et al. 2010). Continuing on from this, several studies tried to measure and analyse the wing-beats frequencies in *An. gambiae s.s.* and *An. coluzzii* that could cause different wing tones. There was no difference detected when the measurements were done separately for the individuals (Tripet et al. 2004; Gibson et al. 2010). However, when the measurements of flight tone frequencies were done for male-female pairs for the same and different species, the mosquitoes start matching their flight tones significantly in the same species pairs more than the different species pairs (Pennetier et al. 2010). The differences in the flight tones could be also due to the wing shape differences in *An. gambiae s.s.* and *An. coluzzii* species (Sanford et al. 2011).

Sanford and his colleagues (2011) measured the wing width and length of females of *An. gambiae s.s.* and *An .coluzzii* species collected from West Africa, they found that there are no wing length differences, whereas the species differed significantly in wing width (Sanford et al. 2011). These results seemed to support the wing-beat hypothesis, which suggested that sibling species in *An. gambiae* complex could recognize their own type of mates during the mating by harmonizing their wing-beat frequency to create specific flight tones (Lanzaro & Tripet 2003; Gibson et al. 2010).

Moreover, Lanzaro and Tripet (2003) assumed that contact pheromones may be involved in mating recognition process (Lanzaro & Tripet 2003). However, pheromones have volatile nature which exclude them as a conspecific recognition mechanism in single and mixed swarms, but they could serve as a short range recognition mechanisms (Lanzaro & Tripet 2003).This could occur through the contact between the males and females using the pheromone that present on the females' legs and receptors located on the males tarsi; e.g. this contact was reported in *Culiseta inornata* (Lang & Woodbridge 1976) and *Aedes albopictus* (Nijhout & Craig 1971). However, there is no evidence of the pheromone role in the conspecific recognition within *An. gambiae* populations. Generally, the chemical and the physiological mechanisms that are involved in mating process and recognition are still unknown and further studies are required to explain the reproductive isolation mechanisms.

1.12. Genetic structure and models of speciation

Several studies tried to outline the speciation models and the circumstances that create two species from one original taxon (Via 2012; Singh 2012; Feder et al. 2012). The speciation process is the key for understanding biodiversity and evolution in species (Singh 2012).

To understand the speciation process, different models of speciation were explained based on the speciation mechanisms (Via 2012). In allopatric speciation, when the populations are isolated geographically, the physical factors prevent the recombination between species. Therefore, gene flow is blocked between the whole genome in these species (Via 2012). On the other hand, sympatric model is controversial because the speciation occurs despite the on-going gene flow between species as the new species remain in the same area (Via 2012). Therefore, the sympatric speciation can be explained by the occurrence of reproductive isolation between the parental species in the same area, which leads to create new species in this area (Singh 2012). Sympatric speciation occurs during the adaptation to ecological changes, this leads to a reduction in the random mating and recombination within population (Via 2012). Subsequently, a divergent selection against the adapted alleles takes place and continues to the nearby genes. This process limits recombination in the selected adapted genes between the two emerging populations and somewhat protects these genes from disruption (Via 2012). The limited recombination between the species could lead to assortative mating within these species. For this to happen mate choice needs to be associated with a specific trait under divergent selection, which leads to within species mate behaviour and facilitates the creation of two different species from one taxon (Via 2012). This process can be greatly facilitated if genes of assortative mating are genetically linked with those of divergent selection (Via 2012).

Recently some cases of sympatric speciation were observed and studied in different populations of animals. In Santa Cruz Island, two populations of *Geospiza fortis* birds were

recognised based on beak morphs that fall into large and small size, with some individuals with intermediate size. These patterns were the results of selection imposed by the size and hardness of seeds. The difference of beak size was found to affect mating songs of the males, therefore they were considered as mating barriers between the populations suggesting early stage of sympatric speciation in *Geospiza fortis* population (Huber et al. 2007). Another well-studied example of sympatric speciation was introduced when some insects can feed on different species of host plant. *Rhagoletis pomonella* (the apple maggot) species used to feed on hawthorn. Later, apples trees were introduced into North Africa, which made *Rhagoletis pomonella* shift from hawthorns to apples during the past 200 years. The host preference traits act as a barrier to gene flow because the mate selection associated with host-plant recognition. Genetic studies showed genetic differentiation between sympatric hawthorn and apple populations in the field, which confirms that these populations are partially reproductively isolated and they could be representative models of sympatric speciation (Feder et al. 1988). Further example for sympatric speciation was observed in cichlid species complex (*Amphilophus*) in crater lakes and islands. Genetic and morphological comparison studies showed that Crater Lake Apoyo was found to include *Amphilophus citrinellus* as common cichlid species in the area. Later, *Amphilophus zaliosus* species were evolved from *Amphilophus citrinellus*. These two species were found to be morphologically different and reproductively isolated (Barluenga et al. 2006). Furthermore, in California another example of sympatric speciation was presented through studying killer whales. Comparative genomic studies revealed two populations: residents and transients. These populations inhabit the same water but they have different behavior, diet and morphology. Moreover, these populations are reproductively isolated (Morin et al. 2008).

In our study we focused on the sympatric speciation in *An. gambiae* complex. The species of *Anopheles gambiae* s.l. have a mitotic karyotype with two pairs of autosomes and one pair of sex chromosomes (Coluzzi et al. 2002). The ecological adaptations, genetic structure was thought to play a main role in speciation process within *An. gambiae* population. In this regard, early pre-genomic studies of genetic differentiation in continental range using isozyme and microsatellite data detected very low levels of genetic differentiation between samples of *An. gambiae s.s.* and *An coluzzii* species collected from Kenya and Senegal (Lehmann et al. 1996). This lack of genetic differences supported the occurrence of contemporary gene flow between both species (Tripet et al. 2001). In addition, some genetic differentiation studies suggested that small genomic regions are responsible for the reduction of gene flow between these species (Wu 2001). These findings are consistent with the hypothesis that selection can prevent gene flow in regions of chromosomes that includes isolation genes, whereas the other regions of this chromosomes can flow freely between the sibling species (Slotman et al. 2006). Therefore, genomic regions that are highly different between the sibling species that undergo speciation with gene flow are likely to include isolation genes (Slotman et al. 2006).

Gentile (2002) suggested that *An. coluzzii* and *An. gambiae s.s.* have mosaic genome structure that include completely different regions of the genome with no gene flow, whereas the rest of the genome pass freely between species (Gentile et al. 2002). Genome comparison studies in *An. coluzzii* and *An. gambiae s.s.* showed that the genetic differences between them were limited to three regions that were 2L, 3L and X chromosome near the centromere regions (Turner et al. 2005; Turner & Hahn 2007; White et al. 2010). These divergence regions represented only 3% of the total genome (Lanzaro & Lee 2013) (Fig.1.11). They were called islands of speciation because they were considered to contain the genes responsible for the reproductive isolation between *An. gambiae s.s.* and *An.*

coluzzii (Slotman et al. 2006). These islands are believed to be formed by selection that occurs on small linked gene regions, and later it extended by hitchhiking acting on next regions of these genes (Turner et al. 2005; Smadja et al. 2008; Feder & Nosil 2010; Turner & Hahn 2010). Interestingly, recent study of Clarkson in (2014) confirmed that the introgression of insecticide resistance mutation of the voltage-gated sodium channel (Vgsc-1014F) from *An. gambiae s.s.* into *An. coluzzii* led to the homogenization of the whole 2L island of speciation (1.5 % of the genome). Despite increase the rate of the Vgsc1014F mutation in the *An. coluzzii* populations in Ghana, the hybridization rates between *An. gambiae s.s.* and *An. coluzzi* remain stable (Clarkson et al. 2014). These results confirm that the 2L island does not have a direct effect on the reproductive isolation between both species (Clarkson et al. 2014).

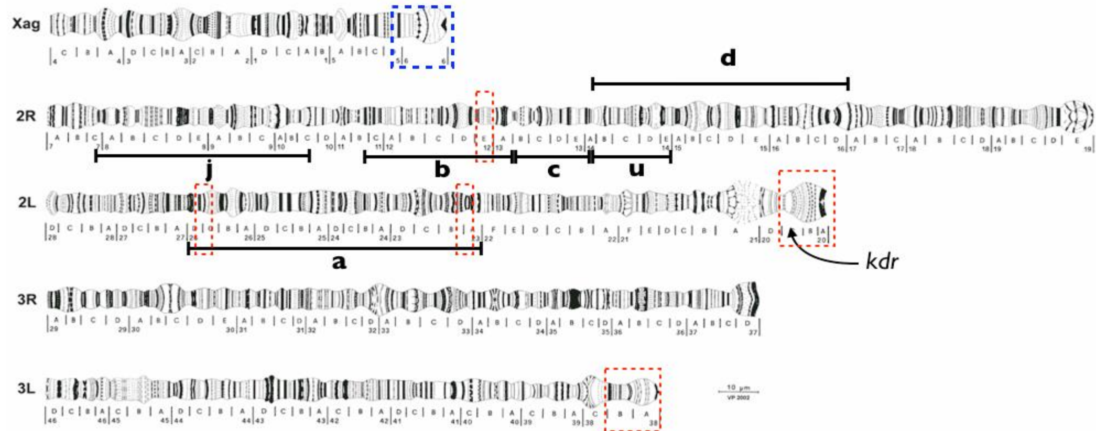


Figure 1.11. Schematic of the polytene chromosomes of *Anopheles gambiae sensu stricto* showing the band-numbering scheme - This chromosome is in the standard arrangement in terms of inversion polymorphism (i.e. Forest chromosomal form). The positions of the 6 inversion sites are indicated in boldface (Coluzzi et al. 1985). The site of the diagnostic rDNA on the X chromosome (Xag) is indicated in blue. Putative "genomic islands of speciation" other than X island (blue) are indicated in red (Turner et al. 2005; White et al. 2010). The approximate position of the kdr allele of the para gene on chromosome 2L is also indicated. The image is adapted from the standard chromosome map for *Anopheles gambiae* available from VectorBase.org, a NIAID bioinformatics resource centre for invertebrate vectors of human pathogens (cited in Doug Paton thesis).

There were attempts to explain the speciation based on the islands of speciation, and two explanations were concluded (Lanzaro & Lee 2013). The first explanation accepted that the recombination between *An. coluzzii* and *An. gambiae s.s.* is high in whole the genome with no recombination in islands of speciation regions (Lanzaro et al. 1998; Wang et al. 2001; Turner et al. 2005; Turner & Hahn 2007). These islands of speciation are located near the centromere regions of the chromosome that are known for low recombination (Turner et al. 2005; White et al. 2010). Therefore, this explanation is consistent with the

conception that divergence is driven by the genes that exist in sites with reduced recombination within divergence islands (Noor et al. 2001; Rieseberg 2001).

The second explanation was suggested by white (2010) study that includes observations of *An. gambiae s.s.* and *An. coluzzii* from Mali, Burkina Faso, Cameroon and Kenya. In this study, complete linkage disequilibrium was found between these islands (White et al. 2010). In addition, genotyping whole genome of *An. coluzzii* and *An. gambiae s.s.* presented that the low recombination regions spread widely in the genome in the sibling species (Lawniczak et al. 2010). Therefore, it was suggested that there is no gene flow between both species based on these results, and these species are completely isolated, and the differences in the divergence islands are incidental (White et al. 2010). As a result, the genes responsible for the reproductive isolation are not necessarily included in these regions and the species isolation is a result of the ancestral segregation and not a result of contemporary gene flow (Turner & Hahn 2010; Hahn et al. 2012). However, subsequent studies showed that the linkage disequilibrium breaks down between the divergence islands in areas where higher introgression occurs between *An. gambiae s.s.* and *An. coluzzii* species (Caputo et al. 2011; Lee et al. 2013). These results in addition to comparative genomics studies would support the 'islands of speciation' model suggesting that these islands could play a major role in sympatric speciation in the face of varying levels of gene flow (Weetman et al. 2012; Lee et al. 2013).

1.13. Research objectives

In my thesis, I tried to fill the gaps in the presented explanations of speciation process in *An. gambiae s.s.* and *An. coluzzii* by focusing on the mechanisms of assortative mating and process of mate choice. Understanding the mate choice process is also important for vector control strategies success that based on releasing GM/sterile laboratory mosquitoes to the field to test their ability to mate with the target species. In addition, understanding the mating behaviour in these species could contribute in developing new approaches of vector control. Therefore, there were two main categories of objectives in my thesis:

Molecular biology objectives

1. Genotyping the X, 2L and 3L islands of speciation in recombinants and parental strains to identify the islands of speciation genotypes. In addition, testing the association between the genotype of X-island of speciation and the mating preference of recombinants and parental strains.

2. Identify the size of the pericentromeric region that associated with the assortative mating behaviour, and finding the protein-coding differences distinguishing recombinant strains, and conserved differences putatively relevant to speciation.

Behavioural ecology objectives

3. Investigating the behavioural differences in swarming and mating process between the recombinants to explain the potential mechanisms that the sibling species use in conspecific recognition within mixed swarms.

4. Describe potential morphological differences that could lead to assortative mating through differences in flight-tones between the *An. gambiae s.s.* and *An. coluzzii* species.

CHAPTER TWO

General material and methods

In this chapter, the general methods that we used to maintain the different strains in the insectary will be explained in addition to the other methods repeatedly used across several experimental chapters, and the methods for creating recombinants in the laboratory.

2.1. Insectary conditions

Two representative M and S strains were maintained:

Kisumu: was brought from Kenya and colonised in the laboratory over 25 years. It has inversion on 2R, therefore it was considered as Savanna chromosomal form, as well as, it characterised as S type molecular form on X island that is representative for *An. gambiae s.s*.

Mopti: was brought from the village of N'Gabacoro near Bamako, Mali from 11 years ago. This strain was characterised as Mopti chromosomal form because of the existence of bc and u inversions on 2R, also it was M molecular form on X island that is representative for *An. coluzzii*.

Colonization was the first step when these mosquitoes were brought from the field into the laboratory and got adapted to the lab After several generations, the colonies or laboratory population were 'maintained' as 'colonies' or 'strains' in separated insectaries in Keele University. The temperature was kept at 25±1˚C and humidity at 70-75% with the strains exposed to a cycle of 12h dark/light using a digital timer. Usually, the adults from each strain were put in separate cages. Standard cages were made from 5L white polypropelene buckets covered with a white mesh, which were provided with wet pieces of cotton on the top of the cage to provide water for the mosquitoes. The cages were provided

with sleeves on the side to introduce mosquitoes as well as cups with pupae or oviposition cups (see below). A small bottle of 5% glucose solution was used for feeding. The sugar solution was prepared by adding 50g Glucose (Acrose Organics, New Jersey, USA) to 1L of dH_2O, and the mix was sieved using a Y funnel with filter paper (Grade 1, Whatmann), then placed in bottles and stored at -20˚C.

2.2. *An. gambiae* strains breeding

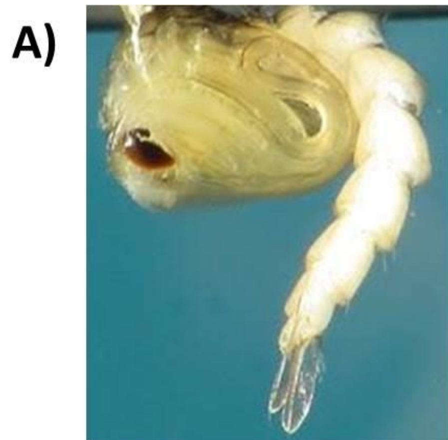
After putting the males and females of each strain in their cage, they are ready to mate. They were fed twice a week using horse blood (Fig. 2.2), which was warmed by the feeder system (Hemotek membrane feeding system) to 37˚C. After 3 days of mating, gravid females were ready to lay eggs (we usually leave the females 48h for eggs development at 25˚C and one extra day so they are all ready to lay eggs.), therefore cages were provided with polystyrene cups (Dart, Insulated foam containers, Kent, UK) and lines with a cone of filter paper (grade1, Whatmann) and filled to half with dH2O to create a slope on which eggs can be kept wet on the cups walls. The cups were removed from the cages after two nights when eggs start hatching to give the first instar larvae. These larvae were placed in clean plastic trays (35cm length x25cm width), in a way that each tray included 200 larvae and 1000ml dH_2O (Fig. 2.1). First instar larvae were supplied with one drop per tray for a first day of liquifry (Interpet Ltd., Dorking, UK), then the larvae were fed on ground fish food (Tetramin, Blacksburg, USA). The larvae stage consists of four larvae instars stages that take about one week to develop to the next pupae stage, then the pupae were transferred by aspiration system and placed in small polystyrene cups (Dart, Insulated foam containers, KENT, UK), that in turn, placed in the adults new clean cages for next generation.

Figure 2.1. The clean plastic trays (approximately 35cm length x25cm width) - each tray included 200 larvae and 1000ml of dH2O.

Figure 2.2. The feeder system (Hemotek membrane feeding system) – Used to feed the mosquitoes twice a week using horse blood, which was warmed to 37˚C and presented to mosquitoes for 30min..

2.3. Sexing the pupae

In some experiments, we needed to do sexing in order to separate males and females in the same day (same age males and females) for each strain. Therefore, the pupae were collected in one day (same age pupae) by an aspirator and placed in a polystyrene cup (Dart, Insulated foam containers, Kent, UK), which filled to half with dH_2O. Thereafter, the pupae were placed in a glass petri plate to be examined under a light microscope using micro pins to separate the pupae. Differentiation between male and female pupae was based on the 10th segment of the abdomen, where the genital lobe is located near the paddles. In females the segment is short and not separated, whereas the lobe splits into two and is much longer in males (Fig. 2.3). Once the females and males pupae were sexed, the pupae were pooled usually with other same-aged males and females into separated polystyrene cups that were put in separate emergence cages to allow them to emerge. The cages were provided with glucose solution (5%) and cotton pads soaked in dH2O.



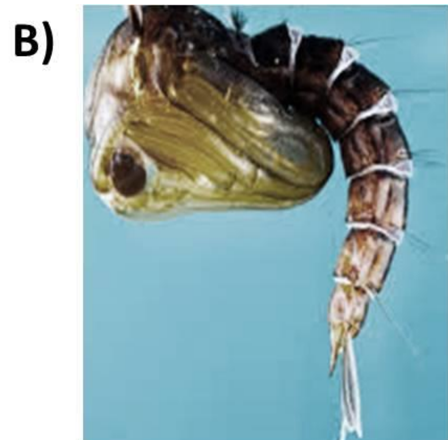


Figure 2.3. Differences between male and female pupae was seen at the 10th segment of the abdomen, where the genital lobe located near the paddles – A) In the females the segment is short and not separated. B) In the males the lobe splits into two and is much longer than the females.

2.4. Females dissection and obtaining the sperm bundle

We dissect the females either to know whether they mated or not, or to obtain the sperm bundle to know if they mated with their own type. To do so, females were placed in plastic 5ml tube (Fisher brand, Leicestershire, UK), which included 3ml of 70% alcohol, and kept in 4°C. In the day of dissection, the females were placed in Petri plate (Sterilin, Massachusetts, USA), and then one female was transferred to a drop of PBS solution (Phosphate Buffer Saline) (Fisher Scientific, Leicestershire, UK) on the slide and placed on the microscope to start dissection. The reproductive system was removed, including the spermatheca which was easily identified as a brown pigmented, spherical organ. The spermatheca was transferred to another clean slide on drop of dH2O and opened to get the sperm bundle using the dissection pins. The sperm was placed in the 1.5ml tube (Fisher brand, Leicestershire, UK) that included 500µl of extraction lysis buffer from ChargeSwitch kit for DNA extraction (see 2.8. paragraph).

2.5. Standard amplification for differentiation between S and M genotype in X-divergence-island

The extracted DNA samples and two positive controls: one is M form and the other is S form were all amplified using the polymerase chain reaction (PCR) molecular form developed by Fanello et al. (2002). The reaction mix for each sample consisted of 2.5µl taq buffer, 0.5µl dNTP (10mM each) (Fermentas, Leicestershire, UK), 0.3µl primer GA (5-CTG GTT TGG TGG GCA CGT TT-3), 0.3µl primer UN (5-GTG TGC CCC TTC CTC GAT GT-3) (Eurofins Genomics, Wolverhampton, UK), 0.125µl Dream Taq polymerase (Thermo Scientific, Massachusetts, United States), 2µl genomic DNA template (5-50ng/µl) and 18.975µl dH2O for a total reaction volume of 25µl. Thereafter, the PCR was carried out using a PTC-200 DNA-Engine thermocycler (BioRad, California, USA) using the following steps: initial denaturation 3min at 95°C, and then 35 cycles (15s at 95°C, 30s

at 62°C and 1min at 72°C). Once these 35 cycles were completed final step was 10min at 72°C then kept it at 4°C forever (G Favia et al. 2001; Fanello et al. 2002).

2.6. DNA restriction reaction for differentiation between S and M genotype in X-divergence-island and gel separation

Once the PCR reaction was completed, we could start digestion reaction to identify S and M genotype in X-divergence-island (Fanello et al. 2002). For this purpose, 10µl of each amplified samples was mixed with 10µl of a restriction endonuclease digestion mix that included: 0.125µl Hhal restriction endonuclease (Promega, Southampton, UK), 2µl 10x Buffer C and 7.875µl dH2O. Thereafter, the restriction digest mix plate was incubated in incubator at 37°C for at least 3 hours.

Following digestion, the DNA fragments that presented in the reaction mix were separated by gel electrophoresis that was prepared by mixing 1.8g of agarose (Fisher Scientific, Leicestershire, UK) and 1µl of Ethidium Bromide (Fisher Scientific, Leicestershire, UK) per 100ml of Tris-Boric Acid-EDTA (TBE) (Invitrogen, Leicestershire, UK). The gel was then run at 80 volt for 1:30 hour. The fractionated DNA, which stained with Ethidium Bromide, was visualized under UV light. In the digestion process, the Hhal enzyme cuts only S genotypes in the intergenic spacer (IGS) region of the rDNA region of the X-divergence-island of each samples, producing two visible bands: first one is 257bp and another one is 110bp, whereas M form stays as only one band which is 367bp-long (Fanello et al. 2002).

2.7. Wings measurements

The wing length of the samples was used in some experiments as a proxy to body size instead of total body length to avoid bias due to changes in the abdomen linked to blood and sugar feeding and gravidity (Renshaw et al. 1994; Mwangangi et al. 2004). Males and

females were collected usually at the end of experiments and kept in 5ml tubes (Fisher brand, Leicestershire, UK) in 70% ethanol until measured. Thereafter, mosquitoes were placed on a microscope slide in a drop of PBS (Fisher Scientific, Leicestershire, UK) and the wings from one side were removed carefully and each moved to a clean slide and small drop of PBS (Fisher Scientific, Leicestershire, UK). Each slide was placed on an Olympus dissecting microscope equipped with an Olympus (model E-520) camera, and digital pictures were taken of each wing individually. Thereafter, the wing images were measured using ImageJ software (Thévenaz 2003). The length of the wings was determined as the distance from the posterior anal cell margin to the tip of radial vein 3(R3) (Sanford et al. 2011). All measurements were converted from pixels (px) to mm using a standard stage micrometer, which is 1mm long. This ruler was measured using ImageJ software, and the following equation was used to convert pixels to mm:

The wing measurement (mm) = Wing measurement px/ Ruler measurement in px.

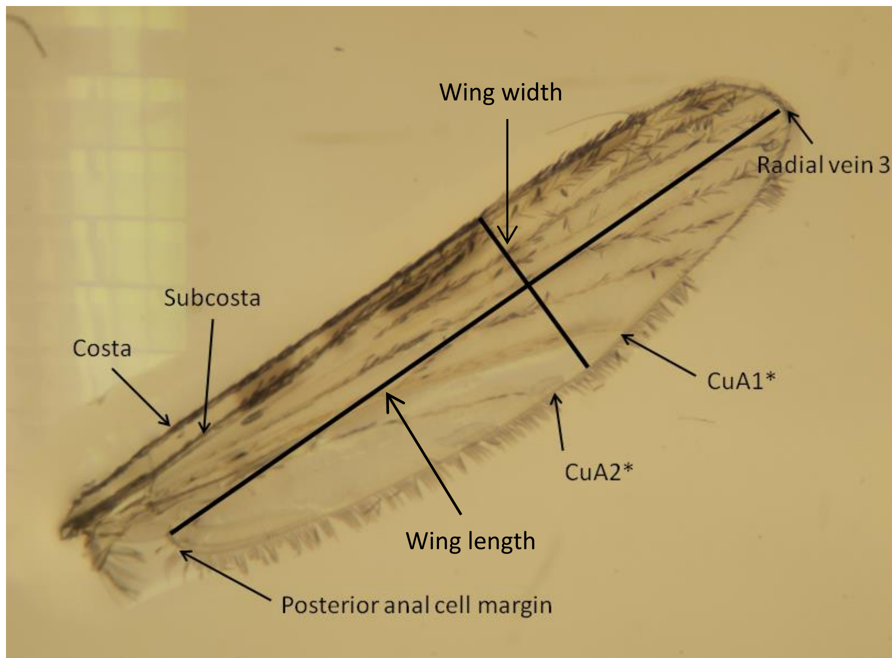


Figure 2.4. The length of wings – length is determined as the distance from the posterior anal cell margin to the tip of radial vein 3(R3).

2.8. ChargeSwitch method for DNA extraction

Usually, this method is accurate to get clean DNA that exclude proteins and RNA using ChargeSwitch kit (Invitrogen, Leicestershire, UK). Sometimes we use this method for the sperm bundle because it includes small amounts of DNA or for full bodies of mosquitoes when we need highly concentrated pure DNA from old bodies. Based on the sample size, we follow the manufacturer's guidelines with modifications (for example for the sperm we use half of all amounts of reagents and more vortexing and short spin steps which is explained here). In this method, the samples of extracted DNA from sperm bundle were placed in 1.5ml centrifuge tubes (Fisher brand, Leicestershire, UK) that include 500µl lysis buffer and 5µl Proteinase K, and then the tubes were vortex and the RNA removed by adding RNase. Thereafter, 100µl purification buffer were added to the samples, and then 20µl of magnetic beads to bind to the DNA. The tubes were placed in the MagnaRack (Fisher Scientific, Leicestershire, UK) and the supernatants were carefully discarded from these tubes. To wash the DNA, 500µl of wash buffer was added to the tubes that were replaced in MagnaRack (Fisher Scientific, Leicestershire, UK) and the supernatants were removed. At the end, a 75µl of Elution buffer (E5) was added to the tubes that were sited in the MagnaRack and the supernatants that include the pure DNA were removed and placed in storage tubes, and these were kept in -20°C.

2.9. DNAzol method for DNA extraction

In this method we also followed the manufacturer's guidelines with some modifications. The adults or fourth instar larvae were placed individually in 1.5ml centrifuge tube (Fisher brand, Leicestershire, UK), and 100µl DNAzol (Fisher Scientific, Leicestershire, UK) was added. The samples then were homogenized with a pestle until there are no recognizable body parts, and then they were centrifuged for 10min at 10000g and the supernatants were

transferred to new 1.5ml centrifuge tubes. 50µl of Ethanol (Sigma-Aldrich Company Ltd., Dorset, UK) 100% were added to each tube and were mixed by shaking gently for 5-8 times and were held at room temperature for 3min. Thereafter, they were centrifuged for 10min at 10000g and the supernatants were removed. The DNA pellet was washed by adding 750µl of 75% Ethanol and was shake to re-suspend pellet. After that, they were centrifuged again and the supernatants were removed. The tubes were placed in an incubator at 25°C overnight to dry. Next day, 200µl of TE (Tris EDTA) were added and the tubes were vortex, and then the DNA was transferred to storage tubes that stored at -20°C.

2.10. Creating recombinants

Repeated backcrosses were processed in order to create 95% genetically identical recombinant strains differing only in the molecular type of their X chromosome speciation island. In this regard, 100 M homozygous type males obtained from Mopti strain and 100 homozygous SS females obtained from Kisumu strain, were bred in our insectaries and cross-mated to create hybrid MS progeny. Thereafter, 100 females of the resulting MS genotype were bred and mated with 100 M-genotype Mopti males in the first backcross, and the resulting MS1 progeny were identified and isolated. After that, 100 virgin MS1 females were backcrossed with 100 M males to produce MS2 progeny and this step was repeated twice to get MS4 progeny of both sexes. Finally, MS and MM females and M and S males from the 4th backcross were randomly mated with one another in a mixed cage (100 males and 100 females) resulting in MM or MS 6th generation families. Those progenies of families identified as being mixed M and S by larval genotyping which featured all possible male and female genotypes were randomly mated to one another in a mixed cage (100 males and 100 females) for a 7th cross in order to generate MM, MS and SS families. The resulting MM and SS families were pooled together in order to obtain a

pair of MM and SS strains, sharing 96.88% of the Mopti genetic background and theoretically 99.61% genetically identical to one another but differing at their rDNA locus and linked X- chromosome island of speciation (Fig. 2.5). This backcrossing process was replicated twice in order to produce two pairs of recombinant differing only at the X island locus. These recombinants are called AMM and ASS and recombinant strains BMM and BSS. All the process was conducted in the standard breeding system conditions in our insectaries (Fig. 2.5).

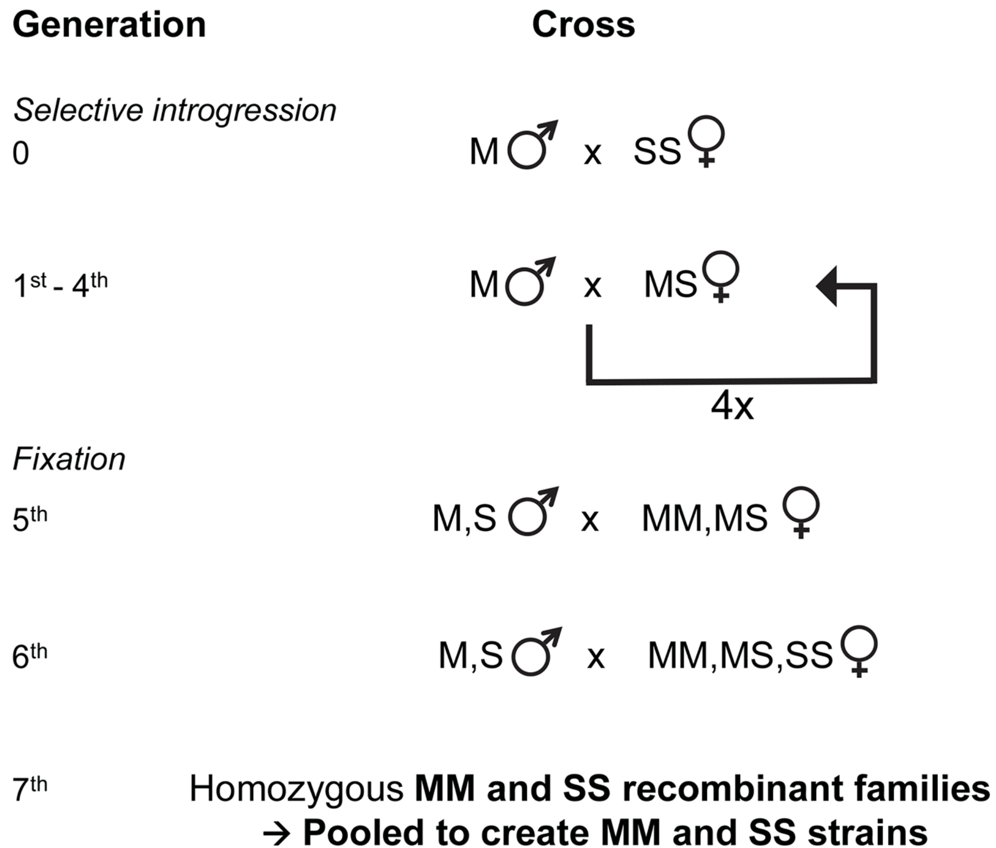


Figure 2.5 Genetic crossing design used for selective introgression of the X-island of divergence in recombinant strains - Following the creation of hybrid females at the X-island diagnostic rDNA locus, 4 generations of backcrosses and 2 generations of fixation were used to introgress the S-type X-island into an M molecular form Mopti genetic background. Here, 'MM', 'MS' and 'SS' refer to the female genotypes at the rDNA marker locus in the X-island and 'M' and 'S' refers to the male genotype at the same locus. The backcrossing design ensured that the resulting recombinants strains shared a high genetic identity but differed at the S or M-type X- chromosome islands of divergence.

CHAPTER THREE

Genomic characterization of X, 2L, 3L divergence islands and assortative mating in females and males of X-island recombinant strains, and parental strains

3.1. Introduction

The population structure of *An. gambiae* was a subject for intensive studies because it is considered as a representative model for studying and understanding speciation process at early stages (Sawadogo et al. 2013; Sawadogo et al. 2014; Diabate & Tripet 2015; Aboagye-Antwi et al. 2015). *An. gambiae* complex is known to have high level of genetic polymorphisms that could facilitate speciation (White et al. 2010). The member species of *An. gambiae* complex are the major vector of malaria and they distribute widely throughout Africa (Lehmann & Diabate 2008). Some of the sibling species of the complex are known for their high ability to adapt with human-made environment changes (Lehmann & Diabate 2008). These new species that are formed by speciation process could be characterised by general traits that affect vectorial capacity (Wu & Ting 2004; Mallet 2006). Therefore, studying *An. gambiae* complex provides fundamental information about speciation process and public health (Lehmann & Diabate 2008). Extensive cytogenetic studies of the *An. gambiae* complex revealed five sympatric chromosomal forms that were named Mopti, Savanna, Bamako, Forest and Bissau (Coluzzi et al. 1979; Coluzzi et al. 1978; Coluzzi et al. 1985; Touré et al. 1998). These forms were characterized by different inversion arrangements on the right arm of the second chromosome (Coluzzi et al. 1979; Coluzzi et al. 1978; Coluzzi et al. 1985; Touré et al. 1998). The inversions facilitate ecological flexibility in these forms and this gave them ability to exploit different

ecological zones (Coluzzi et al. 1985; Coluzzi et al. 2002; Coetzee et al. 2013). Subsequent studies using allozymes and microsatellites genomic based methods found genetic difference on the rDNA intergenic spacer on the X chromosome between Mopti form and both Savanna and Bamako forms in Mali (Favia et al. 1997). Based on these findings Mopti was named M molecular form and both Savanna and Bamako were named S molecular form ((Favia et al. 1997; Della Torre et al. 2001; Favia et al 2001). Later, these two molecular forms were elevated as two different species that are *An. coluzzii* for M molecular form and *An. gambiae s.s.* for the S molecular forms (Coetzee et al. 2013). *An. gambiae s.s.* species distribute throughout Sub-Saharan Africa, whereas *An. coluzzii* could be found strictly in west and central Africa (Lehmann & Diabate 2008).

Some studies reported phenotypic differences between both species in Sahelian regions (Dao et al. 2014). During the dry seasons, *An. coluzzii* usually undergo aestivation, whereas *An. gambiae s.s.* populations migrate away (Dao et al. 2014). Therefore, *An. gambiae s.s.* populations grow up comparatively slower than *An. coluzzii* in the next rainy season (Dao et al. 2014). In addition, *An. coluzzii* larvae can be found in more permanent freshwater habitats which make it able to transmit plasmodium all over the year (della Torre et al. 2005), whereas *An. gambiae s.s.* larvae exploit temporary water bodies which makes these species disappear in the dry season (Roux et al. 2013). Moreover, field studies showed that *An. coluzzii* larvae developed faster than those of *An. gambiae s.s.* in water bodies that include predators (Roux et al. 2013). These adaptations could lead to ecological divergence between these species (Gimonneau et al. 2010; Gimonneau et al. 2012). Therefore, it could play a role in speciation between *An. gambiae s.s.* and *An. coluzzii* populations (Gimonneau et al. 2010; Gimonneau et al. 2012). Despite the importance of the above-mentioned phenotypic differences, reproductive isolation between both species is another important process in the speciation. Therefore, studying mating behaviour and

mate choice in these mosquitoes is considered as key to explain the reproductive isolation and speciation in these two sibling species (Dao et al. 2008).

Studies showed that *An. gambiae* mating occurs at dusk when the males fly together and move as a single unit forming a swarm (Charlwood & Jones 1979; Charlwood et al. 2003). Different species usually swarm over different markers: *An. coluzzii* individuals swarm over visual markers, whereas *An. gambiae s.s.* swarms are found mostly over bare ground (Diabaté et al. 2009; Manoukis et al. 2009; Manoukis et al. 2014). These swarms attract females that visit them to find mates (Charlwood et al. 2003; Manoukis et al. 2014). When the female inter the swarm, it comes near several males, then it chooses the mate based on harmonic frequency matching process (Cator et al. 2010; Pennetier et al. 2010). During the mating process, the male passes his accessory glands secretions and sperm into the female (Clements 1992; Howell & Knols 2009). These secretions induce refractoriness for further mating in females that cannot mate again at the same night (Shutt et al. 2010), whereas the male re-joins the swarm to try to find another female for mating (Clements 1992; Howell & Knols 2009; Sawadogo et al. 2014).

In Mali, the swarms of *An. gambiae s.s.* and *An. coluzzii* were found to be spatially segregated and no mixed swarms were reported (Diabaté et al. 2009; Diabaté et al. 2011). However, other studies in Burkina Faso showed that there are some mixed swarms in both Soumousso and Valle´e du Kou (VK7) areas between 2005 and 2006 (Dabire et al. 2013; Sawadogo et al. 2014). Despite the occurrence of mixed swarms, low frequencies of hybrids were found, suggesting strong assortative mating between both species (Diabaté et al. 2006; Dabire et al. 2013). This suggests that other mechanisms could involve in assortative mating. However, laboratory studies found that hybrids between the two sibling species are viable and fertile (Diabaté et al. 2007). Therefore, the low rate of hybridization in field populations cannot be due to hybrid inviability. Consequently, strong pre-mating

barriers are responsible for the reproductive isolation between *An. gambiae s.s.* and *An. coluzzii* in absence of intrinsic post-mating barriers (Lehmann & Diabate 2008). In addition to spatial segregation some processes like mate choice could be associated with assortative mating that could be driven by either males or females in swarms. Therefore, knowing how males and females choose their mate within mixed swarms is important for understanding the mechanisms of assortative mating.

For speciation process to occur, ecological divergence and assortative mating are crucial, but there also needs to be genetic structure that allows these traits to be genetically associated (Turner et al. 2005; White et al. 2010). In 2005, Affymetrix GeneChip microarray hybridizations of DNA from *An. gambiae s.s.* and *An. coluzzii* were used to differentiate two small genomic areas adjacent to the centromere of chromosome X and 2L that were significantly differentiated between these sibling species contained 50 and 12 predicted genes respectively (Turner et al. 2005). Later, a third pericentromeric divergence region was found on the chromosome 3L (White et al. 2010). These areas were named ‘‘speciation islands’’(Turner et al. 2005). They were highly different between *An. gambiae s.s.* and *An. coluzzii* species, whereas no fixed differences were found across the rest of the genome (Gentile et al. 2001; Slotman et al. 2007). The low recombination across most of the genome is believed to be due to on-going gene flow between both species that lead to homogenize genomic regions that are not directly involved in the speciation process (della Torre et al. 2002; Slotman et al. 2007). Therefore, the divergence islands model implies the genes responsible for pre-mating isolation and ecological divergence inside limited regions of the genome that are protected from recombination between *An. gambiae s.s.* and *An. coluzzii* (Turner et al. 2005). On the other hand, an alternative model called “incidental island” model suggested that the islands of speciation do not contribute to the reproductive isolation and the lack of differences in the rest of the genome due to the segregation of the

ancestral species (White et al. 2010). According to this model there is no gene flow between *An. gambiae s.s.* and *An. coluzzii* (White et al. 2010). This model was based on study for sympatric populations *An. coluzzii* and *An. gambiae* s.s. in Central West Africa showed that X, 2L and 3L were in strict linkage (White et al. 2010). However, subsequent studies have shown the linkage disequilibrium between the islands of speciation breaks down in some areas of Africa that characterised by higher introgression between *An. gambiae s.s.* and *An. coluzzii* (Caputo et al. 2011; Lee et al. 2013).

Since the models that explain sympatric speciation are controversial subjects and limited genomic studies were conducted in this area of study, we tried to reveal the role of the islands in speciation and to identify the candidate island that would include the genes responsible for reproductive isolation. In addition, we tried to reveal the role of males and females of *An. gambiae s.s.* and *An. coluzzi* in driving the assortative mating behaviour in mixed swarms. In the present study, we created recombinant strains by introgressing X-island of speciation of S molecular form (Kisumu from Kenya) into the M molecular form (Mopti from Mali) through 5 generations of backcrossing. Followed by, two generations of crosses to achieve strain homozygosity (Chapter 2). At the end of this process, we obtained two recombinant strains with either M- or S- form in the island of speciation on X chromosome. The main aim of the present study was genotyping the X, 2L and 3L islands of speciation in recombinant and parental strains to check the islands of speciation genotypes. Thereafter, assortative mating experiments were conducted to test the association between the genotype of X-island of speciation and the mating preference of recombinant and parental strains.

When the females were given the choice to mate with matched and non-matched X-island males, we found that the recombinant females mated with matching X-island-type

males. This confirmed that assortative mating genes are located in the X-island of divergence.

These studies pave the way for understanding the role of islands of speciation in assortative mating and could provide information about the importance of the males or females in assortative mating process.

3.2. Materials and methods

3.2.1. Genotyping X, 2L and 3L

DNAzol extraction

Ten males and ten females were collected separately from recombinant strains ASS, AMM, BSS and BMM, and parental strains Mopti and Kisumu. DNA was extracted individually using DNAzol methods (Chapter 2). Thereafter, the extractions were amplified separately based on the required island.

a) X chromosome: DNA was amplified using standard method and primers (Chapter 2).

b) 2L chromosome: One reaction mix consists of 5µl taq buffer, 2.5µl dNTP (Fermentas, Leicestershire, UK), 0.1µl primer Fwd (5-GCATGGCAGAAAGCTGGTAT-3) (100pmol/µl) (Eurofins Genomics, Wolverhampton, UK), 0.1µl primer Rev (5-GGTCAATGCCTTCCACTGTT-3) (100pmol/µl) (Eurofins Genomics, Wolverhampton, UK), 0.5µl Dream Taq polymerase (Thermo Scientific, Massachusetts, United States), 2µl genomic DNA template (5-50ng/µl) and 14.8µl dH2O, the total reaction volume 25µl. The PCR was carried out using a PTC-200 DNA-Engine thermocycler (BioRad, California, USA) with the following steps: initial denaturation 3min at 95°C; followed by 30 cycles of 15sec at 95°C, 30sec at 56°C and 1min at 72°C; once these 30 cycles were completed final step was 10min at 72°C then -4°C forever.

c) 3L chromosome: We have used 5µl taq buffer, 2.5µl dNTP (Fermentas, Leicestershire, UK), 0.1µl primer Fwd (5-CACAGTTTGAATGGCGAAGA-3) (Eurofins Genomics, Wolverhampton, UK), 0.1µl primer Rev (5-CCTAGTCGGTACAGCGGTTCT-3) (Eurofins Genomics, Wolverhampton, UK), 0.5µl Dream Taq polymerase (Thermo Scientific, Massachusetts, United States), 2µl genomic DNA template (5-50ng/µl) and 14.8µl dH2O for one reaction and the total reaction volume 25µl. Thereafter, the PCR was carried with the same steps in (b).

DNA restriction and gel separation

The restriction endonuclease digestion mix was prepared differently for each chromosome using the following proportions:

a) X Chromosome: The restriction mix and gel separation was carried out using the standard steps (Chapter 2)

b) 2L Chromosome: The restriction mix consists of 2µl HpaI restriction endonuclease (Thermo Scientific, Massachusetts, United States), 2µl of 10xBuffer, 16µl dH2O and 10µl of each amplified sample. Reaction tubes were incubated overnight (12hours) at 37°C. Thereafter, the reaction mix was fractionated by gel electrophoresis using a mix of 1.8% agarose (Fisher brand, Leicestershire, UK), 10µl of 1x Tris-Boric Acid-EDTA (TBE) (Invitrogen, Leicestershire, UK), 1µl Ethidium bromide (Fisher Scientific, Leicestershire, UK) and 1L of dH2O. This was run at 60 volt for 2 hours. Fractionated DNA stained with Ethidium Bromide was visualized under UV light. HpaI enzyme cuts the DNA strands only in M molecular form samples, producing a visible two bands first one is 248bp and another is 151bp, whereas S form shows only one band that is 399bp.

c) 3L Chromosome: The restriction mix consists of 2µl EcoRV restriction endonuclease (Thermo Scientific, Massachusetts, United States), 2µl 10xBuffer, 16µl dH2O and 10µl of

each amplified sample. The mix-tubes were incubated overnight (12hours) at 37°C. Then, the reaction mixes were fractionated by gel electrophoresis using the same way in (b). Fractionated DNA stained with Ethidium Bromide was visualized under UV light. EcoRV cuts the DNA strands only in M form samples showing visible two bands, first one is 175bp and other is 160bp, whereas S form stays as only one band, which is 335bp.

Inversions characterization

Polytene chromosome preparations were made from the ovaries of 10–20 semi-gravid females per strain using established protocols (Coluzzi 1968; Hunt 1973). Inversions present on chromosomes 2L and 2R were scored from chromosome spread under the light microscope.

3.2.2. Recombinant and parental strains mating preference

After genotyping the X, 2L and 3L islands of speciation, we found that BMM and ASS recombinants were contaminated therefore we excluded them from further experiments. AMM, BSS, Kisumu and Mopti were bred in the laboratory standard system to get the pupae. The pupae were sexed using standard sexing methods (Chapter 2), and then they were transferred to separate emergence cages and left to emerge. For each replicate, the 4-days 30 females and 30 males from each strain were placed in the experiment cages (standard 5L rearing cages) with access to 5% glucose (Acrose Organics, New Jersey, USA) solution and given 24 hours to mate (6pm start). They were given a choice to mate with the match and opposite X-island adults.

Dissection of mated females and genetic analysis of sperm

After one night, females from each cage were collected and kept in 70% ethanol. The females were dissected to get the sperm bundle. The dissection process was carried out using standard methods (Chapter 2). The DNA was extracted from the sperm using ChargeSwitch gDNA Micro Tissue Kit (Life Technologies, USA) following the manufacturer's instructions (Chapter 2). The PCR for X-island and Fanello Digestion was carried out for all the samples, and then the gel was run to test the samples (Chapter 2).

3.3. Results

Two sets of studies were conducted to test the association of the X-island of divergence with assortative mating behaviour in *An. gambiae s.s.* and *An. coluzzii*. The first set of tests was genotyping X, 2L, 3L islands of speciation in the recombinants and parental strains to reveal the candidate island that is responsible for the assortative mating behaviour. The second set of tests was analyzing the assortative mating behaviour in the recombinants and parental strains.

3.3.1. Genotyping X, 2L, 3L islands in recombinant and parental strains

In parental strains, Mopti strain was M-genotype in all the three islands of speciation. Kisumu strain showed S-genotype only in X-island, and it was polymorphic in both 2L and 3L islands suggesting that this long-established strain could be contaminated with the M-forms strains in the laboratory. Within the recombinant strains, the analyses showed that AMM strain was M-genotype in all 3 islands, whereas BSS was M-genotype at 3L and 2L and S-genotype at X island. These results suggested that the X-island was swapped succefully between Kisumu and Mopti in the backcrossing process. Both recombinant strains were also polymorphic for *a* on 2L and *u* on the 2R chromosome (Table 3.1,2).

Thus their genotypes and karyotypes were those expected from successful backcrossing and crossing steps.

Table 3.1. Recombinant and parental strains sample size and genotypes at the X, 2L and 3L divergence islands.

		Genotypic frequencies		
Strain	**N**	**X**	**2L**	**3L**
M Mopti	14	MM	MM	MM
S Kisumu	14	SS	0.2MM, 0.4MS, 0.4SS	0.6MM, 0.4MS
AMM	20	MM	MM	MM
BSS	21	SS	0.9MM, 0.1MS	MM

Table 3.2. Recombinant and parental strains sample size, and 2L and 2R inversion karyotypes.

		Inversion frequencies	
Strain	**N**	**2L**	**2R**
M Mopti	20	a/a	+/u
S Kisumu	20	+/a	+
AMM	20	0.15a/a, 0.85+/a	0.25+/+, 0.75+/u
BSS	10	0.2a/a, 0.8+/a	0.4+/+, 0.6+/u

3.3.2. Analysing assortative mating behaviour

The females from AMM, BSS, Kisumu and Mopti strains were given a choice to mate with the matched and non-matched X-island virgin males. This test was repeated for the males that also were given a choice to mate with the matched and non-matched X-island females.

Preliminary assortative mating experiments to test the recombinant females preference

The genetic analyses of the sperm that extracted from the spermathecae of mated females showed that females mated preferentially with recombinant males with an X-island matching their own (Table 3.3). Recombinant females with M-type X-islands mated with matching M-type X-island males on average 77% of the times (Chi-square likelihood-

ratios: n=52, x^2=15.9, P< 0.001) (Fig 3.1.). Females with S-type X-island mated with recombinant males with matching S-type X-island on average 81% of the times (Chi-square likelihood-ratios: n= 47, x^2= 19.2, P< 0.001) (Fig 3.1.).

Table 3.3. Number of females mating assortatively and disassortatively in preliminary behavioural assays among recombinant strains - Recombinant females (X-island genotypes MM or SS) were given a choice between recombinant males with X-chromosome speciation island matching their own or not. The number of replicates, mating combinations, numbers and percentages (in brackets) of mating, and level of significance (Chi-square Likelihood-ratios) are indicated.

Replicate	Mating combination		Mating Type (%)		Chi-square	*P*-value
	Females	Males	Assortative	Disassortative		
1	MM	M+S	12	2	7.9	0.005
2	MM	M+S	7	4	0.8	0.362
3	MM	M+S	10	3	4	0.046
4	MM	M+S	11	3	4.9	0.028
All			*40 (76.92)*	*12 (23.08)*	*15.9*	*< 0.001*
1	SS	M+S	9	2	4.8	0.028
2	SS	M+S	10	6	1	0.314
3	SS	M+S	13	1	12.2	0.001
4	SS	M+S	6	0	8.3	0.004
All			*38 (80.85)*	*9 (19.15)*	*19.2*	*< 0.001*

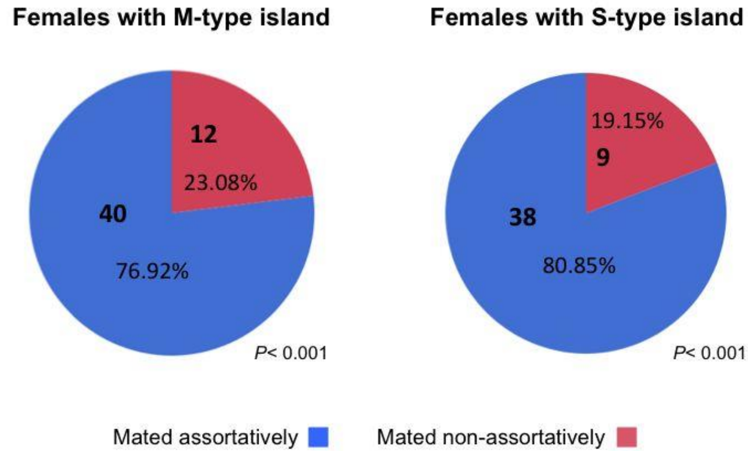


Figure 3.1. Overall percentage assortative mating in females from recombinant strains carrying M or S type X-chromosome islands in preliminary mating experiments - Thirty 2–5 day-old virgin females where presented with a mixture of 2-5-day-old recombinant males matching and non-matching their own X-island molecular types in standardized overnight mating assay.

Assortative mating analyses within parental strains

No previous studies tried to test assortative mating behaviour for laboratory forms of *An. coluzzii* and *An. gambiae s.s.*. Therefore, we tried to test the assortative mating in Kisumu that is S-like form and Mopti which is M-genotype form. The females of both Kisumu and Mopti mated significantly assortatively (P< 0.001 in both strains), whereas the males of both strains mated with both matched and non-matched X-island females (P= 0.073 and 0.163) (Fig. 3.2.b and Table 3.4).

Table 3.4. Number of females and males mating assortatively in reciprocal behavioural assays among the Kisumu and Mopti parental strains - Females (X-island genotypes MM or SS) were given a choice between males from their own strain and molecular type or not (top half of table). The reciprocal experiments were also conducted with males (X-island genotypes M or S) choosing females (bottom half of table). The number of replicates, mating combinations, numbers and percentages (in brackets) of mating, and level of significance (Chi-square Likelihood-ratios) are indicated.

Replicate	Mating combination		Mating Type (%)		Chi-square	*P*-value
	Females	Males	Assortative	Disassortative		
1	MM	M+S	18	5	7.8	0.005
2	MM	M+S	19	2	15.9	< 0.001
3	MM	M+S	20	1	27.1	< 0.001
All			*57 (87.69)*	*8 (12.30)*	*41.6*	*< 0.001*
1	SS	M+S	20	5	9.6	0.002
2	SS	M+S	28	0	-	< 0.001
3	SS	M+S	14	6	3.3	0.07
All			*62 (84.93)*	*11 (15.07)*	*39.3*	*< 0.001*
Replicate	**Mating combination**		**Mating Type (%)**		**Chi-square**	***P*-value**
	Males	**Females**	**Assortative**	**Disassortative**		
1	M	MM+SS	15	21	1	0.316
2	M	MM+SS	10	15	1	0.316
3	M	MM+SS	12	14	0.2	0.695
All			*37 (42.53)*	*50 (57.47)*	*1.9*	*0.163*
1	S	MM+SS	22	9	5.6	0.018
2	S	MM+SS	11	13	0.2	0.683
3	S	MM+SS	15	10	1	0.316
All			*48 (60.00)*	*32 (40.00)*	*3.2*	*0.073*

Assortative mating analyses within recombinant strains

The results showed that the females from AMM strain almost mated assortatively (P< 0.001) (Fig. 3.2.a). The BSS females entirely mated assortatively with the matched X-island males (P< 0.001) suggesting that their mating preference effectively swapped when the X-island swapped during the backcrossing process (Fig. 3.2.a). However, the males from both AMM and BSS recombinant strains were not choosy, and mated with both matching and non-matching X-island females (P= 0.284 and 0.611 respectively) (Table 3.5).

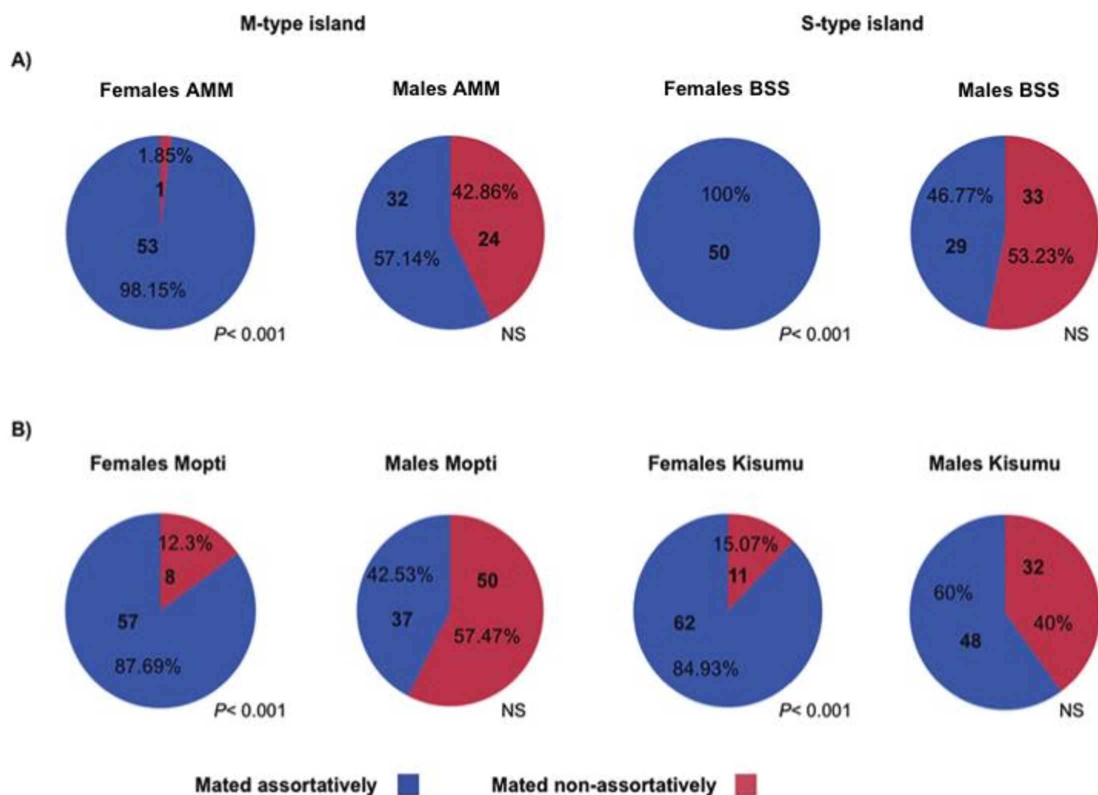


Figure 3.2. Percentage assortative mating in females and males carrying M or S-type X-chromosome islands - A) virgin females or males from the AMM and BSS recombinant strains where presented with a mixture of recombinant individuals of the opposite sex and with X-islands matching and non-matching their own X-island molecular type; B) virgin females or males from the M-form Mopti and S-form Kisumu strains used for creating the recombinant lines were given a choice between equal numbers of potential mates from both strains.

Table 3.5. Number of females and males mating assortatively in reciprocal behavioural assays among the AMM and BSS recombinants strains – Recombinant females (X- island genotypes MM or SS) were given a choice between recombinant males with X- chromosome speciation island matching their own or not (top half of table), The reciprocal experiments were also conducted with recombinant males (X-island genotype M or S) choosing recombinant females (bottom half of table). The number of replicates, mating combinations, numbers and percentages (in brackets) of mating, and level of significance (Chi-square Likelihood-ratios) are indicated. For each mating combinations, 2 replicates were conducted using 5-day-old mosquitoes reared from independent mosquito cohorts.

Replicate	Mating combination		Mating Type (%)		Chi-square	*P*-value
	Females	**Males**	**Assortative**	**Disassortative**		
1	MM	M+S	29	1	32.8	<0.001
2	MM	M+S	24	0	33.3	<0.001
Both			53 (98.15)	1 (1.85)	64.9	<0.001
1	SS	M+S	25	0	34.6	<0.001
2	SS	M+S	25	0	34.6	<0.001
Both			50 (100)	0 (0)	69.3	<0.001
Replicate	**Mating combination**		**Mating Type (%)**		**Chi-square**	***P*-value**
	Males	**Females**	**Assortative**	**Disassortative**		
1	M	MM+SS	15	16	0.1	0.857
2	M	MM+SS	17	8	3.3	0.069
Both			32 (57.14)	24 (42.86)	1.1	0.284
1	S	MM+SS	18	13	0.8	0.368
2	S	MM+SS	11	20	2.7	0.103
Both			29 (46.77)	33 (53.23)	0.26	0.611

3.4. Discussion

Explaining the role of genomic structure of speciation process in sympatric species *An. coluzzii* and *An. gambiae s.s.* was thought to be challenging due to their wide distribution in Africa that accompanied with variety of introgression and selection observed between them within these areas (Weetman et al. 2012; Coetzee et al. 2013; Lee et al. 2013). In this study, the role of X, 2L and 3L islands in the speciation process was tested, and it was found the recombinant strain BSS and the parental strain Kisumu were S genotype only in X-island. These findings suggested that the pericentromeric X-island associated with maintaining the genetic integrity for sympatric species in Central and Eastern areas of Africa where they mate assortatively (Tripet et al. 2001; Lee et al. 2013), and in West Africa where the gene flow is common between these species (Caputo et al. 2011; Lee et al. 2013). These results are compatible with the speciation islands model suggesting that the X-island plays the main role in speciation process (Feder et al. 2012; Via 2012; Seehausen et al 2014). On the other hand, this study showed that the recombinant strains were almost homozygous M-type in 2L and 3L islands of the genome. Therefore, 2L and 3L are not directly associated with the assortative mating behaviour in these recombinants, and further studies are required to reveal their role in speciation process. In addition, the recombinant strains had polymorphic inversions suggesting that there is no evidence for the inversions role in evolution of reproductive isolation between the forms (Lanzaro & Tripet 2003). Therefore, the inversions could be involved in the ecological adaptation rather than assortative mating behaviour in the sibling species.

In addition, the assortative mating was tested in recombinants and parental strains. The results showed that the assortative mating occurred with higher frequencies in recombinant strains compared with parental strains. Given that the recombinants differed only in the X-island, the results suggested that the X-island directly involved in a perfect

assortative mating behaviour in recombinants, whereas the interaction between the X-island and other genomic regions in parental strains could lead to more variable mating phenotype. In addition, the females showed strict assortative mating behaviour in our study, which indicate to a perfect phenotypic expression of conspecific recognition mechanisms in the females and perfect cues in the males. Furthermore, the males did not mate significantly assortatively in our study. This could be due to the limit space was given to these males in the cages that affect their swarming behaviour comparing to the field. However, field studies showed that males could involve in the assortative mating behaviour by forming separated swarms in different areas based on the species (Diabaté et al. 2006; Dabire et al. 2013). Moreover, other similar outdoors preference mate experiments, where virgin females and males of *An. coluzzii* were given a choice to mate with conspecific and interspecific individuals, showed that both sexes mated assortatively. According to these studies, although females drive the assortative mating behaviour in mixed swarms, males contribute directly or indirectly to assortative mating and their role cannot be denied.

Creating recombinant strains that mate perfectly assortatively is promising step for further behavioural and genomic studies that test the conspecific recognition mechanisms that the females and males of *An. gambiae s.s.* and *An. coluzzii* species use in mating process. Moreover, understanding these mechanisms and revealing reproductive isolation genes are important steps for understanding the sympatric speciation process.

CHAPTER FOUR

Genome sequencing of recombinant and parental strains, and comparative genomic with field sympatric populations

4.1. Introduction

Different studies tried to explain the speciation process that facilitates creating two species from one original taxon (Via 2012; Singh 2012). Understanding speciation process is the key for understanding biodiversity and evolution within population (Singh 2012). To understand speciation process, searching for the genetic and ecological conditions is essential step for determining the speciation causes (Via 2012).

Sympatric speciation is considered one of the controversial and challenging speciation models because the speciation occurs despite the on-going gene flow between the species (Via 2012). However, sympatric speciation could be explained by the occurrence of reproductive isolation and ecological adaptation between populations in an area, which leads to create new species in this area (Via 2012). For sympatric speciation to occur, divergent selection against adapted alleles occurs and continues to the nearby genes (Via 2012; Feder et al. 2012; Butlin et al. 2012). This process limits introgression in the selected adapted genes within the population and protects these genes from disruption (Via 2012; Feder et al. 2012; Butlin et al. 2012). This could lead to assortative mating when the mate choice is associated with a specific trait under divergent selection, which leads to intraspecific mating within species and facilitates the creation of two different species from one taxon (Via 2012). Theoretical models of sympatric speciation have long recognised that the genomic architecture often plays a major role (Feder et al. 2012; Via 2012; Seehausen et al. 2014) predicting that chromosomal inversions and peri-centromeric

regions suppress recombination and link together genes of pre-mating isolation and ecological adaptation genes (Rieseberg 2001; Coyne 2004; Servedio 2009; Feder et al. 2012). Moreover, hemizygosity and lower recombination rates in the sex chromosomes are thought to cause rapid accumulation of genes of pre- and post-mating isolation (Qvarnstrom & Bailey 2009). The current studies concern almost species already separated by both pre-mating and intrinsic post-mating reproductive isolation which make it difficult to differentiate between the genes responsible for assortative mating and post-mating in speciation process (Feder et al. 2012; Via 2012; Seehausen et al. 2014).

Anopheles gambiae complex is considered as challenging model for studying and understanding sympatric speciation process. Its importance lies in its complicated structure that presents early stages of speciation (Turner et al. 2005; Turner & Hahn 2007; Weetman et al. 2012; Sawadogo et al. 2013). In addition, *Anopheles gambiae* is known for its ability to transmit *Plasmodium* parasite to human and cause malaria infection throughout Africa (Coetzee et al. 2013). They differentiated two sibling species *An. coluzzii* and *An. gambiae s.s.* were until recently known as the 'M and S' molecular forms of *An. gambiae* based on diagnostic genetic differences in the ribosomal DNA regions (Coetzee et al. 2013). Despite *An. gambiae s.s.* and *An. coluzzii* species overlap in many areas throughout Africa, *An. coluzzii* is mostly found in flooded areas and some areas of Central Africa, whereas *An. gambiae s.s.* can be found in rainy locations widely throughout Africa (Lanzaro & Tripet 2003). Across much of their sympatric range their integrity is maintained by strong assortative mating (Tripet et al. 2001; Dabire et al. 2013). Hybrids frequencies varied in different areas of Africa. In Mali, the hybrids frequency was approximately 1% (Tripet et al. 2001), whereas no hybrids were found in Cameroon (Diabaté et al. 2006; Diabaté et al. 2007). However, in the Western coastal countries of Guinea Bissau and Senegal hybrids can occur locally at much higher frequencies (Caputo et al. 2011), resulting in a large

hybrid zone between the incipient sibling species and high levels of genetic introgression (Coluzzi et al. 2002; Lee et al. 2009). Hybrids in laboratory were found to be fully viable and fertile (Diabaté et al. 2006; Diabaté et al. 2007). These results indicate that there are no intrinsic post-mating barriers, but there are incomplete pre-mating barriers responsible for the reproductive isolation between *An. gambiae s.s.* and *An. coluzzii* species in field (Lanzaro & Lee 2013).

Divergence between these species is thought to be driven by larval adaptation to different types of water bodies (Diabaté et al. 2008; Kamdem et al. 2012). Field studies showed that *An. coluzzii* larvae survived better than those of *An. gambiae s.s.* in presence of aquatic predators (Diabaté et al. 2007; Gimonneau et al. 2012; Roux et al. 2013). In another study, it was reported that in dry season, *An. coluzzii* usually undergo aestivation in Sahelian areas, which helps its population to peak faster when the rainy season starts, whereas *An. gambiae s.s.* individuals migrate away during the dry seasons (Dao et al 2014). However, adults of both species have similar feeding and resting habits and mate in swarms at dusk in villages. Swarm site segregation is thought to contribute to assortative mating (Diabaté et al. 2006), but the occurrence of mixed swarms at various frequencies (Diabaté et al. 2006; Dabire et al. 2013) indicates to additional conspecific recognition mechanisms that the mosquitoes use in assortative mating, possibly based on flight tones (Pennetier et al. 2010; Gibson et al. 2010).

Several studies attempt to explain the speciation process and two explanations were concluded (White et al. 2010; Lanzaro & Lee 2013). The first explanation accepted that pre-mating isolation between these two incipient species probably is caused by selected adapted genes "Islands of speciation" located in low recombination regions of the genome (Slotman et al. 2006; Slotman et al. 2007; Via 2012; Lanzaro & Lee 2013). Genomic comparison studies in *An. coluzzii* and *An. gambiae s.s.* showed that the genetic differences

between the incipient species located in three regions 2L, 3L and X chromosome near the centromere (Turner et al. 2005; White et al. 2010). These pericentromeric islands of divergence were only 3% of the total genome (Lanzaro & Lee 2013) in addition to smaller islands located in the vicinity of inversion breakpoints (Turner et al. 2005; White et al. 2010). Perfect linkage disequilibrium between the X, 2L and 3L islands was found in samples from sympatric populations of *An. coluzzii* and *An. gambiae s.s.* from Central West Africa (White et al. 2010). This leads to the second explanation that suggested very low gene flow between the sibling species and the possibility that pericentromeric islands of divergence were 'incidental rather than instrumental' to the speciation process (White et al. 2010). In addition, genotyping whole genome of *An. coluzzii* and *An. gambiae s.s.* presented that the divergence sites spreads widely in the genome between them (Lawniczak et al. 2010). However, current studies showed that the linkage disequilibrium breaks down between the divergence islands in areas where higher introgression occurs between *An. gambiae s.s.* and *An. coluzzii* species (Caputo et al. 2011; Lanzaro & Lee 2013). These results in addition to comparative genomics studies would support "islands of speciation" model suggesting that these islands could play a major role in sympatric speciation in the face of varying levels of gene flow (Weetman et al. 2012; Lanzaro & Lee 2013).

Since this view is currently subject to debate, we set out to demonstrate the role of divergence islands in the sympatric speciation process using an experimental functional genomics approach. In Chapter 3, genotyping X, 2L and 3L islands of speciation for the recombinants that mated assortatively and differed only in genes of divergence, showed that they only differed in X-island (Chapter 3). These findings revealed a strong association between the X-island and assortative mating behaviour in recombinant strains. These results suggested that the X-island would include the genes responsible for

assortative mating and ecological adaptation genes in the face of on-going gene flow between *An. gambiae s.s.* and *An. coluzzii* because it combines pericentromeric recombination suppression with the hemizygosity and decreased recombination typically associated with the X chromosome. In chapter 2, we selectively introgressed the S-form X-island of divergence of *An. gambiae* s.s. into the M-form *An. coluzzii* genetic background to create recombinant strains that shared *An. coluzzii* genetic background but only differed at their X-islands of speciation. Here, full-genome strain comparisons for parental and recombinant strains, and targeted genome sequencing for sympatric populations were done. This strategy enabled us to identify the size of the pericentromeric region introgressed from *An. gambiae* s.s. into the *An. coluzzii* background. In addition, fixed protein-coding differences were found distinguishing recombinant strains, and conserved differences putatively relevant to speciation.

Revealing the genomic fixed differences between the *An. gambiae s.s.* and *An. coluzzii* is important step to pave the way for identifying genes responsible for reproductive isolation mechanisms and speciation process.

4.2. Materials and Methods

4.2.1. DNA preparation of recombinant and parental strains

Genotyping X, 2L and 3L in AMM, BSS, Kisumu and Mopti strains showed that Kisumu was contaminated in 2L and 3L islands. Therefore, it was excluded from our analyses.

DNA extraction

The DNA was extracted using ChargeSwitch (Chapter 2) to get pure DNA. For this step, 24 individuals from AMM, 23 individuals from BSS and 13 individuals from Mopti were used in our analyses.

Genome amplification

To get the required DNA amount that is equivalent to ≈ 20000 ng/ml for sequencing in Liverpool DNA centre, whole genome amplification was done for all the samples using the Illustra GenomiPhiV2 DNA Amplification kit (GE Healthcare Bio-sciences, Piscataway, NJ). We added 8µl sample buffer into 2µl of extracted DNA for each sample to get final volume 10µl, then it was heated to 95°C for 3min and cooled to 4°C. Thereafter, 9µl of reaction buffer was added to 1µl of enzyme mix for each sample to create master mix. Thereafter, 10µl from master mix was added to every sample. The samples tubes were heated to 30°C for 1:30min, and then to 65°C for 10min. At the end, the samples tubes were cooled to 4°C forever on PCR machine.

Genome purification

The genome was purified using a MinElute Reaction Cleanup Kit (Qiagen, Hilden, Germany). For this purpose, the samples were transferred to 1.5ml tubes (Fisher brand, Leicestershire, UK), and their volume must be between 20µl -100µl. However, when it was

less than 20µl we added dH2O. After that, 300µl ERC (enzymatic reaction cleanup buffer) was added and the mix was applied to min elute columns, and centrifuged for 1min. To wash the DNA, 750µl PE (wash buffer) buffer was added to the samples and centrifuged for 1min and the flow throw was discarded. To elute the DNA, the min elute column was placed in a clean 1.5ml tube and 10µl of EB (elution buffer) buffer was applied to the centre of the membrane, and left stand for 1min, then centrifuged for 1min. The pure DNA concentration was measured for each sample using Nano drop (Thermo Scientific, Massachusetts, United States) and stored in -20°C.

Picogreen assay (DNA Concentration)

For accurate measurements we used the Picogreen assay to measure the DNA concentration. The samples were prepared using standard Picogreen kit method (Invitrogen™ Quant-iT™, Leicestershire, UK): 1xTE (Fisher Scientific, Leicestershire, UK) was prepared from the 20xTE stock by adding 50µl of 20xTE to 950µl dH2O. This diluted stock will be used for diluting the samples. Thereafter, we diluted standard DNA from 100µg/ml to 2µg/ml (to make 300µl final volume 6µl of standard DNA was added to 294µl of 1xTE) for creating standard curve (Table 4.1).

Table 4.1. Series of standard DNA dilutions for creating standard curve to measure the DNA concentration in the samples using Picogreen assay – The table shows the required volumes of the Stock diluted DNA and the TE buffer to get the final required DNA concentrations.

Plate Well	Final DNA (ng/ml)	Vol.(µl) 2µg/ml DNA diluted stock	Vol.(µl) 1xTE buffer	Final total DNA ng/well
Any	0	0	100	0
A1 & A2	200	20	80	40
B1 & B2	300	30	70	60
C1 & C2	400	40	60	80
D1 & D2	500	50	50	100
E1 & E2	1000	100	0	200
F1 & F2	2000	4µl (100µg/ml) stock	96	4µg/ml
G1 & G2	4000	8µl (100µg/ml) stock	96	8µg/ml

For measuring the DNA in unknown samples, 1µl of each unknown sample was added to 99µl of 1xTE in the microplate wells. Thereafter, we diluted Picogreen reagent by adding 1µl into 200µl of 1xTE, and then we added 100µl of diluted Picogreen to each standard and unknown sample. The sample plates were covered with foil and incubated in room temperature for 2-5min. Thereafter, the plates were read by the fluorescence reader and the concentrations were multiplied to the dilution number (x200).

The samples with low concentration were mixed and re-amplified, purified and measured until we got the required concentrations. All the samples of each strain were mixed together and placed in storage tubes, and the purity of DNA was checked using Nano drop (Thermo Scientific, Massachusetts, United States) (Table 4.2).

Table 4.2. DNA concentration and the purity ratios – The diluted DNA concentrations in the Picogreen samples multiplied to the dilution ratio (x200) to get the DNA concentration in the samples, and the purity was measured using nucleic acid purity ratios: the ratio of absorbance at 260/280 ~1.8 is generally accepted as "pure" for DNA. If the ratio is appreciably lower, it may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280nm. The purity was measured also using the second measure of nucleic acid purity 260/230. Expected 260/230 values are commonly in the range of 2.0-2.2. If the ratio is appreciably lower than expected, it may indicate the presence of contaminants that absorb at 230nm (Thermo scientific, Nanodrop 1000 & 8000).

Sample	DNA concentration in the samples	260/280	260/230
Mopti	36948 ng/µl	1.87	2.61
BSS	78482 ng/µl	1.95	2.45
AMM	95477 ng/µl	1.85	2.4

4.2.2. Genome sequencing of recombinant and parental strains

The DNA pools were sent to the Liverpool Centre for Genomic Research (CGR) for sequencing. DNA libraries were prepared according to the Illumina TruSeq DNA protocol (Illumina, San Diego, CA), multiplexed and sequenced on two lanes of an Illumina HiSeq 2000 sequencer.

Base-calling of indexed reads was performed with the program CASAVA 1.8.2 (Illumina). The reads were trimmed using the software Cutadapt 1.2.1 (Martin 2011) and Sickle 1.200 (Joshi 2011) and mapped to the *An. gambiae* (PEST) reference sequence (assembly AgamP3) using Bowtie 2.1.0 (Langmead 2012). Alignments were filtered to remove low mapping quality reads and redundant duplicate reads were filtered out using the Picard MarkDuplicates Tool 1.85 (http://picard.sourceforge.net). Mapped reads were locally re-aligned around indels using the Genome Analysis Tool Kit (GATK) version 2.1.13 (McKenna et al. 2010; DePristo et al. 2011). The mean coverage depth after local re-alignment and duplicate removal of reads was equal to 25.3x for BSS, 30.9x for AMM

and 34.1x for the Mopti parental strain. Variant detection was performed using the GATK 'UnifiedGenotyper' package (McKenna et al. 2010; DePristo et al. 2011) with an expected SNP heterozygosity of 0.01. An expected ploidy of 20 was used (i.e. allele frequencies calculated in increments of 5%) in order to best balance accurate sample representation and computational efficiency. Variants were further filtered using the GATK 'VariantFiltration' package (McKenna et al. 2010; DePristo et al. 2011). This resulted in the characterization of ~6 million SNPs (~4.8 million passing all filters) and 900,000 indels in each of the sequenced strains. All variants were annotated using snpEff 3.1 (Cingolani et al. 2012). Visual alignment inspections were performed using the Integrative Genomics Viewer (IGV) (Thorvaldsdottir et al. 2013).

Estimates of genetic differentiation F_{ST} between two populations *a* and *b* were calculated based on SNPs satisfying GATK's most stringent 'pass' criteria and using the formula:

$F_{ST} = 1 - H_s/H_t$ where H_s is the mean heterozygosity across populations *a* and *b* and H_t the total heterozygosity across all populations (Nei 1987).

$H_s = 1 - \Sigma pi^2$ where *pi* are the mean frequencies of the major (most frequent) and minor (least frequent) alleles calculated from the *a* or *b* populations

$H_t = 1 - \Sigma pi^2$ with *pi* being SNP frequencies calculated across all three populations.

Pair-wise F_{ST} estimates of genetic differentiation were used for generating scans of genetic differentiation across chromosomes with the software JMP 10 (SAS Institute, Inc). In order to best outline the genomic region(s) introgressed from *An. gambiae s.s.* into *An. coluzzii* in the BSS and AMM strains, the 'spline' function was fitted over every high-confidence SNPs with $\delta = 2.7216$.

Separate data analyses identified all unique protein-coding differences between the BSS, AMM and Mopti strains. These differences were checked by visual inspection and comparison of their genomes using the software IGV.

4.2.3. Deep targeted sequencing of field sympatric populations

Anopheles gambiae s.l. larvae were collected in Akoti-Chirano (Lat. 6° 6' 17.08" N; Long. 2° 19' 0.45" W) in the Bibiani-Anhwiaso-Bekwai district of the Western region of Ghana, West Africa. Populations of *An. coluzzii* and *An. gambiae s.s.* co-occur in this deciduous forested area and are both of the 'Forest chromosomal form' characterized by standard karyotypic arrangements (no paracentric inversions) (Appawu et al. 1994). Larvae were reared to adulthood at the Department of Animal Biology and Conservation Science, University of Ghana, Legon, West Africa, stored in ethanol, and shipped to Keele University. The samples were then individually characterized as *An. coluzzii* (M molecular form) and *An. gambiae s.s.* (S molecular form) as described above. The DNA from 30 individuals of each sibling species was pooled and purified using a MinElute Reaction Cleanup Kit (Qiagen, Hilden, Germany) and DNA pools were sent to the Liverpool Centre for Genomic Research (CGR) for sequence capture and sequencing.

SureSelect RNA oligomer baits (Agilent, Santa Clara, CA) were designed to cover coding regions from position 17Mb to the centromere of the X-chromosome based on the (PEST) reference sequence (assembly AgamP3). Prior to the amplification of pre-capture libraries, DNA fragments larger than ~700bp were removed from DNA pools using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA). Following amplification and adapter-ligation, 750ng of pre-capture libraries were hybridised to 2µl of RNA oligomer baits for ~24 hours at 65°C. Captured libraries were amplified, indexed, pooled and sequenced on 1 lane of an Illumina HiSeq 2000. All other procedures were as described

above. Within the targeted pericentric region of the X chromosome, the mean coverage depth after local re-alignment and duplicate removal of low mapping quality and redundant reads was 200x for *An. gambiae s.s.,* leading to the identification of 26,974 SNPs and 2,593 indels. In *An. coluzzii*, coverage depth was 255x and 18,772 SNPs and 1,738 indels were identified.

4.3. Results

4.3.1. Sequencing the full-genome of recombinants and parental strains

Full-genome sequencing of Mopti, AMM and BSS strains was conducted to determine the X-island and nearby regions size that are used to differentiate AMM and BSS strains. The size was determined using genome-wide genetic differentiation (F_{ST}) scans, and the occurrence of fixed coding differences between AMM and BSS. The genetic differentiation was analysed and calculated for 3,743,318 SNPs marker loci across the sex, second and third chromosomes. The genome-wide scans showed that the BSS differs from the AMM and Mopti strains, by high F_{ST} values in the region that extended from site ~14.8Mb to the centromere on the chromosome X from (Fig. 4.1).

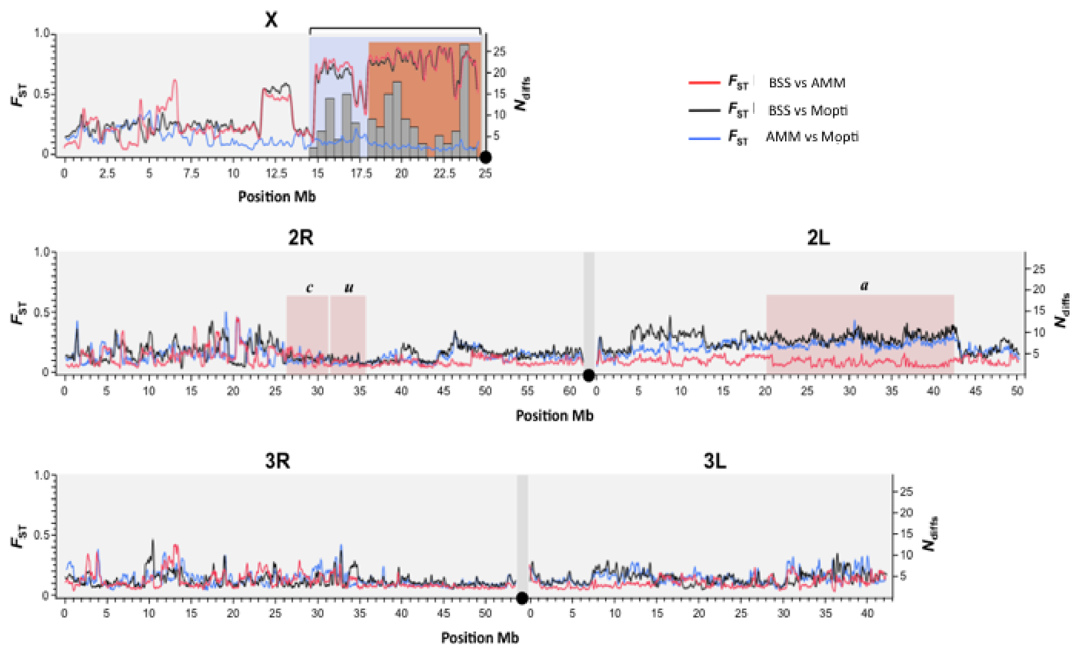


Figure 4.1. Genomic structure of recombinant strains - The genomes of the assortatively-mating AMM, BSS and parental Mopti strains were compared using F_{ST} estimates at ~3x10^{6} SNP marker loci (left Y-axis and red, blue and black lines). The genomic region introgressed from Kisumu into the Mopti genetic background and differing between the AMM and BSS recombinant strains is characterized by high F_{ST} values (blue shade) and extends from position ~14.5Mb to the centromere on chromosome X. The AMM and BSS differed at 160 protein-changing positions all of which located within the introgressed island and flanking region (right Y-axis, grey histogram bars). The pericentromeric region sharing conserved fixed differences with the field *An. coluzzii and An. gambiae s.s* populations starts at position ~18.1Mb (orange shade). The position of inversions *c*, *u* and *a* on chromosome 2 is indicated (pink shade).

This different region included whole X island of divergence and a big nearby region. Additionally, the genome-wide comparison between the BSS and AMM strains showed that there was additional region that is genetically different between them at a ~2Mb S-form fragment, that extended approximately from sites 11.5- 13.5Mb of the X chromosome (Fig. 4.1). However, there were no other genetically different regions detected between the Mopti, AMM and BSS. These findings indicated that the selective introgression -for creating the recombinant strains- between the *An. coluzzii* and *An. gambiae s.s.* species worked as hoped for, that the differences between the recombinant strains only occurred on the X-island of speciation. Importantly, The AMM and BSS differed at 160 non-synonymous SNPs positions that change the protein-coding sequence, that were found only in the introgressed pericentromeric region between AMM and BSS (Fig. 4.1). However, the other fixed differences that occurred elsewhere in the genome were either coding synonymous changes of the protein-coding sequence or non-coding regions.

4.3.2. Comparison with X-island from sympatric field populations

Kisumu and Mopti strains were colonized from allopatric populations for over 25 and 7 years in laboratory. Because of colonisation effects, the differences that were observed between the AMM and BSS recombinant strains -that were obtained from backcrossing colonised Kisumu and Mopti strains- could be due to either genetic divergence of the original populations or genetic drift and inbreeding (Baeshen et al. 2014). Therefore, the protein-coding differences that were different between AMM and BSS were compared with those identified between two sympatric *An. gambiae s.s.* and *An. coluzzii* populations collected from Southern Ghana. Deep-pooled-targeted exon re-sequencing was performed for the region extending from centromere to the 17Mb position on the X chromosome. The analyses showed that the M-form allele was fixed (freq> 0.95) in the field strain *An.*

coluzzii in 114 of the 117 coding differences that differentiate BSS from AMM. In addition, the alternate S-form allele was found (freq> 0.8) in the sympatric strain *An. gambiae s.s.* in 61 of the 114 differences, and 20 of those were fixed (freq >0.95). The genetic differences extended from site ~18.1Mb, and its frequency increased when it approach the centromere and affect 12 genes on X chromosome (Fig. 4.2, Table 4.3).

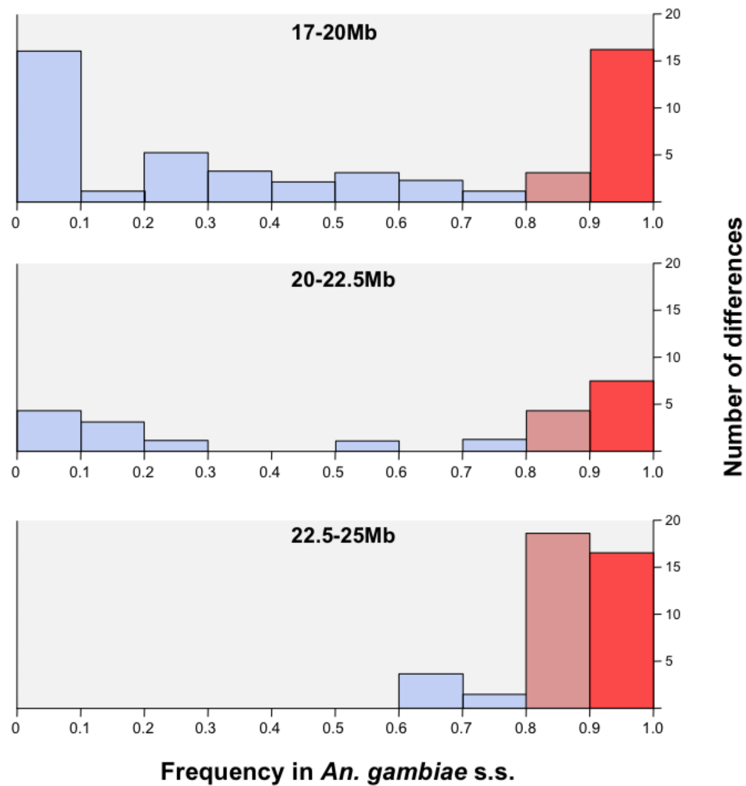


Figure 4.2. Field population frequency distribution of protein coding SNPs identified in BSS - A region covering the X-island and flanking region up to reference position 17Mbp was captured and re-sequenced in sympatric *An. gambiae s.s.* and *An. coluzzii* populations from Ghana. The frequency of alleles coding for unique protein differences in the BSS recombinant strain was measured in the field *An. gambiae s.s.* population. The proportion of alleles occurring at high 0.8 (orange bars) and very high (freq >0.95: red bars) frequency increased towards the centromere suggesting a potential role in speciation whilst other alleles (blue bars) were not conserved.

Table 4.3. Protein coding changes identified between the AMM and BSS strains confirmed in sympatric field populations of *An. coluzzii* and *An. gambiae s.s.* – Twenty differences identified between the AMM and BSS recombinant strains were also fixed or nearly-fixed (freq> 0.95) in sympatric field populations. These were distributed over 12 genes located within the ~6MB pericentromeric island of speciation. These genes cause changes in specific proteins codes, hence they change the proteins function; e.g: the change in genes code from gGg in M-form to gAg in S-form causes a change in the protein 74 code from G to E.

Reference position	M-form allele	S-form allele	Codon (protein change)	Gene (putative function)
18314527	G	A	gGg/gAg (G74E)	AGAP013136
19052774	T	G	gAc/aCc (N25T)	AGAP001002 (Toll Protein)
19114172	C	G	gGg/gCg (G232A)	AGAP001009
19114460	A	C	Tgg/Ggg (W160G)	AGAP001009
19114646	G	C	Ctg/Gtg (L98V)	AGAP001009
19116606	T	C	cAc/cGc (H197R)	AGAP013526
19636043	A	G	aAg/aGg (K272R)	AGAP001022 (gastrin/cholecystokinin receptor)
19636265	C	T	gCc/gTc (A346V)	AGAP001022
19637489	T	C	Tcg/Ccg (S355P)	AGAP001022
19714995	C	T	cGc/cAc (R491H)	AGAP001025 (Protein msta)
19815325	A	T	Acg/Tcg (T339S)	AGAP001033 (mab-21 like protein)
20955148	G	A	aCg/aTg (T281M)	AGAP001050 (chondroitin polymerizing factor)
21093789	A	C	Ttt/Gtt (F288V)	AGAP001052 (ubiquitin carboxyl-terminal hydrolase)
22104989	T	C	Aat/Gat (N1296D)	AGAP001061
22159522	T	C	Atg/Gtg (M514V)	AGAP001061
22751398	A	G	Act/Gct (T252A)	AGAP001073
23799338	T	A	gaA/gaT (E879D)	AGAP0013341
23799431	T	G	gaA/gaC (E848D)	AGAP0013341
23799525	T	G	aAt/aCt (N817T)	AGAP0013341
23799541	T	G	Att/Ctt (I812L)	AGAP0013341

4.4. Discussion

Analysing the genetic differences between *An. gambiae s.s.* and *An. coluzzii* is fundamental for understanding and explaining the sympatric speciation process. There were consensus that the main genetical differences between them occurred in three pericentromeric islands of speciation X, 2L and 3L of the chromosomes (Turner et al. 2005; White et al. 2010). However, recent study of Clarkson in (2014) confirmed that the introgression of insecticide resistance mutation from *An. gambiae s.s.* into *An. coluzzii* resulted in homogenization the whole 2L island of speciation between them suggesting a limit role of 2L in speciation process in *An. gambiae s.s.* and *An. coluzzii* species (Clarkson et al. 2014). In addition, the X-island was the largest amongst the islands of divergence (Turner et al. 2005a; Stump et al. 2005). Previous researches reported that this island might spans 3-5Mb and includes 75-200 genes (Stump et al. 2005; Slotman et al. 2006). Other studies reported a reduction in the recombination by 16and 35-fold in the pericentromiric regions compared to other regions on the X chromosome (Slotman et al. 2006; Pombi et al. 2006). These studies suggested that the X-island could play main role in assortative mating and it could contain the genes responsible for reproductive isolation between *An. gambiae s.s.* and *An. coluzzii* species.

This is the first genomic study focused on the X-island function in reproductive isolation, and its importance in sympatric speciation between *An. gambiae s.s.* and *An. coluzzi*. In our study, a comparison between *An. gambiae s.s.* and *An. coluzzii* sympatric species collected from Ghana and lab recombinant strains showed that the X-island covered 6Mb of the X chromosome and spans from site ~18.1 to 24.2Mb. In addition, we identified 20 protein-coding changes between *An. gambiae s.s.* and *An. coluzzii*. These protein-coding changes affect 12 genes, 6 of them have biological functions that are: AGAP001002 (Toll protein) is involved in development and immunity; and AGAP001033

(mab-21 like protein) is involved in neural and sensory organ development; AGAP001050 (chondroitin polymerizing factor); and AGAP001052 (ubiquitin carboxyl-terminal hydrolase) are involved in protein secretion and proteolysis; AGAP001022 (gastrin/cholecystokinin receptor) is a receptor for peptides in the brain and gastrointestinal tract; and AGAP001025 (protein MSTA) is involved in negative regulation of gene expression. Moreover, some of these genes could directly involve in the mating process or indirectly through interaction with other mating genes that could be elsewhere in the genome. However, some of these genes might affect directly or indirectly genes responsible for ecological speciation of the *An. gambiae s.s.* and *An. coluzzii,* e.g. those are involved in distribution of the larvae in different habitats (Kamdem et al. 2012; Gimonneau et al. 2012) and the response of the larvae to the predators presence (Diabaté et al. 2008; Gimonneau et al. 2012) .

In addition, the non-synonymous coding differences between the recombinants strains were found only within introgressed pericentromeric X-island region, whereas non-coding differences were found in other regions of the genome. These differences could be due to the differences between the original Kisumu and Mopti strains. On the other hand, they could be a result of the fixation that occurs during the introgression process when we create the recombinants. Moreover, the non-coding differences could also be a result of Kisumu and Mopti laboratory strains assortative mating mechanisms and the resulting recombinant strains model design.

Interestingly, further re-sequencing studies could consider intergenic regions that might include elements with cis and trans effects on genes expression within the X, 2L and 3L islands. Trans-acting effects could explain the generally poor correspondence between differentially-expressed genes and islands of divergence observed in some *An. coluzzii* and *An. gambiae s.s.* populations (Cassone et al. 2008; Cassone et al. 2014). Genomic

comparisons studies between the sibling species and finding the fixed SNPs differences that code different proteins, and revealing the proteins functions are fundamental steps for understanding the association between the genomic differentiation in specific islands with the sympatric speciation process.

CHAPTER FIVE

Analysing the stages of mating behaviour and sperm quantification in *An. coluzzii* and *An. gambiae s.s.*

5.1. Introduction

An. coluzzii and *An. gambiae s.s.* were both previously known as forms of *An. gambiae.* They are considered as two of the most extensively studied mosquitoes due to their importance in transmitting malaria parasites and their genetic complexity (WHO 1999; Howell & Knols 2009). Understanding the biology of these mosquitoes is an important point for successfully finding alternative controlling malaria methods (Benedict 2003). It could also be important to develop control programmes that depend on releasing genetically-modified or sterile laboratory males in the field, and to check their mating compatibility and ability to compete with the field ones (Benedict 2003). Furthermore, the forms of *An. gambiae* are believed to be undergoing speciation and different researchers try to evaluate speciation models based on its genetic structure (Lehmann & Diabate 2008).

An. gambiae mosquitoes distribute and survive in very different habitats (Touré et al. 1998; Lee et al. 2009). Therefore, researchers were interested to know how these mosquitoes are able to exploit different environments (Touré et al. 1998). In this regard, it was believed that *An. gambiae* mosquitoes did not consist of a single taxonomic unit (Touré et al. 1998). Before the development of PCR, cytogenetic was used to describe five paracentric inversions on the right arm and one inversion on the left arm of the second chromosome (della Torre et al. 2002; della Torre et al. 2005). Based on these inversions and analyses of their frequencies, they identified and named five chromosomal forms

Mopti, Savanna, Bamako, Forest, and Bissau in reference to the places they were collected from (della Torre et al. 2002; della Torre et al. 2005). Due to the different ecological and seasonal distribution for these forms and the low number of hybrids found in the field, Toure et al. (1998) believed that these inversions could lead to ecological speciation in *An. gambiae* (Touré et al. 1998; Manoukis et al. 2009). However, laboratory studies showed that there was no unviability or infertility of the hybrid progeny resulting from crosses between the different forms (Touré et al. 1998). When PCR-based taxonomical tools became more common, attempts to distinguish the chromosomal forms based on the ribosomal DNA locus near the centromere of X chromosome led to the description of the M and S molecular forms of *An. gambiae* (Favia et al. 1997). Based on these analyses, they genotyped the chromosomal forms in different locations and found that Mopti chromosomal form was characterized as M molecular form, Bamako and Bissau chromosomal forms were S molecular form, whereas Savanna and Forest forms can be either S or M molecular forms (della Torre et al. 2005). The M and S molecular forms were recently elevated to the sibling species level due to a variety of ecological and genomic considerations (see below) and called *An. coluzzii* for the M molecular form and *An. gambiae s.s.* now limited to the S molecular form (Coetzee et al. 2013) .

Although the sympatric *An. coluzzii* and *An. gambiae s.s.* incipient species overlap in different areas of West Africa, a genetic analysis of transferred sperm in mated females showed that the two cryptic taxa were separated by strong assortative mating (Tripet et al. 2001). In terms of the mechanisms of assortative mating, it was believed that it could be linked to recognition processes as part of mate choice (Diabaté et al. 2006) . The mating process in *An. coluzzii and An. gambiae s.s.* starts when dusk comes (Howell & Knols 2009; Sawadogo et al. 2013). At this time the males' antennal fibrillae erect and after about one hour they leave their resting areas (Charlwood & Jones 1979). Swarms are initially

started by one or two males who move with zigzag-like flight patterns over an open area with minimum angle of view of the sky (Diabate et al. 2003; Dao et al. 2008). Then more males join them to make a loose cloud, which becomes more dense, and starts moving as a single unit (Reisen & Aslamkhan 1975; Howell & Knols 2009). The exact swarming starting time is determined by the relationship between swarming stimulation, circadian rhythm and light level, and it was proposed that it could play a role as pre-mating barrier between the different species (Howell & Knols 2009). In this regard, field studies found that *An. gambiae s.s.* starts swarming a few minutes earlier than *An. coluzzii* but because swarming time largely overlap, this small difference does not constitute a strong reproductive barrier (Sawadogo et al. 2013). Later on, the females join the swarms (Reisen & Aslamkhan 1975; Diabate& Baldet 2003). Usually, males prefer to fly in the centre of the swarm in a way that possibly increases their chances to mate (Manoukis et al. 2009; Howell & Knols 2009). It is thought that during this time, the wing beat frequencies of females attract the males, and the individuals try to harmonize their wing beats with other nearby mosquitoes to check if they are the same or opposite sex (Gibson & Russell 2006). However, these mechanisms could include recognition of flight tones specific to the two sibling species, possibly due to morphological differences in wings shape (Pennetier et al. 2010; Sanford et al. 2011). The courtship continues when the male approaches the female and grips one of her legs with his tarsal claws (Clements 1992), after establishing formal grip with female legs then he swings himself under her, bends his abdomen and his genitalia to interlock them with female's ones (Clements 1992; Howell & Knols 2009). Next, the male starts immediately releases the female's legs to take an tandem position and they fall on the ground or leave the swarm in copula (Clements 1992; Howell & Knols 2009). After the male finishes insemination, he opens his claspers and shakes himself free and leaves her to re-join the swarm again to seek another female for mating (Clements

1992; Howell & Knols 2009). During the mating process the male passes his accessory gland secretions (MAG) that include combinations of proteins and steroid hormone 20-hydroxyecdysone (20E). Usually, the MAG peptides make post-mating physiological and behavioural changes in the females (Gabrieli 2014). The main effects include preventing further mating in these females by forming gelatinous mating plug that can be digested in 1-2 days, and triggering egg-laying after blood feeding (Gabrieli 2014). Therefore, the females do not mate again at the same night. After about 20min of swarming, the number of couples decreases in the swarm, and the males leave to rest and restore energy via nectar feeding for flight and daily survival (Charlwood & Jones 1979; Howell & Knols 2009). The females, on the other hand, are known to start seeking a blood meal which enables them to produce eggs (Cook & Zumla 2003).

An. gambiae species are believed to determine the swarm place differently. Usually, *An. coluzzii* individuals are attracted to the markers with an angle of sky view, whereas *An. gambiae s.s.* swarms over ground regardless to the markers presence (Alexander 1975; Gibson 1983; Gibson 1986; Diabaté et al. 2011). The swarms themselves can be found mixed at various frequencies (Diabaté et al. 2006; Dabire et al. 2013). Relatively high frequencies of mixed swarms were found in the villages of Soumousso (Savannah habitat) and Vallée du Kou (VK7) in Burkina Faso, West Africa (Dabire et al. 2013). However, only 4 out of 33 collected copulae were mixed-form pairs (Dabire et al. 2013). This suggested that assortative mating behaviour occurs even within mixed-form swarms but that segregation is not perfect because some hybrids are still found (Diabaté et al. 2006).

There were attempts to find the molecular type of the sperm in the mated females in mixed swarms. In this regard, Tripet and his colleagues in 2001 suggested that in some cases, if females mated with males of another form, they might re-mate with their own type males (Tripet et al. 2001; Lanzaro & Tripet 2003; Tripet et al. 2005). This is because when

interspecific sperm was in females' spermatheca, conspecific sperm was also found. Therefore, males MAG proteins are likely to not induce re-mating in reciprocal crosses in sibling forms in *An. gambiae s.s.* and *An. coluzzii* (Tripet et al. 2005). On the other hand, the results of Tripet study in 2001 could suggest that when a male mates with a wrong form female, the amount of sperm could be less than when he mates with his own type female, thus explaining the female re-mating behaviour. However, neither Tripet et al (2005) nor Shutt et al (2010) laboratory experiments found that cross-inseminations of MAG extracts between *An. gambiae s.s.* and *An. coluzzii*, were less likely to caused monogamy (Tripet et al. 2005; Shutt et al. 2010).

As discussed above, studies have shown that pre-mating barriers play the main role in reproductive isolation between the two sibling species and that intrinsic post-mating barriers are absent (Diabaté et al. 2007). Currently, the idea that extrinsic barriers to reproduction occur in the form of fitness costs to hybrids is supported by observations of divergent ecological adaptation between *An. coluzzii* and *An. gambiae s.s.* . Several studies have reported notable differences between the sibling species. For example, the larval stage of the two species are adapted to different types of water (Diabaté et al. 2007). Field transplant experiments using first instar larvae showed that *An. coluzzii* larvae are able to survive better than those of *An. gambiae s.s.* in habitats with aquatic predators (Diabaté et al. 2007; Gimonneau et al. 2012; Roux et al. 2013). In addition, laboratory studies reported that the larvae of the *An. coluzzii* developed slower than those of *An. gambiae s.s.* (Mouline et al. 2012). Moreover, newly emerged *An. coluzzii* females were heavier and had higher proteins and lipids reserves than *An. gambiae s.s.* adult females of similar body size (Mouline et al. 2012). As a consequence, *An. coluzzii* had higher ability to develop and mature eggs after a single blood meal than *An. gambiae s.s.* (Mouline et al. 2012). Furthermore, *An. coluzzii* is thought to remain in Sahelian regions to undergo aestivation in

the dry season, which would help its populations to build-up faster when the next rainy season starts. In contrast, *An. gambiae s.s.* is thought to possibly migrate away from Sahelian areas during the same dry seasons (Dao et al 2014).

These ecological differences at the adult and larval stages between the two species are believed to drive the speciation process in *An. gambiae* (Touré et al. 1998). However, for sympatric speciation with on-going gene flow between sibling species to happen, the genomic structure of these species could also play an important role (Touré et al. 1998). In this regard, it was first thought that the chromosomal forms were undergoing ecotypic speciation (Touré et al. 1998; Manoukis et al. 2009), and this process was believed to be driven by chromosomal inversions that are distributed non-randomly in the chromosomal forms and contribute to their adaptation to different ecological zones (Touré et al. 1998). Microsatellite loci studies were conducted by Tripet et al (2005) to compare patterns of genetic divergence between the Bamako and Mopti forms within the *j* inversion on chromosome 2, and microsatellites on chromosome 3 where no inversions exist. The results showed high degree of genetic differentiation between forms for loci within the *j* inversion comparing with the chromosome 3 loci. This seemed to confirm the role of inversions in protecting gene complexes from recombination (Tripet et al. 2005). On the other hand, other studies that used microsatellite distributed throughout genome to compare different populations in Africa, suggested that the isolation between these two species occurs because of the reduced recombination in small regions of the genome, whereas gene flow occurs in other parts of the genome leading to a so-called 'mosaic genome structure' (Slotman et al. 2006). The regions of low recombination are believed to include clusters of isolation genes that cause reproductive isolation between species (Slotman et al. 2006). Later comparative genome-wide studies revealed that *An. gambiae s.s.* and *An. coluzzii* differed genetically at 3 small pericentromeric regions on

chromosomes 2L, 3L and X (Turner et al. 2005; Turner & Hahn 2007; White et al. 2010). These low recombination genomic regions are named ‘islands of speciation’ because of their hypothesized importance in protecting genes of assortative mating and ecological adaptations that are fundamental to the sympatric speciation process.

In the present study, we created recombinant strains by introgressing the island of divergence located in the pericentric region of the X chromosome of S form (Kisumu from Kenya) into the M molecular form (Mopti from Mali) through 5 generations of backcrossing followed by two generations of crosses to fix the X island (Chapter 2). At the end of that process, we obtained two recombinant strains with either M or S molecular form island of speciation on chromosome X. We found in a series of experiments that recombinant females mated with matching island-type males (Chapter 3). This confirmed that assortative mating genes are located in the X-island of divergence. Thereafter, we sequenced the genetic code of the recombinant strains and field populations, to identify putative assortative mating genes (Chapter 4). These results revealed the importance of speciation islands in X chromosome for the genetic and behavioural isolation in sympatric speciation process.

Although the exact function of candidate genes is not known yet, they could encode behavioural traits involved in conspecific recognition. In this regard, the main aim of the present study was to investigate the behavioural differences in swarming and copulation between the recombinant strains. For this purpose, the swarming activity was recorded overnight using infrared video recording under standard laboratory conditions. Additionally, the quantity of sperm transferred in the spermatheca of the mated females was measured by using QPCR. These observations and measurements were made in mating experiments involving males and females with matching X-island type and experiments where they differed at the pericentromeric islands.

These comparisons constitute the first step for understanding the role of X-island of speciation in assortative mating and can provides crucial information about mechanisms involved in mating recognition.

5.2. Materials and methods

5.2.1. Sexing the pupae

After the recombinant strains AMM and BSS were blood fed and the eggs hatched, the larvae were trayed out in 12 trays for each group and were bred in the insectary under the standard regime (Chapter 2). The pupae were sexed using standard method (Chapter 2). After that, they were transferred to separate emergence cages for each day to let them emerge.

5.2.2. Set the experiment cages

The 4-days aged adults were placed in the experiment cages that have small windows on the sides. The window was covered with wight papers to dissipate the light (Table 5.1).

Table 5.1. The mating type, number of replicates and number of individuals in every combination - The combinations were placed in the experiment cages that were set under the video cameras.

Mating Type	Number of replicates	Females (50 individuals)	Males (100 individuals)
Matched	1+2+3	BSS	BSS
island	1+2+3	AMM	AMM
Mating Type	**Number of replicates**	**Females (30 individuals)**	**Males (60 individuals)**
Not-matched	1+2+3	AMM	BSS
island	1+2+3	BSS	AMM

The cages were placed on stands. The infrared red lights (48LED IR Illuminator lamp, Amazon) were used for recording the videos because they are invisible for the mosquitoes

to prevent any effect on their behaviour. These lights were situated underneath the cages. The cages themselves were covered on the top with colourless invisible plastic sheets that allow the light goes through. We provided small bottles of sugar solution 5% with a filter papers to keep the mosquitoes alive overnight. The video cameras (Handycam HDR-SR12E) were fixed on the holders that fixed on the tables nearby the cages to let the cameras' lens face the bottom of the cage (Fig 5.1). The females were released in the cages firstly. The males were collected with the mosquitos' vacuum and kept aside until the light of the insectary was switched off. Thereafter, the recording was started and the males were released at the same time into the cages to avoid the differences in starting time of the swarming activity. The videos were recorded overnight (6pm-9am). The records were downloaded to the computer, then next day the next record was started for the next replicate. The recording was carried out for one night to prevent the females from further mating in the next day. However, This experiment was set up in the standard conditions in the lab to prevent the effect of humidity, temperature and light intensity on the swarming time and activity.

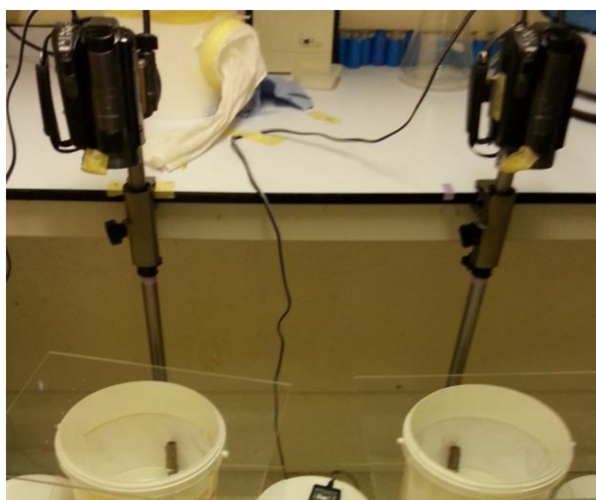


Figure 5.1. The experiment cages on the stands - The cages were placed on stands and the lights were placed underneath the cages, the cameras situated on the top with fixed distance from the cage.

5.2.3. Videos recording analyses

All the videos were converted to IMovie or IDVD extension by using Wondershare Video Converter Ultimate software. Thereafter, the videos were analysed using IMovie software. For this purpose, every video clip was imported to IMovie and played at 50% slow motion. The timer was used to register the time for every stage of mating for every couple of mosquitoes. In addition, the number of the mosquitoes in top wall, down wall and bottom was registered every 10min.

As was mentioned previously, the mating process in *An. gambiae* consists of stages that start when the male approaches the female and grips her tarsal claws on one of her legs with his forelegs. We called this stage in our study leg grips. Next stage begins when he swings himself under her, bends his abdomen and his genitalia to interlock them with female's ones (Clements 1992). Then the male starts immediately releasing the female legs to take a tandem position and they fall on the ground or leave the swarm in copula (Clements 1992; Howell & Knols 2009). After the male finishes insemination, he opens his

claspers and shakes himself free and leaves her (Clements 1992; Howell & Knols 2009) and this is the end of the tandem stage. The variables were measured and compared among the mating combinations and between the mating types:

1. Initiation of leg grips is the period from beginning of swarming to griping tarsal claws in mating process.
2. Initiation of tandem is the period from beginning of swarming to the beginning of tandem position.
3. Duration of leg grips is the period from griping tarsal claws to the beginning of tandem.
4. Duration of tandem is the period from beginning the tandem to the releasing (finishing mating).
5. Proportion of swarming individuals is the number of swarming individuals to the total number of mosquitoes.
6. Proportion of resting positions is the number of non-swarming individuals to the total number of mosquitoes.
7. Proportions of resting individuals in different positions are the numbers of resting individuals in upper wall, lower wall and the bottom of the cage to the total number of mosquitoes.
8. Sperm quantification is estimating the amount of the sperm that male inseminate into female's spermatheca.

5.2.4. Wings measurements

The rest of the sexed pupae that were not used in the experiment cages were situated in emergence cages. Once they emerge and maturated, 40 males and 40 females random

samples from control group and experiment group were collected and placed in 70% alcohol and the wings were measured using standard method (Chapter 2).

5.2.5. Females dissection

After one night of mating, the females were collected from the experiment cages and placed in 70% alcohol for dissection to get the sperm bundle using the standard methods (Chapter 2). Thereafter, the DNA was extracted from the sperm bundle samples using ChargeSwitch kit standard methods (Invitrogen, Leicestershire, UK) (Chapter 2).

5.2.6. Creating gDNA standard curve for QPCR

The sperm that the males pass into the females from the same and different type was quantified and compared using the QPCR. Standard curve for the QPCR was created using known concentration of DNA sample through several steps:

A) Identifying the genome size of the *Anopheles gambiae*: The size of *An. gambiae* is 278,253,050bp= $0.278e^{9}$bp per haploid according to the vector base website (https://www.vectorbase.org/organisms/anopheles-gambiae).

B) Identifying the mass of DNA per genome: The mass of the genome was calculated using the equation

$m= (n)(1.096e^{-21} g/bp)$

Where: n= genome size bp, m= mass, $e^{-21} = \times 10^{-21}$

$m= (0.278e^{9} bp)(1.096e^{-21} g/bp)= (0.278e^{9})(1.096e^{-21} g)= 0.3e^{-12} g$

This was converted to pictogram units by multiplying to constant number:

$(0.3e^{-12} g)(1e^{12} pg/g) = 0.3pg.$

C) The mass of the genome was divided by the copy number of the gene of interest per haploid genome, but the sperm include a single copy gene, therefore we divided to (1):

0.3pg/genome ÷ 1copy of genome= (0.3pg/genome)(genome/1copy) = 0.3pg/ 1copy of genome

Therefor, 0.3pg of mosquito genome includes 1 copy of genome.

D) Calculating the mass of gDNA containing the copy of interest, that is 10,000 to 80 copies:

Copy of interest x mass of haploid genome= mass of required gDNA

Copy of interest	Mass of required DNA= Copy of interest x 0.3pg
10,000	3000
2000	600
400	120
80	24

E) Calculate the concentrations of gDNA that is needed to achieve the copies of interest by dividing the required mass by the volume to be pipetted into each reaction (2µl in this reaction).

Copy	Mass of DNA	Final Concentration of gDNA (pg/ul)= Mass of DNA /2ul
10,000	3,000	1,500
2,500	600	300
500	120	60
100	24	12

F) The DNA extractions from of 3 males' bodies were prepared using ChargeSwitch kit (Invitrogen, Leicestershire, UK), and then 3 extractions were mixed together. The whole procedure was done twice to get two pools of gDNA extractions.

G) The gDNA concentration was measured using Nanodrop for each pool 3 times and the averages were calculated for these pools.

Measurements	gDNA pool1	gDNA pool2
Measured 1	18.1 ng/µl	25 ng/µl
Measured 2	17.8 ng/µl	24.8 ng/µl
Measured 3	18.7 ng/µl	24.9 ng/µl
Average	18.2 ng/µl = 18,200 pg/µl	24.9 ng/µl = 24,900 pg/µl

H) Prepare a serial dilution of gDNA using the formula:

C1V1 = C2V2

C1: Concentration in 1µl of gDNA pool (was measured by Nanodrop)

C2: The final concentration.

V1: The volume we need to add from gDNA pool.

V2: The final volume for each dilution (100µl).

I) The required volume of each gDNA pools was calculated for the first dilution in the dilution serial tubes.

gDNA pool1:

C1= 18,200pg/µl

C2= 1,500pg/µl

V2= 100µl

V1= (C2xV2)/C1= 8.2µl (The volume of gDNA pools)

100-8.2= 91.8µl (The volume of T.E. dilution buffer)

First dilution tube (D1): 8.2µl (gDNA pool)+91.8µl (T.E. buffer)

gDNA pool 2:

C1= 24,900pg/µl

C2= 1,500pg/µl

V2= 100µl

V1= (C2xV2)/C1= 6µl (The volume of the gDNA pools)

100-6= 94µl (The volume of T.E. dilution buffer)

First dilution tube (D1): 6µl (gDNA pool)+94µl (T.E. buffer)

J) Six PCR tubes were labelled for each gDNA pool D2, …..D6, and 95µl T.E. buffer were added to these tubes.

K) We took 5µl from each tube and added it to the next tube and mixed it to make the serial dilution for each gDNA pool, and they were kept in -20°C.

5.2.7. Sperm quantification

The DNA was extracted from the samples using ChargeSwitch (Invitrogen, Leicestershire, UK) standard method (Chapter 2). The PCR reaction mix contains: 10µl of Taq Man master mix (Applied biosystems, California, USA), 1.5µl forward primer (Fd Y) (5'-CGT GCA ACA GCT CGT GAT G-3') (Eurofins Genomics, Wolverhampton, UK), 1.5µl Reverse primer (Rv Y) (Eurofins Genomics) (5'-TTA CCA CGC TGG CAA ATG C-3') (Eurofins Genomics, Wolverhampton, UK), 1.25µl Probe (5'-AGA TGG ATG CGG CGT-3') (Eurofins Genomics, Wolverhampton, UK) and 3.75µl dH2O to complete the volume to 18µl, then 2µl of the template was added to the mix to make the final volume 20µl for each reaction. Thereafter, the PCR plate was placed in the QPCR machine (Applied biosystems, California, USA) and was run using standard quantitative protocol. Once finished, the results were exported as Excel file then the CT values were transferred to JMP software (SAS Institute, Inc) for analyses.

5.2.8. Data analyses

All statistical analyses were done using the JMP 11.0 (SAS Institute, Inc) software. All data were also checked for normality by using Shapiro-Wilk test and heterogeneity of variances, and they were analysed accordingly using parametric or non-parametric procedures. The replicate effect was tested in all the analyses but only reported when it had significant effect. In addition, the non-significant interactions were removed from the multivariate models unless otherwise indicated. Pearson Chi-squares (likelihood ratios) and logistic regression was also used to analyse the frequencies.

5.3. Results

5.3.1. Phenotypic quality of individuals for mating experiment

Random samples of experiment mosquitoes were collected from each strain and the wings were removed, measured and analysed as a representative method to compare the body size of the strains (Renshaw et al. 1994; Mwangangi et al. 2004) to avoid the effect of mosquitos size on the mating process. There was no significant difference between the wing lengths for the males (T-test: n= 40, df= 1, T= -0.10, P= 0.92) (Fig 5.2). The means and the confidence intervals of the wing length were 3.04 (3.02-3.07CI) for AMM strain, 3.04 (3.02-3.06CI) for BSS strain. There was no significant difference between the wing lengths for the females (T-test: n= 40, df= 1, T= 1.68, P= 0.10) (Fig 5.2). The means and the confidence intervals of the wings length were 3.0 (2.98-3.02CI) for AMM strain, 3.03 (3.0-3.05CI) for BSS strain.

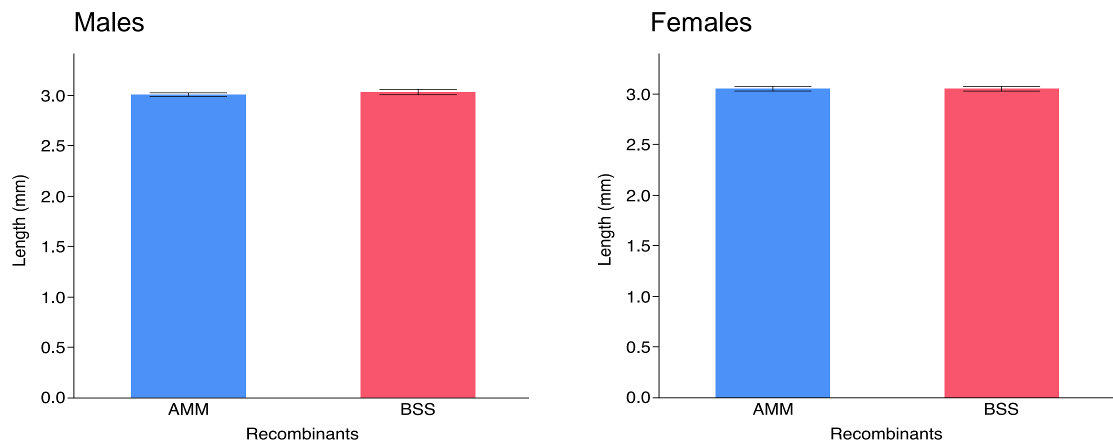


Figure 5.2. The wings' size of random sample from the mosquitoes of mating experiment for males and females - The bars showed the means and confidence intervals 95% values. (P >0.05) there is no significant difference between the groups body size.

5.3.2. Matched islands combinations and not-matched islands combinations analyses

5.3.2.1. Mating behaviour

The recombinant males (100 individuals) in this type of combination were placed with the matched speciation X-island females (50 individuals) in standard laboratory conditions at the same time. Additionally, 60 individuals of recombinants males were placed with 30 individuals of the opposite type of females in the laboratory standard conditions at the same time.

Leg grips and tandems frequencies

The frequencies of leg grips and tandems were tested every 10min in matched islands combinations AMM♀+AMM♂ and BSS♀+BSS♂ combinations. The frequency of leg grips and tandems dropped gradually in AMM♀+AMM♂ through time. The mean and confidence intervals were 72.7 (69.2-76.1CI) for a total number of 922 observations of leg grips and 53.2 (42.3-64.1CI) for a total number of 96 observations of tandems. The numbers of leg grips and tandems also decreased gradually through the time in BSS♀+BSS♂ combination. The mean and confidence intervals were 77.0 (74.0-80.0CI) for a total number of 1102 observations of leg grips and 58.2 (46.4-69.9CI) for a total number of 72 observations of tandems (Fig 5.3 & 5.4).

AMM♀+AMM♂

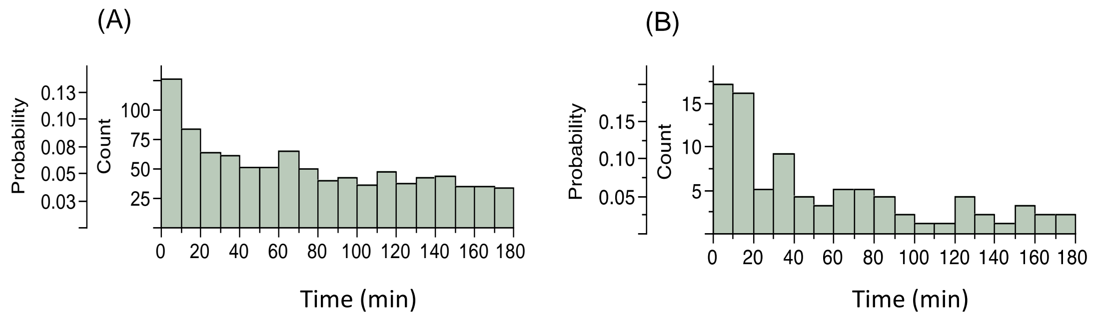


Figure 5.3. The frequency of leg grips and tandems in AMM♀+AMM♂ combination every 10min. (A) Frequency of leg grips (B) Frequency of tandems. The count is the sum of frequency of leg grips and tandems in the time intervals, and the probability is the frequency of leg grips and tandems to the total number of mosquitoes.

BSS♀+BSS♂

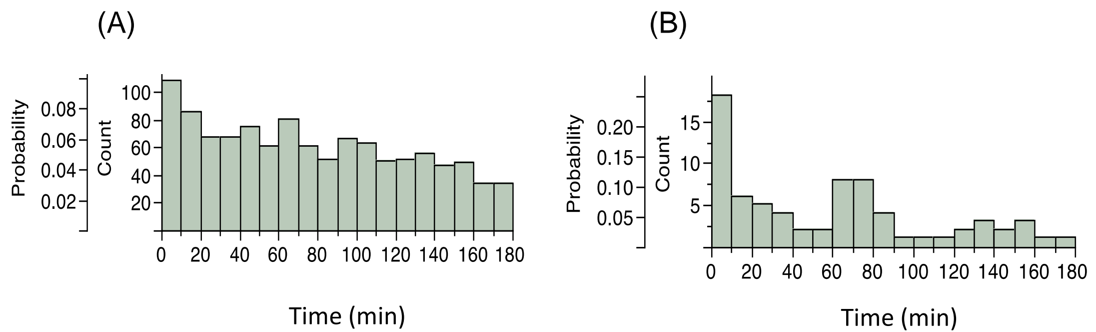


Figure 5.4. The frequency of leg grips and tandems in BSS♀+BSS♂ combination every 10min. (A) Frequency of leg grips (B) Frequency of tandems. The count is the sum of frequency of leg grips and tandems in the time intervals, and the probability is the frequency of leg grips and tandems to the total number of mosquitoes.

The frequency of leg grips and tandems was also checked every 10min in AMM♀+ BSS♂ and BSS♀+ AMM♂ combinations. The frequency of leg grips and tandems dropped gradually in AMM♀+ BSS♂ through time. The means and confidence intervals were 63.9 (60.8-67.1CI) for total number of 892 observations of leg grips and 30.0 (20.4-39.7CI) for total number of 45 observations of tandems. The numbers of leg grips and tandems also decreased gradually through the time in BSS♀+ AMM♂ combination. The means and confidence intervals were 72.5 (69.2-75.8CI) for total number of 972 observations of leg grips and 37.3 (24.2-50.4CI) for total number of 44 observations of tandem (Fig 5.5 & 5.6).

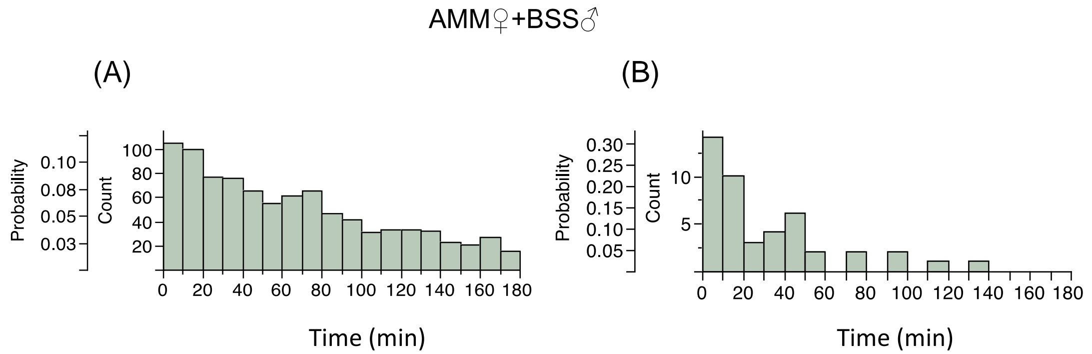


Figure 5.5. The frequency of leg grips and tandems in AMM♀+BSS♂ combination every 10min. (A) Frequency of leg grips (B) Frequency of tandem. The count is the sum of frequency of leg grips and tandems in the time intervals, and the probability is the frequency of leg grips and tandems to the total number of mosquitoes.

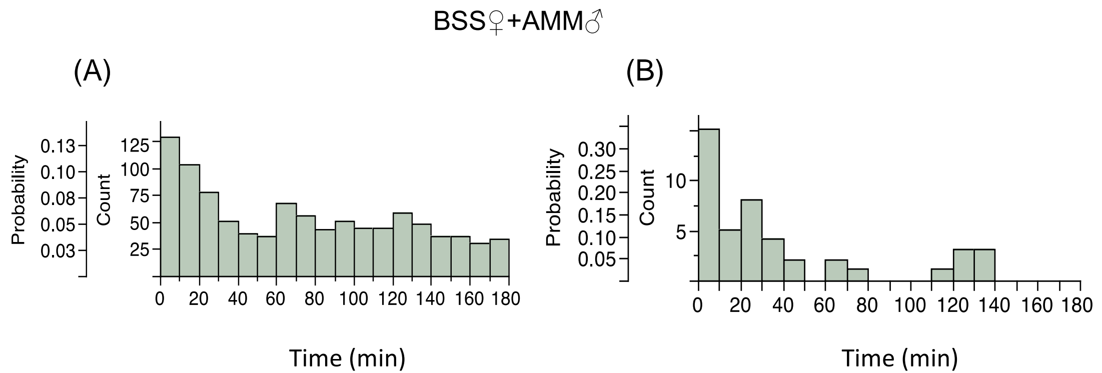


Figure 5.6. The frequency of leg grips and tandems in BSS♀+AMM♂ combination every 10 min. (A) Frequency of leg grips (B) Frequency of tandem. The count is the sum of frequency of leg grips and tandems in the time intervals, and the probability is the frequency of leg grips and tandems to the total number of mosquitoes.

Initiation time of leg grips

The effect of mating combination AMM♀+AMM♂, BSS♀+BSS♂ on the sqrt (square root) initiation time of leg grips was tested. There was a small significant difference in overall the replicates (Wilcoxon: n= 1888, df= 1, Z= 2.46, P= 0.033) (Fig 5.7). The initiation time of leg grips was also tested for every replicate. BSS started leg grips significantly earlier than AMM for replicate 1 (Wilcoxon: n= 514, df= 1, Z= 3.78, P< 0.001) and for replicate 2 (Wilcoxon: n= 636, df= 1, Z= 4.43, P< 0.001). In contrast, AMM started significantly before BSS in replicate 3 (Wilcoxon: n=738, df= 1, Z= 3.95, P< 0.001). The overall medians and (25-75%) quartiles of initiation time of leg grips in overall replicates were 28.8 (20.5-35.5) for BSS♀+BSS♂ and 29.6 (20.1-37.7) for AMM♀+AMM♂. However, the BSS♀+BSS♂ combination was found to have a mean initiation time of legs grips of only 1 sec longer than AMM based on 1888 observations over three hours, which is considered comparatively a very small difference.

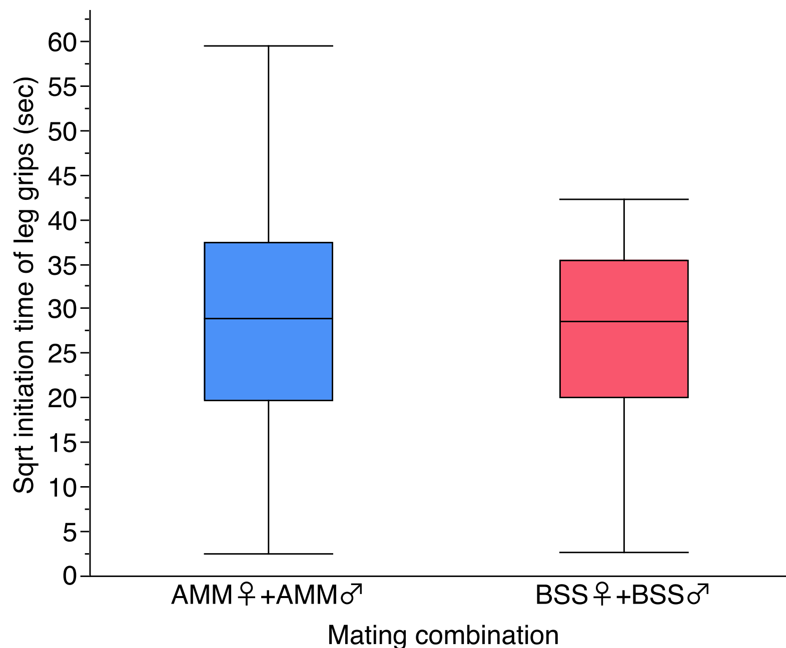


Figure 5.7. Comparison of the initiation square root (sqrt) time of leg grips between AMM♀+AMM♂ and BSS♀+BSS♂ mating combinations - Males and females of the two recombinant strains were placed in mating cages overnight (6pm-9am) and their swarming behaviour was recorded. Boxplots show the median and quartiles. (P< 0.001) there is a significant difference in the initiation time of leg grips.

In addition, we analysed the effect of mating combination BSS♀+AMM♂, AMM♀+BSS♂ on sqrt initiation time of leg grips. There was a significant difference between the combinations. BSS females started before AMM for overall replicates (Wilcoxon: n= 1979, df= 1, Z= 3.95, P< 0.001) (Fig 5.8). This significant difference due to the difference in replicate1 (Wilcoxon: n= 758, df= 1, Z= 6.45, P< 0.001), whereas there was no significant difference in replicate 2 (Wilcoxon: n= 559, df= 1, Z= 0.71, P= 0.475) and replicate 3 (Wilcoxon: n= 662, df= 1, Z= -1.45, P= 0.147). The overall medians and (25-75%) quartiles of initiation of leg grips in overall replicates were 30.1 (20.8-37.0) for AMM♀+BSS♂, and 27.6 (18.4-35.3) for BSS♀+AMM♂.

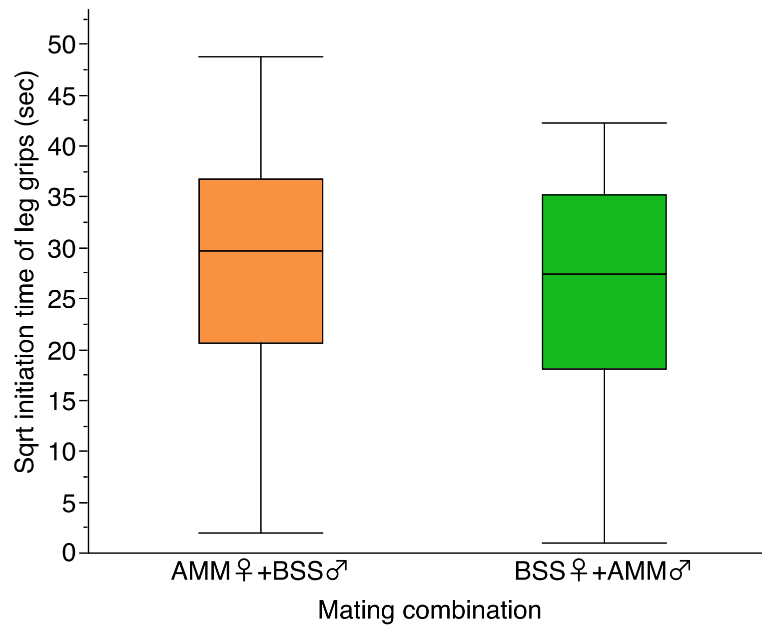


Figure 5.8. Comparison of initiation square root (Sqrt) time of leg grips between AMM♀+ BSS♂ and BSS♀+AMM♂ mating combinations - Males and females of the two recombinant strains were placed in mating cages overnight (6pm-9am) and their swarming behaviour was recorded. Boxplots show the median and quartiles. ($P< 0.001$) there is a significant difference only in replicate 1. ($P> 0.05$) in both replicate 2 and 3.

Initiation time and duration time of tandem

There was no significant difference when we analysed the effect of mating combinations on the stating time of tandem in overall the replicates (T-test: n= 159, df= 1, T= 0.27, P= 0.790). The stating time of tandem was also tested in every replicate. For replicate 1 (T-test: n= 40, df= 1, T= 0.64, P= 0.526). For replicate 2 (T-test: n= 46, df= 1, T= 1.41, P= 0.166). For replicate 3 (T-test: n= 73, df= 1, T= 1.32, P= 0.193). The mean time and confidence intervals of initiation of tandem for overall replicates were 25.2 (27.5-22.7CI) for AMM♀+AMM♂, and 25.6 (27.8-23.4CI) for BSS♀+BSS♂. Moreover, the duration time of tandem was compared between the matched islands combinations. There was no significant difference between the combinations for overall the replicates (T-test: n= 159, df= 1, T= -0.84, P=0.404) (Fig 5.9). The duration of tandem was also tested in each replicate. For replicate 1 (T-test: n= 40, df= 1, T= 0.30, P= 0.764), for replicate 2 (T-test: n= 46, df= 1, T= -1.16, P= 0.251), and for replicate 3 (T-test: n= 73, df= 1, T= -1.11, P= 0.269). The means and confidence interval values for the duration of tandem for overall replicates were 4.1 (3.9-4.3CI) for AMM♀+AMM♂, and 4.0 (3.8- 4.2CI) for BSS♀+BSS♂.

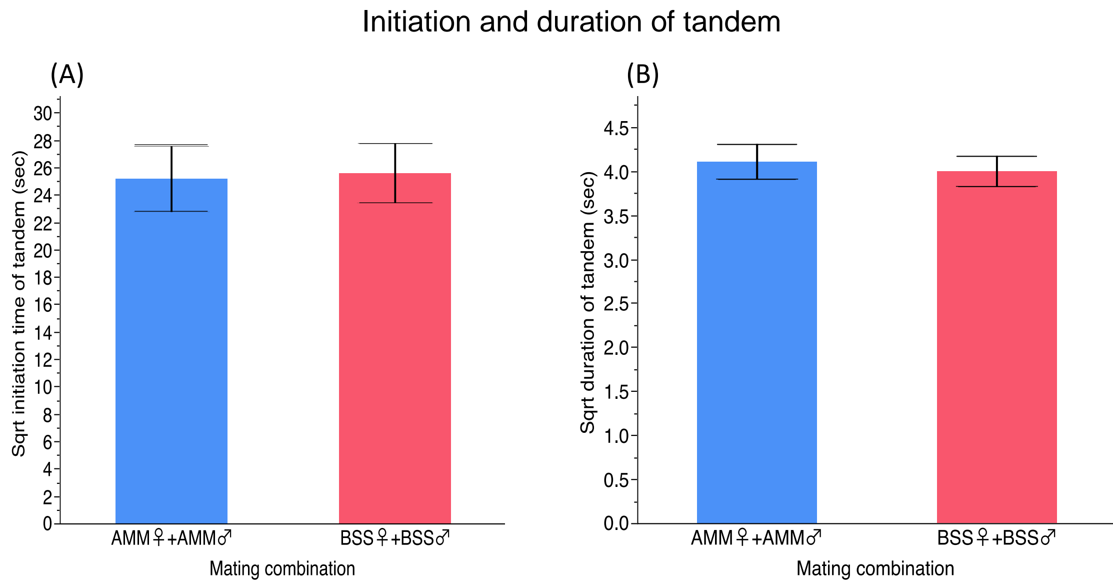


Figure 5.9. Comparison of the initiation and duration square root (sqrt) time of tandem position between AMM♀+AMM♂ and BSS♀+BSS♂ combinations - (A) Initiation of tandem, (B) Duration of tandem. The bars showed the means and confidence intervals 95%. Blue bars indicate to AMM♀+AMM♂ combination and the red bars indicate to BSS♀+BSS♂ combination. ($P > 0.05$) there is no significant difference in the initiation time and duration of tandem.

We found no significant difference when we analysed the effect of mating combination on sqrt start time of tandems for overall the replicates in not-matched combinations (T-test: $n = 90$, $df = 1$, $T = -0.89$, $P = 0.377$). The sqrt start time of tandem was also tested in different replicates. For replicate 1 (T-test: $n = 25$, $df = 1$, $T = -0.97$, $P = 0.337$), for replicate 2 (T-test: $n = 28$, $df = 1$, $T = 0.8$, $P = 0.432$), and for replicate 3 (T-test: $n = 27$, $df = 1$, $T = -1.18$, $P = 0.25$). The means of initiation of tandems for overall replicates were 25.0 (21.8-28.2CI) for AMM♀+BSS♂, and 23.0 (19.8-26.3CI) for BSS♀+AMM♂. Furthermore, there was no significant difference of sqrt duration of tandems in the overall replicates (T-test: $n = 90$, $df = 1$, $T = 0.51$, $P = 0.61$) (Fig. 5.10). The sqrt duration of tandems was tested in different replicates. For replicate 1 (T-test: $n = 25$, $df = 1$, $T = 0.57$, $P = 0.57$), for replicate 2 (T-test: $n = 28$, $df = 1$, $T = -0.51$, $P = 0.62$), and for replicate 3 (T-test: $n = 27$, $df = 1$, $T = 0.32$, $P =$

0.76). The mean values of sqrt duration for overall replicates were 3.6 (3.3-4.0CI) for AMM♀+BSS♂, 3.8 (3.5-4.1CI) for BSS♀+AMM♂.

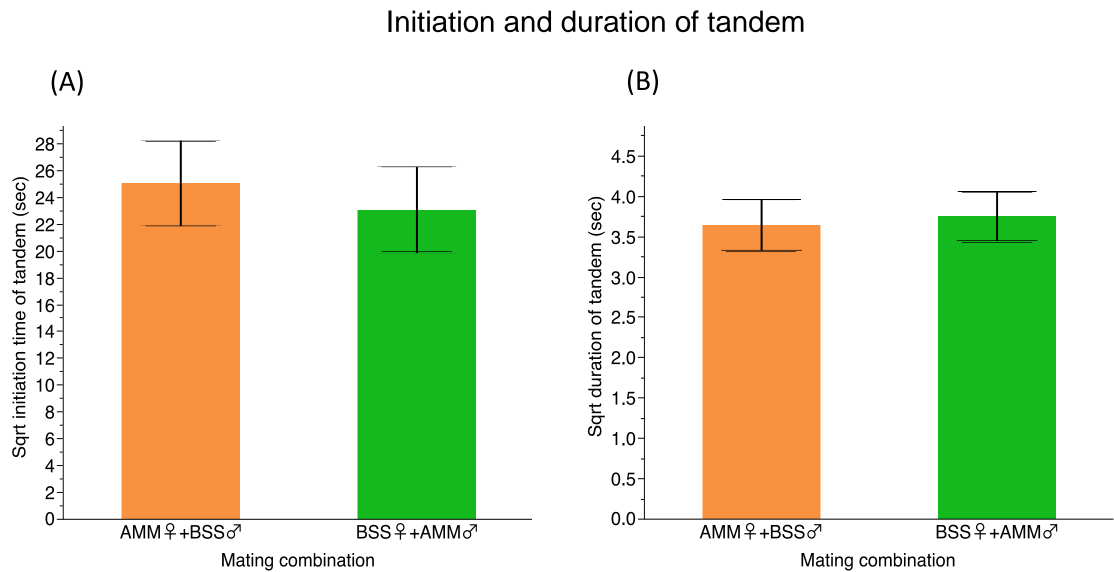


Figure 5.10. Comparison of the initiation and duration square root (sqrt) time of tandem between AMM♀+ BSS♂ and BSS♀+AMM♂ combinations - (A) Initiation of tandem, (B) Duration of tandem. The graph showed the means and confidence intervals 95% values. Orang bar indicates to AMM♀+ BSS♂ combination and the green bar indicates to BSS♀+AMM♂ combination. (P> 0.05) there is no significant difference in the initiation and duration time of tandem between the combinations.

Tandem to leg grips ratio

The proportions of tandems to leg grips were tested in both matched islands combinations. The results indicate a significant difference between them when we analysed the overall replicates (ChiSquare likelihood-ratio: n= 2047, df= 1, X^2= 4.32, P= 0.02). This small difference presented only in replicate 3 (ChiSquare likelihood-ratio: n= 811, df= 1, X^2= 5.86, P= 0.02), whereas no significant difference appeared in replicate 1 (ChiSquare likelihood-ratio: n= 554, df= 1, X^2= 0.002, P=0.962) and replicate 2 (ChiSquare likelihood-ratio: n= 682, df= 1, X^2= 0.99, P= 0.29).

The proportion of tandems to leg grips was analysed for total time of 180min in AMM♀+AMM♂ and BSS♀+BSS♂ every hour. There were significant differences between the two mating combinations. Moreover, the proportion of tandem reduced significantly through time in the first hour in the both combinations (ChiSquare likelihood-ratio: df= 2, X^2= 5.31, P= 0.021), whereas there was no significant difference in the second hour (ChiSquare likelihood-ratio: df= 2, X^2= 0.53, P= 0.46) (Table 5.2) (Fig 5.11).

Table 5.2. Nominal logistic regression effect of mating combination of the proportion of tandems to leg grips and time hours.

Source	***df***	***L-R* ChiSquare**	***P*-value**
Mating combination	1	3.99	0.045*
Hours	2	10.61	0.005*

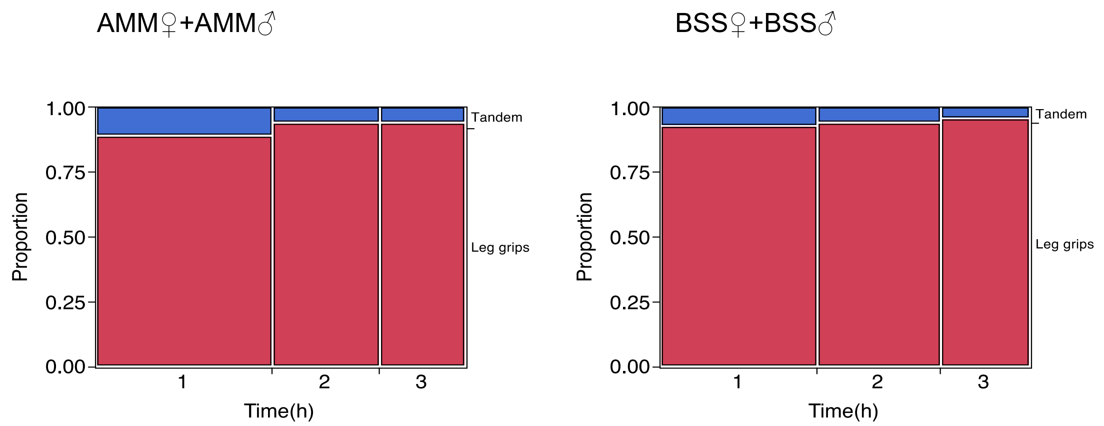


Figure 5.11. Change in proportion of tandem to leg grips in relation to time during 3 hours of the record for AMM♀+AMM♂ and BSS♀+BSS♂ combinations - The blue bars indicate to the tandem ratios and the red bars indicate to the leg grips ratios every hour. (*P*< 0.045) there is significant difference between the combinations.

The proportion of tandems to leg grips was analysed in both not-matched islands combinations. There was no significant difference between them in overall replicates (ChiSquare likelihood-ratio: n= 1890, df= 1, X^2= 21.08, P= 0.5472). The proportion was also tested in different replicates. For replicate 1 (ChiSquare likelihood-ratio: n= 723, df= 1, X^2= 3.42, P= 0.052), for replicate 2 (ChiSquare likelihood-ratio: n= 531, df= 1, X^2= 0.04, P= 0.99), and for replicate 3 (ChiSquare likelihood-ratio: n= 636, df= 1, X^2= 0.52, P= 0.33).

The proportion of tandems to leg grips was analysed for total time of 180min in AMM♀+ BSS♂ and BSS♀+AMM♂ every hour. There were no differences between the two mating combinations. However, the proportion of tandems reduced significantly through time in the first hour in both combinations (ChiSquare likelihood-ratio: df= 2, X^2= 21.08, P< 0.001), whereas there was no significant difference in the second hour (ChiSquare likelihood-ratio: df= 2, X^2= 0.09, P= 0.76) (Table 5.3) (Fig. 5.12).

Table 5.3. Nominal logistic regretion effect of mating combination of the proportion of tandem to leg grips and time hours.

Source	*df*	*L-R* ChiSquare	*P*-value
Mating combination	1	0.00054171	0.9814
Hours	2	41.0525211	<0001*

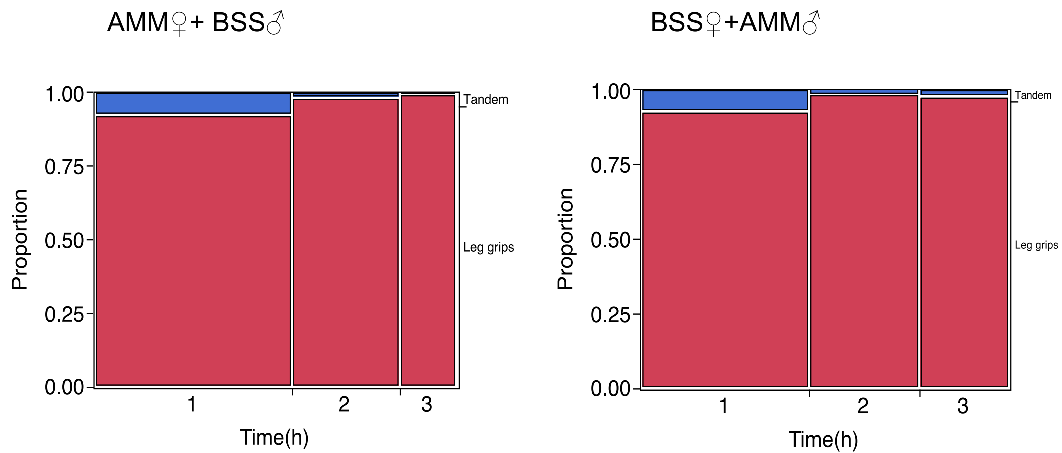


Figure 5.12. Change in proportion of tandems to leg grips in relation to time during 3 hours of the record - The blue bars indicate to the tandem ratios and the red bars indicate to the leg grips ratios every 30min.

5.3.2.2. Proportion of swarming and non-swarming individuals

The proportion of the active and resting individuals was analysed, in addition to the proportions of resting individuals in different locations of the cage were also analysed to check the differences between BSS and AMM recombinants in matched and not-matched islands combinations.

Proportions of swarming individuals

The proportions of the swarming and non-swarming individuals were analysed to check any differences in the activity when they mate with matched X-island of males/females. To limit pseudoreplication, the effect of mating combination, replicate and time on swarming and non-swarming numbers of individuals were analysed at four time points (10min, 60min, 120min, 180min). The analyses showed that the mating combination had no effect on the number of swarming individuals in general but it had effect between different replicates. However, the number of swarming individuals differed based on the replicate and time (Table 5.4). The means and the confidence intervals of swarming and non-swarming individuals in both combinations at the four time points were shown in (Table 5.5).

Table 5.4. The effect of mating combination, time and replicate on the number of swarming individuals in matched islands combinations at four time points (10min, 60min, 120min, 180min).

Source	***df***	***L-R* ChiSquare**	***P*-value**
Mating combination	1	0.01	0.925
Replicate	2	122.0	<0.001*
Time points	3	56.6	<0.001*
Mating combination*Replicate	2	11.0	0.004*
Replicate*Time points	6	112.9	<0.001*

Table 5.5. The means and confidence intervals of swarming and non-swarming individuals in matched islands combinations - brackets indicate the minimum and maximum values of confidence intervals for the three replicates in the different combinations at four time intervals (10min, 60min, 120min, 180min).

AMM♀+AMM♂				
Swarming				
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	2.78 (1.1-36.9)	26 (19.6-33.7)	34.9 (27.2-43.4)	20.8 (17.2-24.9)
60min	19 (13.6-25.8)	14.6 (09.7-21.3)	27.7 (20.5-36.4)	20 (16.4-24.0)
120min	21 (15.3-28.2)	6.4 (3.5-11.4)	7.5 (4.1-13.3)	11.8 (9.1-15.1)
180min	9.5 (5.5-15.9)	4.8 (2.2-10.1)	5.5 (2.5-11.4)	6.6 (4.5-9.7)
Non-swarming				
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	97.2 (93.1-98.9)	74 (66.3-80.4)	65.1 (56.6-72.8)	79.2 (75.1-82.8)
60min	81 (74.2-86.4)	85.4 (78.7-90.3)	72.3 (63.6-79.5)	80.0 (76.0-83.6)
120min	78.9 (71.8-84.7)	93.6 (88.6-96.5)	92.5 (86.7-95.9)	88.2 (84.9-90.9)
180min	90.5 (84.1-94.5)	95.2 (89.9-97.8)	94.5 (88.6-97.5)	93.4 (90.3-95.5)
BSS♀+BSS♂				
Swarming				
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	2.2 (0.7-6.2)	21.7 (15.7-29.3)	37.9 (30.1-46.4)	20.3 (16.7-24.4)
60min	13.8 (9.1-20.3)	17.2 (11.4-25.1)	28.1 (20.9-36.7)	19.4 (15.7-23.6)
120min	10.6 (6.5-16.7)	23.9 (17.7-31.5)	8.2 (4.5-14.4)	14.5 (11.4-18.3)
180min	7.9 (4.4-14)	5.3 (2.3-11.9)	7.3 (4-12.9)	7.0 (4.8-10.1)
Non-swarming				
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	97.8 (93.8-99.3)	78.3 (70.7-84.3)	62.1 (53.6-69.9)	79.7 (75.5-83.3)
60min	86.2 (79.7-90.9)	82.8 (74.9-88.6)	71.9 (63.3-79.1)	80.6 (76.4-84.3)
120min	89.4 (83.3-93.5)	76.1 (68.4-82.3)	91.8 (85.6-95.5)	85.5 (81.7-88.6)
180min	92.1 (86-95.6)	94.7 (88-97.7)	92.7 (87.1-96)	93.0 (89.9-95.2)

In this regard, we found that the proportion of swarming individuals reduced through time, so the individuals left the swarm to rest in different parts of the cage. Therefore the resting individuals proportion increased through time (Fig 5.13).

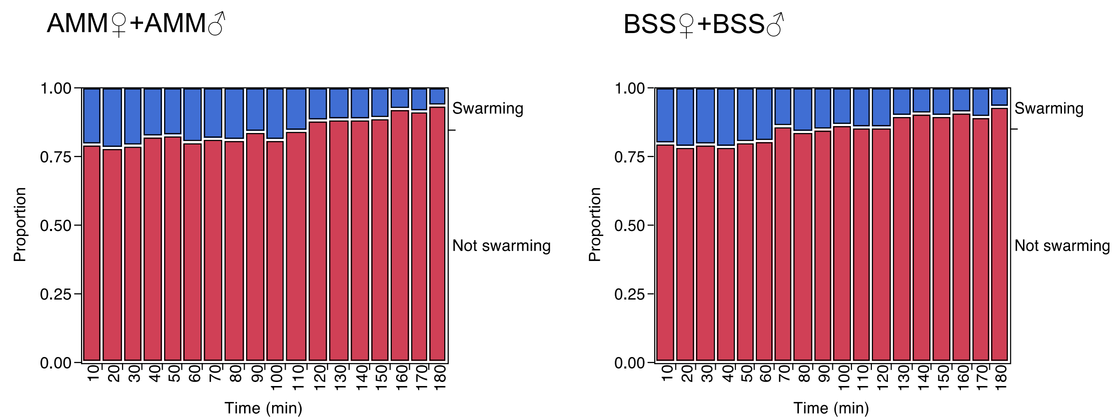


Figure 5.13. Change in proportion of swarming in relation to time during 3 hours of the record for AMM♀+AMM♂ and BSS♀+BSS♂ combinations - The blue bars showed the swarming ratios and the red bars showed the non-swarming ratios every 10min. ($P > 0.05$) there is no significant difference between the combinations.

There was no difference between the means of swarming and non-swarming numbers in AMM♂+AMM♀ and BSS♀+BSS♂ combinations when we compare them in total time of three hours for each replicate and in the mean of overall replicates (as discussed previously) (Fig 5.14).

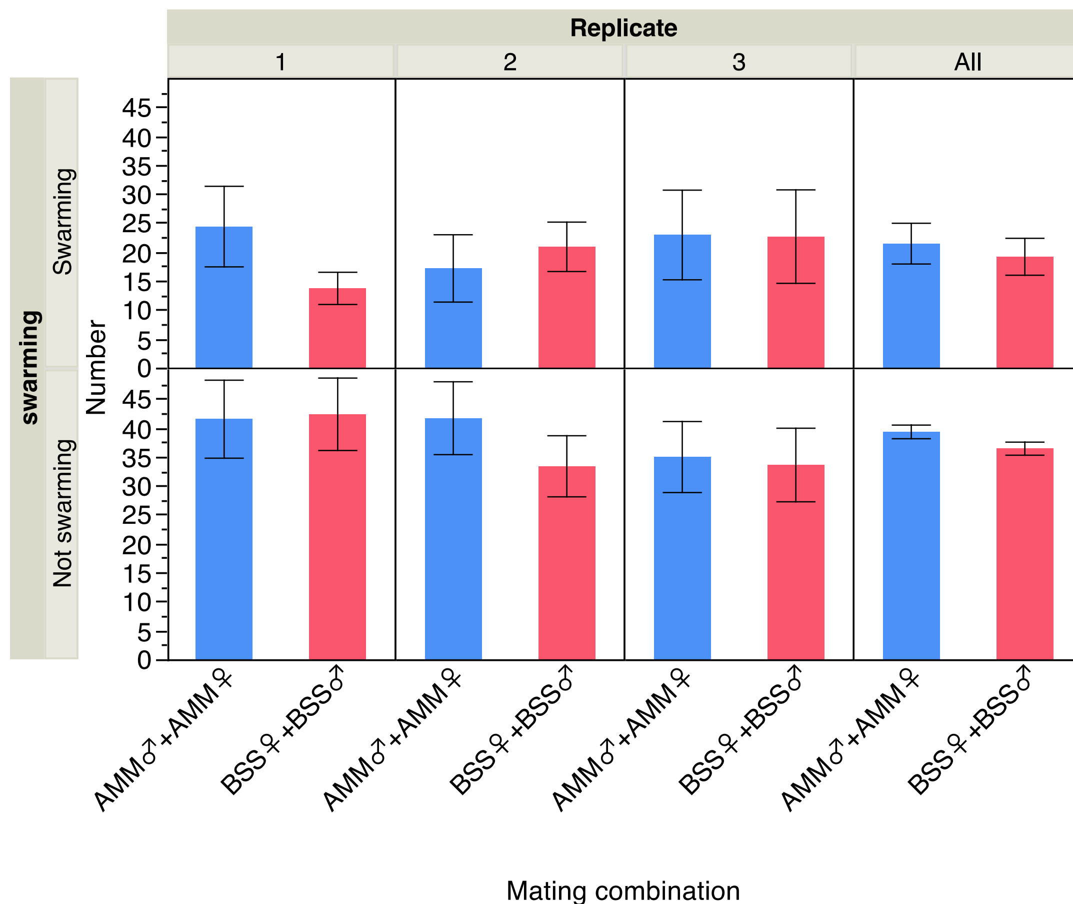


Figure 5.14. The means of swarming and non-swarming numbers for each replicate and overall the replicates for AMM♂+AMM♀ and BSS♀+BSS♂ combinations in different replicates. The bars showed the mean and confidence intervals 95% values. The blue bars showed AMM♀+AMM♂ combination and the red bars showed BSS♀+BSS♂ combination.

The proportions of active and non-swarming individuals, in addition to the resting individuals in different locations of the cage were analysed to check the differences between BSS and AMM recombinants in not-matched X-islands combinations. To limit pseudoreplication, the effect of mating combination, replicate and time on swarming and non-swarming numbers was analysed at four time points (10min, 60min, 120min, 180min). We found that the mating combination and replicate have no effect on the number of swarming individuals, whereas the numbers differed through time (Table 5.6). Means and

confidence intervals of the swarming and non-swarming individuals in the different combinations were shown (Table 5.7).

Table 5.6. The effect of mating combination, time and replicate on the number of swarming individuals in non-matched islands combinations at four time points (10min, 60min, 120min, 180min).

Source	***df***	***L-R* ChiSquare**	***P*-value**
Mating combination	1	1.05	0.305
Replicate	2	0.26	0.879
Time points	3	173.6	<0.001*

Table 5.7. The means and confidence intervals of the swarming and non-swarming individuals in non-matched islands combinations - brackets indicate the minimum and maximum values of confidence intervals for the three replicates in the different combinations at four time intervals (10min, 60min, 120min, 180min).

AMM♀+BSS♂				
Swarming				
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	25.6 (17.3-36.3)	39.3 (29.8-49.7)	38.5 (29.1-48.7)	34.9 (29.3-40.9)
60min	28.7 (20.3-39)	21.1 (13.2-32)	24.8 (17.4-34)	25.1 (20.2-30.7)
120min	12.8 (7.5-21)	8.2 (3.6-17.8)	9.6 (4.7-18.5)	10.5 (7.2-15.2)
180min	6 (2.6-13.3)	10.8 (05.6-19.9)	10.4 (5.4-19.2)	9.0 (5.9-13.3)
Non swarming				
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	74.4 (63.7-82.7)	60.7 (50.3-70.2)	61.5 (51.3-70.9)	65.1 (59.1-70.7)
60min	71.3 (61-79.7)	78.9 (68-86.8)	75.2 (66-82.6)	74.9 (69.3-79.8)
120min	87.2 (79-92.5)	91.8 (82.2-96.4)	90.4 (81.5-95.3)	89.5 (84.8-92.8)
180min	94 (86.7-97.4)	89.2 (80.1-94.4)	89.6 (80.8-94.6)	91.0 (86.7-94.1)
BSS♀+AMM♂				
Swarming				
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	43.2 (33-54.1)	37.7 (29.1-47.2)	41.6 (32.5-51.3)	40.6 (35.1-46.4)
60min	30.1 (21.3-40.7)	23.8 (16.7-32.8)	24.7 (16.8-34.8)	26.0 (21.2-31.5)
120min	10.1 (5-19.5)	15.7(9.4-25)	9.2 (04.7-17.1)	11.7 (8.2-16.4)
180min	3.4 (1.2-09.4)	11.1 (6.1-19.3)	9.9 (4.9-19)	8.0 (5.3-12.0)
Non swarming				
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	56.8 (45.9-67)	62.3 (52.8-70.9)	58.4 (48.7-67.5)	59.4 (53.6-64.9)
60min	69.9 (59.3-78.7)	76.2 (67.2-83.3)	75.3 (65.2-83.2)	74.0 (68.5-78.8)
120min	89.9 (80.5-95)	84.3(75-90.6)	90.8 (82.9-95.3)	88.3 (83.6-91.8)
180min	96.6 (90.6-98.8)	88.9 (80.7-93.9)	90.1 (81-95.1)	92.0 (88.0-94.8)

However, we found that the proportion of swarming individuals reduced through time, so the individuals left the swarm to rest in different parts of the cage. Therefore the proportion of resting individuals increased through time (Fig. 5.15)

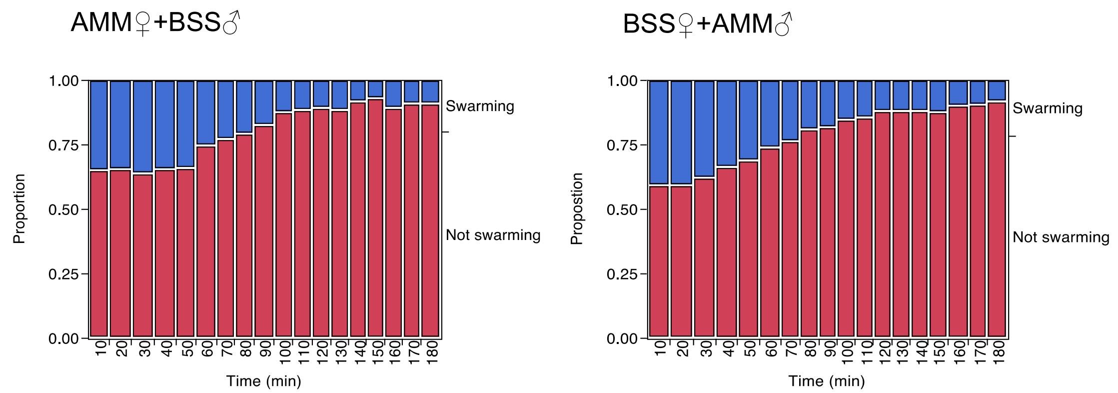


Figure 5.15. Change in proportion of swarming in relation to time during 3 hours in the cages for AMM♀+BSS♂ and BSS♀+AMM♂ combinations - The blue colour bars indicate to the swarming ratios and the red colour bars indicate to the non-swarming ratios every 10min. (P> 0.05) there is no significant difference between the combinations.

There was no significant difference in number of swarming and non-swarming individuals for AMM♂+ BSS♂ and BSS♀+AMM♀ combinations in different replicates (Fig. 5.16).

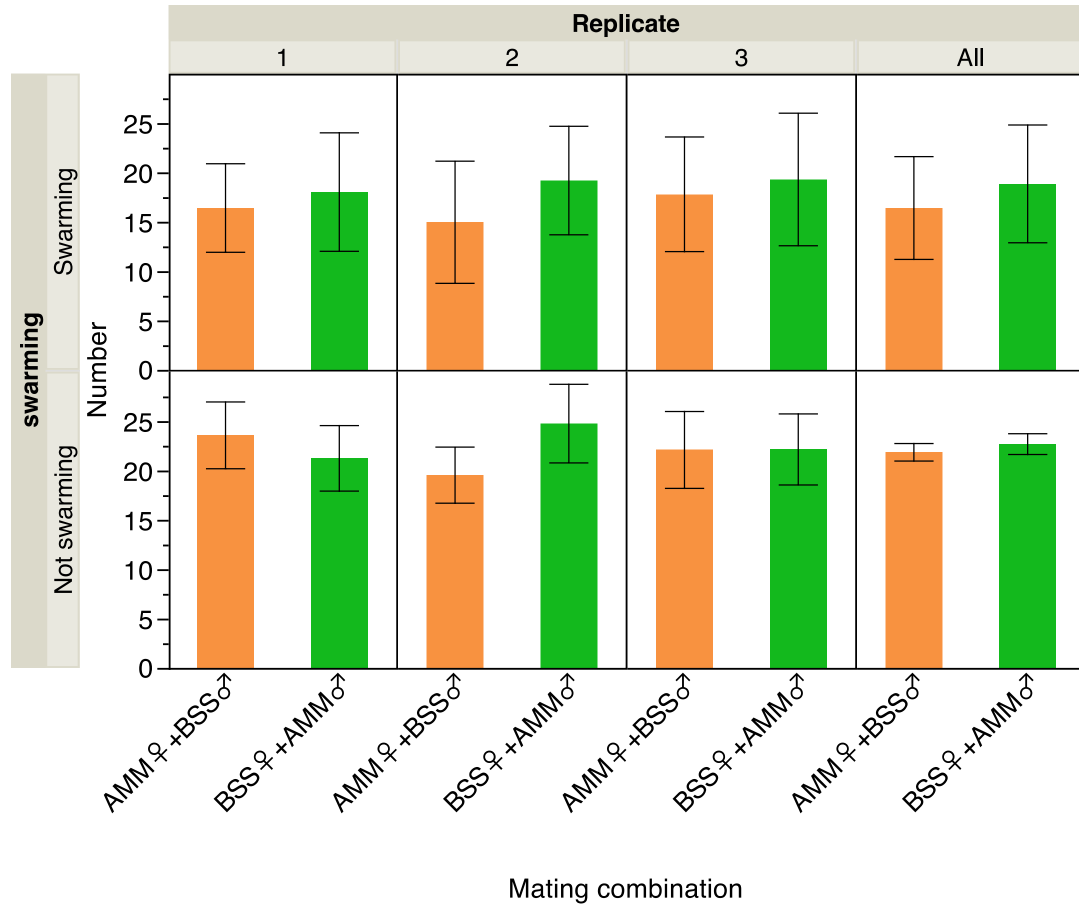


Figure 5.16. The means of swarming and non-swarming numbers for each replicate and overall the replicates for AMM♂+ BSS♂ and BSS♀+AMM♀ combinations in different replicates. The bars showed the means and confidence intervals 95% values. The orang colour bars indicate to AMM♀+ BSS♂ combination and the purple colour bars indicate to BSS♀+AMM♂ combination.

Proportion of non-swarming individuals

In matched islands combinations, to limit the pseudoreplicate, the effect of mating combination, time, replicate and the interaction between them on the number of resting individuals in different locations in the cages were analysed at four time intervals (10min, 60min, 120min, 180min). The resting individuals numbers changed based on the mating combination, replicate and time (Table 5.8). The means and confidence intervals of the different replicates were shown (Table 5.9).

Table 5.8. The effect of mating combination, time and replicate on the number of mosquitoes in different locations in matched islands combinations at four time intervals (10min, 60min, 120min, 180min).

Source	*df*	*L-R* ChiSquare	*P-value*
Mating combination	2	13.86	0.001*
Replicate	4	16.02	0.003*
Time points	2	7.6	0.022*
Replicate* Time points	4	11.63	0.02*

Table 5.9. The means and confidence intervals of resting individuals in matched islands combinations – brackets indicate the minimum and maximum values of confidence intervals for the three replicates in the different combinations at four time intervals (10min, 60min, 120min, 180min).

AMM♀+AMM♂				
Top wall				
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	58.6 (50.3-66.4)	51.9 (42.5-61)	45.2 (35-55.9)	53.2 (43.9-62.2)
60min	53.1 (44.5-61.6)	43.1 (34.7-51.9)	41.9 (32-52.4)	46.4 (37.5-55.6)
120min	40.0 (31.7-48.9)	34.8 (27.3-43)	45.5 (37-54.3)	40.2 (32-48.9)
180min	42.1 (33.4-51.3)	35.3 (27.3-44.2)	48.1 (38.7-57.6)	41.6 (32.9-50.8)
Down wall				
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	37.1(29.6-45.4)	44.4 (35.4-53.8)	47.6 (37.3-58.2)	42.3 (33.6-51.6)

60min	42.2 (34-50.8)	48.8 (40.1-57.5)	51.2 (40.8-61.4)	47.3 (38.3-56.5)
120min	51.7 (42.8-60.4)	53.6 (45.3-61.7)	53.6 (44.9-62.2)	52.8 (44.1-61.2)
180min	50.9 (41.8-59.9)	54.6 (45.7-63.3)	46.2 (36.9-55.7)	50.4 (41.4-59.5)
		Bottom		
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	4.3 (2.0-9.0)	3.7 (1.4-9.1)	7.1 (3.3-14.7)	4.5 (1.9-10.1)
60min	4.7 (2.2-9.8)	8.1 (4.5-14.3)	7.0 (3.2-14.4)	6.3 (3.1-12.3)
120min	08.3 (4.6-14.7)	11.6 (7.3-18)	0.8(0.1-4.5)	7.1 (3.8-12.9)
180min	7.0 (3.6-13.2)	10.1 (5.9-16.8)	5.8 (2.7-12)	8.0 (4.2-14.4)
		BSS♀+BSS♂		
		Top wall		
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	54.4 (46-62.5)	50 (40.7-59.3)	53.7 (42.9-64)	52.3 (43-61.4)
60min	56.8 (48-65.2)	55.2 (45.3-64.8)	50.6 (40.3-60.8)	54.4 (44.8-63.7)
120min	53.5 (44.9-62)	46.3 (37.2-55.7)	55.4 (46.1-64.2)	52.2 (43.1-61.1)
180min	44.8 (36-53.9)	46.1 (36.1-56.4)	55.1 (46.4-63.5)	49.1 (39.9-58.3)
		Down wall		
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	39.7 (31.9-48.1)	42.6 (33.7-52)	41.5 (31.4-52.3)	41.3 (32.5-50.7)
60min	34.4 (26.6-43.1)	37.5 (28.5-47.5)	46 (35.9-56.4)	38.8 (30-48.5)
120min	36.2 (28.4-44.9)	48.1 (39-57.5)	39.3 (30.7-48.5)	40.9 (32.3-50)
180min	43.1 (34.5-52.2)	41.6 (31.9-52)	42.5 (34.3-51.2)	42.7 (33.9-52.1)
		Bottom		
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	5.9 (3-11.2)	7.4 (3.8-14)	4.9 (1.9-11.9)	6.4 (3.1-12.7)
60min	8.8 (5-15.1)	7.3 (3.6-14.3)	3.4 (1.2-9.7)	6.8 (3.3-13.4)
120min	10.2 (6.1-16.7)	5.6 (2.6-11.6)	5.4 (2.5-11.2)	7 (3.6-13.1)
180min	12.1 (7.3-19.2)	12.4 (7-20.8)	2.4 (0.8-6.7)	8.2 (4.4-14.8)

The number of resting individuals in total three hours was higher in the upper wall for BSS individuals than AMM individuals in the first, third replicates and overall means of replicates. In contrast, the number of the resting individuals in the down wall of the cage was higher for the AMM individuals than BSS individuals in all the replicates. However, the numbers changed differently in the bottom of the cage for both combinations (Fig 5.17 & 5.18).

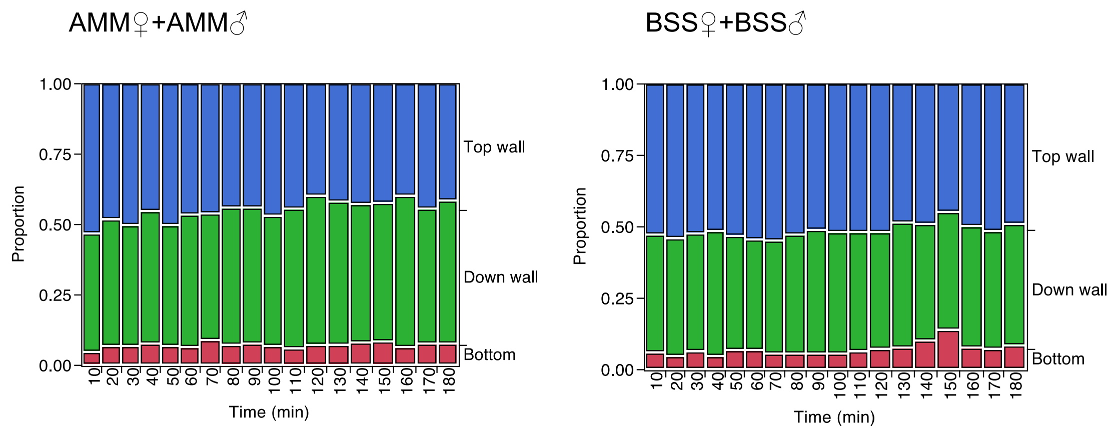


Figure 5.17. Change in proportion of resting individuals in relation to time for AMM♀+AMM♂ and BSS♀+BSS♂ combinations - Proportion of mosquitos' numbers in top wall, down wall and bottom during 3 hours in the cages. The blue bars indicate to the top wall ratios, the green bars indicate to the down wall ratios and the red bars indicate to the bottom ratios every 10min. ($P<$ 0.001) there is a significant difference between mating combinations.

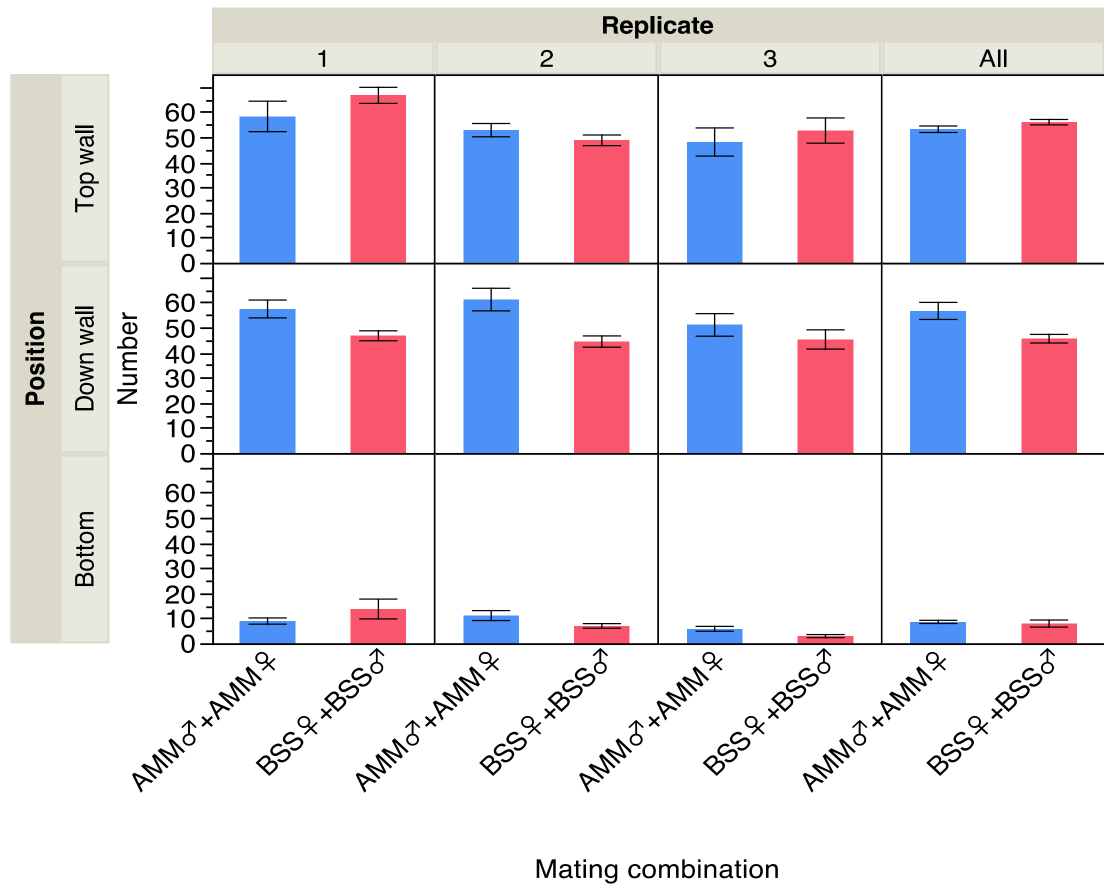


Figure 5.18. The means of resting individuals for each replicate and means of overall replicates for AMM♂+AMM♀ and BSS♀+BSS♂ combinations in different replicates. The bars showed to the means and confidence intervals 95% values. The blue bars indicate to AMM♀+AMM♂ combination and the red bars indicate to BSS♀+BSS♂ combination.

In non-matched islands combinations comparisons, to limit the pseudoreplication, the effect of mating combination, time, replicate and interaction between them on number of resting mosquitoes in different locations of the cages were analysed at four time points (10min, 60min, 120min, 180min). There was no effect of mating combination and time, whereas the number of individuals in different locations of the cage differed based on replicate (Table 5.10). The means and confidence intervals for the different replicates were shown (Table 5.11).

Table 5.10. The effect of mating combination, time and replicate on number of resting individuals in non-matched islands combinations at four time points (10min, 60min, 120min, 180min).

Source	*df*	*L-R* ChiSquare	*P*-value
Mating combination	2	0.24	0.886
Replicate	4	10.54	0.032*
Hour	2	1.06	0.59
Mating combination*Replicate	4	13.7	0.008*

Table 5.11. The means and confidence intervals of resting individuals in non-matched islands combinations- brackets showed the minimum and maximum values of confidence intervals for the three replicates in the different combinations at four time intervals (10min, 60min, 120min, 180min).

AMM♂+ BSS♀				
Top wall				
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	51.7 (39.2-64.1)	40.7 (28.7-54)	42.9 (30.8-55.9)	44.6 (32.4-57.6)
60min	58.1 (45.7-69.5)	39.3 (27.6-52.4)	26.3 (17.7-37.2)	40.6 (29.4-52.9)
120min	41.5 (31.4-52.3)	55.4 (42.4-67.6)	42.4 (31.2-54.4)	45.6 (34.3-57.3)
180min	46.2 (35.5-57.1)	45.5 (34-57.4)	46.4 (35.1-58)	46.5 (35.4-58)
Down wall				
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	41.4 (29.6-54.2)	44.4 (32-57.6)	46.4 (34-59.3)	44.6 (32.4-57.6)

60min	32.3 (22-44.6)	46.4 (34-59.3)	71 (60-80)	51.6 (39.6-63.4)
120min	41.5 (31.4-52.3)	42.9 (30.8-55.9)	51.5 (39.7-63.1)	45.6 (34.3-57.3)
180min	41 (30.8-52.1)	42.4 (31.2-54.4)	52.2 (40.6-63.5)	45.1 (34-56.6)
		Bottom		
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	6.9 (2.7-16.4)	14.8 (7.7-26.6)	10.7 (5-21.5)	10.7 (5-21.5)
60min	9.7 (4.5-19.5)	14.3 (7.4-25.7)	2.6 (0.7-9.1)	7.8 (3.4-17)
120min	17.1 (10.5-26.6)	1.8 (0.3-9.4)	6.1 (2.4-14.6)	8.8 (4.1-17.9)
180min	12.8 (7.1-22)	12.1 (06.3-22.1)	1.4 (0.3-7.8)	8.5 (3.9-17.2)
		BSS♀+ AMM♂		
		Top wall		
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	47.8 (34.1-61.9)	27.3 (18-39)	47.5 (35.3-60)	40.4 (28.6-53.3)
60min	44.8 (32.7-57.5)	40(30-51)	51.6 (39.6-63.4)	44.8 (33.5-56.6)
120min	48.4 (36.4-60.6)	51.4 (40-62.8)	44.3 (33.9-55.3)	47.9 (36.7-59.3)
180min	39.5 (29.9-50.1)	42.5 (32.3-53.4)	48.4 (36.6-60.4)	42.9 (32.4-54)
		Down wall		
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	47.8 (34.1-61.9)	57.6 (45.6-68.8)	35.6 (24.6-48.3)	47.4 (35-60.1)
60min	48.2 (35.9-60.8)	47.5 (36.9-58.3)	46.9 (35.2-58.9)	47.8 (36.3-59.5)
120min	45.2 (33.4-57.5)	40 (29.3-51.7)	46.8 (36.2-57.7)	43.7 (32.7-55.2)
180min	44.2 (34.2-54.7)	52.5 (41.7-63.1)	43.8 (32.3-55.9)	46.8 (36-57.8)
		Bottom		
Time	**Replicate1**	**Replicate2**	**Replicate3**	**All**
10min	4.3 (1.2-14.5)	15.2 (8.4-25.7)	16.9 (9.5-28.5)	12.3 (6-23.2)
60min	06.9 (2.7-16.4)	12.5 (6.9-21.5)	1.6 (0.3-8.3)	7.5 (3.2-16.3)
120min	6.5 (2.5-15.4)	8.6 (4-17.5)	8.9 (4.4-17.2)	8.5 (3.9-17.2)
180min	16.3 (10-25.5)	5 (2-12.2)	7.9 (3.4-17)	10.4 (5.4-19.2)

The number of resting individuals changed differently through time in different locations of the cages and there was no difference between AMM♂+ BSS♂ and BSS♀+AMM♀ combinations (Fig 5.19).

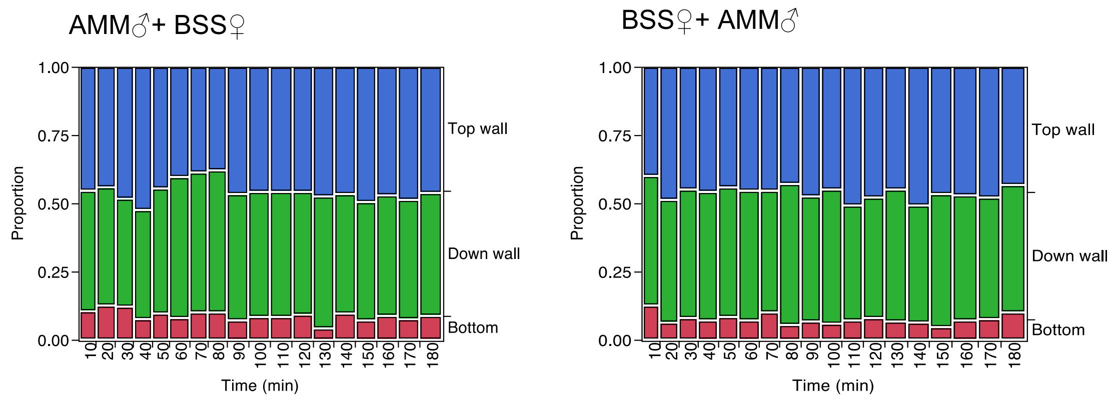


Figure 5.19. Change in proportion of non-swarming individuals in relation to time for AMM♂+ BSS♀ and BSS♀+ AMM♂ combinations - Proportion of mosquitoes' numbers in top wall, down wall and bottom during 3 hours in the cages. The blue colour bars indicate to the top wall ratios, the green bars indicate to the down wall ratios and the red colour bars indicate to the bottom ratios every 10 minutes. ($P > 0.05$) there is no significant difference between the combinations.

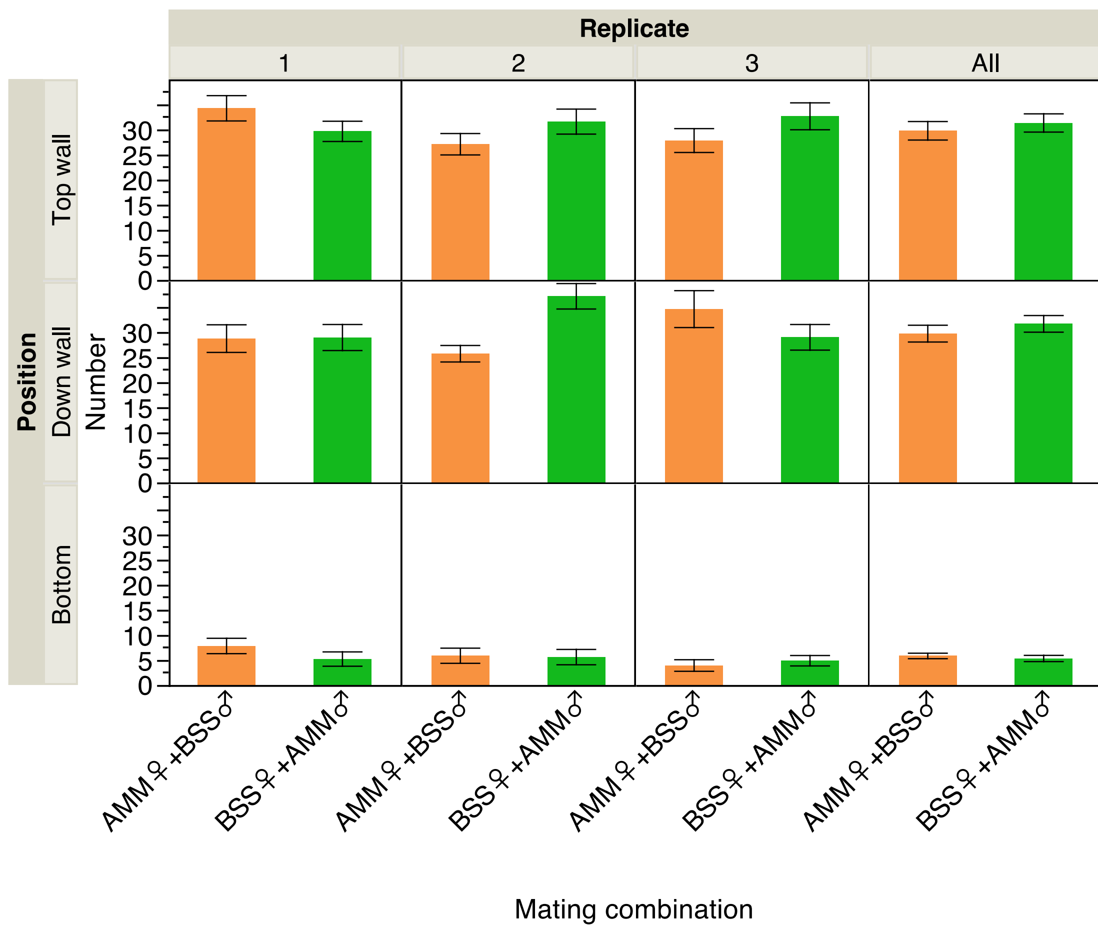


Figure 5.20. The means of resting individuals for each replicate and overall replicates for AMM♂+ BSS♂ and BSS♀+AMM♀ combinations in different replicates. The bars showed to the mean and confidence intervals 95% values. The orang colour bars indicate to AMM♀+ BSS♂ combination and the purple colour bars indicate to BSS♀+AMM♂ combination.

5.3.3. Matched and not-matched X-island comparisons

5.3.3.1. Mating behavior

Different analyses amongst the mating combinations (AMM♂+ BSS♂, BSS♀+AMM♀, AMM♂+AMM♀ and BSS♀+BSS♂) and mating types (Matched and not-matched X-island combinations) were conducted to check if there was any difference between the AMM and BSS recombinants mating behaviour.

Duration of mating stages

There was no significant difference in leg grips duration amongst different combinations (Fisher's test: $F_{3,32}$= 0.33, P= 0.80). Moreover, No difference was found in tandems duration amongst the different combinations (Fisher's test: $F_{3,243}$= 2.35, P= 0.07) and also no difference in total mating duration between them (Fisher's test: $F_{3,243}$= 2.30, P= 0.078) (Fig. 5.21) (Table 5.12).

Table 5.12. The means and confidence intervals of the leg grips, tandems and total mating durations (sec) in the different mating combinations - The mating time (sec) includes leg grips and tandems periods of total of three replicates.

Mating combinations	Leg grips duration (sec)	Tandem duration (sec)	Total duration (sec)
AMM♀+AMM♂	2.1 (1.9-2.3CI)	17.7 (16-19.3CI)	19.5 (17.9-21.2CI)
BSS♀+BSS♂	2.1 (1.9-2.3CI)	16.5 (14.7-18.3CI)	18.0 (16.2-19.8CI)
AMM♀+BSS♂	2.1 (1.9-2.3CI)	14.2 (12-16.5CI)	16.2 (13.9-18.5CI)
BSS♀+AMM♂	2.0 (1.8-2.2CI)	15.0 (12.7-17.4CI)	16.7 (14.4-19.0CI)

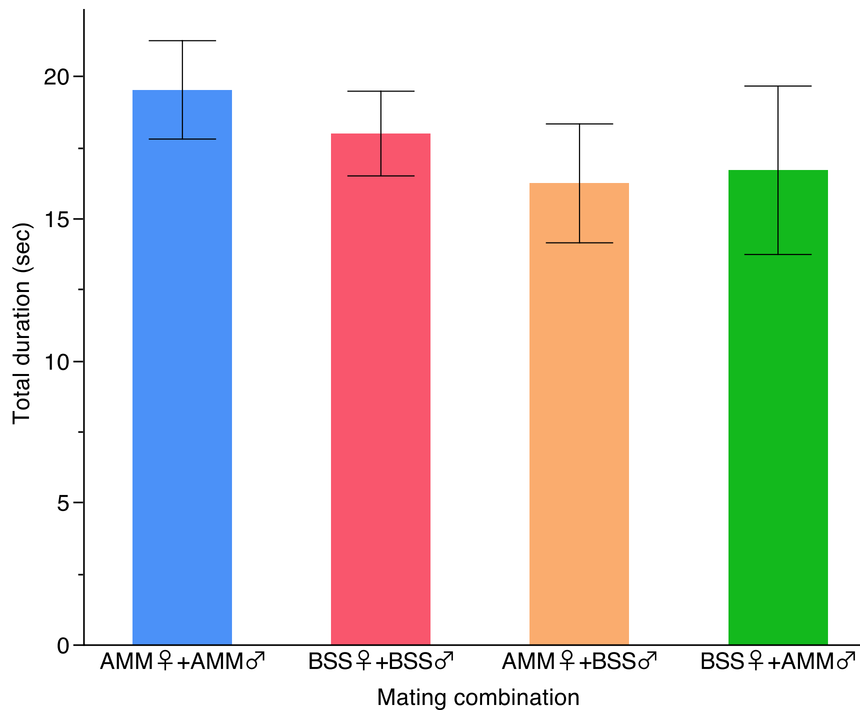


Figure 5.21. The means of total duration of mating time (sec) – The mating time includes leg grips and tandems periods for total of three replicates. ($P > 0.05$) there is no significant difference between the combinations.

The effect of mating combination on sqrt duration of tandems for overall number of replicates was analysed. There was a significant difference between AMM♀+AMM♂ and AMM♀+BSS♂ combinations (Tukey: $P = 0.0184$), and between AMM♀+AMM♂ and BSS♀+AMM♂ combinations (Tukey: $P = 0.0233$), but there was no significant difference between the other combinations (Tukey: $P = 1.0-0.1$) (Fig. 5.22).

In addition, the effect of mating type on sqrt duration of tandem was analysed for overall number of replicates. There was a significant difference between tandems duration time between different mating types (T-test: $n = 248$, $T = 3.62$, $P = 0.0004$). The mean time of duration of tandems was 17.2 (16.0-18.3CI) for matched speciation X-island mating type, and 14.1 (12.8-15.4CI) for not-matched speciation X-island mating type (Fig. 5.22).

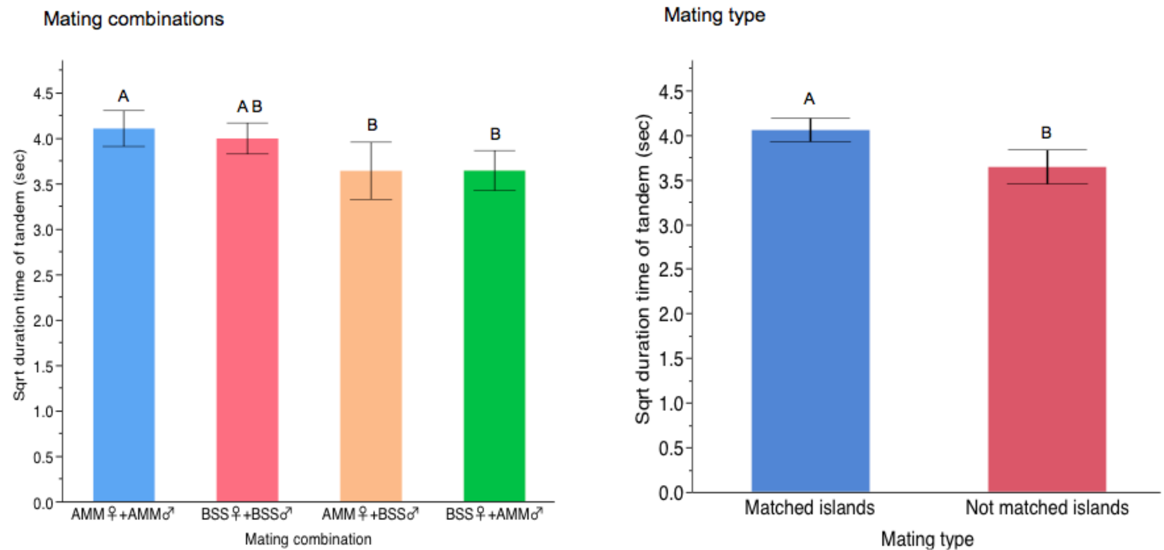


Figure 5.22. The sqrt duration (time sec) of tandem of overall mean of three replicates for mating combinations and mating types - The different colour bars showed the mean of sqrt duration of tandems and confidence intervals 95% values in different mating combinations and mating types. A and B groups are significantly different, AB is not significantly different from both A and B groups.

Proportion of tandems to leg grips

The differences in the proportion of tandems to leg grips across all the replicates was analysed in different mating combinations to check any difference when they mate with their matched/not-matched X-island combinations (ChiSquare likelihood-ratio: n= 4135, df= 3, X^2= 18.31, P< 0.001) (Fig 5.23).

The results showed a significant difference between the different mating types when party the data into the two categories, matched and not-matched X-island types. The tandem proportion was significantly more in matched islands mating type than the not-matched island mating type (ChiSquare likelihood-ratio: n= 4135, df= 1, X^2= 13.432, P< 0.001) (Fig 5.23).

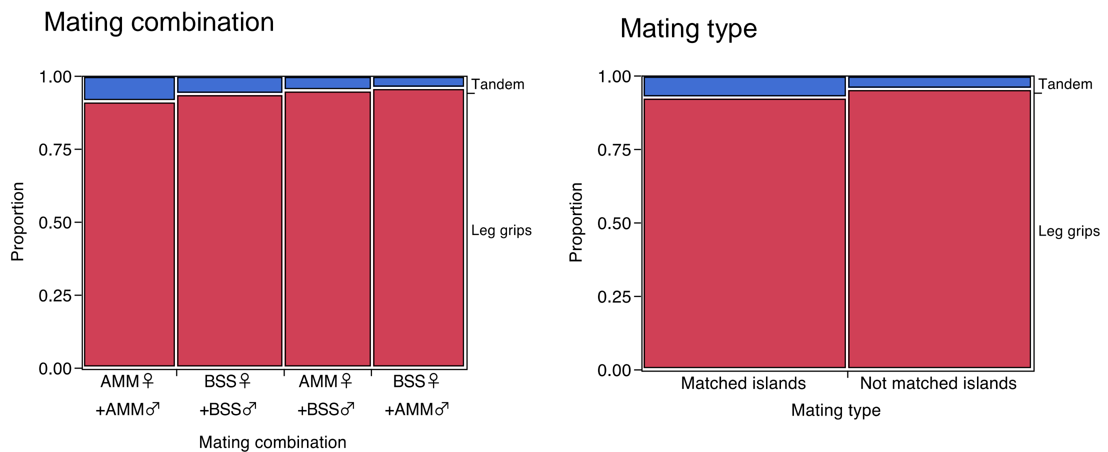


Figure 5.23. The proportion of tandems to leg grips of the mean numbers of overall replicates for mating combination and mating type - The blue bars indicate to tandem ratios and the red bars indicate to leg grips. (P<0.001) there is a significant difference among the mating combinations, and (P<0.001) there is a significant difference between the mating types.

Furthermore, The ratio of tandems to leg grips for the mean number of overall replicates was analysed in different mating types in every hour to know if it was changing through time. In the first hour, there was no significant difference in ratio of tandems to leg grips between mating types (ChiSquare likelihood-ratio: n= 1960, df= 1, X^2= 2.04, P= 0.153), whereas there was a significant difference in the second hour (ChiSquare likelihood-ratio: n= 1269, df= 1, X^2= 18.0, P< 0.001), and the third hour (ChiSquare likelihood-ratio: n= 906, df= 1, X^2= 7.34, P= 0.007). However this number of tandems dropped faster in not-matched islands type, whereas it comparatively decreased gradually in matched islands type (Table 5.13) (Fig 5.24).

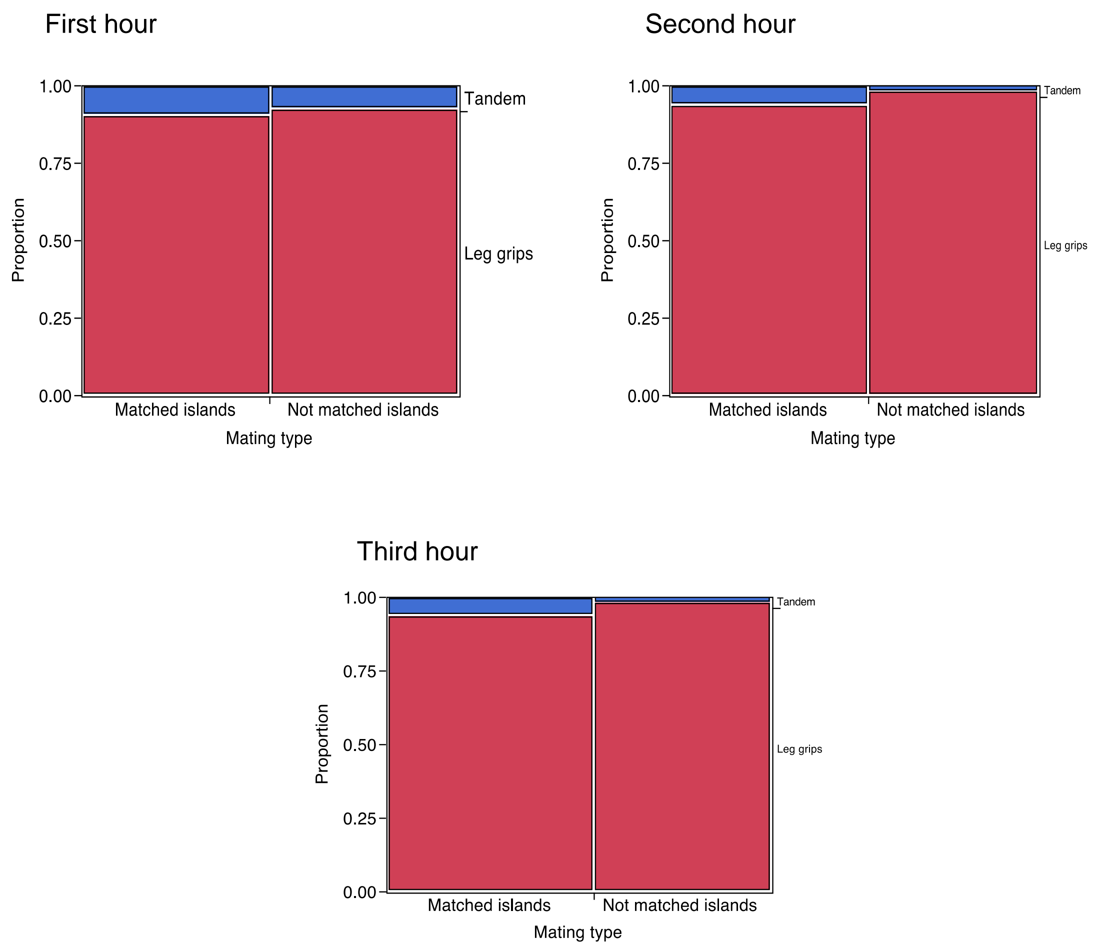


Figure 5.24. The proportion of tandems to leg grips for overall three replicates amongst different mating types for the first hour, second hour and third hour - The blue bars indicate to tandems ratios and the red bars indicate to leg grips between the couples. (P=0.153) there was no significant difference in the first hour, but (P<0.01) there was a significant difference in the second and third hours between the mating types.

Table 5.13. The effect of mating Type and time (sec) on the leg grips and tandems' frequencies.

Source	***df***	***L-R* ChiSquare**	***P*-value**
Mating type	1	2.04	0.153
Hours	2	50.72	<0.001*
Mating type*Hours	2	14.37	0.026*

Proportions of swarming individuals

To limit pseudoreplication, the effect of mating combination and time on swarming and non-swarming numbers was analysed at four time points (10min, 60min, 120min, 180min). We found that the mating combinations have effect on the number of swarming individuals, and the numbers differed through time (Table 5.14).

Table 5.14. The effect of mating combination and time on the number of swarming individuals in the cages at four time points (10min, 60min, 120min, 180min).

Source	***df***	***L-R* ChiSquare**	***P*-value**
Mating combination	3	266.07	<0.001*
Hours	3	58.74	<0.001*

The proportion of swarming and non-swarming individuals was analysed in matched and non-matched island types at four time points (10min, 60min, 120min, 180min). There was a significant difference between them (Table 5.15).

Table 5.15. The effect of mating type and time on the number of swarming individuals in the cages at four time points (10min, 60min, 120min, 180min).

Source	***df***	***L-R* ChiSquare**	***P*-value**
Mating type	1	265.73	<0.001*
Hours	3	58.78	<0.001*

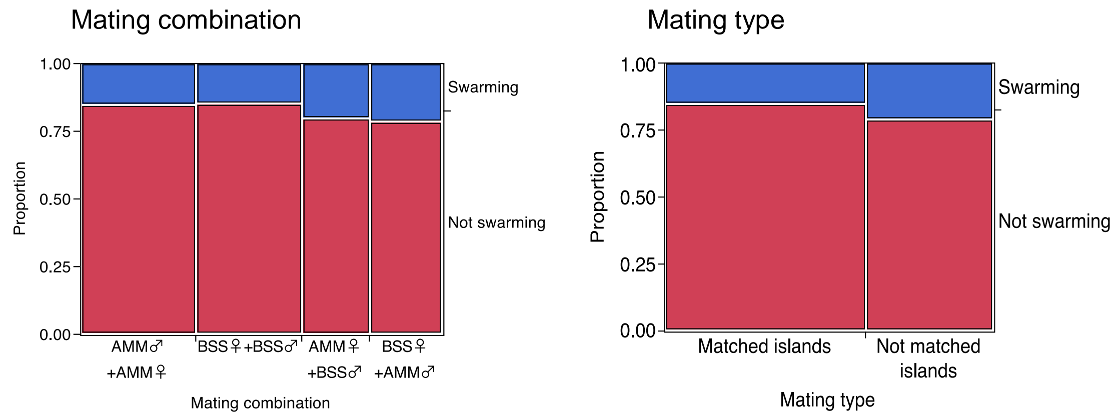


Figure 5.25. Proportion of swarming individuals in overall replicates for mating combination and mating type - The blue bars indicate to tandems ratio and the red bars indicate to leg grips ratio between the couples. ($P<0.001$) there was a significant difference among the mating combinations, and ($P<0.001$) there was a significant difference between the mating types.

Proportion of non-swarming individuals

To limit the pseudoreplication, the effect of mating combination and time on the number of resting mosquitoes in different locations of the cages were analysed at four time points (10min, 60min, 120min, 180min). There was an effect of mating combination whereas no effect of time (Table 5.16).

Table 5.16. The effect of mating combination and time on the number of resting individuals in the cages at four time points (10min, 60min, 120min, 180min).

Source	*df*	*L-R* ChiSquare	*P*-value
Mating combination	6	22.46	0.001*
Time min	2	3.12	0.21

The proportion of non-swarming individuals was analysed in different locations of the cages between matched and non-matched island types. There was a significant difference between them through time (Table 5.17).

Table 5.17. The effect of mating island type and time on the number of resting individuals in the cages at four time points (10min, 60min, 120min, 180min).

Source	*df*	*L-R* ChiSquare	*P*-value
Mating type	3	269.62	<.0001*
Time min	9	60.53	<.0001*

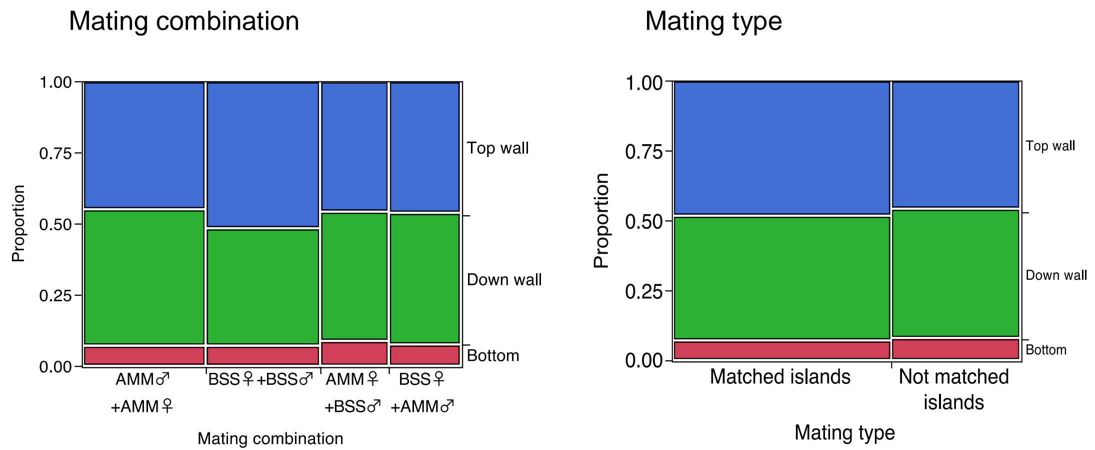


Figure 5.26. Proportion of resting individuals for overall replicates amongst mating combinations and mating types - The blue bars indicate to the individuals' ratio on the top walls, the green bars indicate to the individuals' ratio on the down walls and the red bars indicate to individual ratio in the bottom of the cages. ($P<0.001$) there was a significant difference among the mating combinations, and ($P<0.001$) there was a significant difference between the mating types.

Table 5.18. Summary table for the analyses and comparisons (*P-Values*) between the matched (AMM♂+AMM♀ and BSS♀+BSS♂) and non-matched (AMM♂+ BSS♂ and BSS♀+AMM♀) X-islands combinations of recombinants.

Variable	Matched combinations	Non-matched combinations
Initiation of leg grips	P= 0.03 SD	P< 0.001 SD
Initiation of tandem	P= 0.79 NSD	P= 0.37 NSD
Duration of tandem	P= 0.4 NSD	P= 0.61 NSD
Proportion of swarming individuals	P= 0.92 NSD	P= 0.3 NSD
Proportion of resting positions	P= 0.001 SD	P= 0.88 NSD

Table 5.19. Summary table for the analyses and comparisons (*P-Values*) between the (AMM♂+AMM♀, BSS♀+BSS♂, AMM♂+ BSS♂ and BSS♀+AMM♀) combinations and mating types (matched and non- matched islands types)of recombinants.

Variable	Mating combinations	Mating type
Duration of leg grips	P= 0.8 NSD	P< 0.001 SD
Duration of tandem	P= 0.07 NSD	P< 0.001 SD
Proportion of swarming individuals	P< 0.001 SD	P< 0.001 SD
Proportion of resting positions	P< 0.001 SD	P< 0.001 SD
Proportion resting individuals positions	P< 0.001 SD	P< 0.001 SD

5.3.4. Sperm quantification

The sperm was extracted from the mated females in the different mating combinations and quantified by using QPCR method to find if there was any difference in the inseminated sperm amount between matching and not-matching X-island type.

5.3.4.1. Phenotypic quality of individuals for the sperm quantification samples

Random samples of the mosquitoes of sperm quantification experiment were collected from each strain and the wings were removed, measured and analysed as a representative method to compare the body size of the strains (Renshaw et al. 1994; Mwangangi et al. 2004) to avoid the effect of mosquitos' size on the mating process. There was no significant difference in the size between the males (T-test: n= 40, df= 1, T= -0.35, P> 0.5) (Fig 5.27). The means and confidence interval values for the wing lengths were 3.09 (3.05-3.14CI) for AMM strain, and 3.08 (3.04-3.13CI) for BSS strain. There was also no significant difference in the size between the females (T-test: n= 40, df= 1, T= 1.70, P= 0.09) (Fig 5.27). The means and confidence intervals values for the wing length were 2.97 (2.93-3.01CI) for AMM strain, and 3.01 (2.98-3.05 CI) for BSS strain.

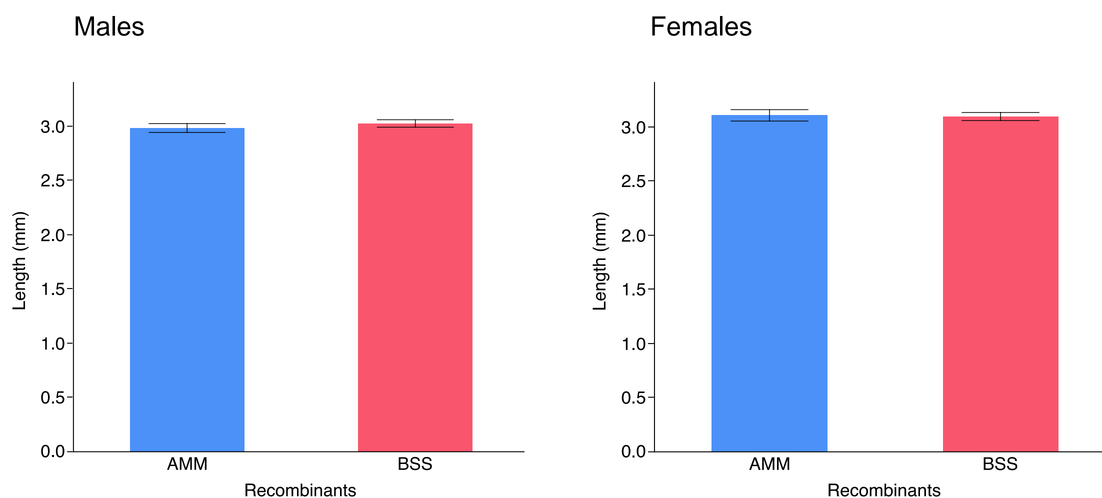


Figure 5.27. Means of wing lengths in males and females of random samples for sperm quantification experiment - The bars showed the means and confidence intervals 95% values. (P>0.5) there is no significant difference between the male groups, and (P= 0.09) there is no significant difference between the female groups.

5.3.4.2. The sperm quantification in different mating combinations

The results showed a significant difference in estimated number of sperm that was inseminated into the females of different mating combinations (Wilcoxon/Kruskal-wallis: n= 125, df= 3, X^2= 16.5, P< 0.001). Therefore, Dunn test for joint ranking was used to compare each pair of the combinations. There was a significant difference between BSS♀+AMM♂ and AMM♀+AMM♂ (Dunn for joint ranking: Z= 3.48, P= 0.003), and between BSS♀+AMM♂ and AMM♀+BSS♂ (Dunn for joint ranking: Z= 3.13, P= 0.01). The medians and (25-75%) quartiles values were 8687.4 (17588.5-6316.1) for AMM♀+AMM♂, and 21125.2 (27056.3-16174) for BSS♀+AMM♂, and 7868.5 (24913.8-5879.4) for AMM♀+BSS♂, and 18851.5 (25499-5904.19) for BSS♀+ BSS♂. There was no significant difference between the estimated sperm number in the other combinations (Fig. 5.28)

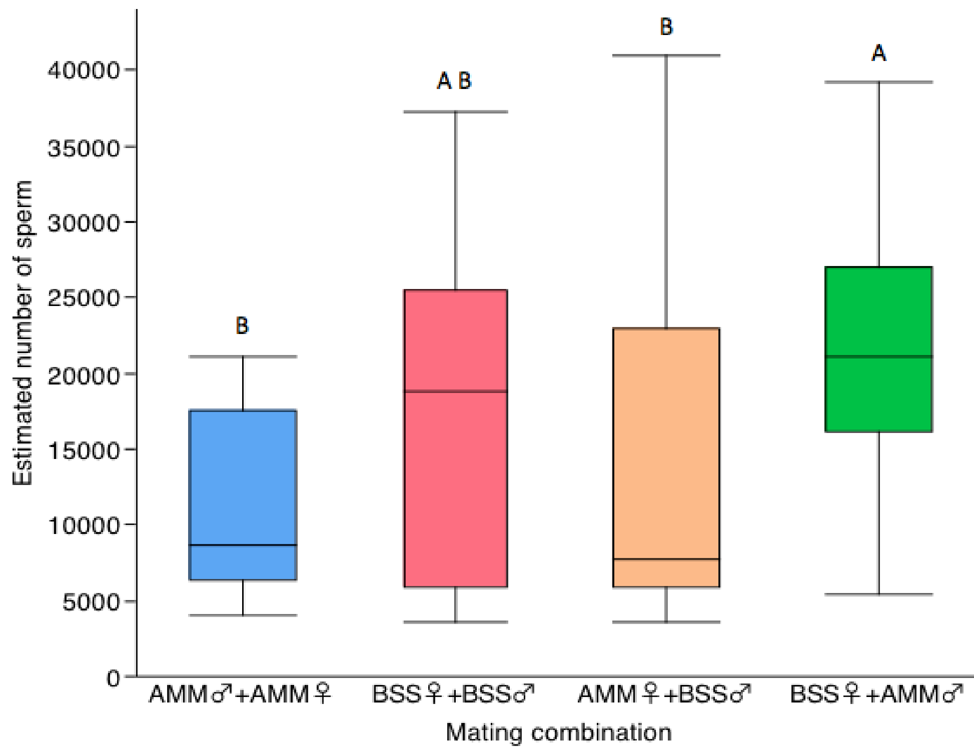


Figure 5.28. The estimated number of sperm that was inseminated in females of different combinations - Boxplots shows the median, quartiles and brackets indicate the minimum and maximum values. A and B groups are significantly different, whereas AB group is not significantly different from both A and B groups.

5.3.4.3. The sperm quantification in different females types

The estimated number of sperm was compared based on the type of females in different combinations. The results showed a difference between the estimated numbers of sperm that was inseminated in the different types of females (Wilcoxon: n= 125, Z= -3.3, P= 0.001) (Fig. 5.29). The medians and (25-75%) quartiles values were 7986 (5900.217-20202.45) for AMM females, and 20126.52 (13134.78-26805.03) for BSS females.

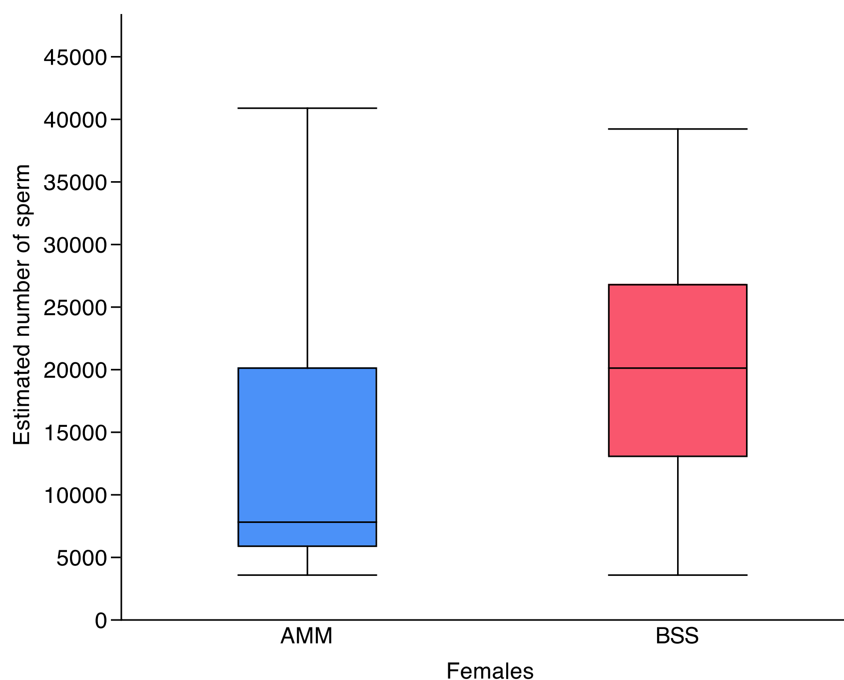


Figure 5.29. The estimated number of sperm based on the females type in different combinations - Boxplots show the median, quartiles and brackets indicate the minimum and maximum values. (P= 0.001) there is a significant difference between the AMM and BSS females uptake of sperm.

5.4. Discussion

Several studies tried to describe the dynamics of the swarms and mating behaviour of *An. gambiae s.s.* and *An. coluzzii*, especially in field (Charlwood & Jones 1979; Gibson 1985; Diabaté et al. 2009; Manoukis et al. 2009; Diabaté et al. 2011; Dabire et al. 2013). It is believed that the females and males could use specific mechanisms to attract conspecific individuals even in mixed swarms (Manoukis et al. 2009; Pennetier et al. 2010). However, no full explanation for these mechanisms was found. These ways could include behavioural or sound differences between the sibling species that are still not completely unravelled and understood (Charlwood & Jones 1979; Howell & Knols 2009; Manoukis et al. 2009; Pennetier et al. 2010; Sawadogo et al. 2013). Therefore, we tried to fill some gaps in our understanding of how males and females of *An. gambiae s.s.* and *An. coluzzii* mate assortatively by comparing the swarming behaviour in the laboratory.

Here, comparisons were made between AMM and BSS recombinant strains that differ at their X-island genotype and have been shown to mate strictly assortatively in the previous experiments (Chapter 3). This is also the first laboratory study describing and comparing time and duration of the main mating stages (leg grips and tandems) that occur during swarming. Moreover, the proportion of males and females participating in swarms and resting in different sites of the cages was also described. In addition, the quantity of sperm that have been transferred to spermatheca of both types of females when they mated with males with matching or non-matching island type was compared.

In general, the copulation in all the types of our lab combinations lasts 16.7-17.5sec including the periods of leg grip and tandem stages and this result was consistent with the copulation period in the field studies in *An. gambiae* s.l. that lasts 15-20sec (Charlwood & Jones 1979). In experimental swarms involving females and males with matching islands, leg grips and tandems started immediately after the light was dimmed, occurred over the

first 3 hours and their frequency decreased gradually with time. This activity closely matched the changes observed in overall number of individuals involved in swarming as opposed to resting over the 3 first hours. Similar observations were also reported in the field studies in *An. gambiae s.s.* and *An. coluzzii* species, albeit over much shorter total swarm duration 22min (Sawadogo et al. 2013). The effect of light intensity on mating activity of *Anopheles gambiae* s.l. was examined long time ago (Jones & Gubbins 1978). These studies suggested that it is associated with the circadian rhythm of mosquitoes' activity that allow them to adjust their sexual activity based on day length changes in different periods of the year (Jones & Gubbins 1978; Sawadogo et al. 2013). Field swarm studies suggested that the high mating activity shortly after swarming starts is due to high males receptivity to females flight tones (Charlwood & Jones 1979). However, swarming is energetically costly, it was reported that the males of both species spend ~50% of reserved sugar and glycogen when they swarm for 25min (Maïga, Niang, Simon P. Sawadogo, et al. 2014). Therefore, males leave swarms to rest and feed to replenish the energy reserves that they used during the mating process (Howell & Knols 2009; Maïga, Niang, Simon P. Sawadogo, et al. 2014). Moreover, mated females leave the swarms because they cannot mate again due to mating plugs that the males insert during the mating (Gillott 2003). Therefore, the number of virgin females also decreases in the swarms and so do the chances of males in finding a new virgin female for mating (Gillott 2003).

In a previous study, it was found that males *An. gambiae s.s.* start swarming 3.5min and end 5min earlier than those of *An. coluzzii,* for a total period of swarming equivalent 22min in both species (Sawadogo et al. 2013). This led the authors to hypothesize that this 15-20% difference in swarming time between both species could play a small role in the reproductive isolation between them. Interestingly, here we found in the matched islands combinations that BSS strain initiated leg grips earlier than AMM by a mean of ~1sec over

three hours, which is a very tiny difference. In addition, this difference was only observed in 2 out of 3 replicates. Therefore, this small difference cannot explain the strong reproductive isolation observed between recombinant strains in our studies that suggest a limited role of swarming time in assortative mating. This is further supported by the absence of differences between the matched islands combinations in start and duration time of tandem stage.

When the mosquitoes left the swarms to rest, we found that BSS individuals preferred to rest in upper level of the cage compared to AMM individuals. This is the first study observed the differences in resting positions between the sibling species and there is no previous data suggesting differences in resting positions heights between the two sibling species in field or laboratory studies.

When we combined non-matched islands in mating cages, the mating occurred between them when we did not give them the chance to choose their mates freely. The results in non-matched islands combinations were similar to the matched islands combinations regarding to the frequency of leg grips, tandems and swarming activity that also decreased gradually with time regardless of the male and female types. In addition, there was no significant difference detected in non-matched islands combinations mating time, mating ratios and the number of resting mosquitoes in different sites of the cages.

Interestingly, there was no significant difference in the frequency of leg grips and tandems between matched and not-matched islands combinations. This suggested that the mosquitoes could mate with heterospecific type with the same frequencies when they mate with conspecific type when we do not give them another choice. It has been already reported in indoors mating behaviour between *An. gambiae s.s.* and *An. coluzzii* that the assortative mating breaks down in some cases (Dao et al. 2008). Moreover, It was found

that the males also abandon the choosy behaviour in some outdoor mating cases when the females are limiting factor and the costs of mate finding are high (Howell & Knols 2009).

Interestingly, when we measured and analysed the period of tandem in matched and non-matched islands combinations, we found that BSS and AMM recombinants spent longer time in the tandem position when they mated with their own type than when they mated with the opposite type. In addition, the proportion of tandems to leg grips was high and similar in the first hour in both matched and non-matched mating types, but the ratio decreased significantly in non-matched islands combinations, whereas it decreased slowly in matched island mating type. This also supported the previous suggestion that the mosquitoes could mate with heterospecific type when they do not have another choice in the first hour. This behaviour did not continue in the lab and we found that the tandems to leg grips ratio reduced sharply and significantly in the second and third hour when they mated with the wrong type of mosquitoes. However, the mating ratio still little in the heterospecific mating type comparing with the conspecific mating type that continued with high ratio over time during the three hours. This behaviour could be due to their ability in detecting wing beats of conspecific individuals in the nearby cage that included their own type. Therefore, they rejected the heterospecific individuals that exist with them in the same cage. In overall, we found that the mating rejection did not happen after mosquitoes arrive to tandem stage in the mating process, whereas the mating rejection could happen in leg grips stage. These findings suggest that the mosquitoes could use olfactory receptors that either located directly in the forelegs or elsewhere in the body that they use to recognize conspecific mosquitoes.

Another interesting finding is that the number of swarming individuals was approximately twice higher when the mosquitoes mated with non-matched individuals compared to their numbers when they mated with the matched ones. The potential

explanation that when they were with the wrong mates they kept trying to harmonize their wing beats to recognize the nearby mosquitoes type and try to find their own type of mates. Possibly they had difficulties in harmonizing their wing beats with the non-matching island types but these difficulties did not stop them from attaching their legs in leg grips position as above-motioned analyses that showed no differences in the frequencies of leg grips and tandems when we compared matched and non-matched combinations.

The long tandem duration in matched islands mating experiments suggested that the males might inseminate more sperm when they mate with the same type of females. In this regard, we tested and analysed the amount of sperm transferred to spermatheca in both mating types. We found that BSS females stored more sperm than AMM ones from the same body size individuals regardless to the male type. Interestingly, BSS females received more than twice higher amount of sperm when they mated with AMM males. Based on findings of Shutt et al (2010) this cannot be explained by higher frequency of re-mating, because cross-injections of MAG extracts between *An. gambiae s.s.* and *An. coluzzii*, also caused monogamy in the mated females. Therefore, it is unlikely for the females in our experiment to mate more than ones in the same night. Moreover, the high sperm insemination of AMM males in spermatheca of BSS females could be due to the BSS spermatheca size and the interaction between seminal fluids and spermatheca receptors that could affect the sperm uptake. Thus, it was identified about 100 MAG genes that include the genes responsible for encoding the proteins and sex peptides that form mating plug and cause after mating behavioural changes in females (Baldini et al. 2012). In addition, it was suggested that the mating plug has a role in correcting the sperm storage by females after the mating process (Baldini et al. 2012). In this regard, it was shown that females mated with males that unable to transfer the plug, failed to store sperm in their spermatheca (Baldini et al. 2012). These results can be supported by further studies about

the seminal products that may interact with the receptors of the females that cause post-mating changes between *An. coluzzii* and *An. gambiae s.s.*.

We tried in this study to unravel the mechanisms of conspecific recognition between the males and females in the incipient species. These differences were found in mating time, swarming activity, resting location in the cages and sperm quantity in the mated females for both recombinants when they mated with the same or opposite type of mosquitoes. Although these results did not give clear explanation about the differences that can lead to assortative mating and reproductive isolation between *An. gambiae s.s.* and *An. coluzzii*, they suggested that the mate rejection takes place before the mosquitoes get into tandem position. This led us to think about the mosquitos' behaviour before tandem stage that most likely involves leg-attaching behaviour or contact pheromones and also this behaviour could be related to harmonizing their wing beats with the nearby mosquitoes. However, the design of the experiment did not enable us to assess the flight tone harmonizing in assortative mating.

The results still need to be supported by further studies that would provide both types of males in the same experiment cages with one type of females to test the harmonizing activity and the receptors interaction on the legs of the mosquitoes. Next studies also could include comparisons amongst laboratory parental strains Mopti and Kisumu, and field strains *An. coluzzii* and *An. gambiae s.s.*. These studies are essential steps for the control programmes that based on releasing sterile or genetically modified males of malaria vector in Africa. These programmes to be successful, lab males have to compete with wild males, locate the swarms and mate with the target females (Manoukis et al. 2014; Diabate et al. 2015). Understanding the mechanisms involved in mating behaviour in these vector species is fundamental in any control programme and could help

in creating alternative control methods to reduce the infection. In addition, it is essential for understanding the sympatric speciation process in other species complexes.

CHAPTER SIX

A role of wing morphology in the wing-tone hypothesis of assortative mating?

6.1. Introduction

An. gambiae complex is undergoing a process of sympatric speciation which makes it a promising model to explain diversification mechanisms (Rundle & Nosil 2005; Nosil et al 2009). The conditions that cause separation and speciation in *An. gambiae* s.l. in general are not fully revealed. It has been suggested that human-made modifications in the environment in Africa created new habitats, and this lead to the mosquitos adaptation to the new regions and environments (Coluzzi et al. 1985; Coluzzi 1999; Coluzzi et al. 2002).

Based on PCR taxonomical tools, a difference was found on the rDNA intergenic spacer on the X chromosome that led them to define and name the M and S molecular forms (Favia et al. 1997; della Torre et al. 2001; Favia et al. 2001). Later, These molecular forms were elevated and named as two sibling species that are *An. coluzzii* for M molecular form and *An. gambiae s.s.* for S molecular form due to the variety of ecological and genomic considerations (Coetzee et al. 2013). Subsequent studies based on Affymetrix GeneChip microarrays analyses revealed high differences in three regions adjacent to centromeres, on chromosomes 2L, 3L and X that were named “speciation islands” (Turner et al. 2005; White et al. 2010). The X island is the most highly diverged and the largest island in *An. gambiae s.s.* and *An. coluzzii* species, which lead to assumption that this island includes the genes responsible for reproductive isolation between both species (Turner et al. 2005; Slotman et al. 2007; Turner & Hahn 2007; White et al. 2010). In contrast, there were low or no differences elsewhere on the genome (Gentile et al. 2001;

Wondji et al. 2002; Turner et al. 2005; Turner & Hahn 2007). The lack of differences in most of the genome is thought to be a result of on-going gene flow between the two species, which leads to homogenization in most regions of the genome (della Torre et al. 2002). Although significant gene flow was observed between the two species, the speciation islands regions were highly different suggesting a 'mosaic genome structure' (Turner et al. 2005; Slotman et al. 2007). This structure was created due to the ecological and behavioral differences between the species that caused divergence in phenotypes that became favoured in different environments (Schluter 2001). The different environments could favour different alleles that could vary hugely between species, despite the gene flow in the rest places of the genome (Schluter 2001). The ecological adaptation genes and the premating isolation genes effects combined to facilitate the speciation with gene flow process, which occur under restricted conditions (Benedict 2003; Coetzee et al. 2013; Diabate & Tripet 2015).

Several studies tried to reveal the ecological and biological roles in the speciation process between *An. gambiae* species (Diabaté et al. 2008). Ecological differences were observed between *An. coluzzii* and *An. gambiae s.s.* in Sahelian regions. In dry seasons, *An. coluzzii* undergo aestivation, which helps it to build up faster when rainy season starts. In contrast, *An. gambiae s.s.* populations migrate away in dry seasons, therefor its population numbers grow up comparatively slower in next rainy seasons (Dao et al. 2014). Furthermore, laboratory studies showed that *An. coluzzii* larvae developed slower than *An. gambiae s.s.* larvae, and females dry weight of *An. coluzzii* was heavier than that of *An. gambiae s.s.* from the same body size (Mouline et al. 2012). This indicates to a higher reservoir of proteins and lipids in *An. coluzzii* adults, which refer to its ability to exploit resources at the larval stages better than *An. gambiae s.s.* (Mouline et al. 2012). In addition, field studies showed that *An. coluzzii* larvae developed faster than *An. gambiae*

s.s. ones when they were transplanted in water collections that include predators (Diabaté et al. 2006; Diabaté et al. 2008). This could lead to a larval divergence between these species (Gimonneau et al. 2010). Moreover, the predation was considered as a selective force for the development of behavioral and adaptive morphological traits in different animals (Lima& Dill 1990; Vamosi 2005). Hence, it can play a role in speciation between *An. gambiae s.s.* and *An. coluzzii* populations (Gimonneau et al. 2010).

Some other studies focused on the role of mating behavior differences between *An. gambiae* sibling species in reproductive isolation (Charlwood et al. 2003; Dao et al. 2008; Diabate & Tripet 2015). In general, *An. gambiae* s.l. individuals mate within swarms that can be formed by the males at dusk, then the females start locating the swarms and join them to start mating process (Charlwood et al. 2003). The ability of the males to recognize the opposite sex by harmonizing wing beats with nearby mosquitoes was reported in different species of *An. gambiae* s.l. (Gibson & Russell 2006; Howell & Knols 2009). The copulating pair usually leaves the swarm to finish the mating process (Howell & Knols 2009). *An. gambiae* sibling species swarm over different markers, for example: *An. coluzzii* individuals swarm over visible markers, whereas *An. gambiae s.s.* mosquitoes usually swarm over bare ground (Diabaté et al. 2009b; Manoukis et al. 2009; Manoukis et al. 2014). Despite the differences in the swarms markers nature, the occurrence of mixed swarms has been reported in Soumousso (a typical savannah village of Burkina Faso) and Valle ́e du Kou (VK7) (Diabaté et al. 2006; Diabaté et al. 2007; Dabire et al. 2013). However, the hybrids frequency was below 1% in the field (della Torre et al. 2005), even though hybrids are fully viable and fertile in laboratory studies (Diabaté et al. 2007). Moreover, the analyzed sperm that collected from mated females in field belonged to the same type of males even in mixed swarms (Tripet et al. 2001). These results suggested extreme assortative mating within species (Tripet et al. 2001; Dabiré et al. 2014). The

frequency of mixed swarms is high compared to the frequency of hybrids in the field (Diabaté et al. 2006). Therefore it was believed that the conspecific recognition within swarms could play a key role in the assortative mating between different species (Diabaté et al. 2006). The assortative mating behavior in turn plays important role as a pre-mating barrier for reproductive isolation between these species (Tripet et al. 2001). However, although there is no evidence for intrinsic post-mating barriers, it is thought there might be a lack in the hybrids fitness in different ecological zones, which leads to post-mating isolation (Weetman et al. 2012; Lee et al. 2013).

The conspecific recognition within swarms was considered one of the main factors affecting assortative mating between *An. gambiae s.s.* and *An. coluzzii* populations. Different types of conspecific recognition mechanisms were proposed in different species of mosquitoes. The pheromones are one of the methods that were used for conspecific recognition in some species of mosquitoes, but it has not been proved in *An. gambiae s.s.* and *An. coluzzii* (Diabate & Tripet 2015). Therefore, there were more attempts to focus on flight tone hypothesis in these species to explain the reproductive isolation between them especially in the mixed swarms (Diabate & Tripet 2015).

Hearing organ in mosquitoes includes two parts: Johnston's organ which is sensory organ located in the second segment of the mosquito antenna, and the antennae that are movement receivers that respond to oscillations of air particles in the sound field (Göpfert et al. 1999). The sound oscillations cause hair vibrations that are transmitted to the flagellar shaft and thus to the Johnston's organ at the base of the flagellum (Roth 1948). In the males, the antennae hold many long hairs; and it is known to be plumose. Therefore, the sensitivity of the male antennae exceeds the sensitivity of the female ones. The best frequency of the male antenna is around 380Hz, which corresponds to the frequency of female flight tones. However, the best frequency of the female antenna is around

230Hz(Göpfert et al. 1999). These auditory organs help the male and female mosquitoes match their flight tone frequencies. Hence, the mosquitoes use auditory interactions for the sex recognition (Charlwood & Jones 1979; Clements 1992). Similarly, The flight tone hypothesis was suggested to explain the conspecific recognition mechanism between the different species (Tripet et al. 2004). In this regard, several studies tried to measure and analyse the wing-beat frequencies that could cause different wing tones. There was no difference detected when the measurements were done separately for the individuals of *An. gambiae s.s.* and *An. coluzzii* (Tripet et al. 2004; Gibson et al. 2010). However, when the measurements of flight tone frequencies were done for male-female pairs for the same and different species, the mosquitoes start matching their flight tones significantly in the same species pairs more than the different species pairs (Pennetier et al. 2010).

On the other hand, the differences in flight tones could be also due to wing shape differences in *An. gambiae s.s.* and *An. coluzzii* species (Sanford et al. 2011). Sanford (2011) did her study in Mali and Guinea-Bissau in Africa. These areas differ with respect to the level of assortative mating between both species. In Mali, where the species experience strong assortative mating with low frequency of hybrids, the wing lengths and widths were larger than those in Guinea-Bissau. Moreover, the wing widths differed significantly between *An. gambiae s.s.* and *An. coluzzii* in Mali. However, In Guinea-Bissau, where the assortative mating is not strong with comparatively higher frequency of hybrids, the wing lengths and widths did not present any differences between *An. gambiae s.s.* and *An. coluzzii*. The wing lengths of the hybrids in Guinea-Bissau area were similar to *An. coluzzii* whereas the wing widths were similar to *An. gambiae s.s.*. The differences in the wing widths between both species in Mali explained the strong assortative mating comparing to Guinea-Bissau. The differences in the wings shape could produce different wing tones that attract conspecific mate in the swarms. These findings supported the wing

tones hypothesis that leads to a strong assortative mating between *An. gambiae s.s.* and *An. coluzzii* species in some areas of Africa (Sanford et al. 2011). Therefore, several studies reported the importance of harmonic divergence as a conspecific recognition mechanism in *An. gambiae s.s.* and *An. coluzzii* (Tripet et al. 2004; Pennetier et al. 2010; Cator et al. 2010). This led us to hypothesize that the different species could have different wings size or shape.

In this study, we examine the wing-shape differences and its role in the wing-tones hypothesis in *An. coluzzii* and *An. gambiae s.s.* species, that play the main role in reproductive isolation between these species. For this purpose, we created recombinants by introgressing the island of speciation located in pericentric region of X chromosome of S molecular form (Kisumu from Kenya) into M molecular form (Mopti from Mali) through 5 generations of backcrossing followed by two generations of crosses to fix the X island (Chapter 2). We obtained two recombinant strains that are either M or S of X-island of speciation. In previous experiments (Chapter 3,4), we showed that these recombinants mated assortatively with matched type of X island. This confirmed that the genes responsible for assortative mating are located in X-island of divergence. Their function has not been revealed yet but they could play a role in encoding morphological traits that help in conspecific recognition between the species. The main aim for the present study is finding the differences between the parental and recombinant strains wing size by analysing the wing widths and lengths using JMP. In addition, analysing different landmarks on their wings using the powerful tools available in ImageJ software and the morphological analyses program MorphoJ to compare the subtle wing-shape differences between the recombinant and parental strains.

The results were discussed in relation to recent recognition mechanisms hypothesis and speciation process in *An. coluzzii* and *An. gambiae s.s.* species. Understanding the

mechanisms involved in mating recognition in these main vector species is important in any control programme to reduce malaria and could help in creating alternative control strategies. In addition, it is an important step to pave the way for understanding the sympatric speciation process in other species.

6.2. Materials and Methods

The measurements and analyses were made for wings that were taken from females and males from the recombinant and parental strains. The first set of analyses focused on comparing the wing widths, lengths and their ratio amongst the different strains using simple ANOVAs. A second set of analysis was conducted by identifying different landmarks on the veins of the wings, and elaborate morphometric analyses.

6.2.1. Preparing the wings for studies

Thirty females and thirty males were collected from the strains: AMM, BSS, Kisumu and Mopti, and the wings were prepared for both sets of analyses. Every group that includes thirty mosquitos was placed in 1.5ml tubes (Fisher brand, Leicestershire, UK) that include 1ml of 10% Potassium hydroxide (Sigma-Aldrich Company Ltd., Dorset, UK). After 1 hour, every group was placed in 1.5ml tubes include 1ml of deionised water (dH2O) for 15 minutes. After that, the mosquitoes were transferred to 1.5ml tubes include 1ml of 80% Ethanol (Sigma-Aldrich Company Ltd., Dorset, UK). Thereafter, wings were removed carefully from mosquitoes with a mesonotum and mesopleuron parts of the thorax to handle the wings easily without damaging them. The wings were placed on a microscope slides without coverslip. Digital pictures were taken under an Olympus (SZ-PT) dissecting microscope equipped with an Olympus (model E-520) camera.

6.2.2. Comparisons wing length and width

Wing measurements were taken using ImageJ software (Thévenaz 2003). For this purpose, wing length was measured as the distance from posterior anal cell margin to the tip of radial vein 3 (R3) (Sanford et al. 2011) (Fig. 6.1). Wing width was measured as the distance from sub-costal junction with costa, at the leading edge to the wing, to half way between junctions of anterior branches of cubitus veins CuA1 and CuA2 on the trailing edge of the wing (Sanford et al. 2011) (Fig. 6.1). Only one wing from either side of the mosquito is required for the analysis.

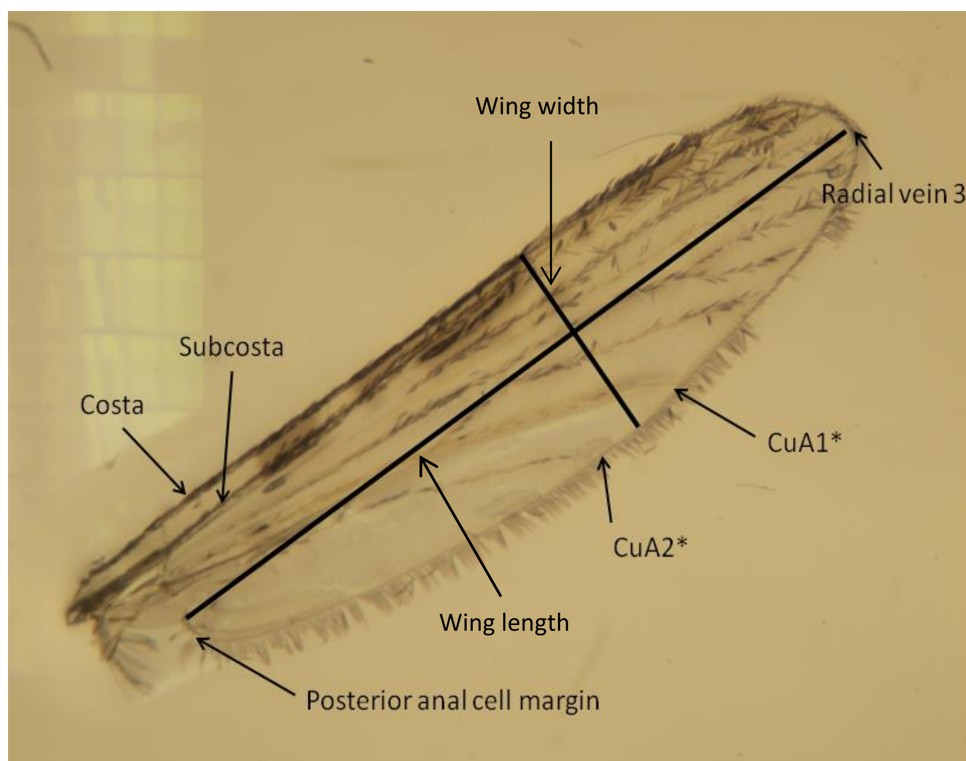


Figure 6.1. The wing measurements made by photomicroscopy and imaging – The wing length is the distance from the posterior anal cell margin to the tip of radial vein 3, and width is the distance from the sub-costal junction with the costa, at the leading edge to the wing, to half way between the junctions of the anterior branches of the cubitus veins CuA1 and CuA2 on the trailing edge of the wing.

Wing length and width analyses

The wing length, width and width/length ratio were measured using ImageJ (Thévenaz 2003), and compared between AMM, BSS, Kisumu and Mopti strains. The ANOVAs analyses were conducted using the JMP 11.0 (SAS Institute, Inc).

6.2.3. Detailed wing morphometry study

Defining landmarks on the wings

Three lines (a,b,c) were added to the photos using Power Point software to create additional landmarks on the wings (Fig. 6.2). The first line (a) runs along the Costa of the wing. The second line (b) is perpendicular to line (a) and crosses the junction between Cubitus1 and Cubitus2. The third line (c) started from the junction between Costa and Sub-costa and crosses the junction point between Cubitus1 and Cubitus2. To convert the landmarks into X, Y coordinates; the 'point picker' plugin was used in the ImageJ software (Thévenaz 2003). The total of 21 landmarks were recorded using existing feature of the wing and junctions of the lines with the wing veins (e.g. Cubitus1, Cubitus2, Costa, Sub-costa, Radius1, Radius2…) (Fig. 6.2).

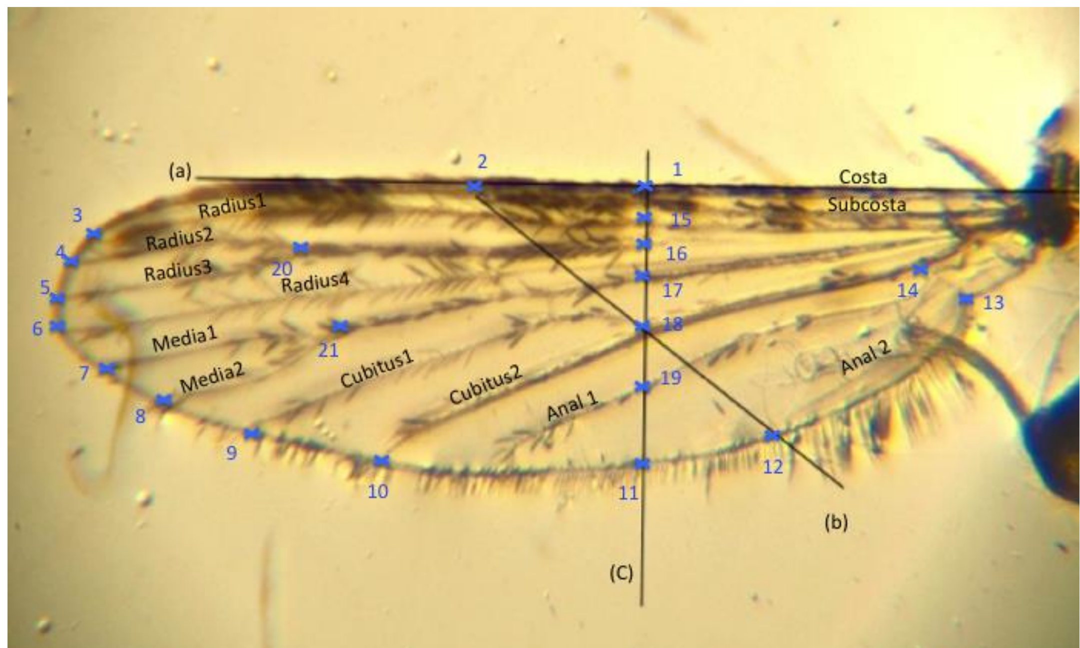


Figure 6.2. The wings landmarks determination - The landmarks were defined on the veins and angles respectively for the recombinant and parental strains.

Formatting landmark coordinates files for MorphoJ

Two files were required for MorphoJ analyses, the first one was prepared by translating the landmarks into X, Y coordinates using "Point Picker" plugin in the ImageJ software. Thereafter, the data were organized as TXT file included the wing ID. The file was formatted as Thin-plate spline (TPS) format which is a mathematical model that produce a grid graph that shows the changes of shape (Sheets 2010). The individual males and females were given continues names to become compatible with MorphoJ software (Table 6.1). The second file is the identifier TXT file that included the ID, group and sex of the mosquitoes (Table 6.2).

Table 6.1. The 21 wings landmarks were translated into coordinates X, Y and formatted in TPS format for using MorphoJ – The X and Y coordinates of each landmark are listed in the second and third columns, respectively. LM is the number of landmarks, and ID is the individual name.

	X	Y
	LM=21	
1	400	184
2	286	186
3	69	212
4	49	232
5	43	244
6	40	264
7	60	287
8	117	305
9	161	316
10	239	324
11	395	320
12	476	306
13	594	235
14	579	224
15	399	198
16	399	214
17	398	233
18	397	258
19	397	288
20	211	220
21	223	260
	ID=Moptif1	
	LM=21	

Table 6.2. The identifier file was required for MorphoJ – Including the ID of the individuals and their group and sex.

ID	Group	Sex
Mopti1	Mopti	Female
Mopti2	Mopti	Female
Mopti3	Mopti	Female
…Table continued	…	…
Mopti31	Mopti	Male
Mopti32	Mopti	Male
Mopti33	Mopti	Male
…Table continued	…	…
Kisumu1	Kisumu	Female
Kisumu2	Kisumu	Female
Kisumu3	Kisumu	Female
…Table continued	…	…
Kisumu31	Kisumu	Male
Kisumu32	Kisumu	Male
Kisumu33	Kisumu	Male
…Table continued	…	…
AMM1	AMM	Female
AMM2	AMM	Female
AMM3	AMM	Female
…Table continued	…	…
AMM31	AMM	Male
AMM32	AMM	Male
AMM33	AMM	Male
…Table continued	…	…
BSS1	BSS	Female
BSS2	BSS	Female
BSS3	BSS	Female
…Table continued	…	…
BSS31	BSS	Male
BSS32	BSS	Male
BSS33	BSS	Male
…Table continued	…	…

The wing morphometric analyses

The statistical analyses of the shape were conducted using Geometric morphometric (MorphoJ) that has been used succefully in previous studies (Klingenberg 2011; Jaramillo-O et al. 2015). The measurements file was opened by MorphoJ software and the identifier file was imported. A new database was created and subdivided into females and males datasets, and then subdivided into Mopti, Kisumu, AMM and BSS groups. Analysing the shape data included three steps.

a) A Procrustes fit was performed for both male and female groups and for the strain groups to align the landmarks and center them on the same point. The differences in size were removed and the landmarks were rotated until they fit together into a common position in this step. This operation is called superimposition that includes calculating the centroid of each configuration, and then making the centroid the origin of a new coordinate system, e.g.: to remove the differences in orientation between two configurations, one configuration that is called target form was rotated around its centroid until it shows minimal offset in location relative to the other configuration that is called the reference form (Sheets 2010). After this step is achieved, any differences between configurations by the next analyses must be the result of differences in shape between them (Sheets 2010).

b) Covariance matrix was created for both male and female groups and for each strain group. This step is important for doing the Principal components analyses.

c) Principal component analysis was conducted to test main varied features of the shape. Discriminant variation analyses were conducted, included Discriminant scores for the identified data, and Cross-validation test by omitting each observation one at time, recalculating the classification function using the remaining data, and then classifying the omitted observation. A Canonical variation analysis was also performed to find the shape changes role as a discriminant between the groups (SAS Institute Inc 2008).

6.3. Results

6.3.1. Wing length and width ANOVAs

6.3.1.1. Dimorphism between females and males

The effects of the sex, strain and the interaction between them were tested on the wing length, width and width/length ratio. The female wings were bigger than the males regardless of the strain. There was a significant difference also between the different strains wing width and the width/length ratio. In some cases, the strain showed indirect effect on the wing width and width/length ratio through the interaction with the sex type of the mosquitoes (Table 6.3).

Table 6.3. 2-way ANOVA of the effect of sex and strain and the interaction between them on the length, width and width/length ratio in recombinants and parental strains

Variable	Source	*df*	*F*-ratio	*P*-value
Wings length	Sex	1	13.51	<0.001
	Strains	3	1.55	0.203
Wings width	Sex	1	122.8	< 0.001
	Strains	3	6.96	<0.001
	Strains*sex	3	3.14	0.026
Width/length	Sex	1	158.45	< 0.001
	Strains	3	8.77	<0.001
	Strains*sex	3	6.56	<0.001

6.3.1.2. Variation between strains

Within the females group, there were no significant statistical differences in the wing length (ANOVA: n= 120, $F_{3,116}$= 1.56, P= 0.201), also in the male group there were no significant statistical differences in the wing length (ANOVA: n= 120, $F_{3,116}$= 0.416, P= 0.742) among the different strains. The females group analyses also showed that there is no significant difference in the wing width (ANOVA: n= 120, $F_{3,116}$= 1.40, P= 0.244), and width/length ratio (ANOVA: n= 120, $F_{3,116}$= 0.27, P= 0.847), whereas in the males group

there was a significant difference in the wing width (ANOVA: n= 120, $F_{3,116}$= 9.34, P> 0.001), and the width/length ratio (ANOVA: n= 120, $F_{3,116}$= 16.03, P>0.001) amongst AMM, BSS, Mopti and Kisumu individuals.

These results explained the interaction between strain and sex in dimorphism analyses and showed that Kisumu male wings were on average wider than AMM and Mopti wings (Tukey: P< 0.01). This difference appeared also in width/length ratio of Kisumu wings that were bigger than Mopti and AMM wings (Tukey: P< 0.001) (Fig. 6.3) (Table 6.4). However, BSS males wing-width was in the average between Mopti, AMM and Kisumu wing-width size and did not differ significantly from these strains wings widths. In conclusion, Kisumu males wing width ≥ BSS males wing width≥ AMM males wing width ≈ Mopti males wing width (Fig. 6.3)(Table 6.4).

Table 6.4. The raw data ±95% confidence intervals values for length, width and width/length ratio for males and females wings of recombinants and parental strains.

		Female	
Strain	**Length (mm)**	**Width (mm)**	**Width/Length**
Mopti	2.91 (2.84-2.98)	0.79 (0.76-0.81)	0.27 (0.27-0.27)
Kisumu	2.93 (2.88-2.98)	0.8 (0.78-0.87)	0.27 (0.27-0.28)
AMM	2.87 (2.81-2.93)	0.78 (0.76-0.8)	0.27 (0.27-0.28)
BSS	2.96 (2.89-3.03)	0.81 (0.79-0.83)	0.27 (0.27-0.28)
		Male	
Strain	**Length**	**Width**	**Width/Length**
Mopti	2.82 (2.75-2.9)	0.7 (0.67-0.72)	0.25 (0.24-0.25)
Kisumu	2.86 (2.82-2.91)	0.76 (0.74-0.78)	0.26 (0.26-0.27)
AMM	2.82 (2.77-2.88)	0.71 (0.69-0.72)	0.25 (0.25-0.25)
BSS	2.85 (2.78-2.92)	0.72 (0.70-0.74)	0.25 (0.25-0.26)

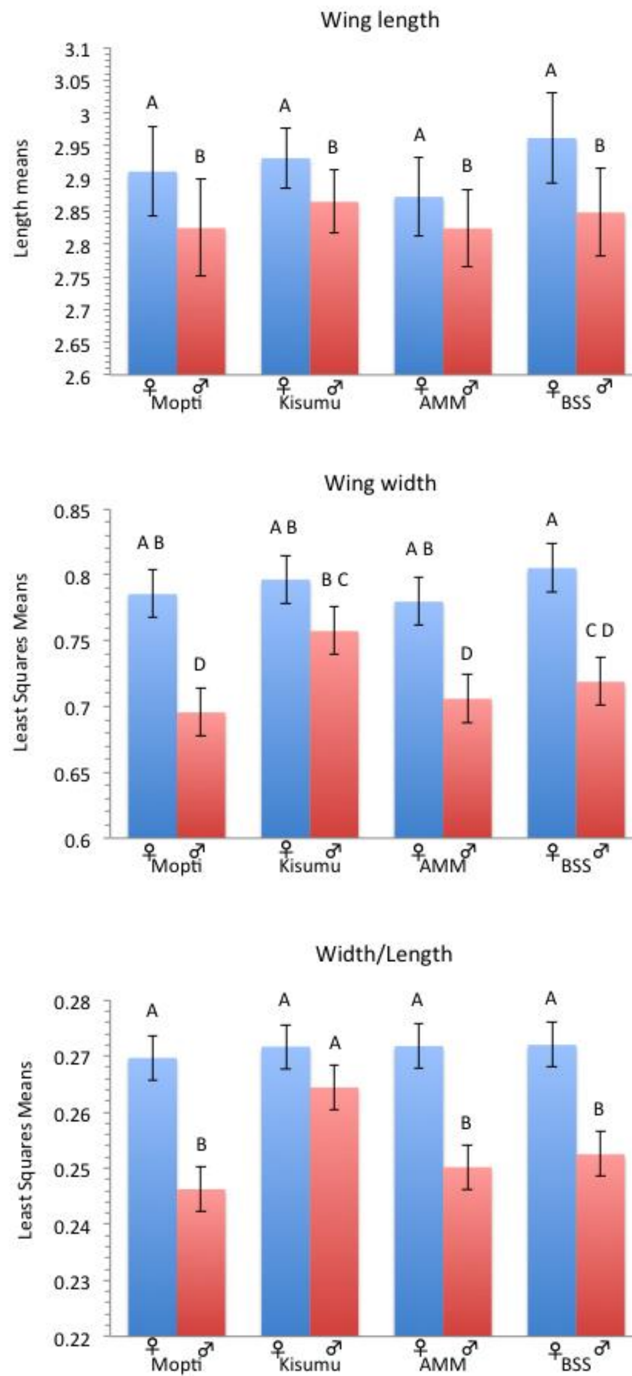


Figure 6.3. Comparison of least squares means between the strains – The differences of the wings length, width and width/length ratio least squares means between recombinants and parental colonies. A,B,C and D groups are significantly different from each other; (e.g: AB group is not significantly different from both A and B groups, whereas BC is not different from both B and C groups…).

6.3.2. Morphometrics analyses of wings

6.3.2.1. Dimorphism between females and males

Principal components analyses (PCA) within the strains groups

Principal components analyses (PCA) were conducted to find the most important axes in the space that explain the maximal variation of the shape. This analysis would show the similarities and differences between the landmarks in "morphospace". This procedure explores the outliers and reveals the clusters of groups with a similar morphology (Sheets 2010).

The most important PCs were PC1 that explained more than 50% of total variance, PC2 that explained more than 10% of total variance and PC3 that explained more than 5% of total variance between the females and males within the strain groups (Table 6.5) (Fig. 6.4). These axes provided information about the changes of the shape over the total set of landmarks, and gave graphs that showed the changes in landmarks positions (Sheets 2010). The scatterplot graphs showed that the females and males groups can appear separately based on the changes of the shape for PC1 versus PC2 in all the strains (Fig. 6.4).

Table 6.5. The first three PCs that contributed for the wings shape variance between females and males within the strains – The eigenvalues showed the amount of variance shared the linear combination of variables, so the first one explain the majority of variance in the relationship. The percentage of variance shows the importance of the PCA. The Cumulative percentage shows the percentage of variance as each function is added to the table (Maydeu-Olivares & McArdle 2005).

Strain	PCs	Eigenvalues	% Variance	% Cumulative
Mopti	1	$10.3426e^{-4}$	57.715	57.715
	2	$1.8763e^{-4}$	10.47	68.185
	3	$1.2841e^{-4}$	7.166	75.351
Kisumu	1	$9.0362e^{-4}$	55.978	55.978
	2	$2.2578e^{-4}$	13.987	69.965
	3	$1.356e^{-4}$	8.4	78.365
AMM	1	$8.2197e^{-4}$	52.248	52.248
	2	$1.7914e^{-4}$	11.387	63.635
	3	$1.4744e^{-4}$	9.372	73.007
BSS	1	$9.0319e^{-4}$	55.747	55.747
	2	$2.3164e^{-4}$	14.297	70.045
	3	$1.414e^{-4}$	8.728	78.772

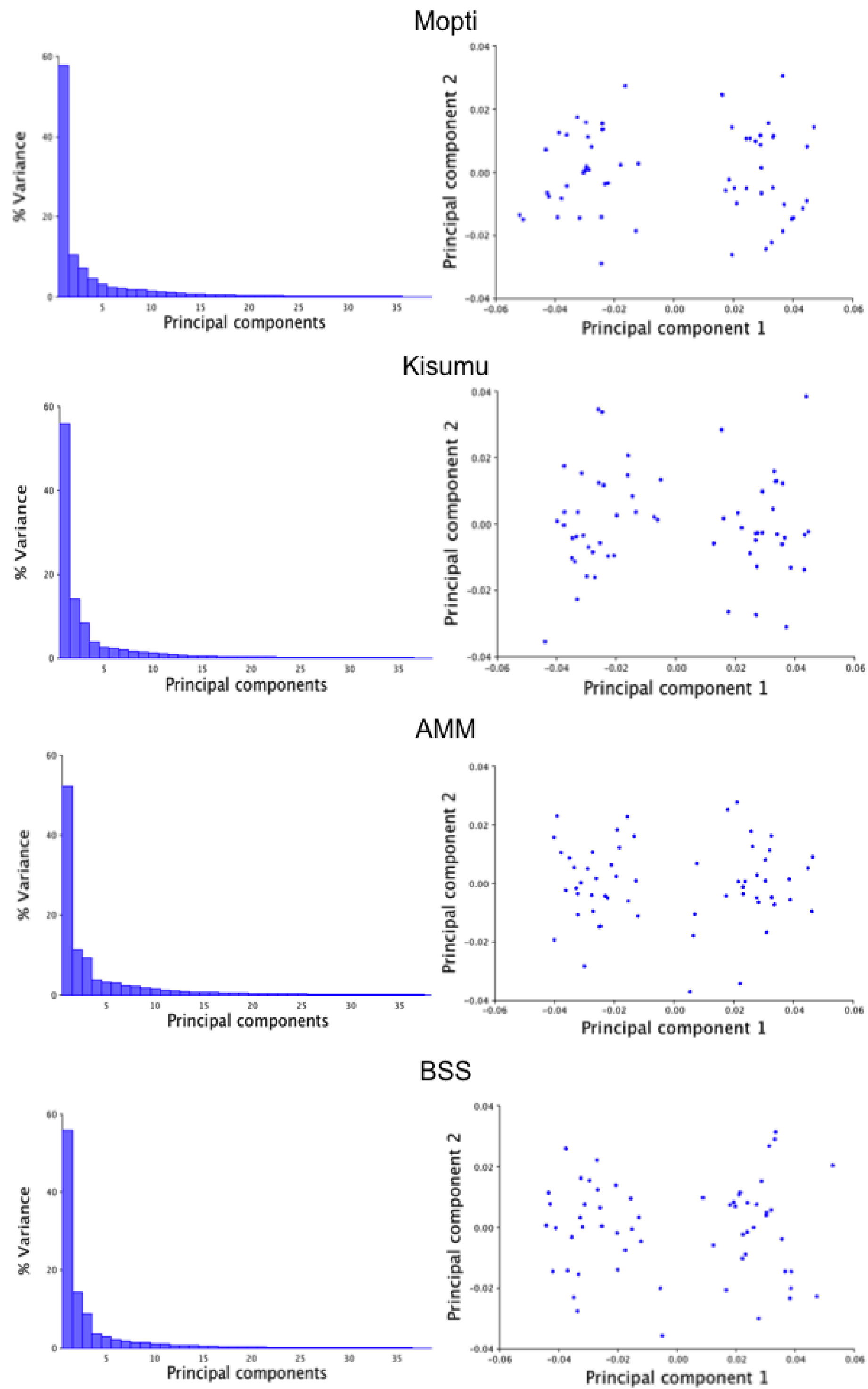


Figure 6.4. Percentage of variance explained by principal components, and the scatterplot graphs explained the results of PCA in PC1 versus PC2 showing the variance between the females and males for recombinant and parental strains.

The PCA showed that in Mopti strain the landmarks that contributed to principal component were 2, 20 and 21 landmarks for PC1, and 2, 12 and 14 landmarks for PC2, and 2, 20 and 21 landmarks for PC3. The PCA for Kisumu strain showed that the landmarks that contributed to principal component were 13, 20 and 21 landmarks for PC1, and 2, 20 and 21 landmarks for PC2, and 12, 13 and 14 landmarks for PC3. The PCA for AMM strain showed that, the landmarks that contributed to principal component were 13, 20 and 21 landmarks for PC1, and 12, 13 and 14 landmarks for PC2, and 2, 20 and 21 landmarks for PC3. The PCA for BSS strain showed that, the landmarks that contributed to principal component were 14, 20 and 21 landmarks for PC1, and 2, 12 and 14 landmarks for PC2, and 2, 20 and 21 landmarks for PC3 (Fig. 6.5).

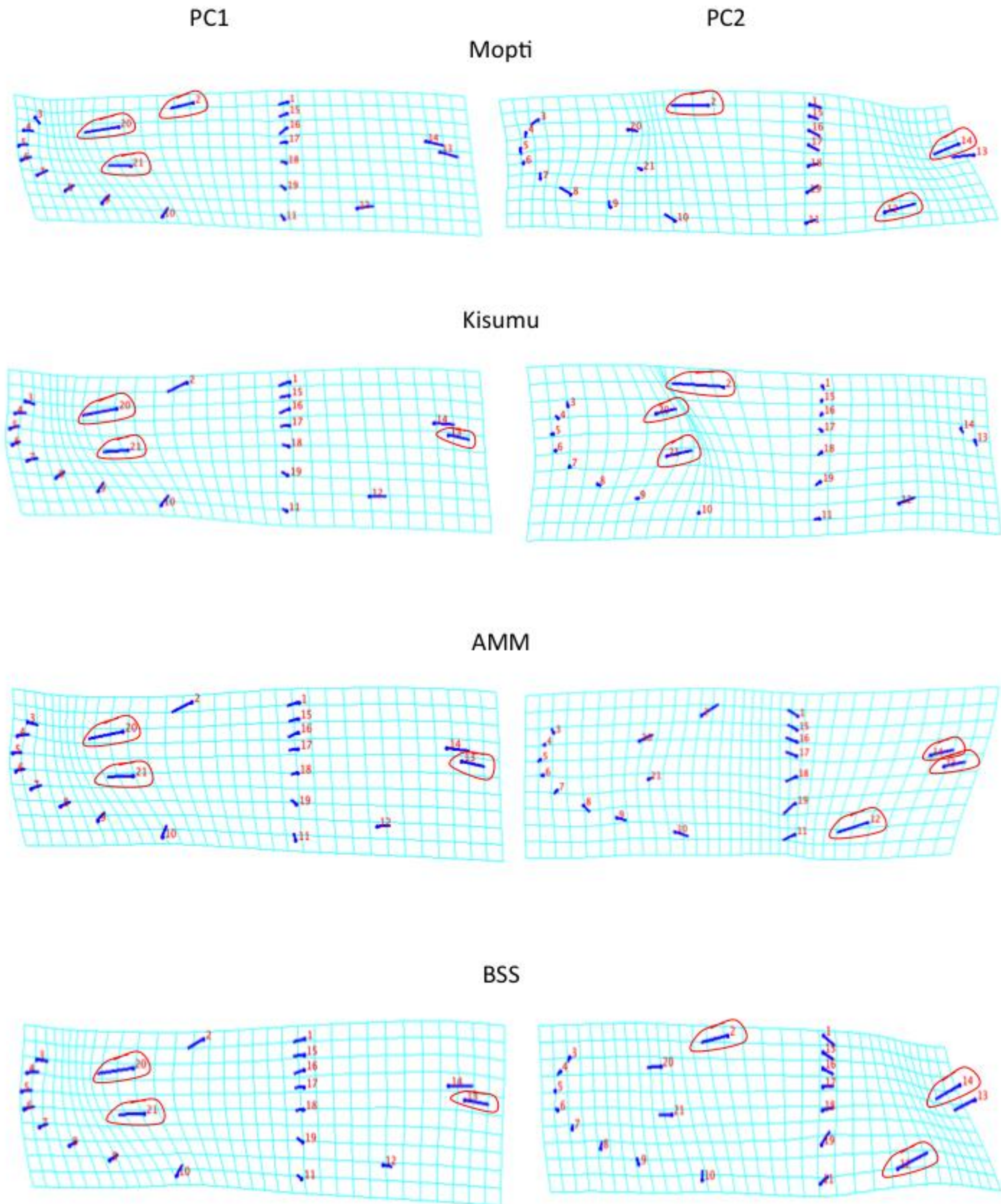


Figure 6.5. Thin-plate spline deformation grid of shape variation along PC1 and PC2 for the sexual dimorphisms within the strains - Three landmarks whose variance contributed more for PC1 and PC2 are highlighted for each recombinant and parental strains.

Pairwise discriminant analyses between the females and males within strains

Comparison between females and males was conducted using discriminate function analyses to predict mosquitoes sex on the basis of morphological measurements within each strain. The Mahalanobis distance -that measures the variation among groups relative to within-sample variation (Claude 2009; Oksanen et al. 2011)- was high between the female and male groups within the strains in BSS and Mopti, followed by AMM respectively, whereas there was comparatively low Mahalanobis distance and higher P-value between the females and males in Kisumu strain which could be due to the high variation between Kisumu individuals and existence of some outliers (Table 6.6) (Fig. 6.7). However, the means for the variables were significantly different between the females and males within strains, therefore the shape change can separate them when it was given identified females and males groups (Table 6.6) (Fig. 6.6). On the other hand, the Cross-validation scores showed that the shape could not identify the omitted individuals sex based on shape features and differences between females and males groups. Therefore, the shape differences cannot be a discriminant between the females and males groups (Fig. 6.6).

Table 6.6. Discriminant components for the variance between females and males within the groups - Mahalanobis distances measures the variation between groups relative to the within-sample variation. Procrustes distance measured the distance between means of the two data sets.

Females-Males comparisons				
Groups	**Procrustes distance**	**Mahalanobis distance**	***T*-square**	***P*-value**
Mopti	$610.8113e^{-4}$	13.3318	2666.0531	<0.001
Kisumu	$313.3773e^{-4}$	4.1842	262.6123	0.014
AMM	$535.6626e^{-4}$	11.6071	2020.8691	<0.001
BSS	$560.7382e^{-4}$	13.5562	2756.5591	<0.001

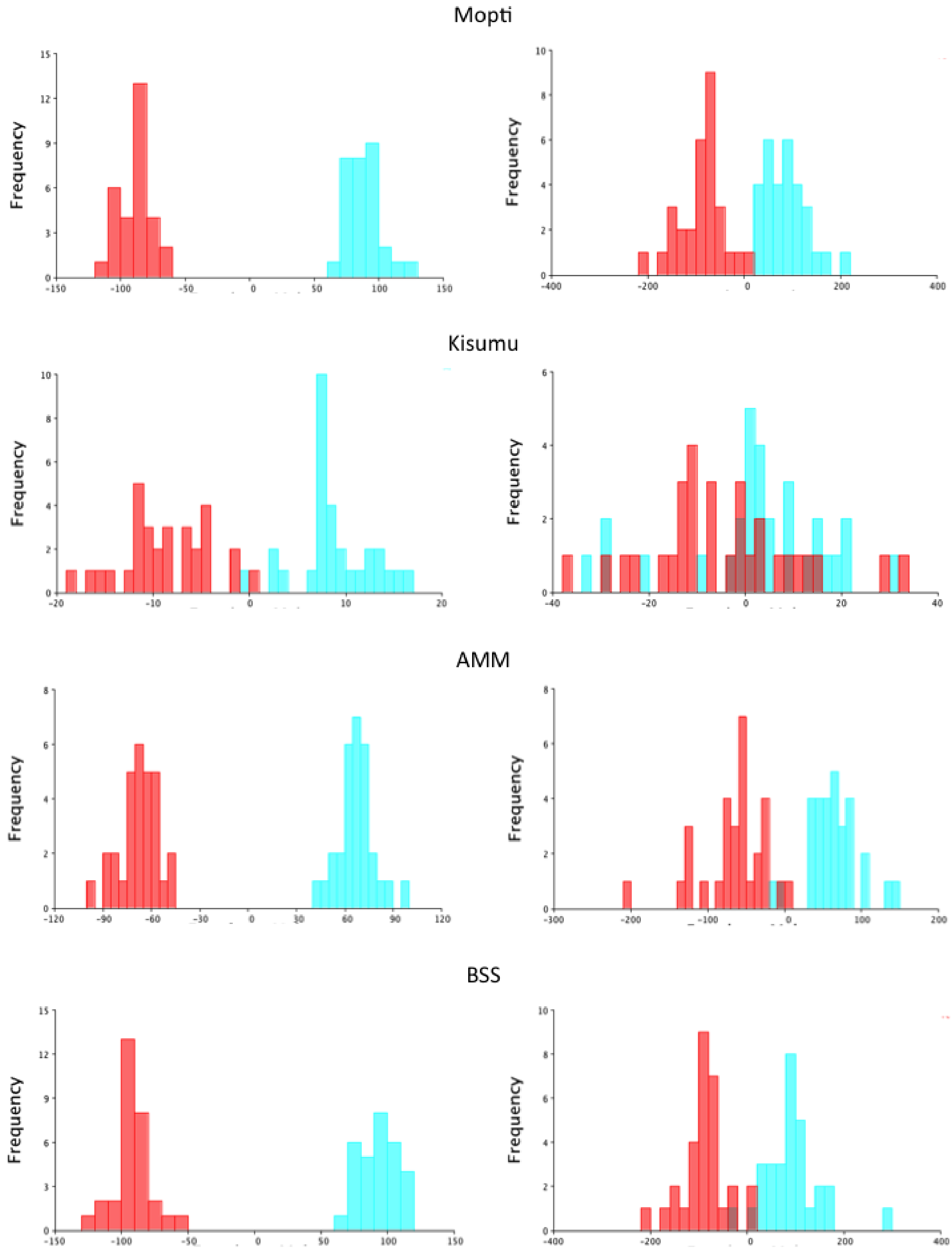


Figure 6.6. Pairwise comparison in the strains groups – the left figures show frequency of males and females groups that were descriminated and separated based on the wing discribtion that were given to the software. The right figures show discriminant scores for all data and cross-validation when omitted each observation one at time and recalculate the classification function using the remaining data for the males and females.

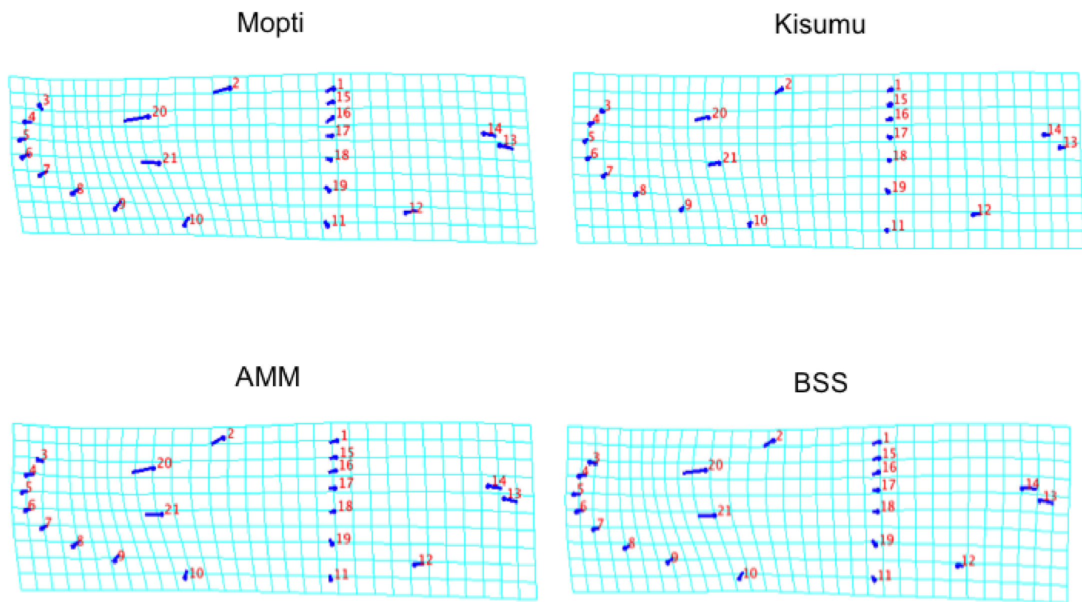


Figure 6.7. Thin-plate spline deformation grid for discriminant analyses showing sexual dimorphism within the recombinant and parental strains.

6.3.2.2. Variation between strains

Principals components analyses (PCA) for all the strains within the sexes

Principal components analyses (PCA) were conducted to find most important axes in the space that explain the maximal variation of shape. The scatterplot graphs did not show defined distribution for the different strains based on the changes of shape for PC1 versus PC2 (Fig. 6.8).

The most important PCs were the PC1 that explained more than 25% of total variance, PC2 that explained more than 15% of total variance and PC3 that explained more than 10% of total variance amongst the strains in the females and males groups (Table 6.7) (Fig. 6.8).

Table 6.7. The first three PCs that contributed for wings shape variance between the strains within the females and males groups – The eigenvalues showed the amount of variance shared the linear combination of variables, so the first one explains the majority of variance in the relationship. The percentage of variance shows the importance of the PCA. The Cumulative percentage shows the percentage of variance as each function is added to the table (Maydeu-Olivares & McArdle 2005).

Sex	PCs	Eigenvalues	% Variance	% Cumulative
Females	1	$2.76e^{-4}$	25.07	25.07
	2	$2.101e^{-4}$	19.087	44.157
	3	$1.5653e^{-4}$	14.221	58.378
Males	1	$3.152e^{-4}$	29.111	29.111
	2	$1.9717e^{-4}$	18.21	47.322
	3	$1.4068e^{-4}$	12.993	60.315

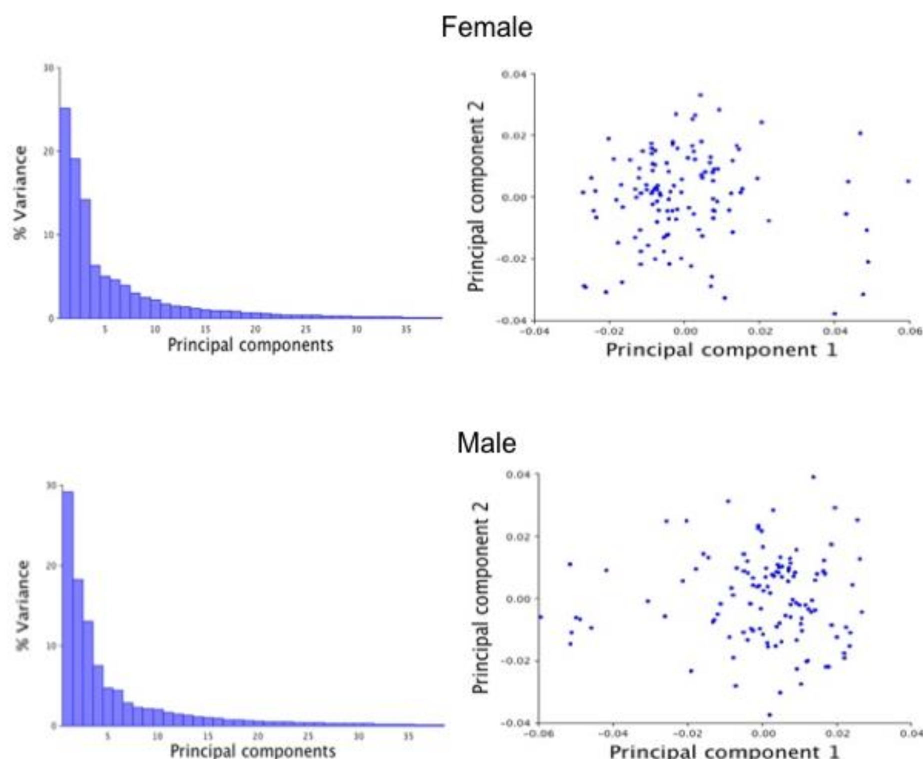


Figure 6.8. Percentage of variance explained by principal components, and the scatterplot graphs explained the results of PCA in PC1 versus PC2 - showing the variance between the recombinants and parental strains in females and males groups.

The PCA showed that within the females group landmarks that contributed to principal components were 2, 14 and 20 landmarks for PC1, and 2,12 and 21 landmarks for PC2, and 14, 20 and 21 for PC3 (Fig. 6.9). Within the males, the landmarks that contributed to principal components were 13, 14 and 20 landmarks for PC1, and 2, 12 and 14 landmarks for PC2, and 2, 20 and 21 for PC3 (Fig. 6.9).

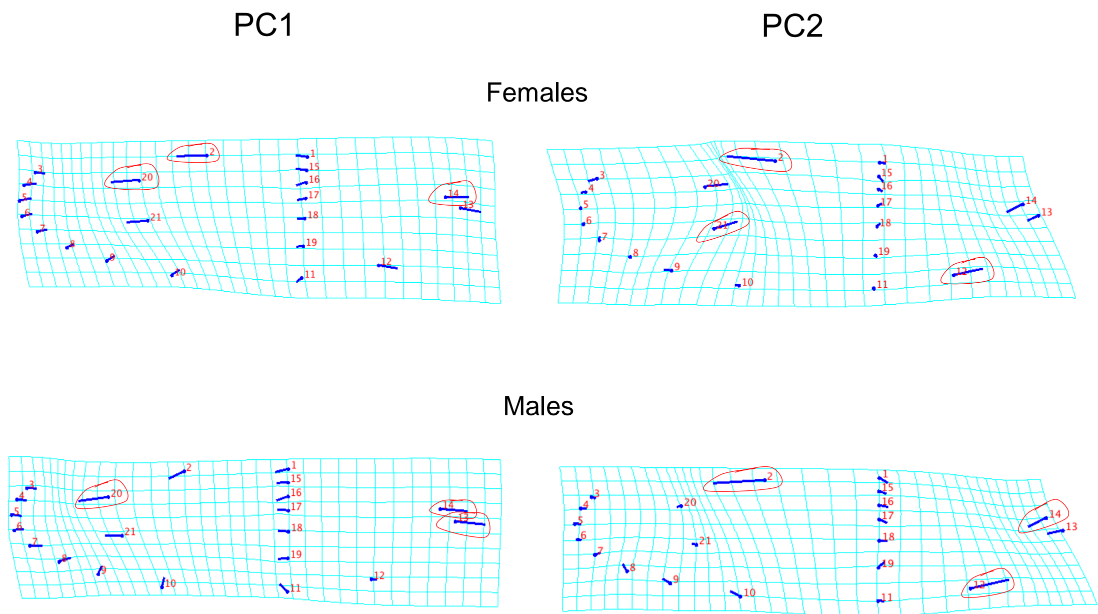


Figure 6.9. Thin-plate spline deformation grid showed the shape variation along PC1 and PC2 for the strains within the females and males groups - Three landmarks whose variance contributed more for PC1 and PC2 are highlighted for recombinant and parental strains.

Centroid and procrustes ANOVA analysis within females and males

In centroid size analyses, no significant difference was found amongst strains within the females and males. Centroid ANOVA analyses revealed no difference in mean centroid position between the strains in females group (ANOVA: $F_{3,116}$= 1.26, P= 0.291) and males group (ANOVA: $F_{3,116}$= 1.51, P= 0.216).

Canonical variation analyses amongst strains within females and males

The canonical variation analysis was performed within females and males groups to classify the mosquitoes within their strains based on differences in the wing shapes. In this test, each CV is oriented to describe the maximal differences between groups relative to variation within groups (Sheets 2010). The shape variance discriminate the mosquitoes in different strains for CV1 and CV2 axis that showed the maximal variation between the mosquitoes shape (Fig. 6.10). However, there is a cross distribution between the strains within the females group more than their distribution in the males group.

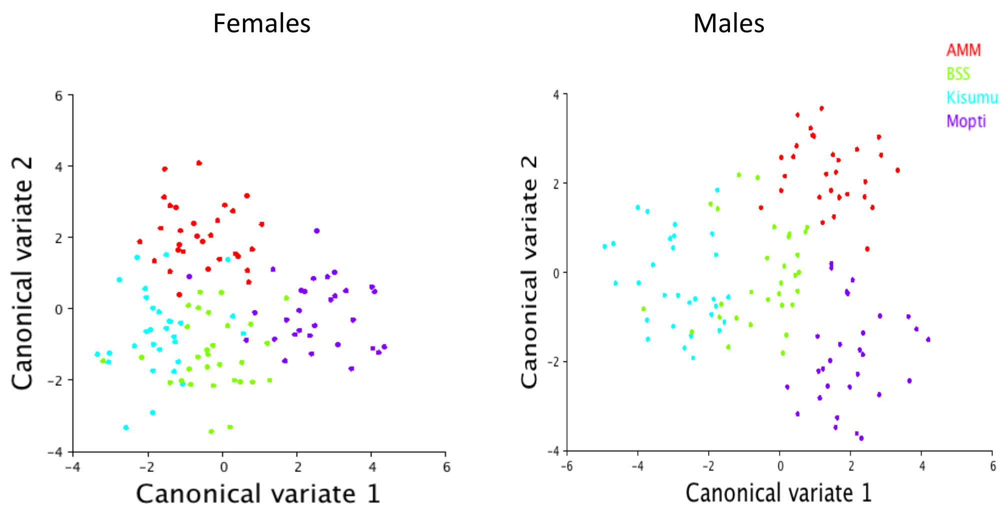


Figure 6.10. The scatterplot graphs explained the results of distinct canonical variates giving the mean shape differences between the recombinant and parental colonies within the females and males groups – CV1 versus CV2 showing the mean variance between the recombinant and parental strains in the females and males groups.

Pairwise discriminant analyses between strains within females and males

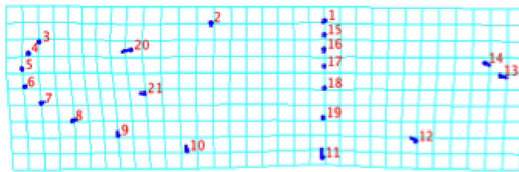
Comparison between each two strains was conducted using discriminate function analyses within the males and females groups to predict the mosquito strain on the basis of morphological measurements. The means for the variables were significantly different between each two strains within the females and males groups in the given identified strains. Therefore the shape changes discriminate the strains within the females and males groups (Table 6.8) (Fig. 6.11 & 6.12). However, the Cross-validation analysis showed that there is a cross distribution between the strains that could not be recognized based on the shape differences within the females and males groups (Fig. 6.13& 6.14). Therefore, the shape differences cannot discriminate the individuals and relate them to the strains based on the shape features when we do not give identified individuals.

Table 6.8. Discriminant components for the variance between the strains within the females and males groups - Mahalanobis distances measured the variation between groups relative to the within-sample variation. Procrustes distance measured the distance between means of the two data sets.

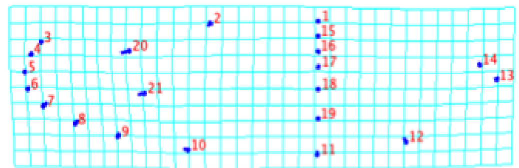
Females				
Groups	**Procrustes distance**	**Mahalanobis distance**	***T*-square**	***P*-value**
Kisumu - Mopti	$220.7056e^{-4}$	5.3093	422.8262	0.0006
AMM - BSS	$140.3922e^{-4}$	5.3012	421.5438	0.0006
AMM - Mopti	$169.4824e^{-4}$	4.9148	362.3337	0.0018
AMM - Kisumu	$170.6714e^{-4}$	4.4085	291.5192	0.0075
BSS - Mopti	$203.1499e^{-4}$	4.7676	340.9478	0.0027
BSS - Kisumu	$207.0334e^{-4}$	3.7315	208.857	0.0478

Males				
Groups	**Procrustes distance**	**Mahalanobis distance**	***T*-square**	***P*-value**
Kisumu - Mopti	$254.6907e^{-4}$	7.6305	873.3665	0.0001
AMM - BSS	$154.6031e^{-4}$	4.4525	297.3714	0.0066
AMM - Mopti	$154.8653e^{-4}$	6.2845	592.4304	0.0001
AMM - Kisumu	$205.1941e^{-4}$	10.0101	1503.0309	0.0001
BSS - Mopti	$167.1003e^{-4}$	5.9758	535.6481	0.0001
BSS - Kisumu	$154.3522e^{-4}$	6.1837	573.5778	0.0001

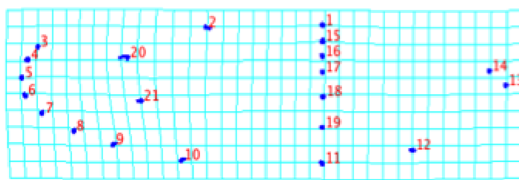
Mopti - Kisumu



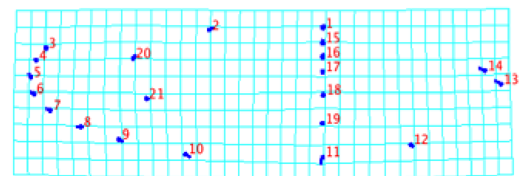
AMM - BSS



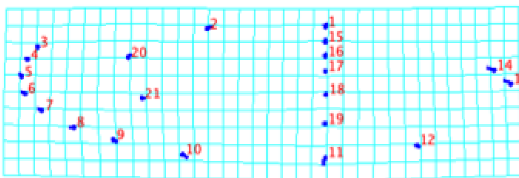
Mopti - AMM



Mopti - BSS



Kisumu - AMM



Kisumu - BSS

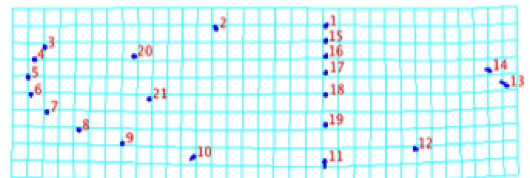


Figure 6.11. Thin-plate spline deformation grid showed the shape variation for the strains within the females group.

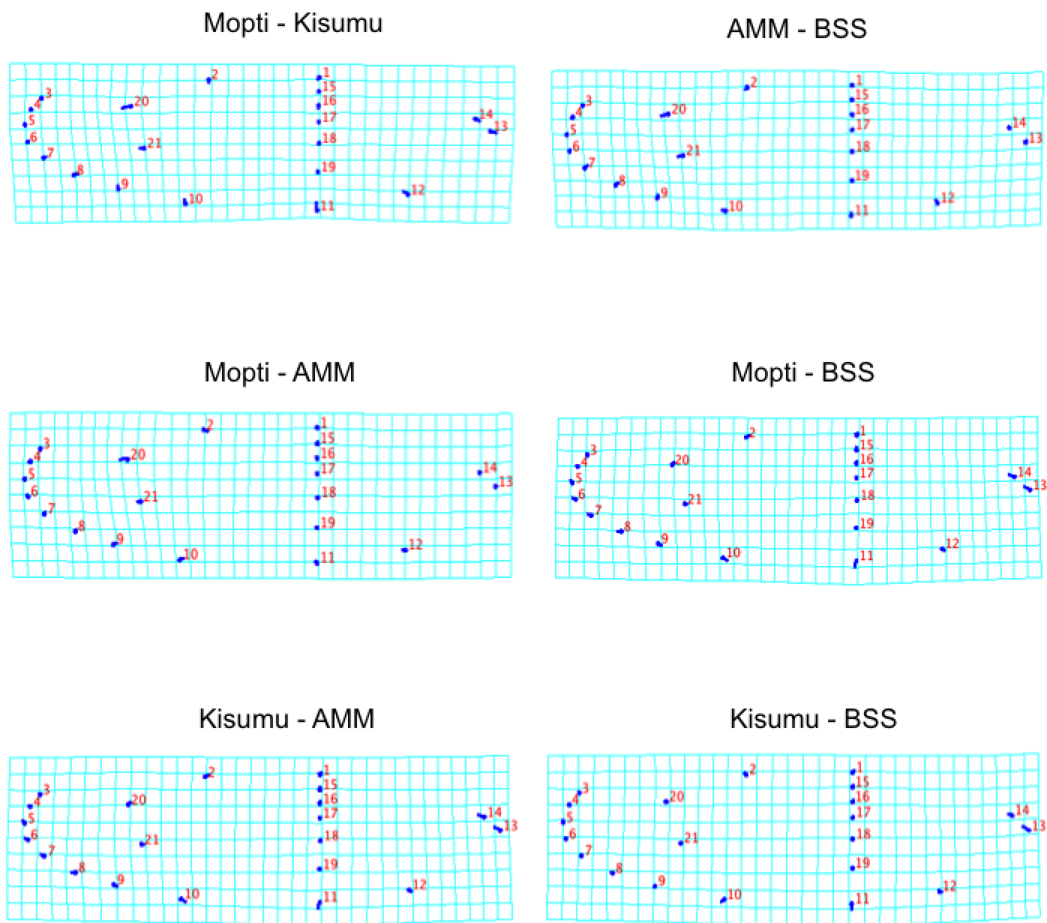


Figure 6.12. Thin-plate spline deformation grid showed the shape variation for the strains within the males group.

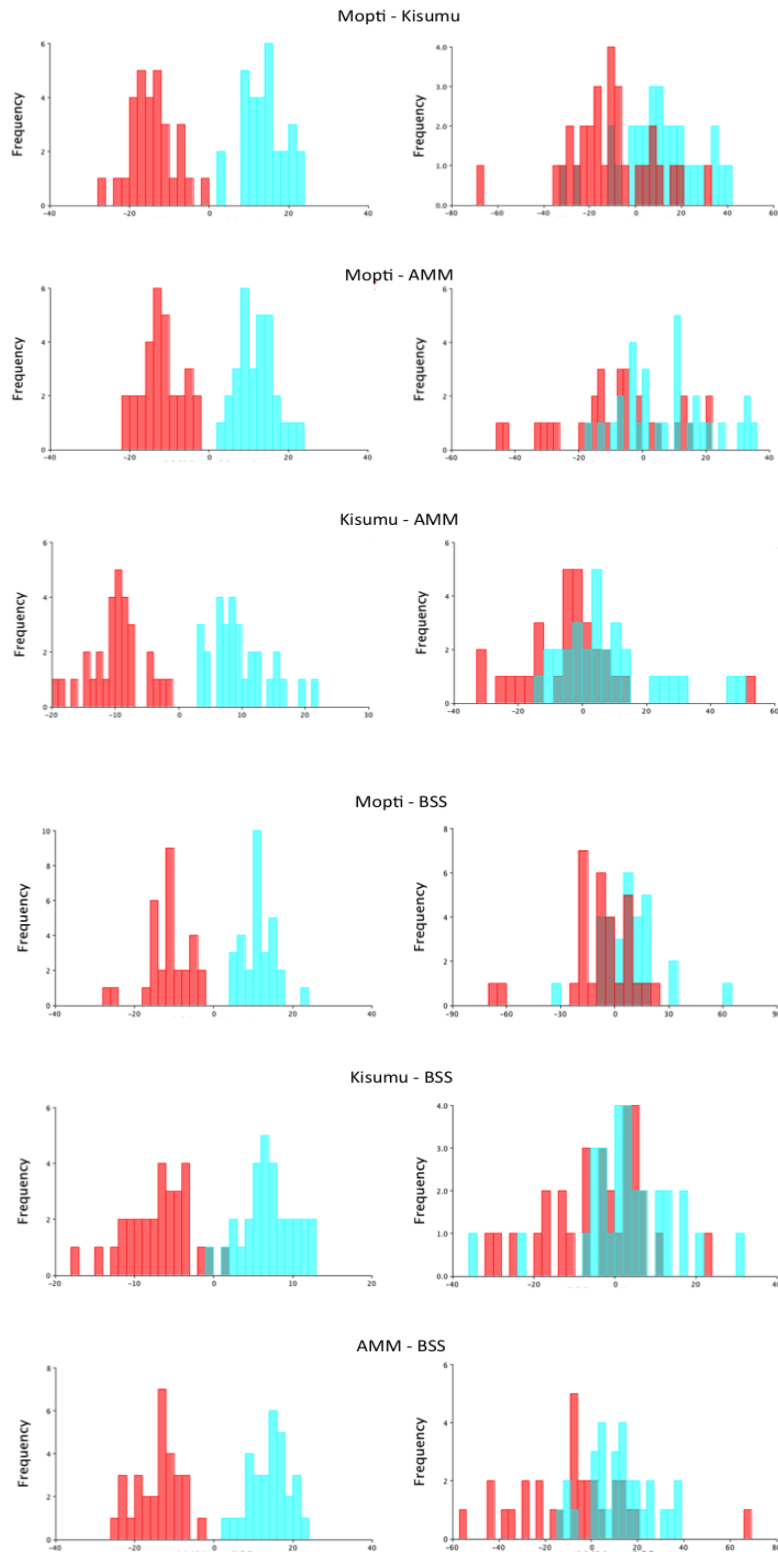


Figure 6.13. Pairwise comparison within the females group - the left figures show frequency of groups that were descriminated and separated based on the wing discribtion that were given to the software. The right figures show discriminant scores for all data and cross-validation when omitted each observation one at time and recalculate the classification function using the remaining data.

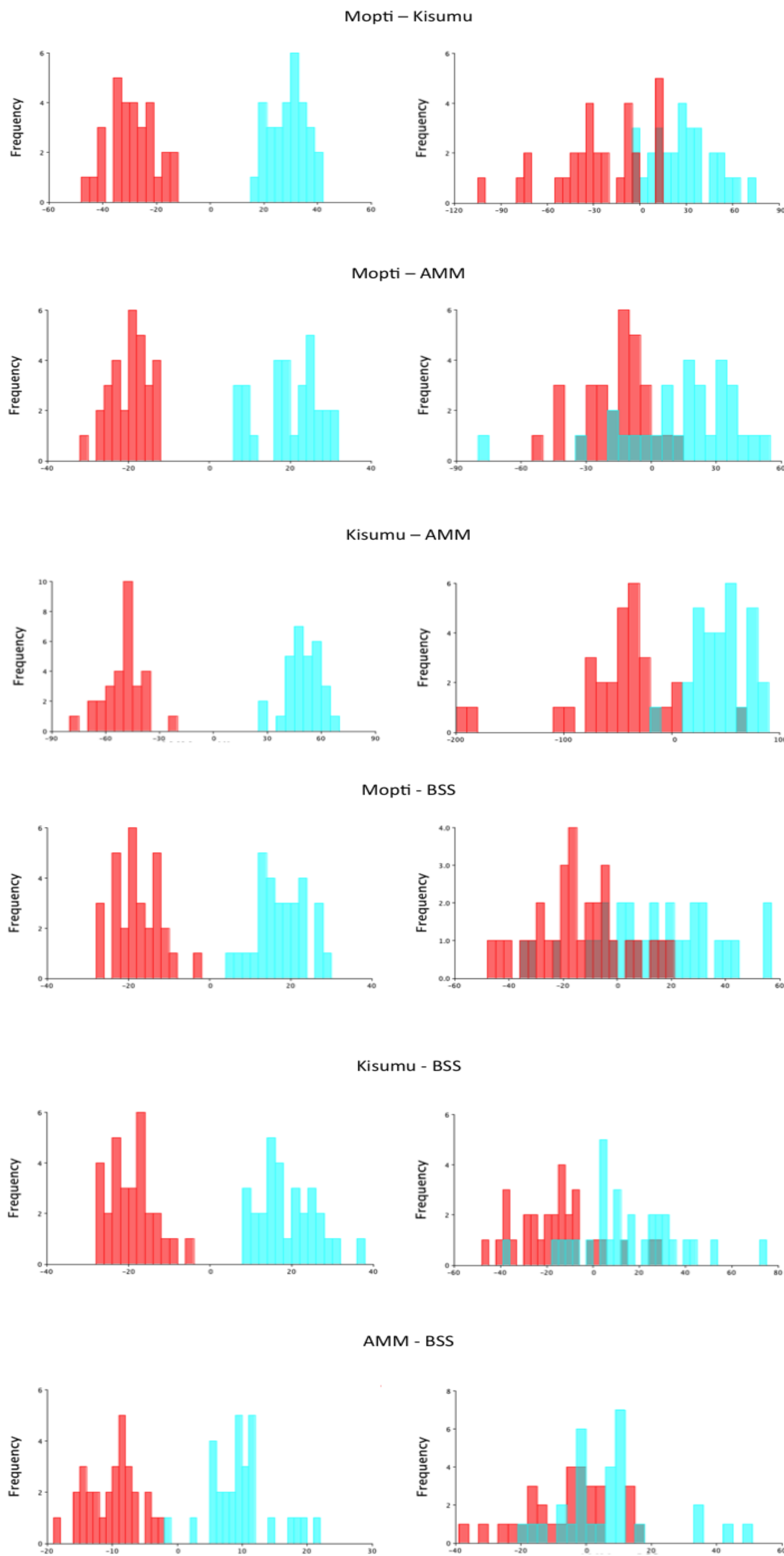


Figure 6.14. Pairwise comparison within the males groups - the left figures show frequency of groups that were descriminated and separated based on the wing discribtion that were given to the software. The right figures show discriminant scores for all data and cross-validation when omitted each observation one at time and recalculate the classification function using the remaining data.

Table 6.9. Summary table for the length, width and width/length ratio *P-values* for males and females wings of recombinants and parental strains.

Variable	Females	Males	Strains
Wing length	P= 0.2 NSD	P= 0.7 NSD	P= 0.2 NSD
Wing width	P= 0.2 NSD	P< 0.001 SD	Kisumu≥ BSS≥ AMM= Mopti
width/length	P= 0.8 NSD	P< 0.001 SD	P< 0.001 NSD

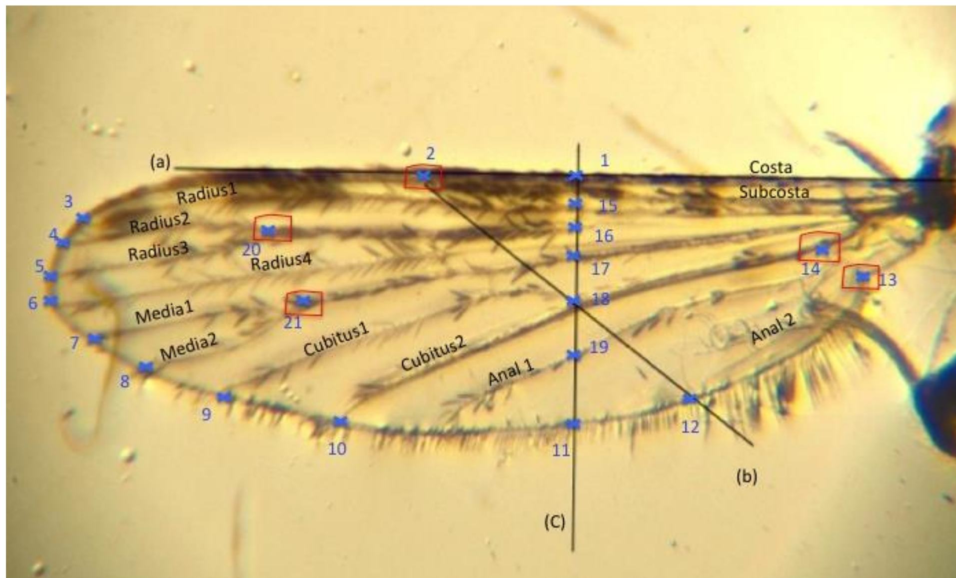


Figure 6.14. The wing variation summary between the sexes and strains – variation included elongation of wings in the red rectangles marked landmarks that are: 2. The Junction between the Costal and Sub-costal vein, 13. The posterior Anal cell margin, 14. The junction between Anal1 and Cubitus2, 20. The junction between Radius branches, 21. The junction between Media branches.

6.4. Discussion

The importance of this study lies in explaining the mechanisms responsible for assortative mating that led to reproductive isolation between *An. gambiae s.s.* and *An. coluzzii* species (Charlwood et al. 2003; Dao et al. 2008; Diabate & Tripet 2015). Several studies registered phenotypic differences between both species in different aspects (Gimonneau et al. 2010; Mouline et al. 2012; Dao et al. 2014). However, the information about the wing shape differences is limited (Sanford et al. 2011). Therefore, this is the first documented study includes comparison for the subtle wing-shape between Mopti and Kisumu that represent M and S molecular forms respectively. The comparison was also conducted between the laboratory recombinants AMM and BSS that are found to mate assortatively and genetically differed only in the X-island of speciation (Chapter 3). The results of detailed morphometric studies demonstrated differences in the amount of sexual dimorphism between the different strains with Kisumu being the least dimorphic and Mopti is the most dimorphic. In addition, significant differences in specific morphometric features of the wings were detected amongst the strains and this could be promising for future studies.

Using simple ANOVAs, significant differences were found between females and males wing length, width and their ratio. The wings of the females were bigger than the males ones in all the observed strains. These results demonstrate the importance of wing size differences between sexes that could result in different wing beats and it could be crucial for flight tones-frequency matching between sexes in the process of mate choice (Gibson & Russell 2006; Howell & Knols 2009).

Looking at variation amongst strains within sexes, wing length did not differ significantly amongst strains in females and males. Interestingly, the main pattern observed was that over interaction between sex and strain in the width and width/length ratio, and this was due to the fact that females did not differed in that variables but males did. The

interaction was mostly explained by the fact that Kisumu have significantly wider and larger width/length ratio than Mopti and AMM and generaly less sexual dimorphic than any other strains. However, BSS wing widths were not significantly different from any other strain. At the present we do not know why the sexual differences is less in Kisumu than in Mopti and recombinant strains, it could be because the Kisumu (*An. gambiae s.s.*) was raised in area that includes less competition compared to the Mopti form (*An. coluzzii*) that shares the habitat with Savanna and Bamako forms of *An. gambiae s.s.*, leading to potentially stronger sexual selection in the male phenotype.

These results bear some similarities with those reported by Sanford et al. (2011) who found larger differences wing width between *An. gambiae s.s.* and *An. coluzzii* females in Mali where they mate strictly assortatively compared with populations in Guinea-Bissau areas where there is a current break-down in premating isolation. However, the study of Sanford included only small sample sizes and focused only on females. Here, wing width and width/length ratio differences between the strains were found only in males supporting the hypothesis that sexual selection is driven by females choice (Chapter 3).

Here, despite finding significant differences between wing width and width/length ratio, the distribution of these variables largely overlapped, which suggests that this differences alone cannot account for accurate species-specific recognition. Moreover there was no significant difference between BSS and AMM in terms of wing width and width/length ratio and this would further confirm that wing size plays no role in recombinant assortative mating in our model. These results did not support Sanford assumptions that different wings size could lead to different wing tone relevant to sexual selection between strains.

Generally, the morphometric analyses of subtle wing-shape differences amongst the sexes and strains confirmed the results of ANOVAs. Wing shape differences were found

between females and males within the strains in particular parts of the wings. Moreover, the cross-validation analyses revealed that the sexual wings shape variation was less in Kisumu compared to the other strains, and Mopti variation was the highest.

The overall comparison for the distribution of individuals based on the shape differences within sexes also revealed the same interaction between sex dimorphism as in ANOVAs analyses. This was again because there was less amongst strains variations in females than those in males. The wing variation responsible for significant variation function between the sexes and strains included elongation of wings in the junction between the Costal and Sub-costal vein, the posterior Anal cell margin, the junction between Anal1 and Cubitus2, the junction between Radius branches, and the junction between Media branches. Although the variation in these particular features of the wings could be promising for further studies, it should be known that cross-validation that is part of these analyses revealed that there was a large overlap in the distribution of these variables between sexes and strains.

In conclusion, the ANOVAs and wing shape analyses showed that the wing shape differences could not be account for the wing-tone differences and within species recognition. Therefore, further studies are required for other aspects of wing structure to reveal the other wing characters that could play role in the wing tone hypothesis: e.g. distribution and color of scales in specific parts of the wings, the wing thickness, etc… . In addition, other mechanisms of assortative mating could be involved such as, color that associated with the vision, the pheromones that associated with olfaction. These studies are important to fill the gaps in our understanding of the mechanisms that the mosquitoes use for intraspecific mating which could lead to reproductive isolation that leads to speciation.

CHAPTER SEVEN

General discussion about behavioural mechanisms of reproductive isolation and sympatric speciation in *An. gambiae* s.l species, and their effects on malaria control strategies

7.1. Introduction

A few decades ago, some studies reported that *An. gambiae* s.l. includes several distinct species (Coluzzi et al. 2002; Slotman et al. 2006; Slotman et al. 2007) . Some of these species are reported as important due to their ability to exploit both temporary and man-made breeding sites in field (Coluzzi et al., 1978; della Torre et al., 2002; della Torre, Tu, & Petrarca, 2005). In addition, some of these species are able to transmit malaria, which makes them wide-open subject for studying to find new methods for controlling malaria.

Several studies reported that malaria causes about a million deaths every year (Knell 1991; WHO 1999), 90% of these deaths are found in sub-Saharan Africa, in the tropical areas that are rainy in most of the year, where malaria transmission is continuing over the year (Knell 1991; WHO 1999). There were attempts to control the mosquitoes to eradicate the malaria disease, e.g. using insecticide-treated bed nets and indoor residual spraying (Howell and Knols 2009). However, using these traditional methods is not likely to meet the expected results due to the evolution of pesticide resistance and the lack of the basic biological information about the mosquitoes (Lounibos & Conn 2000; Shiff 2002). Subsequently, they tried to find new alternative methods: e.g. sterilizing mosquitoes' males using chemicals or X ray, thereafter release them in the field. These males are able to mate and give sterile eggs. These males release programmes gave good results in Aedes mosquitoes but did not meet the expected results in Anopheles mosquitoes (Bellini,

Balestrino, et al. 2013; Bellini, Medici, et al. 2013). Therefore, the success of these malaria control programmes requires a good understanding of genetic diversity of the vectors in field, and how the vector species distribute in time and space and the factors that help maintaining this diversity (Lanzaro & Tripet 2003).

In general, *An. gambiae* s.l species have been shown to be adapted to different environments and this reduces competition between them for limiting resources (della Torre et al. 2005). Different studies attempt to explain the speciation process and the conditions that help in creating two species from one original taxon (Via 2012; Singh 2012; Feder et al. 2012). Sympatric speciation could occur due to the adaptation of sub-populations to different ecological changes resulting in reduced random mating among individuals. This could lead to the occurrence of a divergent selection against the adapted alleles and nearby genes that could reduce the introgression in the selected adapted genes within population (Via 2012). This process could lead to assortative mating behaviour within the population, when genes of mate choice are genetically associated with those of the specific trait under divergent selection. This can also lead to reinforcement of assortative mating and facilitate the creation of two different species from one taxon (Via 2012).

In this regard, two objectives were taken into consideration in this thesis. The first was a molecular and genetic objective that aimed to identify the islands of speciation that are more likely responsible for assortative mating. The second was a behavioural ecological objective that aimed to identify behavioural mechanisms of assortative mating using recombinant and parental strains. These studies were important steps for trying to explain the speciation process within *An. gambiae* s.l. species that is in turn important for malaria control strategies based on vector control.

7.2. The mechanisms of assortative mating in *An. gambiae s.s.* and *An. coluzzii*

The *An. gambiae s.s.* and *An. coluzzii* sibling species have been the subject for extensive researches that revealed complex patterns in gene flow and reproductive isolation. Some studies suggested extreme assortative mating between these species even in mixed swarms (Tripet et al. 2001; Dabiré et al. 2014). However, hybrids still occur in different regions of Africa, and laboratory studies showed that they were fully viable and fertile (Diabaté et al. 2006; Diabaté et al. 2007), which suggested that the pre-mating barriers play the main role in reproductive isolation between *An. gambiae s.s.* and *An. coluzzii*.

These findings were supported by our results, when we found that high assortative mating frequency occurred amongst recombinants and parental strains (Chapter 3). It is believed that males and females of these species use specific mechanisms to attract conspecific individuals even in the mixed swarm (Manoukis et al. 2009; Pennetier et al. 2010). No previous studies gave a full explanation for these mechanisms. Therefore, we tried to fill some gaps in our understanding of how the males and females of *An. gambiae s.s.* and *An. coluzzii* mate assortatively by comparing the swarming behaviour between the recombinants when they were placed with their own type or wrong type of mosquitoes in the laboratory. In general, the mean mating duration in the laboratory was 16.7-17.5sec including leg grips and tandem stages, which was compatible with the 15-20sec copulation period reported for *An. gambiae s.s.* and *An. coluzzii* in the field (Charlwood & Jones 1979; Howell & Knols 2009); thereafter the number of swarming mosquitoes decreased gradually with time (Chapter 5). These results showed that, with regard to that trait, selection for laboratory conditions did not significantly affect the mating process in these mosquitoes, and we can consider our laboratory observations as a representative model of what may happen in the field. Field studies explained that the high mating activity shortly after swarming start is due to high males receptivity to females flight tones (Charlwood &

Jones 1979). In order to explain the mating behaviour differences between these species, previous study suggested that males of *An. gambiae s.s.* start swarming 3.5min and end 5min earlier than those of *An. coluzzii,* that could play a role as a pre-mating barrier between them (Sawadogo et al. 2013). However, we found that the recombinant strain BSS initiated leg grips earlier than AMM by a mean of ~1sec over three hours in 2 out of 3 replicates, which is a very tiny difference and thus cannot play a role in the assortative mating observed amongst recombinant strains and therefore also suggest that other mechanisms are likely more important in the field (Chapter 5). We also found that mating initially seemed to occur freely even when we combined the wrong males and females types together in the experimental cages (Chapter 5). This is consistent with reports that assortative mating may, at times partially break down, as for example indoors as observed between *An. gambiae s.s.* and *An. coluzzii* (Dao et al. 2008). Such observations confirm that pre-mating barriers are incomplete between *An. gambiae s.s.* and *An. coluzzii* (Lanzaro & Lee 2013). However, in our experiment this behaviour did not continue in the laboratory and we found that the tandems to leg grips ratio reduced sharply and significantly in the second and third hour in interspecific mating type combinations comparing to the intraspecific mating type combinations (Chapter 5). This was due to mating rejection occurring at the leg grip stage before the mosquitoes took a tandem position (Chapter 5). Therefore, rejection upon leg grip could be another assortative mating mechanism suggesting that the mosquitoes could use olfactory receptors that either located directly in the forelegs or elsewhere on the body that would be used to recognize conspecific mosquitoes. In addition, our results suggested another assortative mating mechanism when we found that the number of swarming individuals was approximately twice higher when the mosquitoes were combined with wrong individuals compared to their numbers when they were mated to their own ones (Chapter 5). This suggested that the mosquitoes may be

spending more time swarming and using flight tones trying to attract or identify their own type mates during the swarming (Tripet et al. 2004; Pennetier et al. 2010). These findings led us to think generally about the mosquitoes' behaviour before tandem stage that most likely involved leg-grip behaviour, potentially combined with harmonizing wing beats with nearby mosquitoes.

There were some attempts to explain the flight tone hypothesis in *An. gambiae* s.l. sibling species and its role in reproductive isolation between these species (Diabate & Tripet 2015).

The sound oscillations cause vibrations of the antennae hairs and flagellar shaft that are transmitted to the Johnston's organ at the base of the flagellum (Roth 1948). In the males, the antennae hold many long hairs; and it is known to be plumose. Therefore, the sensitivity of the male antennae exceeds the sensitivity of the female ones. The best frequency of the male antenna is around 380Hz, which corresponds to the frequency of female flight tones. However, the best frequency of the female antenna is around 230Hz(Göpfert et al. 1999). In this regard, the flight tone hypothesis assumes that the auditory organs help the male and female mosquitoes match their flight tone frequencies with the nearby individuals to attract their own species mates (Charlwood & Jones 1979; Clements 1992; Cator et al. 2010; Pennetier et al. 2010). Therefore, it is important to describe the acoustic signals that the mosquitoes produce not only to understand the mechanisms that they use for assortative mating within mixed swarms, but also for designing the sound traps for controlling the vectors (Diabate 2015). The differences in the flight tones could be due to the wing-shape differences in *An. gambiae s.s.* and *An. coluzzii* species (Sanford et al. 2011). In this thesis, the wings of the females were bigger than the males ones in all the observed strains (Chapter 6) suggesting that the differences in wings size could cause differences in flight tones-frequencies. This suggested that the flight

tones-frequency matching between the males and females could play main role in sex choice in process of mating (Gibson & Russell 2006; Howell & Knols 2009). However, in regard of the concpecific recognition, significant wing shape variations were found only in the males between the different strains (Chapter 6) supporting the hypothesis that assortative mating process is driven by females choice (Chapter 3). These findings suggested that the females could choose the concpecific males by recognizing their tones that associated with wing shape variation in these males (Chapter 6, Chapter 3). However, the distribution of these variables largely overlapped, which suggests that these differences alone cannot account for accurate species-specific recognition (Chapter 6). In addition, The wing shape differed in specific parts that included elongation of wings in the junction between the Costal and Sub-costal vein, the posterior Anal cell margin, the junction between Anal1 and Cubitus2, the junction between Radius branches, and the junction between Media branches. The variation in these particular features of the wings could be a promising subject for further studies (Chapter 6).

Moreover, our studies showed another important finding when we compared tandem time between recombinants with matched/non-matching island type. The tandem time was significantly shorter in non-matching type recombinants compared with matching island type recombinants suggesting some potential differences in post-mating processes through the interaction between seminal products of males with the receptors of spermatheca of females that could cause post-mating changes between *An. coluzzii* and *An. gambiae s.s.* (Chapter 5). This was tested when we analysed the sperm amount that extracted from spermatheca of mated females in the recombinants that mated with their own and opposite type of males in the laboratory. The results showed that BSS females stored more sperm than AMM ones from the same body size individuals regardless to the male type (Chapter 5). This could be due to differences in the BSS and AMM spermatheca size (can be tested

in future studies) and the interaction between the seminal fluids and spermatheca receptors that could affect the sperm uptake.

Our studies that included comparison of mating time and behavior in non-matching type male swarms and matching-type swarms in the laboratory (Chapter 3, Chapter 5) and testing the wing size and shape (Chapter 6) did not give full explanation for the assortative mating behavior mechanisms in *An. gambiae s.s.* and *An. coluzzii*. The small differences that were detected in these different biological aspects of mating behavior between *An. gambiae s.s.* and *An. coluzzii* could combine together to help in conspecific recognition within mixed swarms in these species. However, other mechanisms of assortative mating could involve other traits such as, color that associated with the vision, and pheromones that are associated with olfaction. These studies are important to fill the gaps in our understanding of the mechanisms that mosquitoes use for intraspecific mating which could lead to reproductive isolation that leads to speciation process between *An. coluzzii* and *An. gambiae s.s.*.

7.3. Explaining sympatric speciation based on the divergence islands

The genes that are responsible for assortative mating are believed to facilitate the speciation process in *An. gambiae* s.l. sibling species. Genomic comparison studies in *An. coluzzii* and *An. gambiae s.s.* showed that the genetic differences between them were limited to three pericentromeric regions 2L, 3L and X chromosomes (Turner et al. 2005; Turner & Hahn 2007; White et al. 2010). These divergence regions represented only 3% of the total genome (Lanzaro & Lee 2013). Explaining the speciation based on the islands of speciation took two directions (Lanzaro & Lee 2013). The first direction explained that the gene flow between *An. coluzzii* and *An. gambiae s.s.* is high throughout the genome except for the low recombination islands of speciation regions (Lanzaro et al. 1998; Wang et al.

2001; Turner et al. 2005; Turner & Hahn 2007). In this context, the divergence is driven by genes exist in sites with reduced recombination within divergence islands (Noor et al. 2001; Rieseberg 2001). The second direction suggested complete linkage disequilibrium between these islands (White et al. 2010). Therefore, it was suggested that there is no gene flow between both species that are completely isolated, and the high differences in the divergence islands is incidental (White et al. 2010). Therefore, genes responsible for assortative mating are not necessarily included in the island of speciation regions (Turner & Hahn 2010; Hahn et al. 2012).

These models of speciation are still controversial subjects. Therefore, in this thesis we tried to reveal the role of the islands in speciation and to identify the candidate island that would include the genes responsible for assortative mating and reproductive isolation between *An. gambiae s.s.* and *An. coluzzii* . Genotyping X, 2L and 3L in recombinant and parental strains in our lab confirmed that BSS and Kisumu were S-genotype in X-island, whereas they were homozygous M-type in 2L and 3L islands of the genome (Chapter 3). Given that BSS recombinant strain mate strictly assortatively (Chapter 3) suggested that the X-island could include the genes that code traits responsible for assortative mating between the sibling species. This concept is consistent with divergence islands model suggesting that the X-island plays the main role in sympatric speciation process (Feder et al. 2012; Via 2012; Seehausen et al 2014). In addition, our results showed that the inversions were polymorphic in the recombinant strains confirming that there is no role of inversions in assortative mating behaviour and they could have indirect effect on the speciation process through the ecological adaptation to new regions and habitats (Chapter 3). Based on these results, genomic comparison was conducted to identify the genomic regions that have the highest differentiation between the sibling sympatric species (Chapter 4). The comparison between *An. gambiae s.s.* and *An. coluzzii* collected from Ghana and

laboratory recombinant strains showed that the highest genomic differences were in X-island that covered 6Mb of the X chromosome, and spans from site ~18.1 to 24.2Mb (Chapter 4). In addition, we identified 20 protein-coding changes between *An. gambiae s.s.* and *An. coluzzii* that affect 12 genes, 6 of these genes have biological functions that are: AGAP001002 (Toll protein) is involved in development and immunity; and AGAP001033 (mab-21 like protein) is involved in neural and sensory organ development; AGAP001050 (chondroitin polymerizing factor); and AGAP001052 (ubiquitin carboxyl-terminal hydrolase) are involved in protein secretion and proteolysis; AGAP001022 (gastrin/cholecystokinin receptor) is a receptor for peptides in the brain and gastrointestinal tract; and AGAP001025 (protein MSTA) is involved in negative regulation of gene expression. Some of these genes could directly involve in the mating process or indirectly through interaction with other mating genes that could be elsewhere in the genome (Chapter 4). However, some of these genes also might affect directly or indirectly genes responsible for ecological adaptation in *An. gambiae s.s.* and *An. coluzzii,:* e.g. those are involved in distribution of the larvae in different habitats (Kamdem et al. 2012; Gimonneau et al. 2012) and the response of the larvae to the predators presence (Diabaté et al. 2008; Gimonneau et al. 2012). Further re-sequencing studies could consider intergenic regions that might include elements with cis and trans effects on genes expression within the X, 2L and 3L islands (Cassone et al. 2008; Cassone et al. 2014).

In conclusion, our studies supported the concept that there is no linkage disequilibrium between X, 2L and 3L islands of speciation, and also the X-island was highly different between the sibling species (Chapter 3, Chapter 4). Therefore, it is more likely that the X-island includes the genes that are responsible for assortative mating between *An. gambiae s.s.* and *An. coluzzii*, even though the phenotypic traits that are coded by some of these genes are not known yet and further studies are required to find their

effect in reproductive isolation between these species. However, they could code different biological traits that interact together and lead to reproductive isolation between the sibling species. Further studies are required to specify the genes on the X-island that responsible for assortative mating, and these genes could be target for genetic modification approaches to create either choosy strain that actually mates with the target population or less choosy modified mosquitoes that could mate with several target species in mosquitoes control programmes.

7.4. Malaria control strategies

The success of malaria control programmes depends on our understanding of vector biology, population structure, speciation process (Lanzaro & Tripet 2003, Diabate & Tripet 2015). Recently, studies focusing on three main aspects: revealing the mechanisms of reproductive isolation between the sibling species, assessing new methods of controlling the mosquitoes, and improving the lab mosquitoes ability to mate and compete with the field individuals (Benedict et al. 2009; Cator et al. 2010; Shutt et al. 2010; Burt 2014; Diabate & Tripet 2015).

Here, our studies focused on understanding the mating behaviour and genetic structure of the *An. gambiae* complex to improve programmes based on releasing sterile or genetically-modified mosquitoes in the field, and to find new approaches for controlling malaria vectors. Understanding the process of swarming and mating behaviour is key role for improving males release programmes through improving the ability of laboratory males to compete and mate with females of the target species. In this regard, we compared the mating behaviour of males and females by giving them the choice between mates bearing the same or the opposite X-island genotype (Chapter 3). In such small cage choice assays, our results suggested that assortative mating process is driven by females in mixed swarms, when we found that the females mate assortatively with higher frequencies than the males when they had a choice between their own X-island type and the reciprocal type of mosquitoes (Chapter 3) suggesting that the males could potentially be used in controlling methods for targeting several sympatric species of *An. gambiae* s.l. mosquitoes (Diabate 2015). However, we cannot deny the role of males in assortative mating behaviour that lies in their ability to form spatially segregated swarms in the field (Diabaté et al. 2006; Dabire et al. 2013). In Chapter 5, we also compared the mating behaviour between intraspecific and interspecific mating type combinations and found that the mating

frequency was lower in interspecific mating type combinations (Chapter 5) and this could again be associated with the ability of the females to mate assortatively with conspecific males (Chapter 3) and therefore could prevent the released sterile males to mate with more than one type of mosquitoes in the control programmes. Therefore, it is important to understand the mechanisms of assortative mating as well as the genes responsible for reproductive isolation between *An. gambiae* s.l. sibling species for males releasing programmes (Aboagye-Antwi et al. 2015). This could help in finding new ways to target several species of vectors, or to make several releasing programmes for the different sibling species separately.

Our findings are considered as step stone for further studies that could reveal more precise behavioural and genomics differences between *An. gambiae s.s*. and *An. coluzzii* that could in turn open new horizons for new strategies and techniques for controlling malaria vectors.

APPENDICES

Appendices 1: Abbreviations

Abbreviations	Full names
SIT	Sterile insect technique
An. gambiae s.l.	Anopheles gambiae senso lato
An. gambiae s.s	Anopheles gambiae senso strecto
H-W	Hardy–Weinberg
IGS	Nucleotide polymorphisms intergenic spacer
MAG	Male accessory gland secretions
Acps	Accessory gland proteins
DDT	Dichlorodiphenyltrichloroethane
PBS	Phosphate-buffered saline
ERC	MinElute cleanup Kit (Enzymatic reaction cleanup buffer)
PE	MinElute cleanup Kit (wash buffer)
EB	MinElute cleanup Kit (elution buffer)
TE buffer	Tris, a common pH buffer, and EDTA, like Mg2+
20E	Steroid hormone 20-hydroxyecdysone
dH2O	Deionised water
TPS	Thin-plate spline
LM	The number of landmarks
ID	Mosquitos' individual name.
PCA	Principal components analyses
CV	Canonical variation

Appendices 2: The tools and devices and their companies

Tools and devises	Company and Country
1.5ml centrifuge tubes	Fisher brand, Leicestershire, UK
Agarose	Fisher Scientific, Leicestershire, UK
Aminobenzoic acid (PABA)	Sigma-Aldrich Company Ltd., Dorset, UK
Bugs vacuum	Backyard Safari, UK
Charge Switch Kit	Invitrogen, Leicestershire, UK
Dissection microscope	Olympus, UK
dNTP	Fermentas, Leicestershire, UK
Dream Taq Polymerase	Thermo Scientific, Massachusetts, United States
Eco321 (EcoRv)	Fermentas, Leicestershire, UK
Ethidium bromide	Fisher Scientific, Leicestershire, UK
Primer FdY (5'-CGT GCA ACA GCT CGT GAT G-3')	Eurofins Genomics, Wolverhampton, UK
Feeder	Hemoteck membrane feeding system
Filter paper	Grade 1, Whatmann
Glass Petri plate	Sterilin, Massachusetts, USA,
Glucose	Acrose Organics, New Jersey, USA
Ground fish food	Tetramin, Blacksburg, USA
Hha1' restriction endonuclease	Promega, Southampton, UK
Horse blood	TCS biosciences LTD, Buckingham, UK
Humidifier	JS, West Sussex, UK
Kspa1 (Hpal)	Thermo Scientific, Massachusetts, United States
Ladder	Thermo Scientific, Massachusetts, United States
Liquifry	interpet ltd., dorking, uk
Digital camera	Olympus (E-520)
ParaFilm	Sigma-Aldrich Company Ltd., Dorset, UK
PBS (Phosphate Buffer Saline)	Fisher Scientific, Leicestershire, UK
Polystrene cups	Dart, Insulated foam containers, KENT, UK
Primers	Eurofins Genomics, Wolverhampton, UK
ProbY (5'-AGA TGG ATG CGG CGT-3')	Eurofins Genomics, Wolverhampton, UK
PTC-200 DNA- Engine thermocycler	BioRad, California, USA
QPCR plates	Applied biosystems, California, USA
Primer RvY (5'-TTA CCA CGC TGG CAA ATG C-3')	Eurofins Genomics, Wolverhampton, UK
Taq Man	Applied biosystems, California, USA
TBE	Invitrogen, Leicestershire, UK
Tris-Boric Acid-EDTA	Sigma-Aldrich Company Ltd., Dorset, UK

primer GA(5'-CTG GTT TGG TGG GCA CGT TT-3́)	Eurofins Genomics, Wolverhampton, UK
primer UN(5'-GTG TGC CCC TTC CTC GAT GT-3́)	Eurofins Genomics, Wolverhampton, UK
Nano drop	Thermo Scientific, Massachusetts, United States
DNAzol	Fisher Scientific, Leicestershire, UK
picogreen kit method	Invitrogen™ Quant-iT™, Leicestershire, UK
Ethanol	Sigma-Aldrich Company Ltd., Dorset, UK
TE	Fisher Scientific, Leicestershire, UK
MagnaRack	Fisher Scientific, Leicestershire, UK
5ml tubes	Fisher brand, Leicestershire, UK
infrared red lights	48LED IR Illuminator lamp, Amazon
Video cameras	Sony (HandyCam HDR-SR12E)
QPCR machine	Applied biosystems, California, USA
Potassium hydroxide	Sigma-Aldrich Company Ltd., Dorset, UK

REFERENCES

Aboagye-Antwi, F. et al., 2015. Experimental Swap of Anopheles gambiae's Assortative Mating Preferences Demonstrates Key Role of X-Chromosome Divergence Island in Incipient Sympatric Speciation. *PLoS genetics*, 11(4), p.e1005141. Available at: http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005141.

Al-Mekhlafi, A.M. et al., 2010. Molecular epidemiology of Plasmodium species prevalent in Yemen based on 18 s rRNA. *Parasites & vectors*, 3(1), p.110. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2997089&tool=pmcentrez&rendertype=abstract [Accessed February 5, 2014].

Alexander, 1975. *Natural selection and specialized chorusing behavior in acoustical insects* P. D. N. Y. A. Press, ed.,

Anitram, 2012. http://www.biographix.cz/portfolio/schemes-models/life-cycle-of-the-mosquito-anopheles-gambiae/.

Appawu, M. et al., 1994. Species composition and inversion polymorphism of the Anopheles gambiae complex in some sites of Ghana, West Africa. *Acta Tropica*, 56, pp.15–23.

Baeshen, R. et al., 2014. Differential effects of inbreeding and selection on male reproductive phenotype associated with the colonization and laboratory maintenance of Anopheles gambiae. *Malaria journal*, (13).

Baldini, F. et al., 2012. Function and composition of male accessory gland secretions in <I>Anopheles gambiae</I>: a comparison with other insect vectors of infectious diseases. *Pathogens and Global Health*, 106(2), pp.82–93.

Bannister, L. & Mitchell, G., 2003. The ins, outs and roundabouts of malaria. *Trends in Parasitology*, 19(5), pp.209–213. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1471492203000862 [Accessed January 20, 2014].

Barluenga, M. et al., 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, 439, pp.719–723.

Bellini, R., Balestrino, F., et al., 2013. Mating Competitiveness of Aedes albopictus Radio-Sterilized Males in Large Enclosures Exposed to Natural Conditions. *Journal of Medical Entomology*, 50(1), pp.94–102. Available at: http://jme.oxfordjournals.org/content/50/1/94.abstract.

Benedict, M., 2003. The first releases of transgenic mosquitoes: an argument for the sterile insect technique.

Trends in Parasitology, 19(8), pp.349–355. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1471492203001442 [Accessed January 18, 2015].

Benedict, M.Q. et al., 2009. Colonisation and mass rearing: learning from others. *Malaria journal*, 8 Suppl 2, p.S4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2777326&tool=pmcentrez&rendertype=abstract [Accessed March 21, 2014].

Boulesteix, M. et al., 2007. Insertion polymorphism of transposable elements and population structure of Anopheles gambiae M and S molecular forms in Cameroon. *Molecular ecology*, 16(2), pp.441–452.

Bradbury, J., Gibson, R. & Tsai, I.M., 1986. Hotspots and the dispersion of leks. *Anim Behav*, 34(6), pp.1694–1709.

Bradbury, J.W. & Gibson, R.M., 1983. Leks and mate choice. , pp.109–138.

Brunhes, J., Le Goff, G. & Geoffroy, B., 1997. Anopheles Afro--- Tropicaux. I— Descriptions d'especes nouvelles et changements to statuts taxonomiques (Diptera: Culicidae). *Annales de la Societe Entomologique de France*, 33, pp.173–183.

Bryan, J. et al., 1982. Inversion polymorphism and incipient speciation in Anopheles gambiae in The Gambia, West Africa. Genetica. *Genetica*, 59, pp.167–76.

Burt, A., 2014. Heritable strategies for controlling insect vectors of disease. *Biological sciences*. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4024225&tool=pmcentrez&rendertype=abstract.

Butlin, R. et al., 2012. What do we need to know about speciation? *Trends in Ecology & Evolution*, 27(1), pp.27–39. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0169534711002618.

Caputo, B. et al., 2008. Anopheles gambiae complex along The Gambia river, with particular reference to the molecular forms of An. gambiae s.s. *Malaria journal*, 7(1), p.182. Available at: http://www.malariajournal.com/content/7/1/182.

Caputo, B. et al., 2011. The "far-west" of Anopheles gambiae molecular forms. *PLoS ONE*, 6(2), pp.1–7.

Cassone, B. et al., 2008. Differential gene expression in incipient species of Anopheles gambiae. *Molecular Ecology*, 17, pp.2491–2504.

Cassone, B. et al., 2014. Gene expression divergence between malaria vector sibling species. *Molecular Ecology*, 23, pp.2242–2259.

Cator, L.J. et al., 2010. Sizing up a mate: Variation in production and response to acoustic signals in

Anopheles gambiae. *Behavioral Ecology*, 21, pp.1033–1039.

Catteruccia, F., 2007. Malaria vector control in the third millennium: progress and perspectives of molecular approaches. *Society of Chemical Industry*, 63, pp.634–640.

Charlwood, D. & Jones, R., 1979. Mating behaviour in the mosquito, Anopheles gambiae s.l. I. Close range and contact behaviour. *Physiological Entomology*, 4, pp.111–120.

Charlwood, J.D., Thompson, R. & Madsen, H., 2003. Observations on the swarming and mating behaviour of Anopheles funestus from southern Mozambique. *Malaria Journal*, 2, p.2.

Cingolani, P. et al., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of of Drosophila melanogaster strain w. *Fly*, 6, pp.80–92.

Clarkson, C.S. et al., 2014. Adaptive introgression between Anopheles sibling species eliminates a major genomic island but not reproductive isolation. *Nature Communications*, 5(4248). Available at: http://www.nature.com/doifinder/10.1038/ncomms5248.

Claude, 2009. *Morphometrics with R*

Clements, 1992. *The biology of mosquitoes, Development, Nutrition and reproduction*, the University of Michigan,London (UK),: Chapman & Hall.

Coetzee, M. et al., 2013. Anopheles coluzzii and Anopheles amharicus, new members of the Anopheles gambiae complex. *Zootaxa*, 3619(3), pp.246–274.

Coluzzi et al., 2002. A Polytene Chromosome Analysis of the Anopheles gambiae Species. *Science (New York, N.Y.)*, 298(5597), pp.1415–1418.

Coluzzi, M. et al., 1978. Chromosomal differentiation and adaptation to human environments in the Anopheles gambiae complex. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 73(5), pp.483–497.

Coluzzi, M., 1968. Cromosomi politenici delle cellule nutrici ovariche nel complesso gambiae del genre Anopheles. *Parassitologia*, 10, pp.179–183.

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

Coluzzi, M., Petrarca, V. & Deco, M.A., 1985. Chromosomal inversion intergradation and incipient speciation in Anopheles gambiae. *Bolletino di zoologia*, 52(1-2), pp.45–63.

Cook, G.. & Zumla, A., 2003. *Manson's Tropical Diseases* 21st ed. :W.B. Saunders Co, ed., London.

Coyne, J. & Orr, H., 2004. *Speciation.*, USA: Sinauer Associates, Inc.: Sinauer Associates,Inc.

Dabire, K.R. et al., 2013. Assortative mating in mixed swarms of the mosquito anopheles gambiae s.s. m and s molecular forms, in burkina faso, west africa. *Medical and Veterinary Entomology*, 27(3), pp.298–312.

Dabiré, R. et al., 2014. Mechanisms of reproductive isolation between incipient species in the Anopheles gambiae species complex. *Malaria Journal*, 13(Suppl 1), p.O4. Available at: http://www.malariajournal.com/content/13/S1/O4.

Dao, A. et al., 2008. Assessment of alternative mating strategies in Anopheles gambiae: Does mating occur indoors? *Journal of medical entomology*, 45(4), pp.643–652. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18714863.

Dao, A. et al., 2014. Signatures of aestivation and migration in Sahelian malaria mosquito populations ´ 1 ,. *Nature*, 516(7531), pp.387–390.

DePristo, M. et al., 2011. A framework for variation discovery and genotyping using next--- generation DNA sequencing data. *Nature Genetics*, (43), pp.491–498.

Diabate, A. et al., 2003. Natural swarming behaviour of the molecular M form of Anophelesgambiae. *TRANSACTIONS OF THE ROY AL SOCIETY OF TROPICAL MEDICINE AND HYGIENE*, 97(6), pp.713–716.

Diabaté, A. et al., 2007. Evaluating the effect of postmating isolation between molecular forms of Anopheles gambiae (Diptera: Culicidae). *Journal of medical entomology*, 44(1), pp.60–64.

Diabaté, A. et al., 2008. Evidence for divergent selection between the molecular forms of Anopheles gambiae: role of predation. *BMC evolutionary biology*, 8, p.5.

Diabaté, A. et al., 2006. Mixed swarms of the molecular M and S forms of Anopheles gambiae (Diptera: Culicidae) in sympatric area from Burkina Faso. *Journal of medical entomology*, 43(3), pp.480–483. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16739404.

Diabaté, A. et al., 2011. Spatial distribution and male mating success of Anopheles gambiae swarms. *BMC evolutionary biology*, 11(1), p.184. Available at: http://www.biomedcentral.com/1471-2148/11/184.

Diabaté, A. et al., 2009. Spatial swarm segregation and reproductive isolation between the molecular forms of Anopheles gambiae. *Proceedings. Biological sciences / The Royal Society*, 276(1676), pp.4215–4222. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2821344&tool=pmcentrez&rendertype=abstract.

Diabate, A. & Tripet, F., 2015. Targeting male mosquito mating behaviour for malaria control. *Parasites & Vectors*, 8(1), p.347. Available at: http://www.parasitesandvectors.com/content/8/1/347.

Fanello, C., Santolamazza, F. & della Torre, A., 2002. Simultaneous identification of species and molecular forms of the Anopheles gambiae complex by PCR-RFLP. *Medical and veterinary entomology*, 16(4), pp.461–464.

Favia, G. et al., 2001. Molecular characterization of ribosomal DNA polymorphisms discriminating among chromosomal forms of Anopheles gambiae. *Insect Molecular Biology*, 10(1), pp.19–23.

Favia, G. et al., 1997. Molecular identi ® cation of sympatric chromosomal forms of Anopheles gambiae and further evidence of their reproductive isolation. *Insect molecular biology*, 6(4), pp.377–383.

Feder, J., Egan, S. & Nosil, P., 2012. The genomics of speciation with gene flow. *Trends in Genetics*, 28(7), pp.342–350.

Feder, J.L., Chilcote, C.A. & Bush, G.L., 1988. Genetic differentiation between sympatric host races of the apple maggot fly Rhagoletis pomonella. *Nature*, 336, pp.61–64.

Feder, J.L. & Nosil, P., 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, 64, pp.1729–1747.

Gentile, G. et al., 2001. Attempts to molecularly distinguish cryptic taxa in Anopheles gambiae s.s. *Insect molecular biology*, 10(1), pp.25–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11240634.

Gentile, G. et al., 2002. Genetic differentiation in the African malaria vector, Anopheles gambiae s.s., and the problem of taxonomic status. *Genetics Society of America*, 161(4), pp.1561–1578.

Gibson, G. & Russell, I., 2006. Flying in tune: sexual recognition in mosquitoes. *Current biology : CB*, 16(13), pp.1311–1316. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16824918 [Accessed January 20, 2014].

Gibson, G., Warren, B. & Russell, I.J., 2010. Humming in tune: sex and species recognition by mosquitoes on the wing. *Journal of the Association for Research in Otolaryngology : JARO*, 11(4), pp.527–540. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2975882&tool=pmcentrez&rendertype=abstract [Accessed June 9, 2014].

Gibson, G.A., 1985. Swarming behavior of the mosquito Culex pipiens-quinquefasciatus—a quantitative analysis. *Physiol. Entomol*, 10, pp.283–296.

Gillott, C., 2003. Male accessory gland secretions: modulators of female reproductive physiology and

behavior. *Annual review of entomology*, 48, pp.163–184. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12208817 [Accessed May 27, 2014].

Gimonneau, G. et al., 2010. A behavioral mechanism underlying ecological divergence in the malaria mosquito Anopheles gambiae. *Behavioral Ecology*, 21(5), pp.1087–1092.

Gimonneau, G. et al., 2012. Larval habitat segregation between the molecular forms of the mosquito Anopheles gambiae in a rice field area of Burkina Faso, West Africa. *Medical and Veterinary Entomology*, 26(1), pp.9–17.

Göpfert, M.C., Briegel, H. & Robert, D., 1999. Mosquito hearing: sound-induced antennal vibrations in male and female Aedes aegypti. *The Journal of experimental biology*, 202, pp.2727–2738.

Hahn, M.W. et al., 2012. No evidence for biased co-transmission of speciation islands in Anopheles gambiae. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), pp.374–384. Available at: http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2011.0188.

Harbach, R.E., 2004. The classification of genus Anopheles (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. *Bulletin of Entomological Research*, 94(06), pp.537–553. Available at: http://www.journals.cambridge.org/abstract_S0007485304000562 [Accessed January 21, 2014].

Helinski, M.E.H. & Knols, B.G.J., 2008. Mating competitiveness of male Anopheles arabiensis mosquitoes irradiated with a partially or fully sterilizing dose in small and large laboratory cages. *Journal of medical entomology*, 45(4), pp.698–705.

Helinski, M.E.H., Parker, A.G. & Knols, B.G.J., 2009. Radiation biology of mosquitoes. *Malaria journal*, 8 Suppl 2, p.S6.

Howell, P.I. & Knols, B.G.J., 2009. Male mating biology. *Malaria journal*, 8 Suppl 2, p.S8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2777330&tool=pmcentrez&rendertype=abstract [Accessed January 20, 2014].

Huber, S.K. et al., 2007. Reproductive isolation of sympatric morphs in a population of Darwin ' s finches. *The Royal Society*, 274, pp.1709–1714.

Huho, B.J. et al., 2006. A reliable morphological method to assess the age of male Anopheles gambiae. *Malaria journal*, 5, p.62. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1570359&tool=pmcentrez&rendertype=abstract [Accessed January 20, 2014].

Hunt, R., 1973. A cytological technique for the study of Anopheles gambiae complex. *Parassitologia*, 15,

pp.137–139.

Hunt, R.H., Coetzee, M. & Fettene, M., 1998. The Anopheles gambiae complex: a new species from Ethiopia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 92(2), pp.231–235.

Jaramillo-O, N. et al., 2015. Geometric morphometrics for the taxonomy of 11 species of *Anopheles* (*Nyssorhynchus*) mosquitoes. *Medical and Veterinary Entomology*, 29(1), pp.26–36. Available at: http://doi.wiley.com/10.1111/mve.12091.

Jones, M.D.R. & Gubbins, S.J., 1978. Changes in the circadian flight activity of the mosquito Anopheles gambiae in relation to insemination, feeding and oviposition. *Physiological Entomology*, 12, pp.213–220. Available at: z:\4481.pdf.

Joshi, N. & Fass, J., 2011. Sickle: A sliding--- window, adaptive, quality--- based trimming tool for FastQ files (Version 1.33)[Software]. Available at https://github.com/najoshi/sickle.

Kamdem, C. et al., 2012. Anthropogenic habitat disturbance and ecological divergence between incipient species of the malaria mosquito Anopheles gambiae. *PloS one*, 7(6), p.e39453. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3382172&tool=pmcentrez&rendertype=abstract.

Klingenberg, C.P., 2011. MorphoJ: an integrated software package for geometric morphometrics. *Mol Ecol Resour*, 11(2), pp.353–357. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=21429143.

Knell, A., 1991. *Malaria : a publication of the tropical programme of the Wellcome Trust* Book: Engl., Oxford ; New York : Oxford University Press.

Lang, T. & Woodbridge, A., 1976. Is There a Female Sex Pheromone in the Mosquito Culiseta inornata. *Environmental Entomology*.

Langmead, B. & Salzberg, S., 2012. Fast gapped--- read alignment with Bowtie 2. *Nature Methods*, 9, pp.354–357.

Lanzaro, G. et al., 1998. Complexities in the genetic structure of Anopheles gambiae populations in west Africa as revealed by microsatellite DNA analysis. *PubMed*, 95(24), pp.14260–14265.

Lanzaro, G. & Lee, Y., 2013. Speciation in Anopheles gambiae -The distribution of genetic polymorphism and patterns of reproductive isolation among natural populations. *Intech*, pp.173–195.

Lanzaro, G.C. & Tripet, F., 2003. Gene flow among populations of Anopheles gambiae: a critical review.

Kluwer Academic Publishers, pp.109–132.

Lawniczak, M.K.N. et al., 2010. Widespread Divergence Between Incipient Anopheles gambiae Species Revealed by Whole Genome Sequences. *Science*, 330(6003), pp.512–514. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1195755.

Lee, Y. et al., 2012. An In- dividual-level Population Genomics Database for Arthropod Disease Vectors.(PopI). Davis, CA2004 [updated 2012; cited 2012]. Available from: https://grassi2.ucdavis.edu/.

Lee, Y. et al., 2009. Ecological and genetic relationships of the Forest-M form among chromosomal and molecular forms of the malaria vector Anopheles gambiae sensu stricto. *Malaria journal*, 8, p.75. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2680901&tool=pmcentrez&rendertype=abstract [Accessed January 20, 2014].

Lee, Y. et al., 2013. Spatiotemporal dynamics of gene flow and hybrid fitness between the M and S forms of the malaria mosquito, Anopheles gambiae. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49), pp.19854–9. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=MarkedList&qid=16&SID=2A9wHTCY5n3UXMDSGPt&page=1&doc=1&colName=WOS.

Lehmann, T. et al., 1996. Genetic differentiation of Anopheles gambiae populations from East and west Africa: comparison of microsatellite and allozyme loci. *Heredity*, 77 (Pt 2), pp.192–200. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8760401.

Lehmann, T. & Diabate, A., 2008. The molecular forms of Anopheles gambiae: a phenotypic perspective. *Infection, genetics and evolution*, 8(5), pp.737–746. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2731232&tool=pmcentrez&rendertype=abstract [Accessed January 20, 2014].

Lounibos, L.P. & Conn, J.E., 2000. Malaria Vector Heterogeneity in South America. *AMERICAN ENTOMOLOGIST*, 46(4), pp.238–249.

Lukashevich, E.D. & Mostovski, M.B., 2003. Hematophagous Insects in the Fossil Record. *Paleontological Journal*, 37(2), pp.153–161.

Mahmood, F. & Reisen, W.K., 1982. Anopheles Stephensi (Diptera: Culcidae): Changes in male mating competence and reproductive system morphology associated with aging and mating. *Journal of*

Medical Entomology, 19(5), pp.573–588.

Maïga, H., Niang, A., Sawadogo, S.P., et al., 2014. Role of nutritional reserves and body size in Anopheles gambiae males mating success. *Acta Tropica*, 132(1), pp.102–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24021933.

Mallet, J., 2006. What does Drosophila genetics tell us about speciation? *Trends in Ecology & Evolution*, 21(7), pp.386–393. Available at: http://linkinghub.elsevier.com/retrieve/pii/S016953470600156X.

Manoukis, N.C. et al., 2014. Stereoscopic video analysis of Anopheles gambiae behavior in the field: Challenges and opportunities. *Acta Tropica*, 132(1), pp.80–85.

Manoukis, N.C. et al., 2009. Structure and Dynamics of Male Swarms of Anopheles gambiae. *J Med Entomol*, 46(2), pp.227–235.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17, pp.10–12. Available at: http://journal.embnet.org/index.php/embnetjournal/article/view/200/479.

Maydeu-Olivares, A. & McArdle, J.J., 2005. Contemporary Psychometrics. , p.582. Available at: http://books.google.com/books?hl=en&lr=&id=y965h25v1VIC&pgis=1.

McKenna, A. et al., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next---generation DNA sequencing data. *Genome Research*, (20), pp.1297–1303.

Morin, P.A. et al., 2008. Complete mitochondrial genome phylogeographic analysis of killer whales (Orcinus orca) indicates multiple species. *Genome Research*, (20), pp.908–916.

Mouline, K. et al., 2012. Physiology and development of the M and S molecular forms of Anopheles gambiae in Burkina Faso (West Africa). *Medical and Veterinary Entomology*, 26(4), pp.447–454.

Mwangangi, M.J. et al., 2004. Relationships between body size of Anopheles mosquitoes and Plasmodium Falciparum sporozoite rates along the Kenya coast. *Journal of the American Mosquito Control Association*, 20(4), pp.390–394.

Nei, M., 1987. *Molecular evolutionary genetics: Columbia University Press, NY, USA.*, USA: Columbia University Press,NY.

Nijhout, H. & Craig, B., 1971. Reproductive Isolation in Stegomyia Mosquitoes. Iii Evidence for a Sexual Pheromone*. *Entomologia Experimentalis et Applicata*, 14(4), pp.399–412. Available at: http://doi.wiley.com/10.1111/j.1570-7458.1971.tb00178.x.

Nishimoto, Y. et al., 2008. Evolution and phylogeny of the heterogeneous cytosolic SSU rRNA genes in the

genus Plasmodium. *Molecular phylogenetics and evolution*, 47(1), pp.45–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18334303 [Accessed February 5, 2014].

Noor, M. a et al., 2001. The genetics of reproductive isolation and the potential for gene exchange between Drosophila pseudoobscura and D. persimilis via backcross hybrid males. *Evolution; international journal of organic evolution*, 55(3), pp.512–521. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11327159.

Nosil, P., Funk, D. & Ortiz-Barrientos, D., 2009. Divergent selection and heterogeneous genomic divergence. *Molecular ecology*, 18, pp.375–402.

Nwakanma, D.C. et al., 2013. Breakdown in the process of incipient speciation in Anopheles gambiae. *Genetics Society of America*, 193, pp.1221–1231. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3606099&tool=pmcentrez&rendertype=abstract [Accessed May 9, 2014].

Oksanen, J. et al., 2011. vegan: Community Ecology Package.

Pennetier, C. et al., 2010. “Singing on the Wing” as a Mechanism for Species Recognition in the Malarial Mosquito Anopheles gambiae. *Current Biology*, 20(2), pp.131–136.

Poinar, G.O. et al., 2000. Paleoculicis minutus (Diptera : Culicidae) n . gen ., n . sp ., from Cretaceous Canadian amber , with a summary of described fossil mosquitoes. *Acta Geologica Hispanica*, 35(1-2), pp.119–128.

Pombi, M. et al., 2006. Variation in recombination rate across the X chromosome of Anopheles gambiae. *American Journal of Tropical Medicine and Hygiene*, 75, pp.901–903.

Qvarnstrom, A. & Bailey, R., 2009. Speciation through evolution of sex--- linked genes. *Heredity*, 102, pp.4–15.

Rai, K.S., 1999. Genetics of mosquitoes. *Journal of Genetics*, 78(3), pp.163–169.

Reisen, W.K. & Aslamkhan, M., 1975. Observations on the swarming and mating behaviour of Anopheles culicifacies Giles in nature *. *wourld health organisation*, 54, pp.155–158.

Renshaw, M., Service, M. & Birley, M., 1994. Size variation and reproductive success in the mosquito Aedes cantans. *Med Vet Entomol*, 8(2), pp.179–186.

Rieseberg, L., 2001. Chromosomal rearrangements and speciation. *Journal of Chemical Information and Modeling*, 16, pp.351–358.

Roth, L.M., 1948. A Study of Mosquito Behavior’ An Experimental Laboratory Study of the Sexual

Behavior of Aedes aegypti (Linnaeus). *American Midland Naturalist*, 40(2), pp.265–352.

Roux, O., Diabaté, A. & Simard, F., 2013. Larvae of cryptic species of Anopheles gambiae respond differently to cues of predation risk. *Freshwater Biology*, 58(6), pp.1178–1189.

Rueda, L.M., 2008. Global diversity of mosquitoes (Insecta: Diptera: Culicidae) in freshwater. *Hydrobiologia*, 595(1), pp.477–487. Available at: http://link.springer.com/10.1007/s10750-007-9037-x [Accessed January 21, 2014].

Rundle, D.H. & Nosil, P., 2005. Ecological speciation. *Ecology Letters*, 8, pp.336–352.

Russell, P.F. & Rozeboom, E., 1943. Keys to the Anopheline mosquitoes. *The American Entomological Society The Academy of Natural Sciences*.

Sanford, M.R. et al., 2011. Morphological differentiation may mediate mate-choice between incipient species of Anopheles gambiae s.s. *PloS one*, 6(11), p.e27920. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3221689&tool=pmcentrez&rendertype=abstract [Accessed June 9, 2014].

SAS Institute Inc, 2008. SAS/STAT 9.2 User's Guide The DISCRIM Procedure.

Sawadogo, P.S. et al., 2014. Swarming behaviour in natural populations of Anopheles gambiae and An. coluzzii: Review of 4 years survey in rural areas of sympatry, Burkina Faso (West Africa). *Acta Tropica*, 130(1), pp.24–34. Available at: http://dx.doi.org/10.1016/j.actatropica.2013.12.011.

Sawadogo, S.P. et al., 2013. Differences in timing of mating swarms in sympatric populations of Anopheles coluzzii and Anopheles gambiae s.s. (formerly An. gambiae M and S molecular forms) in Burkina Faso, West Africa. *Parasites & vectors*, 6, p.275. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3851435&tool=pmcentrez&rendertype=abstract.

Schluter, D., 2001. Ecology and the originof species. *Trends in Ecology & Evolution*, 16(7), pp.372–380.

Seehausen, O. et al., 2014. Genomics and the origin of species. *Nature Reviews Genetics*, 15, pp.176–192.

Servedio, M., 2009. The role of linkage disequilibrium in the evolution of premating isolation. *Heredity*, 102, pp.51–56.

Service, M.W., 2004. *Medical entomology for students* Third edit., Cambridge.

Sheets, D., 2010. A Practical introduction to landmark-based geometric morphometrics. *The Paleontological Society*, 16, pp.163–188.

Shiff, C., 2002. Integrated Approach to Malaria Control. *American Society for Microbilogy*, 15(2), pp.278–

292.

Shutt, B. et al., 2010. Male accessory gland proteins induce female monogamy in anopheline mosquitoes. *Medical and Veterinary Entomology*, 24(1), pp.91–94.

Singh, B.N., 2012. Concepts of species and modes of speciation. *Current Science*, 103(7), pp.784–790.

Sinka, M.E. et al., 2012. A global map of dominant malaria vectors. *Parasites & vectors*, 5(1), p.69. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3349467&tool=pmcentrez&rendertype=abstract [Accessed January 30, 2014].

Slotman, M.A. et al., 2007. Evidence for subdivision within the M molecular form of Anopheles gambiae. *Molecular Ecology*, 16(3), pp.639–649.

Slotman, M.A. et al., 2006. Reduced recombination rate and genetic differentiation between the M and S forms of Anopheles gambiae s.s. *Genetics*, 174(4), pp.2081–93. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1698612&tool=pmcentrez&rendertype=abstract.

Smadja, C., Galindo, J. & Butlin, R., 2008. Hitching a lift on the road to speciation. *Mol. Ecol.*, 17, pp.4177–4180.

Steven, L.L. & Lawrence, M.D., 1990. Behavioral decisions made under the risk of predation: a review and prospectus. *Canadian Journal of Zoology*, 68(4), pp.619–640.

Stump, A. et al., 2005. Centromere--- proximal differentiation and speciation in Anopheles gambiae. *Proceedings of the National Academy of Sciences of the United States of America*, 102, pp.15930–15935.

Takken, W.K., 2009. Malaria vector control: Current and future strategies. *Trends Parasitol*, 25, pp.101–104.

Taylor, C. et al., 2001. Gene flow among populations of the malaria vector, Anopheles gambiae, in Mali,West Africa. *Genetics.*, 157(2), pp.743–750.

Thévenaz, P., 2003. Point Picker (ImageJ).

Thorvaldsdottir, H., Robinson, J. & Mesirov, J., 2013. Integrative Genomics Viewer (IGV): high---performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14, pp.178–192.

della Torre, A. et al., 2001. Molecular evidence of incipient speciation within Anopheles gambiae s . s . in West Africa. *Insect molecular biology*, 10(February 2000), pp.9–18.

della Torre, A. et al., 2002. Speciation Within Anopheles gambiae-- the Glass Is Half Full. *Science*,

298(5591), pp.115–117.

della Torre, A., Tu, Z. & Petrarca, V., 2005. On the distribution and genetic differentiation of Anopheles gambiae s.s. molecular forms. *Insect biochemistry and molecular biology*, 35(7), pp.755–69. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15894192 [Accessed March 28, 2014].

Touré, Y.T. et al., 1998. The distribution and inversion polymorphism of chromosomally recognized taxa of the Anopheles gambiae complex in Mali, West Africa. *Parassitologia*, 40(4), pp.477–511. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10645562.

Tripet, F. et al., 2001. DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of Anopheles gambiae. *Molecular Ecology*, 10, pp.1725–1732.

Tripet, F. et al., 2004. The "wingbeat hypothesis" of reproductive isolation between members of the Anopheles gambiae complex (Diptera: Culicidae) does not fly. *Journal of medical entomology*, 41, pp.375–384.

Tripet, F., Thiemann, T. & Lazaro, G.., 2005. Effect of Seminal Fluids in Mating Between M and S Forms of Anopheles gambiae. *Entomological Society of America*, 42(4), pp.596–603.

Turner, T.L. & Hahn, M.W., 2010. Genomic islands of speciation or genomic islands and speciation? *Mol. Ecol.*, 19, pp.848–850.

Turner, T.L. & Hahn, M.W., 2007. Locus- and Population-Specific Selection and Differentiation between Incipient Species of Anopheles gambiae. *Mol. Biol. Evol.*, 24(9), pp.2132–2138.

Turner, T.L., Hahn, M.W. & Nuzhdin, S. V, 2005. Genomic Islands of Speciation in Anopheles gambiae. *PLoS Biology*, 3(9), pp.1572–1578.

Vamosi, M.S., 2005. On the role of enemies in divergence and diversification of prey: a review and synthesis. *Canadian Journal of Zoology*, 83(7), pp.894–910.

Via, S., 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Royal Society*, 367, pp.451–460.

Vlachou, D. et al., 2006. The developmental migration of Plasmodium in mosquitoes. *Current Opinion in Genetics & Development*, 16, pp.384–391.

Wang, R. et al., 2001. When genetic distance matters: measuring genetic differentiation at microsatellite loci in whole-genome scans of recent and incipient mosquito species. *PubMed*, 19, pp.10769–74.

Weetman, D. et al., 2014. Contemporary gene flow between wild An. gambiae s.s. and An. arabiensis. *Parasites & vectors*, 7(1), p.345. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4124135&tool=pmcentrez&rendertype=abstract.

Weetman, D. et al., 2012. Gene flow-dependent genomic divergence between Anopheles gambiae M and S forms. *Molecular biology and evolution*, 29(1), pp.279–291. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3259608&tool=pmcentrez&rendertype=abstract.

Werren, J.H., Baldo, L. & Clark, M.E., 2008. Wolbachia: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, 6(10), pp.741–751. Available at: http://www.nature.com/doifinder/10.1038/nrmicro1969.

White, B. et al., 2007. Molecular karyotyping of the 2La inversion in Anopheles gambiae . *Am J Trop Med Hyg.*, 76(2), pp.334–339.

White, B. et al., 2009. The population genomics of trans-specific inversion polymorphisms in Anopheles gambiae . *Genetics*, 183(1), pp.275–288.

White, B.J. et al., 2010. Genetic association of physically unlinked islands of genomic divergence in incipient species of Anopheles gambiae. *Molecular Ecology*, 19(5), pp.925–939.

WHO, 1999. Making a Difference. , pp.49–63.

Wondji, C., Simard, F. & Fontenille, D., 2002. Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of Anopheles gambiae in an area of sympatry. *Insect Molecular Biology*, 11(1), pp.11–19.

Wu, C., 2001. The genic view of the process of speciation. *J . EVOL. BIOL.*, 14(6), pp.851–865.

Wu, C.-I. & Ting, C.-T., 2004. Genes and speciation. *Nature Reviews Genetics*, 5(2), pp.114–122. Available at: http://www.nature.com/doifinder/10.1038/nrg1269.