Elucidating Mechanisms of Carbamate Resistance and Carbamate/Pyrethroid Cross Resistance in *An. funestus* in Africa

By

Miranda Ndula

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Liverpool School of Tropical Medicine 2014

Supervisors

Dr. Charles Wondji

Prof. Hilary Ranson

DECLARATION

This work has not been previously accepted in substance for any degree and is not being currently submitted in candidature for any degree.

Signed...MND...(Candidate)

Date....................................21/09/2015..

Statement 1

This thesis is the result of my own investigation, except where otherwise stated. Other sources are acknowledged and bibliography is appended.

Signed...MND...(Candidate)

Date....................................21/09/2015..

Statement 2

I hereby give my consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed...MND...(Candidate)

Date....................................21/09/2015..
.

Abstract

Malaria remains one of the most debilitating tropical diseases with more than 90% of cases in Africa among children under five and pregnant women. Resistance observed against the main insecticides used in public health sector in major vectors such as *Anopheles funestus* is threatening the success of vector control interventions. To improve the design of suitable resistance management strategies, it is crucial to elucidate the underlining molecular basis of resistance or cross-resistance between insecticides and also establish patterns of gene flow between populations to predict the speed and direction of spread of resistance genes. To address these questions, this study has investigated the molecular basis of resistance to carbamates and cross/resistance to carbamates/pyrethroids in a population of *An. funestus* from Malawi. This study has revealed that metabolic resistance is the main mechanism driving carbamate resistance through the over-expression of Cytochrome P450 genes. Genome-wide microarray-based transcription analyses consistently revealed that the duplicated P450 genes *CYP6P9a* and *CYP6P9b* were among the most up-regulated genes (>2-fold; P<*0.01*) in bendiocarb and pyrethroid resistant mosquitoes from Malawi. Other P450 genes were also associated with both carbamates resistance and a carbamates/pyrethroid cross-resistance, including *CYP6Z1, CYP9J11*. Analysis of polymorphism patterns of *CYP6P9a* and *CYP6P9b* genes between resistant and susceptible mosquitoes revealed a correlation with permethrin resistance but not for bendiocarb resistance. In addition a reduced penetration mechanism was shown to be likely involved in carbamate resistance through the elevated expression of several cuticular protein genes. An Africa-wide genetic structure of *An. funestus* populations was established using microsatellite markers. This study revealed that *An. funestus* populations from Malawi and Mozambique from southern Africa are genetically closer to each other than the rest of the other African populations from West, Central and East Africa. Two loci AFUB6 and FUNR that are located around same position as the *CYP6P9a* and *CYP6P9b* genes on the rp1 QTL on the 2R chromosome are under selection possibly because of pyrethroid resistance due to their low allele number and reduced heterozygosity. Statistical analysis revealed that both loci were under possible selection and sequencing analysis also revealed that *CYP6P9a* gene is under selection for the resistant *An. funestus* from southern Africa compared to more susceptible population from Cameroon.

The possible cross-resistance observed in this study highlights the need to further elucidate resistance mechanisms before implementing suitable resistance management strategies such as rotation of insecticides to ensure a continued effectiveness of control interventions.

Dedication

To my parents, Mr / Mrs Denis / Regina Njiforfut and my beloved late sister, Balbine Kuna for their love, sacrifice and support.

Acknowledgement

This stage of my scientific development could never have been possible without the guidance of my primary supervisors Dr. Charles Wondji for his ever readiness to impact me with vital knowledge and also his readiness to listen, help and support me throughout the process and also my secondary supervisor Prof. Hilary Ranson who has given me countless comments and valuable advices leading to noticeable improvement of my presentation and writing style.

To Helen Irving, I say a big thank you for taking me through efficient and speedy laboratory techniques. Her love and trust in me played a very major role in enjoying all my days at LSTM. I won't forget all the gossips and latest updates we shared not only as best of friends but also as sisters.

I would like to thank all the members of the vector group especially my colleagues (Intan, Kayla and Sulaiman) and also Dr. David Weetman and Keith for helping me out with the Beckman machine without complaints during my microsatellite experiment. I will also like to thank John for collecting some of my samples, Jacob and Rudolf for their help.

To Dr.Alade for being my strength and my IT star. Thanks for your unconditional love and patience.

This PhD studies could never have been possible without my family for their continuous love, care and sacrifices they made for me. A big thank you to big bros Dr. Ni Eric, for being a good housemate and pick and drop driver. Many thanks also to all my nieces and nephews. To Dr. Thomas Ndula, Nana and Nino for taking care of the girls while I was away, I say thank you.

Finally, my heartfelt love to Lesiga, Bisona, Sonnah and Monique and Dodley for being patient and understanding as always throughout this process.

Thesis abbreviations

<: less than or equal too

%: Percentage

An: Anopheles

IRS: Indoor Residual Spray

LLINs: Long Lasting Insecticide Nets

P: Plasmodium

&: and

ID: Identification

pers Comm: Personal communication

DDT: dichlorodiphenyltrichloroethane

WHO: World Health Organisation

e.g: for example

HCH: hexachlorocyclohexane

i.e: That is

BHC: benzene hexachloride

GABA: Gamma aminobutyric acid

MNOs: monooxygenases

AChE: acetylcholinesterase

UV: ultra violet

PBO: piperonyl butoxide

M: Metarhizium

Rdl: resistance to dieldrin

CYPs: Cythochromes

COEs: carboxylesterases

GST: glutathione s-transferases

>: greater than or equal to.

NADPH: nicotinamide adenine dinucleotide phosphate-oxidase

DNA: Dioxyribonucleic acid

DEF: S, S, S-tributylphosphorotrithioate

TPP: triphenyl phosphate

α: alpha

ß: beta

CPLC8: Cuticular Protein 8 of Low Complexity Family

F_1: first generation.

QTL: Quantitative trait loci

rp1: resistance to pyrethroid

FANG: Fully laboratory susceptible strain of *An. funestus* from Angola

FUMOZ-R: Fully resistant strain of *An. funestus* from Mozambique

VGSC: Voltage-Gated Sodium Channel

kb: kilo base

mRNA: messenger Ribonucleic acid

Cy: Cyanine

SNP: Single Nucleotide Polymorphism

MRR: mark/release/recapture

MtDNA: mitochondria DNA

F_{IS}: That proportion of the total inbreeding within a population due to inbreeding within sub-populations

F_{st}: That proportion of the totalinbreeding in a population due to differentiation among sub-populations

F_{IT}: The total inbreeding in a population due to both inbreeding within subpopulations, and differentiation among sub-populations.

LD: linkage disequilibrium

qRT-PCR: Real-Time quantitative PCR

Contents

List of Figures

List of Tables

CHAPTER ONE

General background

1.1 Introduction

Malaria is one of the most devastating tropical disease causing the death of more than 660,000 deaths each year with children <5years and pregnant women suffering the most (WHO 2012). This disease is also the leading cause of morbidity in Africa with an estimated 360 million annual cases (Murray *et al*. 2012, Snow *et al*. 2005). More than 80% of global malaria cases occur in Africa (WHO 2011c). Malaria is caused by parasite species of the genus *Plasmodium* which are transmitted to humans by female mosquitoes of the *Anopheles* genus. *Anopheles funestus s.s (An. funestus)* and *Anopheles gambiae s.s (An. gambiae)* are two of the main vectors widely distributed across most of tropical Africa (Michel *et al*. 2005, Gillies and De Meillon 1968).

An. funestus species is distributed from the Sahara desert to northern South Africa and is a major malaria vector in Africa (Gillies and De Meillon 1968). The control of this malaria vector in endemic countries is a major public health challenge and relies heavily on insecticide-based control interventions such as Indoor Residual Spraying (IRS) and Long Lasting Insecticide Nets (LLINs). However, the continuous effectiveness of these control tools is threatened by the increasing reports of resistance to main insecticides used in the public health sector for the major malaria vectors such as *An. funestus* (Coetzee and Koekomoer 2013; Ranson *et al.* 2011). To maintain the efficacy of current insecticides, it is necessary to implement suitable Insecticide Resistance Management (IRM) strategies to limit the impact of such resistance on malaria transmission (WHO, 2012 GPIRM). A prerequisite for the design of such IRM strategies are reliant on the good understanding of the molecular basis of resistance and factors affecting the evolution and spread of resistance genes and mutations between populations of malaria vectors. This study was carried out to fill these important knowledge gaps in *An. funestus* and help improve future control of this malaria vector in Africa.

1.1.1 Malaria

Mosquitoes of the *Anopheles* genus; *An. funestus* and *An. gambiae* transmit the human malaria parasite. Transmission is through the spread of the *Plasmodium* (*P*) parasite namely, *P. falciparum*, (Welch 1897) *P. Malariae and P.vivax* (Grassi and Filetti 1890),

P.knowlesi and P. Ovale (Knowles and Das Gupta 1932) (Lee *et al.* 2011, Milner 2012). Among the recognised human parasites, *P. falciparum* is the only species associated with severe morbidity and mortality. The other species cause milder illness. However, infections with *P. ovale and P. vivax* may relapse months later if appropriate treatment is not provided (Bruce-Chwatt, 1963). Infected *Anophelinae* mosquitoes transmit the sporozoite stage located in their salivary glands to a human during a blood meal (Milner 2012). *P. falciparum* and *P. vivax* takes 48 hours to develop mature red blood schizonts (asexual form). Most released merozoites continue in the asexual cycle and infect new red blood cells, while a few differentiate into male and female gametocytes (sexual forms). These forms show no symptoms of malaria but circulate in the blood stream until they are ingested by a blood feeding *Anophelinae* mosquito. The life cycle is then completed in the mid gut of the *Anophelinae* mosquito when the gametes fertilize each other forming mobile ookinets which then leads to the formation of the oocysts and then the sporozoites that migrate to the salivary glands of the mosquito where they can cause re-infection of a human (Fig. 1.1) (Volkan *et al.* 2007).
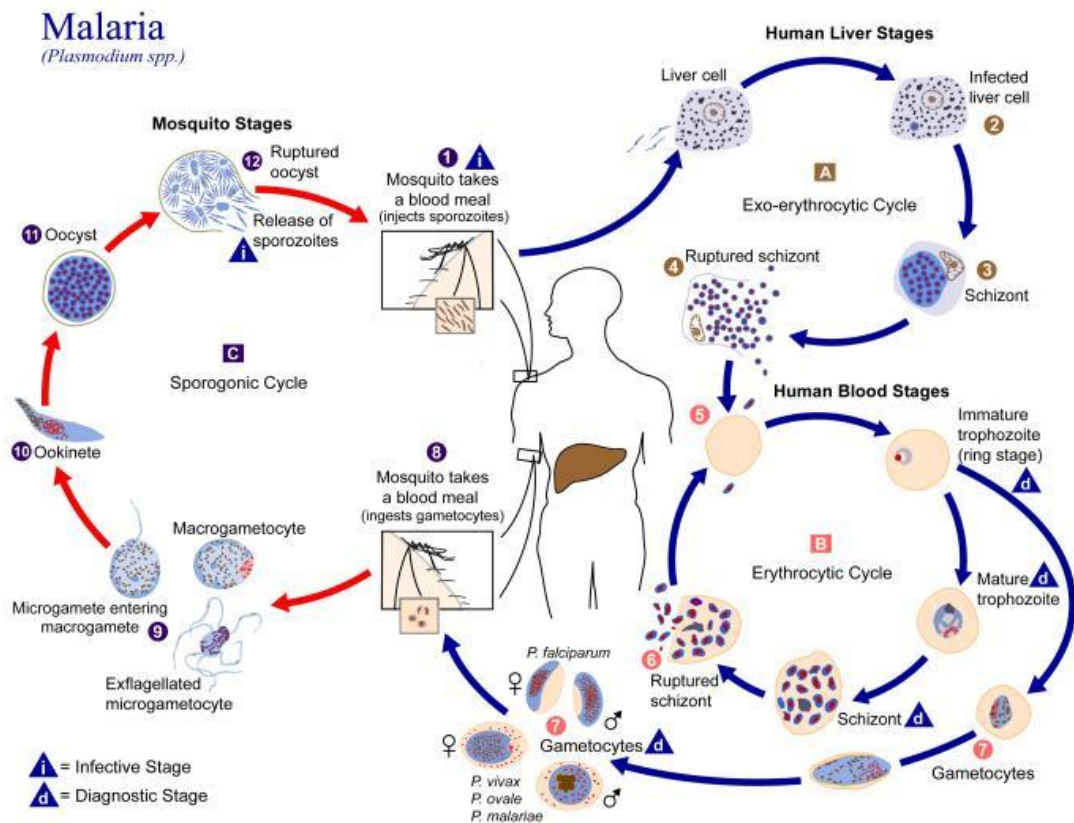
Fig. 1.1: The life cycle of *P. falciparum* (DPDx - Malaria Image Library)

1.2 Taxonomy

Mosquito vectors of the malaria parasites belong to the Arthropoda Phylum, Class *Insecta*, in the order *Diptera*, and Family *Culicidae* (Darsie 2005). The *Culicidae* family is divided into three sub families which comprises of *Anophelinae*, *Culicinae*, and *Toxorhynchitinae*. They are known to have 3450 identifiable mosquito species and 38 genera (Krzywinski and Besansky 2003). *Culicinae* alone is made up of 34 genera, *Anophilinae* 3 genera and *Toxorhynchitinae* a single genus. There are approximately 460 species of the *Anopheles* mosquitoes that have been identified worldwide. Eighty of which are capable of transmitting malaria (Forster and Walker 2002).

1.2.1 Factors affecting *Anopheles* distribution

Climate plays a major role in the distribution and abundance of *Anopheles* but other factors like breeding sites, local climatic conditions and human activities also contribute to the abundance and distribution of mosquitoes. Deforestation, irrigation, agriculture, construction activities are all contributing factors that promote the distribution and abundance of malaria vectors. For example the spread of *An. funestus* across Africa is limited by desert, saltwater (coastal mangrove), high elevation, and dense vegetation of rainforest that shades potential breeding sites (Coluzzi, 1994). Human activities like agriculture and other activities like construction of dams, irrigation and rice cultivation, have supported the distribution of *Anopheles* (Powel *et al*.1999).

1.2.2 Feeding patterns

Adult female mosquitoes bite humans (or animals) and take a blood meal for the maturation of their eggs. Some mosquito species like *An. quadriannulatus* are zoophilic, meaning they feed on animal blood while others such as *An. funestus and An. gambiae* female mosquitoes feed on human blood and are considered anthropophilic (Salvan and Mouchet 1994, Takken and Knols 1999). Some *Anopheles* such as *An. funestus* and *An. gambiae* rest indoors on house walls (endophilic), while others such as *An. arabiensis* preferably rest outdoor (exophilic) after a blood meal. The females normally bite at night and dawn. The malaria vector has the capacity of flying for up to 3km to get a blood meal (Walker 2002). The males feed on nectar and plant juices making them less important since they do not transmit malaria disease.

1.2.3 Developmental stages

The *Anopheles* mosquitoes have four developmental stages; egg, larvae, pupae and adult. The egg hatches into the larvae, which develops to the pupae and the pupae then transforms in to an adult mosquito. The larva has a well developed head and mouth brushes used for feeding. They feed on algae and bacteria and are aquatic at this stage. The pupa is comma shaped and will often come to the surface of water to breathe. The adult mosquito emerges when the dorsal surface of the cephalothorax of the pupa splits. The size of an adult mosquito depends on the density of the larval population and food supply of the breeding site. Mating among adult mosquitoes takes place at dusk within a few days after eclosion, where the males form a huge swarm and the females fly into the swarm to mate (Charlwood and Jones 1979) Parameters like colour, humidity and presence of certain chemical volatiles play a crucial role in the choice of female breeding sites (Serandour *et al.* 2010). In many species, the mosquitoes breed in temporary sites where water depth is very important and limited to predators. Adult mosquito life span varies but the males usually live up to a week while the females can live for up to a month in the wild and in captivity, the adult females can live for more than a month (Koekemoer *et al*. 2002).

1.3 Vector complexity

The high rate of malaria transmission in Africa is due to the presence of very efficient and competent vector system (Brooke *et al.* 2001). Most of the major malaria vector species belong to the *An. gambiae* species complex*, Giles* and the *An. funestus* group*, Giles* (Koekemoer *et al*. 2002, Zahar 1985). These two species complexes are composed of morphologically, genetically and behaviourally separate species, with different ability to transmit malaria (Coluzzi *et al*. 1979, Hunt *et al*. 1998). The vectorial and behavioural variation found within these groups of species or complexes is the major reason why accurate identification of these species is necessary for an assessment of malaria transmission risks (Koekemoer *et al*. 2002, Boccolini *et al.* 1998, Braginets *et al*. 2003).

1.3.1 *An. gambiae s.l* complex

The *An. gambiae* complex, known as *Anopheles gambiae s.l* (*Sensu lato*), until recently consisted of seven recognised species morphologically similar but that differ in behaviour, feeding preferences and breeding requirements (Gillies and De Meillon, 1968,

Gillies and Coetzee 1987, Hunt *et al*. 1998). Species within this complex include the major vectors *An. arabiensis Patton*, *An. gambiae s.s*, *Giles* the minor vectors *An. merus Donitz*, *An. melas Theobald* and *An. bwambae White*, and the non-vectors *An. quadriannulatus* (*Theobald*) species *A* and *B* (White 1972, Zahar 1985, Hunt *et al*. 1998). However, two new members (*An. colluzzi* and *An. amharicus*) have been recently included into the *An. gambiae* complex (Coetzee *et al*. 2013) meaning the *An. gambiae s.l* complex is made up of nine members *An. merus Donitz* in East Africa and the *An. melas Theobold* in West Africa are both salt water breeders and are preferentially zoophilic. In the absence of animals, they will turn and feed on humans (Coetzee *et al*. 2000, Berzosa *et al*. 2002). *An. bwambae White* is halophilic and breeds only in hot mineral water (Davidson and White 1972). This species has been described in Uganda, where it is an important local vector (White and Rosen 1973, Scott *et al*. 1993). *An quadriannulatus B* is named as a fresh water breeder found in Ethiopia (Hunt *et al*. 1998). However, based on chromosomal, cross mating and molecular evidence, this species has recently been renamed in Ethiopia to *An. amharicus* Hunt, Wilkerson and Coetzee, while the name has been retained in the southern African populations of this species (Coetzee *et al*. 2013). Ecological and behavioural diversification shown by *An. gambiae* is associated to fixed chromosomal inversions which can be used to facilitate the species identification (ID) in the field (Coluzzi *et al*. 2002 and 1979). Originally, *An. gambiae s.s.* was classified into various chromosomal forms (Bryan *et al*. 1982, Favia *et al*. 1997, della Torre et *al*. 2001) which was based on inversion karyotypes from intra-specific chromosomal inversions. Five sub-populations were named as chromosomal forms: Forest, Bissau, Bamako, Savannah and Mopti (Bryan *et al*. 1982, Coluzzi *et al*. 1985, della Torre *et al*. 2001). Further genetic complexity has been described in *An. gambiae s.s* to be made up of two molecular forms, the M and the S forms based on differences on the X chromosome (della Torre *et al*. 2001). Recent studies based on molecular and bionomical evidence has re-named the *An. gambiae* molecular "M form" as *An. coluzzii Coetzee* and *Wilkerson*, while the "S form" retains the *An. gambiae Giles* nominotypical name (Coetzee *et al*. 2013).

1.3.2 *An. funestus Giles* group

An. funestus is one of the main sub-Saharan malaria vector with an infection rate sometimes higher than that of *An. gambiae* s.s. (Cohuet *et al*. 2004, Fontenille and Lochouarn, 1999, Mendis *et al*. 2000). It belongs to a group composed of five sub groups

made up of thirteen morphologically similar species that can only be distinguished at specific developmental stages (Dia *et al.* 2013, Harbach 2012, Gillies and Coetzee 1987, Gillies and De Meillon 1968). The group members are *An. funestus s.s* Giles*., An confusus* Evans *and* Leeson*,* and *An. rivulorum* Leeson which were found around 1930s. *An. parensis* Gillies*, An. aruni* Sobti and *An. brucei* Service were found in the 1960s. *An. vaneedeni* Gillies and Coetzee, *An. longipalpis* type C, *An. longipalpis* type A, *An. leesoni* were found in1987 (Gillies *et al.* 1987). *An. fuscivenosus Leeson, An. leesoni Evans,* and *An. rivulorum-like* has also been identified (Cohuet *et al.* 2003). There has also been the identification of *An. funestus*-like in Malawi (Spillings *et al.* 2009). Among these species, *An. leesoni* is the most distinct at both egg and larva stage, while *An. confuses* is easily identified on larva characteristics (Coetzee and Fontenille 2005). *An. rivulorum* and *An. brucei* also have distinctive larvae. *An. fuscivenosus* is known only from the adult stage and by chromosomal banding arrangements that distinguish it from the other group members (Gillies and De Meillon 1968, Green 1982).

To morphologically identify members of the *An. funestus* group, egg batches have to be obtained from wild adult mosquitoes and the progeny should be reared through to adults (apart from *An. leesoni*, whose eggs are distinct). The fourth instar larvae and adults can be used in the identification process (Gillies and Coetzee 1987, Coetzee and Fontenille 2004). In this group, only *An. funestus s.s. and An. rivulorum Leeson* (Gillies and Coetzee 1987, Wilkes *et al*. 1996) play a role in the transmission of malaria. *An. funestus s.s.* has also been shown to transmit other diseases such *as Bancroftian filariasis* and Tanga virus (Gillies and De Meillon 1968, Awolola *et al*. 2005).

An. funestus has the widest distribution among the thirteen species followed by *An. leesoni and An. rivulorum* (Fig. 1.2). It extends across the entire sub-Saharan Africa (Dia *et al.* 2013, Gillies *et al*. 1968, Review Coetzee and Fontenille 2004). *An. confusus* are found in eastern Africa from Kenya and Tanzania in the North to KwaZulu/Natal province in South Africa (Gilies *et al*. 1968, Koekemoer 2002).

The other members of the group only occur locally; *An. vaneedeni* occurs only in Mpumalanga and the northern Province in South Africa and it rarely feeds on humans outdoors (1.22%) even though it has been experimentally infected with *P. falciparum* in the laboratory but not implicated in malaria transmission under natural conditions (De Meillon 1968). *An. aruni* is from Zanzibar in Tanzania, *An. fuscivenosus* from Zimbabwe

and *An. brucei* from Nigeria, *An. parensis* Gillies from Kenya and South Africa, *An. longipalpis* types A (South Africa) and C (Zambia). (Gillies *et al*. 1968, Gillies and Coetzee 1987). *An. funestus s.s.*is highly endophagic and endophilic (Gillies and De Meillon 1968, Gillies and Coetzee 1987) nocturnal feeders with maximum biting taking place between midnight and 4:00 am, but continuing until just after sunrise (Haddow 1943, Gillies and De Meillon 1968, Surtees 1970, Lindsay *et al*. 1989). It can be found in sympatry with other members of the group (Gillies *et al*. 1968 and 1987).
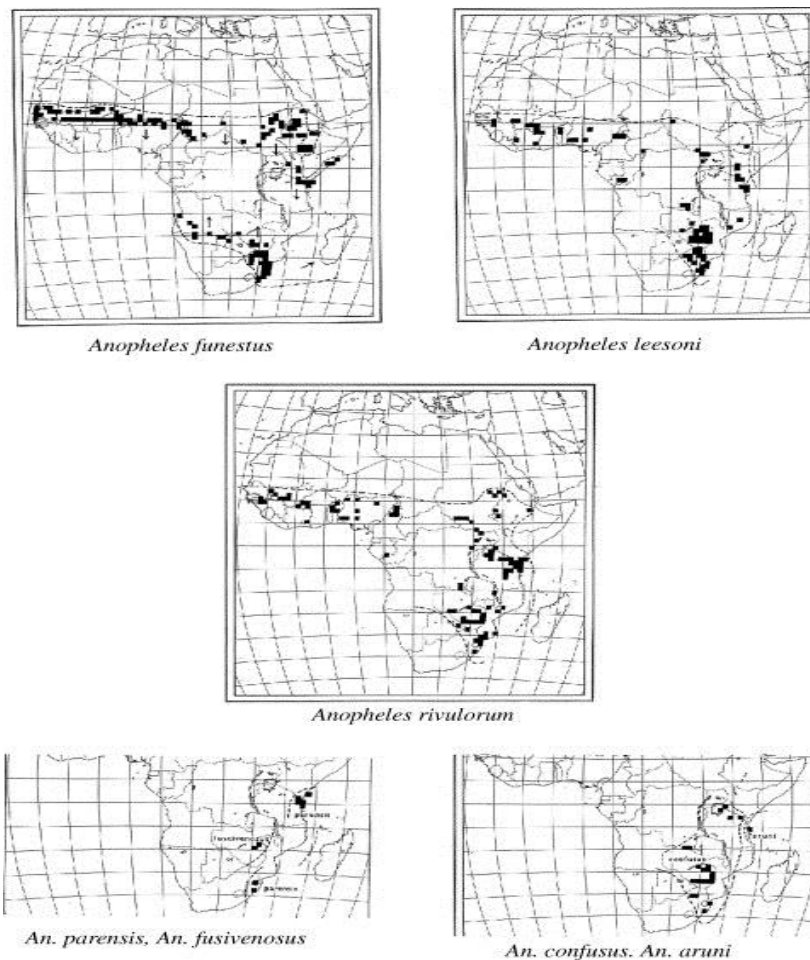
Fig. 1.2: Distribution of *An. funestus* group in Africa (Coetzee and Fontenille 2004)

The other members of the group are mainly zoophilic and only *An. rivulorum* has been implicated as a minor vector of malaria in Tanzania and is mainly zoophilic (77% animal host) (Wilkes *et al*. 1996, Koekemoer *et al*. 2002). Under laboratory conditions, *An. vaneedeni* has shown to be a vector (De Meillon *et al*. 1977), but has not been proven to be associated with malaria transmission in nature (Koekemoer 2002). Although fewer studies have focused on the genetic structure of *An. funestus* populations compared to that

of *An. gambiae*, evidences of population subdivision have been described in this species (Green and Hunt 1980).

1.4 Malaria control

Despite control and management efforts in most malaria affected regions, malaria still remains a major public health problem. Over recent years, through detail understanding of the ecology and epidemiology of malaria vector, there has been great changes and development of new approaches to combat malaria. Malaria like many other vector-borne diseases has no effective vaccine to control it therefore vector control is of great importance since malaria drugs are not always affordable by most people affected (Casimiro *et al*. 2006). Vector control is aimed at significantly reducing the incidences and prevalence of both parasite infection and clinical malaria by controlling the malaria bearing vector and as such altering or reducing transmission (WHO 2011). There are two main control intervention methods against malaria vectors. These are the use of LLINs and IRS (Pluess *et al*. 2010, Yukich *et al*. 2008, WHO 2007). Other vector control methods like larviciding, environmental management and biological control are also used.

1.4.1 Chemical control of adult mosquitoes with insecticides

The usage of chemical insecticides started after the World War II (WW II) with the discovery of DDT by Paul Hermann Muller 1939. This insecticide which belongs to the organochlorine class played a major role in the eradication of malaria in many countries.

1.4.1.1 Available insecticides, classification and mode of action

The first production of insecticide began in 1867 with the development and use of arsenical Paris green (Hemingway and Ranson 2000). The properties of the first insecticide DDT (dichlorodiphenyltrichloroethane) were discovered by Müller in 1939. The use of DDT saw its greatest benefit in the control of malaria where it was used as a pillar of the WHO (World Health Organisation) in the global campaign against malaria (Ware 2010). The first organochlorine insecticides (e.g. DDT-dichlorodiphenyltrichloroethane) were discovered in 1939 followed by the discovery of carbamates (e.g. bendiocarb) in 1953. The discovery of the first synthetic pyrethroids followed in 1960s and permethrin in 1972. These four insecticides form about 90% of the public health insecticide market. No new insecticides have been generated in the past 20

8

years and it is imperative to ensure the continuous effectiveness of the existing insecticides (Ranson *et al.* 2011).

Insecticides are classified according to their chemical structure and compositions. The four most used public health insecticides are presented below.

1.4.1.2 Organochlorines

Chlorinated hydrocarbon insecticides have molecular weights between 291 and 545, and are aryl, carbocyclic or heterocyclic compounds. The insecticidal properties of hexachlorocyclohexane (HCH), synthesized in 1825, were discovered in 1942. It was used seldom with the view of replacing DDT as resistance developed (Matsumura 1975). Cyclodienes, aldrin and dieldrin were synthesized in 1948 and Chlordane in 1945. These compounds function by inhibiting the normal function of the nervous system (Hemingway and Ranson 2000). These insecticides have persistent properties and are stable in soil while relatively stable also to the ultraviolet rays of sunlight. Unlike DDT and HCH, the cyclodienes have a positive temperature correlation, (i.e) they show increasing toxicity with increasing temperature. Dieldrin is more toxic than DDT and HCH to insects, human and animals, while less excito-repellent than DDT (Matsumura 1975). DDT and its allies acts on the sodium channels on the nervous membrane, while benzene hexachloride (BHC) and cyclodienes such as dieldrin act on the GABA receptor. This group of insecticides is chemically very simple thus making them easy to manufacture and cheap (Hemingway and Ranson 2000). Despite their affordability, their environmental persistence and toxicity to wildlife and human have drastically hindered their use since the 1970s.

1.4.1.3 Organophosphates

Insecticides of this group share a common chemical structure but differ in their physical, pharmacological properties and their use. They are a less stable group of insecticides than the organochlorines. Their discovery was made in 1854 but their insecticidal properties were only recognized by Schräder in 1937 (Matsumura 1975). They are ministered as inactive phosphorothiote through the action of monooxygenases (MNOs) in the presence of water and then activated within the insect to the insecticidal organophosphate (Hemingway and Ranson 2005). The organophosphate insecticide class acts by binding the enzyme acetylcholinesterase (*AChE*) at the nerve junction. Once there is binding,

acetylcholine can no longer be removed from the nerve-membrane junction by the enzyme and as a result, the nerves continue to fire in an unregulated manner that leads to paralysis and death. Common organophosphates used in vector control are the temephos, chlorpyrifos, malathion, fenitrothion and pirimiphos methyl (Ware 2010, Hemingway and Ranson 2005).

1.4.1.4 Carbamates

Carbamates have a similar mode of action like the organonophosphates since they both act on the acetylcholinesterase. However, carbamates are used in the insecticidal active form and are derived from carbamic acid. It was first introduced as an insecticide in the form of carbaryl in 1956. They have varied mammalian toxicity but they all have a common structure. They are environmentally less persistent and biodegradable. The most commonly used insecticide in this group for malaria control is propoxur and bendiocarb. It is commonly used in IRS and more so in areas where there are reports of DDT and pyrethroid resistance (Akogbeto *et al.* 2010). The carbamate bendiocarb is presently the sole insecticide used in this group for malaria vector control.

1.4.1.5 Pyrethroids

Pyrethroids have their insecticidal active component pyrethrins from pyrethrum flowers (*Chrysanthemum cinerariaefolium*) (Ray 1991). In the late 1940s and 1950s, many synthetic pyrethrins were made but were all unstable with exposure to ultra violet (UV) light. They are good and safe "knock-down agents" due to their lack of persistence and are still often being used in aerosols with the synergist piperonyl butoxide (PBO) which reduces their cost but increases their insecticidal activity (Hemingway and Ranson 2000). The basic structure of a pyrethroid has an alcohol and acid group derived from pyrethrins. Commercially, the pyrethroids are the most successful insecticidal group due to their extensive development. Some of the group members are stable to air and light making them suitable as residual insecticide (Hassall 1990). They function in very low doses in the insect making them easy and safe to handle. They act on the sodium channel in the same way as DDT and its analogues. They have a more pronounced stimulation than DDT, and represent the most common insecticide class used in the control of medically important insects. Pyrethroids are made up of two types; the type I (e.g. permethrin) and type II (deltamethrin). Examples of pyrethroids used for malaria control are permethrin, deltamethrin and lambda-cyhalothrin. It is the only insecticide class used for LLINs.

The two major ways of using chemical insecticides in vector control are in the form of LLINs and IRS.

1.4.2 Long Lasting Insecticide Nets (LLINs)

Malaria control in most of sub-Saharan Africa currently relies on LLINs (WHO 2011c). All current LLINs (Fig. 1.3) are treated with pyrethroid insecticides which are the only insecticide class recommended by WHO for impregnation of LLINs (WHO 2012). A high coverage of LLINs can result to a high reduction of malaria (Pluess *et al.* 2010, Lengeler 2004). This method has been associated to a decrease in malaria in countries that have high coverage and has been shown to reduce childhood malaria (below five years of age) morbidity by 50% and childhood mortality by 20%-30% (Dia *et al.* 2013, Lengeler *et al.* 2004). There are many ways in which the LLINs act in order to control the vectors. It initially acts by protecting the individual sleeping under the net by preventing the vector from reaching out to the individual as it acts as a physical barrier and causes the death of mosquitoes due to its insecticidal properties. It also reduces the amount of bites that an individual can get during the course of the night and also its excito-repellent properties that causes the mosquitoes to stay away from the nets significantly limiting vector/human contacts (Curtis 2004, Curtis 1998, Lines 1987, Rozendaal 1989a). LLINs play a very good role in protecting the community where there is a high coverage of above 80%. (Killeen and Smith 2007, Sharp *et al.* 2007).

Fig. 1.3: Long lasting indoor net (LLINs) (Google images)

1.4.3 Indoor Residual Spray (IRS)

IRS is another method used to control adult female mosquitoes. This method can be dated as far back as in the 1940's. In this case the inside walls, ceilings, and sometimes the outside eaves, shades and nearby animal sheds of the houses are sprayed with a persistent insecticide (Fig. 1.4) (N'guessan *et al*. 2007). This method targets the Anopheline mosquitoes that rest on the walls before and after a blood meal. Currently around 11% of sub-Saharan Africa relies on its use (N'Guessan *et al*. 2007). The spray is normally made up of an appropriate insecticide that is very toxic to the insects but safe to humans and animals. Many control programs used DDT for IRS for malaria control from 1945 to 1996 (Casimiro *et al*. 2006). The use of DDT started declining in the 1970's due to the introduction of pyrethroids. The DDT even though very effective, was also very toxic not only to the mosquitoes but also left a long lasting toxic effect on the environment. However, pyrethroids, despite being a good insecticide for malaria control, are available in formulations that can last just up to six months after spraying, thus requiring at least 2-3 rounds of IRS per year in endemic areas (Najera *et al*. 2002).

Fig. 1.4: Indoor residual spray (Google images)

1.4.4 Larviciding

This is a control method whereby, mosquito larvae are targeted by spraying the habitat with insecticides to control the larvae. This method is preferable if the larva sites are few and can be easily identified (Devine and Killen 2010). In this case, insecticides in the form of liquid or granules are applied directly into the water (example is the organophosphate temephos). The usage of insecticide can also be in the powder form and applied to the interior of the habitation (Enayati and Hemingway 2010). This method causes contamination of ecosystem and also has a negative effect on native fauna.

1.4.5 Biological control

Vector populations can be controlled using biological approaches such as predation, competition and parasitism. Biological control is generally achieved through the introduction and manipulation of various organisms to suppress the vector population. The principal biological agents that have been used against *Anopheles* are predators especially larvivorous fish like the mosquito fish *Gambusia affinis*, the guppy, *Poecilia reticulata* and *Tilapia* species *Aphyocypris chinensis* (Hemingway 2002), and the bacteria pathogen *Bacillus thuringiensis Isrealiensis* and *Bacillus sphaericus* that attack the larval stages of the mosquito (Hemingway 2002).

Fungi is also been used to control disease vectors. The enthomopathogenic fungus *Metarhizium anisopliae* has been used to target larvae of various adult mosquito species (Daoust *et al*. 2010, Silva *et al*. 2008) like *An. gambiae* and *Culex quinquefasciatus* (Scholte *et al.* 2003). The competence of *An. gambiae* in transmission was decreased by the *Metarhizium anisopliae* which also causes high mortality rates in dengue vectors (Daoust *et al.* 2008, Silva *et al*. 2008). Study carried out by Filliberto *et al*. (2011) also showed how infection of male *Aedes aegypti* with the fungus *M. anisopliae* drastically reduced the fecundity of dengue mosquitoes and was able to cause the death of a large number of females that were infected during mating attempts. The impact could be effective in reducing dengue transmission by reducing the number and longevity of female mosquitoes after releasing fungus-contaminated males in the dwellings. This method has not yet been tested against field populations to confirm its efficacy. The advantage of biological control over chemical control is that it is very safe to humans.

1.4.6 Environmental management

This approach was used to combat malaria over a long period of time. The method involves the alteration of vector breeding sites while also avoiding creation of breeding sites. It also involves modifying natural habitats like filling of marshes and grading, improving human habitation which involves house screening and as a result this will reduce the abundance of a target vector (Utzinger *et al*. 2001, Townson *et al*. 2005). This method of managing the malaria vectors was popularly used before and during World War II but hugely replaced by the use of synthetic insecticides to control vectors and use of drugs to treat the disease during the post war period (Graham 2005).

The difficulty of environmental management is the fact that, malaria poses the greatest problems in the poorest rural areas where traditional houses are common and these houses are also very friendly to the malaria vectors. Modification of human habitat can also be difficult in a situation where people sleep outside. This method can be a very good way of controlling vectors in Africa since it is cheaper and will involve the local communities and will also benefit them without any economic drain (Provost 1972, 1973).

1.5 Problems of insecticide resistance

Insecticides have played a great role in controlling major disease vectors but excessive usage has led to the development of insecticide resistance among vectors (Hemingway and Bates 2003). Resistance is a characteristic that is genetically inherited and increases in frequency in the vector population as a direct result of the effects of the insecticide. It is also a complex issue that comes through the effects of various mechanisms. Mutation or gene duplication causes variation in genes and phenotype that then modifies some of the normal physiological, behavioural or morphological aspects of the phenotype leading to the appearance of resistance (Guillemaud *et al*. 1998).

1.5.1 Types of resistance

The Insecticide Resistance Action Committee (IRAC) defines insecticide resistance as the selection of heritable characteristics in an insect population that results in the repeated failure of an insecticide product to provide the intended level of control when used as recommended. It is the ability in a strain of some organisms to tolerate doses of a toxicant that would prove lethal to a majority of individuals in a normal population of same species. There are different types of resistance which are commonly observed in mosquito populations as described below.

Cross-resistance occurs when a resistance mechanism allows the insects to be resistant to one insecticide as well as been able to confer resistance to compounds within the same class. It may also occur between chemical classes depending on the mechanisms (Lepoivre 2003). Cases are seen between DDT and pyrethroid insecticides which are chemically unrelated but both act on same target site which is the voltage gated sodium channel as observed in *Culex quinquefasciatus* (Coetzee and Koekomoer 2013). There is also cross resistance between carbamates and organophosphates as both acts on the *Ace-1* as seen in *An. culex* (Weil *et al.* 2004). This implies resistance to DDT can also cause

resistance to pyrethroid while resistance to carbamate might also cause resistance to organophosphate.

Multiple resistances are a common phenomenon occurring when several different resistance mechanisms are present simultaneously in resistant insects. The different resistance mechanisms may combine to provide resistance to multiple classes of products (Perera *et al*. 2008). Example has been observed by Edi *et al.* (2012) in *An. gambiae s.l* where DDT/pyrethroids target site (kdr) were found in conjunction with the resistance alleles of *Ace-1^R^*, which is the target site of organophosphates (Weil *et al.* 2004).

Many types of resistance mechanisms have been identified and could be behavioural, reduced penetration of insecticide, target-site insensitivity, increased metabolic detoxification, increased excretion and sequestration.

1.5.1.1 Target site resistance

 Decreased sensitivity of all major insecticide target sites, such as acetylcholinesterases (*AChEs*), the voltage-gated sodium channel gene and gamma aminobutyric acid (GABA) receptors have been reported in insects (McCaffery 1998, Vais*et et al*. 2001). Target site resistance involves the modification of the protein target site thus reducing the binding of insecticide. It is explained by the substitution of amino acids in the protein sequence of the targeted protein through a non-synonymous mutation (Ffrench-Constant *et al*. 2004). Target site insensitivity to an insecticide usually confers cross-resistance to other compounds that share the same target site. For example decreased sensitivity of the sodium channel to DDT confers cross-resistance to pyrethroids, which share the same target site (Brogdon *et al*. 1999, Chandre *et al*. 1999).

1.5.1.1.1 Acetylcholinesterase (*AChE*) modification

AChE is encoded by the acetylcholinesterase 1 (*Ace-1*) gene a target for insecticides of the organophosphate and carbamate classes. *AChE* is a serine esterase that hydrolyses the excitatory neurotransmitter acetylcholine that is found on the post-synaptic nerve membrane. The insecticide fixes on the enzyme that is no more capable of degrading inter-synaptic acetylcholine. The signal transmitted by the neuro-transmitter is not interrupted which then leads to paralysis of the insect and then death. Mutations on the *Ace-1* gene like G119S confer resistance to organophosphates and carbamates in many mosquitoes like *Culex pipiens* (Raymond *et al.* 1985, 1986, Bonning and Hemingway

1991) and *An. gambiae* (Weill *et al.* 2003). In *An. funestus*, evidence was found using a biochemical assay of an altered acetyl-cholinesterase from southern Mozambique but no mutation was detected (Casimiro *et al.* 2006 and 2007, Cuamba *et al.* 2010).

1.5.1.1.2 Modification of sodium gated channel

Several mutations in the sodium channel gene have been associated with the pyrethroids and DDT resistance. Among these mutations, substitutions in the 1014 codon are common in many insects. Two mutations L1014F and L1014S in segment S6 in the domain II of the sodium channel gene confer resistance to pyrethroid and DDT in *An. gambiae* in Africa. Mutation L1014F is mainly seen in *An. gambiae* populations in West Africa (Martinez-Torres *et al.* 1998) while the L1014S mutation is more frequent in *Anopheles* population in East Africa (Ranson *et al.* 2000). Other types of kdr mutations are present in other codons such as the 1575 in *An. gambiae* (Jones *et al.* 2012) and the F1534C in *Aedes ageypti* (Saavedra-Rodriquez *et al.* 2009). In a case where the insect strains are resistant, there will be no knockdown or paralysis but with the susceptible insects, exposure to DDT or pyrethroid leads to a rapid paralysis ("Knock down"). Lack of knock is caused by mutations in the para-gated sodium channel gene whose protein subunits make up the voltage-gated sodium channel on the nerve membrane (Brogdon and McAllister 1988, Hemingway and Ranson 2000). This mechanism leads to the cross-resistance of both pyrethroids and DDT (Miyazaki *et al.* 1996, Williamson *et al.* 1996).

1.5.1.1.3 Modification of GABA

The γ-butyric acid (GABA) receptor is the target for cyclodienes (dieldrin), some organochlorines (lindane) and phenylpyrazols (fipronil). GABA fixation on the receptor regulates the passage of chloride ions across the synaptic membrane. The cyclodienes fix on the GABA and block the passage of chloride ions leading to paralysis of the insect followed by death. An alanine substitution to serine in position 302 on the GABA receptor gene (ffrench-Constant *et al.* 1993) or alanine to glycine (Hosie *et al.* 1997) is associated with resistance to dieldrin (rdl). This mutation was found in *Musca domestica* (Anthony *et al.* 1998), *An. gambiae* (Du *et al.* 2005) and also in *An. funestus* (Wondji *et al.* 2011) where the rdl distribution is high in West/Central Africa, low in East Africa and absent in southern Africa.

1.5.1.2 Metabolic resistance

Metabolic resistance is characterised by an increase in the metabolism of an insecticide generally by a detoxification enzyme. A structural change in the enzyme molecule increases its ability to detoxify or bind to the insecticide and/or an increase in the amount of enzyme produced (Hemingway and Bates 2003). Increase in insecticide metabolism can be as a result of increase in detoxifying enzymes which increases the insecticide tolerance by the insect. In mosquitoes, three families of detoxifying enzymes are frequently implicated in insecticide metabolism; the monooxygenases cytochrome P450 (CYPs), the carboxylesterases (COEs) and the glutathione s-transferases (Hemingway *et al.* 2004). Their role and characteristics are discussed below.

1.5.1.2.1 Monooxygenases

CYP P450 enzymes represent a family of multifunctional enzymes found in every organism (Nelson 2013). They are implicated in a number of metabolic mechanisms including metabolism of certain xenobiotics such as medicine, pesticides, pollutants and plants toxins (Scott 1999). Due to the different reactions that they catalyse, many terms are used to call these enzymes in literature such as monooxygenases or cytochrome P450 (CYPs) with multiple function (Scott 1999). CYPs are hemethiolate membranic enzymes localised in the endoplasmic reticulum and the mitochondria of eukaryotes (Feyereisen 2005).

Mitochondrial CYPs uses a system of transfer that is different from that of the reticulum which is very close to soluble CYPs and the primitive bacteria forms (Scott 1999). Monooxygenases catalyses the transfer of an atom of an oxygen molecule on a substrate and reduces the other atom into water. To function, CYPs requires redox partners (Feyereisen 2005), and NADPH cytochrome P450 reductase which contains two flavine co-factors, adenine and a flavine mononucleotide. It is circled with many CYPs which it forms electrons with. CYPs are classified according to their protein sequence and this proposes that all members of a family share >40% identity at the amino acid level whilst members of a sub family share >55% identity like the CYP4 and CYP6 (Nelson *et al.* 1996). In *Drosophila melanogaster*, there are over 80 cytochrome P450s (Tijet *et al.* 2001), but this number varies between species e.g. *An. gambiae* 111 (Ranson *et al.* 2002a, Strode *et al.* 2008, Arensburger *et al.* 2011).

The main insecticides that these enzymes are known to confer resistance to are the pyrethroids, carbamates and to a lesser extent the organochlorines and organophosphates (Ranson and Hemingway 2000). Brooke *et al*. (2001) published data showing pyrethroids and carbamates resistance in *An. funestus* from Mozambique using the synergist piperonyl butoxide (PBO) was probably due to increase in monooxygenase activity. They also act by activating the phosphorothioate insecticides to the active organophosphate form. High levels of monooxygenases activity are associated with resistance in *An. funestus* (*CYP6P9a* and *CYP6P9b*) (Wondji *et al.* 2009), in *An. gambiae* (e.g. *CYP6P3, CYP6M2*) (Muller *et al*. 2008b) and *Culex quinquefasciatus CY9J10* and *CYP4H34* (Hard-Stone *et al*. 2007). High levels of monooxygenases in resistant insects are mainly as a result of over expression (Carino *et al*. 1994, Tomita and Scott 1995), and a case of P450 amplification has been observed in *Ae. aegypti* (Bariami *et al*. 2012), and also gene duplication in the crop pest *Myzus persicae* (Puinean *et al*. 2010) and *An. funestus* (Wondji *et al*. 2009).

1.5.1.2.2 Esterases

Esterases are mainly involved in organophosphates, carbamates and to a limited extent in pyrethroid resistance by hydrolysing their ester linkages (Hemingway *et al*. 1998, Vulule *et al.* 1999). This enzyme family has been studied in its involvement in insecticide resistance in the *Culex quinquefasciatus* mosquito. In this mosquito, it is reported to be linked to the high activity of one or two esterases. Standard substrates such as α/ß-naphthyl acetate can be used to measure esterase activity in individual insect (Cygler *et. al*. 1993). Esterases act effectively as an insecticide sink thus causing the rapid binding and slow metabolism of the insecticide. Amplification of two esterases takes place on same deoxyribonucleic acid (DNA) piece, a termed known as amplicon and are inherited as a single unit (Vaughan and Hemingway 1995). In bioassay, resistance caused by esterases can be synergized by DEF (S, S, S-tributylphosphorotrithioate). TPP (triphenyl phosphate) can synergize specific esterase-based resistance, which is limited to malathion and related insecticides that share a common carboxylester bond structure. This type of resistance is very common in *Anopheles* mosquitoes and the underlying genetic basis of this resistance is mutation within the active site region rather than amplification (Hemingway 1985).

1.5.1.2.3 Glutathione S-Transferases

Glutathione S-transferases (GSTs) are involved in DDT, pyrethroid and organophosphates resistance (Enayati 2005, Hayes and Wolf 1988). In insects, GSTs have been classified in to six groups (delta, epsilon, zeta, sigma, omega and theta) with an addition of unclassified GSTs. The delta and epsilon class are insect specific (Ranson *et al*. 2002a). The primary metabolites of DDT are thought to have resulted from dehydrochlorination and were related to increase in GSTs activity even though the connection between the two was not known. Vontas *et al.* (2001) detailed the role of GSTs in pyrethroid resistance by demonstrating that GSTs were effective at protecting the cells of the insect against lipid peroxide damage which will otherwise lead to pyrethroid-mediated free-radical activity. GSTe2 has been shown to be responsible for high DDT resistance in *An. funestus* in Benin and there was also the detection of a first DNA-based marker in the *funestus* DDT resistant mosquitoes, the L119F mutation with similar pattern of distribution like the rdl which is high in West/Central Africa, low in East Africa and absent in southern Africa (Riveron *et al*. 2014). Over-expression of GSTe2 has been associated to DDT resistance by resistant strains of *An. gambiae* (Ding *et al.* 2005) and *Ae. Aegypti* (Lumjuan 2005).

1.5.1.3 Reduced penetration or cuticular resistance

Cuticular resistance is conferred by a reduced penetration of insecticide through insects cuticle and it has physio-chemical characteristics, particularly, strong lipophilic properties. Many insecticides are formulated to penetrate the insect through the insect cuticle. Changes to the cuticle lead to reduction in the rate of penetration and as a result, confers resistance to many insecticides. Reduced penetration alongside other resistance mechanisms can lead to increase resistance. When the rate of insecticide penetration to target site of the insect is compromised, other mechanisms can detoxify the insecticide more easily (Brooke *et al.* 2001). Recently, transcription of the cuticular gene *CPLC8* in *An. gambiae* suggested the potential role in the resistance of the cuticle to pyrethroids (Djouaka *et al.* 2008). Another study of a microscopic measurement of the cuticle of *An. funestus* showed that the cuticle of resistant mosquitoes are thicker than those of susceptible ones (Wood *et al*. 2010).

1.5.1.4 Behavioural resistance

Behavioural resistance implies any modification in the behaviour of a mosquito that causes it to avoid surfaces that have been sprayed with insecticides (Hamon 1963). Behavioural changes of vectors to insecticides have been acknowledged previously (Sparks *et al.* 1989, Gould 2010). However, there has been lack of information on behavioural resistance and its medical importance. An example of documented behavioural resistance is the observed change of behaviour in some mosquito populations to feed outdoor in areas where there was extensive indoor use of insecticides like the case observed in Tanzania (Russel *et al.* 2011). This therefore limits the efficacy of LLINs and IRS control with mosquitoes preferentially feeding outdoor been active earlier in the evening before humans go to bed and also resting outside human houses sprayed with insecticides (Gatton *et al.* 2013). There are also other behavioural changes that come from possible intensive control interventions like zoophagy as observed in Senegal (Dia *et al.* 2013). Lack of information about behavioural resistance might be linked to difficulty to investigate using simple methods and difficulty to investigate in the field (Takken 2002, Ferguson *et al.* 2010). *An. funestus* has been observed to avoid contact with insecticide sprayed surfaces by modulating its biting behaviour from late night to the early hours (Moiroux *et al.* 2012). There have been significant increases in exophagy by *An. funestus* in Tanzania due to IRS that has led the mosquitoes to start feeding outside (Russell *et al.* 2011). It was observed that the proportion of *An. gambiae s.l* and *An. funestus* resting indoors and feeding before 2200h increased even though peak feeding still took place after midnight (Mbogo *et al.* 1996, Russell *et al.* 2011), and following the introduction of LLINs, the proportion of mosquitoes feeding early increased. Meaning it was difficult to control early feeders.

Even though *An. funestus s.s* exhibits an almost consistent host feeding (anthropophilic) and resting (endophilic) preferences throughout its range, there have been some behavioural differences associated with chromosomal polymorphisms which have been observed between some populations. One such example is the case reported in Lochouarn *et al.* (1998) where a West-East gradient was observed from human to animal biting preference coinciding with chromosomal polymorphism patterns that also follow this gradient. Different resting and biting activities were linked to different combinations of chromosomal inversions in Burkina Faso (Constantini *et al.* 1999). This study indicated that, those that predominantly feed on humans (anthropophilic) and rest indoors, were

carriers of inverted arrangements on 2R and 3R, while the others exhibited a much higher level of zoophily and exophily (Dia *et al.* 2013 in pers Comm with Guelbeogo). However, the carriers of inverted arrangements 3Ra and 3Rb were less anthropophilic than the standard arrangements carriers in Madagascar (Boccolini *et al.* 1992). The *An. funestus* population with the 3Ra and 3Rb was more zoophilic in Senegal even though heterogeneity to preference of host might be associated to specific local conditions such as availability of host or indoor climatic conditions such as humidity (Dia *et al.* 2000). In Tanzania and Ethiopia (East Africa), there have been reports of human feeding behaviour of *An. longipalpis* collections from indoor and outdoor (Adugna and Petros 1996). Kent *et al.* (2006) also reported that *An. longipalpis* remains mainly zoophilic even when found in huge numbers resting indoors with *An. funestus*.

1.6 Profile of insecticide resistance in *An. funestus* in Africa

There have been increasing reports of insecticide resistance across sub-Saharan Africa in malaria vectors. Resistance to pyrethroid is increasing in *Anopheles* mosquitoes in Africa. First reports about insecticide resistance for *An. funestus* were made from Nigeria where a population of *An. funestus* from the North was not eliminated despite three years of continuous spray with DDT, dieldrin and BHC (Bruce-Chwatt *et al.* as in service 1960). First report of pyrethroid resistance was made when *An. funestus* were found existing in pyrethroid- sprayed homes in the Kwazulu, Natal Province of South Africa which borders Mozambique and later there was the confirmation of pyrethroid resistance (Hargreaves *et al*. 2000, Brooke *et al*. 2001). This led to the re-introduction of DDT for IRS in 2000 to control *An. funestus*. Fig. 1.5 indicates impact of pyrethroid resistance on malaria control in southern Africa. In southern Africa, samples collected in Mozambique from 2002-2006, showed significant resistance to pyrethroid and carbamate (Casimiro *et al*. 2007). In 2009, another evaluation was carried out at another site in Mozambique (Chokwe) showing an increased level of resistance to pyrethroids but also to carbamates (Cuamba *et al.* 2010). For Malawi, only limited knowledge on resistance was available before this study. There have also been pyrethroid resistance reports in Zambia (Chanda *et al.* 2011). There are concerns that this resistance could rapidly spread across the continent posing a threat to pyrethroid-based malaria control programs (Wondji *et al*. 2012). In any case, up to recently there were no reports of DDT resistance in southern Africa but recent work detected such resistance in Zambia (Riveron *et al.* 2014).

In East Africa, Morgan *et al.* (2010) reported pyrethroid resistance of *An. funestus* in Uganda where there was over 60% survivals after exposure to pyrethroid insecticide and also resistance to DDT. These resistances were recently showed to be widespread across Uganda and also extent to Kenya (Mulamba *et al.* 2014) as initially reported in Western Kenya by Kawada *et al.* (2011). In Central Africa (Cameroon), there have been reports of resistance to dieldrin by Wondji *et al.* (2011). There were reports of insecticide resistance in West Africa; Ghana by Okoye *et al.* (2008), in Benin Djouaka *et al.* (2011) where *An. funestus* were seen to be resistant to carbamate, pyrethroid and highly resistant to DDT. In Mali, *An. funestus* also showed resistance to DDT, dieldrin and malathion (Toure 1982). The resistance mechanisms vary and there is also resistance to different insecticide classes for different African countries.
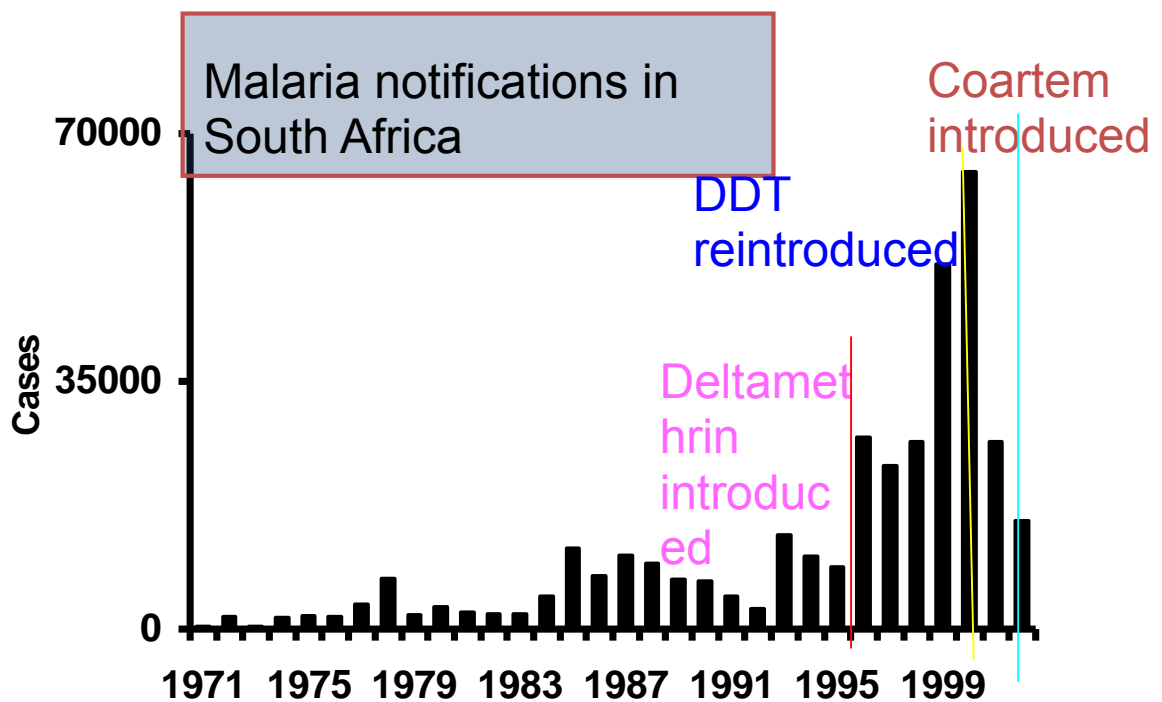
Fig. 1.5: Malaria notifications in South Africa (GPIRM 2012)

1.6.1 Mechanism of pyrethroid resistance in *An. funestus*

In *An. funestus*, pyrethroid resistance is metabolic based. The first data showing pyrethroid and carbamate resistance published by Brooke *et al.* (2001) using F_1 *An. funestus* from Mozambique suggested that this resistance was due to an increased monooxygenase activity as a full recovery was observed after mosquitoes were pre-exposed to the synergist PBO. Wondji *et al.* (2007) identified and chromosomally mapped

Quantitative trait loci (QTL) by using a set of molecular markers associated to pyrethroid resistance with crosses between FUMOZ-R (Resistant *An. funestus* from Mozambique) and FANG (Fully laboratory susceptible strain of *An. funestus* from Angola). The QTL (rp1) was associated to the phenotype resistant to pyrethroid located on chromosome 2R in a region close to the telomere where a cluster of CYP6 P450 genes is located (Wondji *et al.* 2007, 2009). There was further identification of two other QTLs (rp2 in chromosome 2 and rp3 in chromosome 3L) (Wondji *et al*. 2009). Additionally, an over expression of the *CYP6P9* gene was reported in the resistant FUMOZ-R strain compared to the susceptible FANG strain (Amenya *et al.* 2008, Wondji *et al.* 2009). Furthermore, the sequencing of the BAC clone spanning the genomic region of the rp1 revealed the duplication of the *CYP6P9* and *CYP6P4* genes in contrast to *An. gambiae* and the genes were named (*CYP6P9a* and *CP6P9b*, *CYP6P4a* and *CYP6P4b*) (Wondji *et al*. 2009).

At the rp2, Irving *et al.* (2012), using a positional cloning approach, identified candidate genes that were associated with resistance in this QTL by sequencing a 113kb BAC clone spanning a cluster of 15 P450 genes among which *CYP6M7 and CYP6Z3* where shown to be over-expressed in the resistant FUMOZ strain.

Contrary to the case in other mosquito species such as *An. gambiae*, no knock down resistance mutation conferring pyrethroid resistance has been identified yet in *An. funestus* (Djouaka *et al.* 2011, Cuamba *et al.* 2010, Coetzee and Koekemoer 2013).

1.6.2 Mechanism of carbamate resistance in *An. funestus*

There is little available information about the mechanisms associated with *An. funestus* resistance to carbamate despite the existence of this resistance across Africa. Using monooxygenase synergist PBO, Brooke *et al*. (2001), and Kloke *et al*. (2011) showed that, carbamate resistance observed in southern Mozambique is probably caused by cytochrome P450 genes after PBO synergist assays. However, one cannot also rule out the involvement of other gene families such as esterases as it is known that exposure to PBO can also inhibit some esterases (Young *et al.* 2005).

Previous studies using biochemical assay have also found evidence of altered acetylcholinesterase in *An. funestus* from Mozambique (Casimiro *et al*. 2006 and 2007, Cuamba *et al*. 2010) although the G119S *Ace-1* target site mutation was shown to be absent in contrast to *An. gambiae* (Cuamba *et al*. 2010). There has also been evidence

with propoxur of altered acetyl-cholinesterase using biochemical inhibition (Okoye *et al.* 2008).

1.6.3 Methods to characterise molecular basis of resistance mechanisms

Several approaches could be used to investigate the molecular and genetic basis of a trait of interest such as insecticide resistance. Quantitative Trait Loci (QTL) mapping is one of these methods and it has been successfully used to characterise mechanisms of resistance to insecticide in mosquito species such as *An. gambiae* (Ranson *et al.* 2004), *Aedes aegypti* (Saveedra-Rodriguez *et al.* 2010) and in *An. funestus* (Wondji *et al.* 2007 and 2009, Irving *et al.* 2012). Other methods also include genotype-phenotype association studies using genome-wide single nucleotide polymorphism (SNPs) analysis (Pearson *et al.* 2007) and transcriptomic notably through microarray-based transcription profiling of resistant mosquitoes to detect candidate resistance genes.

1.6.3.1 Microarray transcriptome analysis

Microarray was first developed in 1995 as a means of profiling many genes in parallel (Schena *et al.* 1995).

Microarray expressions entail the isolation and fluorescent labelling of mRNA followed by a hybridization process on to an array containing short sequences (probes) derived from information from the transcript. Competitive hybridization is employed by two colour array between two samples labelled with different dyes (Cyanine (cy) 3 and 5) so as to establish fold changes in the gene expression between the samples on laser excitation and scanning (Duggan *et al.* 1999).

Despite advances in microarray technology, microarrays are potentially sensitive to cross-hybridization of probes to non-target sequences (false positive) when the genes are closely related or share sequence similarity (Wren *et al.* 2002). Differences in SNP between targets and sequences employed during the process of designing the probes may also result in signal loss (false positive). These issues can be overcome in part by employing per gene multiple probes that would cover a huge part of the transcript (Roh *et al.* 2010).

1.6.3.2 Microarray studies in *Anopheles*

The first microarray experiment in a mosquito species was carried out on the *An. gambiae* species employing probes that were designed against approximately 6,000 expressed sequence tags (ESTs) in order to investigate gene expression associated with bacteria challenge, injury, oxidative stress and *P. berghei* infection (Dimopoulos *et al.* 2002).

The very first microarray experiments to investigate the insecticide resistance phenotype used a candidate gene approach that employed an array that consisted of 132 metabolic genes that focused on 90 cytochrome (CYP) P450s to investigate the resistance of DDT in *Drosophila* (Le Goff *et al.* 2003). Candidate gene array approach was also used in *An. gambiae* to detect the main resistance genes associated with insecticide resistance. David *et al.* (2005) designed a detoxification chip that contained a 230 *An. gambiae* unique fragments of genes (Cytochrome P450, carboxylesterase, GSTs, redox partners of P450 oxidative metabolic complex and various controls) putatively involved in metabolism of insecticides. This detox chip was used to monitor the expression levels of detoxifying genes in the resistant *An. gambiae* and susceptible laboratory strains. The same detox chip has been used in the sibling species *An. arabiensis* (Muller *et al.* 2008) but also in other species such as *An. funestus* a (Christian *et al.* 2011).

To date, the profiling of insecticide resistant vector populations have provided significant clue into the genetic basis of resistance phenotype such as the role of P450 genes *CYP6P3* and *CYP6M2* in pyrethroid resistance in *An. gambiae* (Djouaka *et al.* 2008, Müller *et al.* 2008) or the role of *CYP6P9a* and *CYP6P9b* in pyrethroid resistance in *An. funestus* (Wondji *et al* 2009).

1.7 Vector Genetics

1.7.1 Population genetics

The study of genetic variation within and among natural populations of a species and the forces (mutation, recombination, geographic distance and migration) that act to differentiate or homogenize populations in nature is known as population genetics (Tabachnick and Black 1995).

The earliest or the most utilised concepts of species "Biological species concept" by Mayr 1945, defined species as "groups of potentially interbreeding natural populations which are reproductively isolated from other such groups". This definition therefore makes the

"population" one of the widest concepts used as a sub specific level and is commonly defined as a group of individuals forming a breeding unit and sharing a particular habitat at a specific time. There are two parts in population genetics; the demographic structure which is associated with life history such as, the mating system, birth, death and dispersal. The second part is the genetic structure that is determined by the actual population structure including mutation, selection and evolution (Slatkin 1994). Population genetics is interested in the analysis of demographic and evolutionary factors that affects the genetic composition of a population (Hartl 1988).

1.7.2   Principles of population genetics

The theoretical population which is one that is large, randomly mating, no occurrence of evolution and maintenance of the gene pool do not exist in nature. Populations do undergo evolution that can be driven by many forces like mutation, gene flow, genetic drift and natural selection (Hoelzel and Dover 1991).

Mutations are changes in the structure of a gene that can result in a variant form which may be transmitted to subsequent generations. The structural change can be at a single nucleotide level in the DNA, or they can involve deletion, or the addition of one or more nucleotides (insertion) or processes such as transposition, unequal crossing-over, slippage, gene conversion and duplication. Mutations can be silent or non-synonymous if they specify for the same amino acid or if they change it, respectively (Hillis *et al.* 1998).

Genetic drift is a change in the allelic frequency of a population by chance and can cause reproductively isolated populations to randomly diverge in terms of their genotype frequencies. The effect of genetic drift on populations depend strongly on their effective population size (number of adults contributing to subsequent generations), which is the most important component of genetic drift, and its variation in time. In populations with smaller sizes, genetic drift can cause strong population divergence and drastic changes in genotype frequencies from generation to generation. If the effective population size is drastically reduced, then genetic drift can increase and diversity of the population decrease (bottleneck effect, Russel (1998)). A similar effect can be detected when a population is established by a small number of breeding individuals (the founder effect; Hoelzel and Dover (1991). In populations with large effective size, the effects of genetic drift are weak and insignificant compared to other evolutionary forces. Genetically, effective migration or gene flow occurs when individuals migrate from a population to another and introduce new alleles changing allelic frequencies within the recipient

26

population. Migration can be considered as a stabilizer of genetic divergence as it increases effective population sizes and reduces genetic drift (Slatkin 1985).

Natural selection can act in different directions and it can eliminate genetic variation or maintain it. In order to understand all the structuring of populations in nature we have to take into consideration all these aspects and analyze them by looking at variation of distinct alleles at defined loci known as molecular or genetic marker, which is a gene or DNA sequence that has a known location on the chromosome and can be used to identify a species (Allendorf *et al.* 2000).

Genetic differentiation and gene flow are assessed in and between populations through direct and indirect approaches. The direct approach is based on direct observation of the population used for assessing the gene flow. This is through methods such as mark/release/recapture (MRR). In this case a species is captured and marked with a dye to ease recognition if captured again. After marking, the species is then released back into the wild and recaptured again. The advantage of this method is that it does help characterize the structure of adult mosquito populations and helps in the understanding of female feeding patterns and the dispersal pattern (Toure *et al*. 1998). The disadvantage is that it is difficult to determine the likelihood of capturing resident and released mosquitoes. MRR is practically impossible during seasons that experience very few mosquitoes like the dry season.

Indirect method is based on the distribution of allele frequencies using genetic markers such as allozymes, DNA sequencing, microsatellite genotyping and mitochondria DNA (mtDNA). It gives the different estimates from direct estimation procedures (Slatkin 1985). It is advantageous in that it takes into account geographical differences assuming gene flow occurs at low levels.

Mathematical tools can be used to assess gene flow with genetic metric such as F_{IS}, F_{ST} and N_m. To better understand population genetics, the most common methodology is Wright's statistics. Wright (1965) introduced the concept of F-statistics or fixation indices which measures the reduction of heterozygosity expected with random mating at any one level of a population hierarchy-relative to another. Gametes drawn from same population or sub-population are more likely to have a common ancestor than gametes drawn from different ones. The level of such genetic differentiation may increase by random genetic drift as mutations accumulate over time and decrease with migration, as input of new

alleles occurs. Wrights F- statistics permits the assessment of this differentiation by describing the hierarchically sub-divided populations as "the correlation between random gametes within a population, relative gametes of the total population (Wright 1965).

Three indices, F_{IS}, F_{IT} and F_{ST} are used in F-statistics and are related according to the following equation (Nei 1977, de Jong *et al*. 1994).

$$F_{ST} = \frac{(F_{IT} - F_{IS})}{(1 - F_{IS})}$$

F_{ST} is a measure of the overall population subdivision and has a value between Zero (no differentiation) and one (complete differentiation).

$$F_{ST} = 1 - \frac{(H_S)}{(H_T)}$$

H_S, the mean expected heterozygosity assuming HWE in the total population and H_T is the mean expected heterozygosity in the total population without sub division (Wright 1969, Nei 1977, de Jong *et al*. 1994). Little genetic differentiation: $0 < F_{ST} < 0.05$ and moderate $0.05 < F_{ST} < 0.1$ and high $0.1 < F_{ST} < 1$

F_{IS} is the fixation index which relates subpopulations with regional groups of inbred organisms to the subpopulation to which they belong, measuring the decrease of heterozygosity between these two levels (Nei 1987). F_{IS} is often referred to as inbreeding coefficient as it measures the correlation of genes used to assess random mating or inbreeding between subpopulation within same sample. When the average frequency of homozygotes among different subpopulations is always greater than the expected with random mating between them, this is known as Wahlund effect (Wahlund 1928) which is one possible cause of high F_{IS} values. H_I measures the observed average heterozygosity in the sub population (Nei 1977, de Jong *et al*. 1994).

$$F_{IS} = \frac{H_S - H_I}{H_S}$$

F_{IT}, measures reduction of heterozygosity of the regional aggregates relative to the total combined population, or in terms of inbreeding, the heterozygosity of the inbred organisms relative to the total population (Nei 1977, de Jong *et al*. 1994).

$$F_{IT} = \frac{H_T - H_I}{H_T}$$

For microsatellites, an alternative to F_{ST}, is the R_{ST} statistics which has been proposed by Slatkin (1995). R_{ST} is based on the assumption that in microsatellites, allele size depends on the size of its ancestor, contrary to the Infinite Allele Model assumed in F_{ST}. The difference is, R_{ST} index is defined in terms of allele size (Slatkin 1995).

$$R_{ST} = \frac{S - S_W}{S}$$

Where S is the variance of allele size considering the total population and S_W the average variance in the allele size within sub-population.

Fixation indices have also been used to estimate gene flow for a long time (Slatkin 1995). Only a small number of migrants are required to prevent genetic divergence among populations (estimated for example by F_{ST}). The relation is translated by:

$$F_{ST} = \frac{1}{1 + 4Nm}$$

1.7.3 Cytogenetics and presence of chromosomal inversions and frequencies across Africa

Cytogenetics is the study of chromosomal rearrangements (Dia *et al.* 2013). The first publication on *An. funestus* group cytogenetics was in 1980 (Green and Hunt, 1980). There was a follow up publication a couple of years later (Green 1982) on chromosome maps for seven of the *An. funestus* members with details of polymorphic inversions occurring within and between species (Coetzee *et al*. 2004). Identification of species using cytogenetics requires considerable skills in interpreting the arrangements of the bands on the polytene chromosomes. Cytological studies in Burkina Faso reported two co-existing taxonomic units of *An. funestus* that differ in the degree of chromosomal polymorphism, host preference and their habits of resting. Variations in inversion frequencies were observed among samples without consistent geographical or temporal clines (Constantini *et al*. 1999). The populations are also reported to be polymorphic for at least four paracentric (2Ra, 3Ra, 3Rb and 3La) chromosomal inversions (Boccolini *et al*. 1994). Based on findings of higher frequencies of inverted arrangements 3Ra and 3Rb found in indoor human-fed samples vs. (versus) corresponding outdoor animal-fed

samples of *An. funestus*, two chromosomal forms were described called Kiribina and Folonzo. Species with inverted karyotypes were correlated to vectorial capacity in Burkina Faso and were also commonly found in higher frequencies around human dwellings (Constantini *et al.* 1999).

Northern populations in Cameroon are related to the Kiribina form while in the south, the populations are related to the Folonzo form (Cohuet *et al.* 2005a andb). Studies carried out in other African countries including Madagascar showed the presence of at least 11 paracentric chromosomal inversions on chromosomes II and III (Green and Hunt 1980, Boccolini *et al.* 1994, 2002, Lochouarn *et al.* 1998, Dia *et al.* 2000, Kamau *et al.* 2002, 2003b). Significant chromosomal differentiation between West and coastal Kenya has been reported in *An. funestus* (Kamau *et al.* 2002).

An. funestus populations with different chromosomal arrangements showed different anthropophilic rates in Senegal (Lochouarn *et al.* 1998). Three chromosomal populations were observed in Senegal with two of the form being sympatric with strong deficit of heterokaryotypes suggesting the presence of two genetically distinct populations (Dia *et al.* 2000). This implies restricted gene flow between chromosomal forms of *An. funestus* (Cohuet *et al*. 2005). In Kenya, *An. funestus* population analysis suggested limited gene flow between western and coastal samples 700km apart with great difference in the chromosomal inversion polymorphisms (Kamau *et al.* 2002). The rift valley might be acting as a barrier to gene flow between populations in Kenya. From previous studies carried out, there is the possibility that the genetic structure of *An. funestus* in West Africa could be different from those of East Africa or southern Africa.

Differences in chromosomal inversions have been reported to be species-specific and some inversions were polymorphic in some species and fixed in others as Green and Hunt (1980) and Green (1982) showed differences in the chromosomal polymorphism within the species of the *An. funestus* group just as in *An. gambiae*. The chromosomal map of *An. funestus* was only established in 2001 by Sharakov *et al.* (2001) which was based on comparisons to the chromosomal map of *An. gambiae* (Sharakhov *et al.* 2004). In the *An. funestus* group, *An. funestus* species is the most studied in the group even though exceeded by *An. gambiae* studies (Coluzzi *et al.* 1979, Coluzzi *et al.* 1982, Pombi *et al.* 2008). In the continent of Africa, there has been the recognition of seventeen chromosomal inversions with specific distribution (Green 1982, Dia *et al.* 2000,

Geulbeogo *et al.* 2005, Boccolini *et al.* 1994, Boccolini *et al.* 1998, Kamau *et al.* 2002). Four of the inversions are found across the continent (2Ra, 3Ra, 3Rb and 3La) while the others are regionally distributed like the 2Rt which is found in West Africa and the very localized 2Rd inversion found in the southern forested areas of Cameroon (Endler 1977). Environmental selection, demographic effects or historic events could be explaining these differences in distribution patterns (Endler 1977).

Deficits of heterozygotes in inversions on the 3R and 3L chromosomal arms were also detected in some regions in Senegal and Cameroon (Dia *et al.* 2000, Cohuet *et al.* 2005, Ayala *et al.* 2011a). However, there was no clear indication of division between the chromosomal forms from Burkina Faso but a distribution of chromosomal inversions and their frequencies through different habitats and environments in a non random manner that suggests that inversion frequencies in *An. funestus* do not follow a neutral pattern (Dia *et al.* 2013). A sharp contrast between population structure measured at chromosomal inversions and neutral markers was observed by Ayala *et al.* (2011a) with only a weak signal of population structure due to distance among geographical zones in Cameroon detected with microsatellite data as previously described (Cohuet *et al.* 2005). This was in contrast to chromosomal inversions which revealed a strong differentiation among habitats suggesting that their distribution might be shaped by environmental selection. In any case, same study did not observe any difference using F_{st} measures between microsatellite loci lying within or outside polymorphic chromosomal inversions (Ayala *et al.* 2011). Based on karyotypes, inversions have been associated with mating behaviour which was assortive in *An. funestus* between 77% and 90% (Ayala *et al* 2012).

1.7.4 Microsatellite

Microsatellites are the most commonly used molecular markers. They consist of Tandem repeats of short sequence motifs, from di-to-hexanucleotides (2-6bp) that can get up to a length of 150bp (Ashley and Dow 1994, Schlötterer 1994, Wright and Bentzen, 1994). Microsatellites are made up of different types including interrupted, uninterrupted or compound microsatellites (Estoup and Angers 1998). Microsatellites are abundant in eukaryotes and spreads throughout the entire genome at 7-100kb intervals (Schlötterer 1994, Wright and Bentzen 1995). There is a wide variation in the mean density of microsatellite within species of different taxonomic group (Estoup and Angers 1998).

Microsatellites loci are a very useful molecular marker for genetic studies because of their mutation rates of sequence repeats which are higher than other types of markers (Eisen 1998). This is thought to be as a result of slipped strand mis-pairing during replication causing insertion or deletion of repeat units (Ashley and Dow 1994). Microsatellite mutations can also be caused by gene conversions or unequal crossing-over during recombination. High mutation rates at these loci leads to extensive allelic variation and high level of heterozygosity (Wright and Bentzen 1995). Microsatellites are inherited in Mendelian fashion, show no functional role and are selectively neutral (Ashley and Dow 1994, Estoup and Angers 1998). Microsatellite loci has many advantages over other markers for example, they exhibit high levels of polymorphism, high heterozygosity, codominant single-locus inheritance, abundance in the genome and are selectively neutral (Estoup and Angers 1998).

Studies conducted in Cameroon (ten different populations) and Senegal (four different populations) using a set of nine microsatellite markers spread over the entire genome of *An. funestus* had results showing that gene flow existed between populations and as well as gene exchanges between chromosomal forms within these two countries respectively. In any case, there was no evidence of population subdivision in populations with strong heterozygote deficits. There was detection of isolation by distance between geographical populations supporting the idea that microsatellite markers can detect population subdivision (Cohuet *et al.* 2005a and b).

Results similar with one above were obtained in Kenya where two coastal populations were compared to two western populations. Differentiation observed was due to geographic distance (Braginets *et al.* 2003). Microsatellite DNA markers were used by Temu *et al.* (2004) to study population structure of *An. funestus* from Kenya, Malawi, Mozambique, South Africa and Uganda showing discontinue populations between southern Africa and East Africa. Malawi samples were different to both southern and East African samples (Coetzee *et al.* 2004). Microsatellite data and chromosomal inversion data do not show same genetic discontinuities. Heterozygote deficits at chromosomal loci can be interpreted from these results to be possibly locus-specific and may be partly dependent on environmental selection of the inversions themselves or the genes they contain rather than population subdivision or incipient speciation (Coetzee *et al.* 2004).

Depending on the problems and available resources, some molecular markers can be appropriate in studying some problems than others. For example microsatellites are more appropriate to study parentage and mating systems and sometimes may not be sufficiently conserved for population-level comparisons (Creasey and Rogers 1999).

1.7.5 Mitochondrial DNA

Mitochondrial DNA (mtDNA) sequences are widely used tools to assess phylogenetic relationships between similar species, populations of same species or even between individuals (Taberlet *et al.* 1996). mtDNA is inherited maternally without recombination and represents only a quarter of the effective population size of nuclear markers. This property makes it more sensitive to detect any reduction in genetic variation (Ferguson and Danzmann 1998). Molecules of mtDNA exist in a high copy number in the mitochondria of cells and have a circular structure. Mitochondria are composed of around 13000bp in *An. gambiae* and *An. funestus*. It has got a non-coding region (+1000bp) responsible for replication known as "control region" or "d-loop", which evolves 4-to-5 times faster than the entire mtDNA which evolves 10 times faster than nuclear DNA (Brown *et al.* 1979). mtDNA sequencing has become the molecular marker of choice when studying closely related species (Stepien and Faber 1998). The disadvantage of mtDNA is that, maternal inheritance does not provide information about males in populations.

mtDNA studies carried out on *An. funestus* population in Burkina Faso by Michel *et al.* (2006), suggested that Kiribina and Folonzo populations are not at neutral equilibrium due to population growth. This implies that the genetic structure of *An. funestus* is due to inherent differences between the two forms and not to physical distance, confirming the hypothesis proposed earlier based on chromosomal inversion evidence that Folonzo and Kiribina are incipient species in Burkina Faso (Costantini *et al.* 1999).

1.8 Rationale of this study

Malaria control relies extensively on the use of insecticides, either as insecticide treated materials or as indoor residual sprays (IRS). Pyrethroid insecticides have been the primary choice for most control programs for the past three decades, but carbamates are being increasingly used, either in rotation with pyrethroids or as an IRS replacement for pyrethroids in areas where pyrethroid resistance is common (Corbel and N'Guessan 2012).

Resistance to pyrethroids and carbamates is developing at an alarming rate in *Anopheles* mosquitoes in Africa and resistance observed in *An. funestus* is disrupting malaria control in southern Africa (Coetzee *et al.* 2006, Casimiro *et al.* 2006, Okoye *et al.* 2008). In Mozambique, *An. funestus* populations have developed multiple resistance mechanisms to both pyrethroids and carbamates and an unusual cross-resistance pattern has been observed between these two insecticides (Casimiro *et al.* 2006) There are fears that this resistance front could spread rapidly if there is no barrier to gene flow between *An. funestus* populations. Such cross-resistance will have a devastating impact on malaria control programmes and will severely limit available resistance management strategies.

This operational problem underpins an urgent need to understand the mechanisms of insecticide resistance in this vector species, to facilitate resistance and disease management by enabling questions on the origin, spread and evolution of resistance to be assessed and characterised. This study was carried out to fill these important knowledge gaps in *An. funestus* and help improve future control of this malaria vector in Africa.

1.8.1 Aim and objectives

1.8.1.1 Aim

Studying the mechanism of metabolic resistance in *An. funestus* was hindered by the fact that there has been less research carried out on *An. funestus* and there is also less available genetics and genomic tools for this species. However, taking advantage of recent progress made, this study aims at elucidating the mechanism of carbamate resistance and carbamate/pyrethroid cross resistance in *An. funestus* population in Africa and also assess patterns of gene flow between *An. funestus* populations in Africa to predict the potential spread of resistance genes.

1.8.1.2. Specific Objectives

The specific objectives are as follows:
1. To investigate the molecular basis of metabolic carbamate resistance and carbamate/pyrethroid cross resistance using Microarray experiments and Real-Time quantitative PCR (qRT-PCR). Analysis of polymorphism patterns of main candidate resistance genes will be performed to support their involvement in the resistance to bendiocarb and permethrin.

2. To establish the genetic structure of *An. funestus* population across Africa in relation to known patterns of insecticide resistance using microsatellite loci in order to estimate the risk of spread of insecticide resistance genes across Africa.

3. To detect potential signatures of selective sweep associated with pyrethroid or carbamate resistance using microsatellite loci.

4. To investigate the extent and scale of selective sweep associated with pyrethroid resistance through a fine-scale sequencing of the major rp1 pyrethroid resistance QTL genomic region between different populations.

CHAPTER TWO

Investigating the molecular basis of carbamate resistance and carbamate/pyrethroid cross resistance in *An. funestus* in Chikwawa Malawi

2.1 Introduction

Insecticide resistance is affecting malaria control programs across Africa (WHO 2012). Pyrethroid insecticides resistance is a major concern for malaria control since pyrethroids are the recommended WHO insecticide for use in LLINs and IRS. Carbamate is one of the alternative insecticides that could be used in replacement of pyrethroids for IRS in resistance management strategies. Organophosphates could also be used as alternatives to pyrethroids since they are also not as toxic as the organochlorines and also have different mode of actions. Resistance to carbamate will seriously limit insecticide choice for IRS across Africa. The resistance developed by the malaria vectors to these two insecticide classes is posing a serious threat to the success of many control programmes (Ranson *et al*. 2011, WHO 2012).

Carbamate and pyrethroid resistance is widespread in southern Africa while susceptibility to DDT is observed (Brooke *et al*. 2001, Wondji *et al.* 2012, Hunt *et al*. 2010). No resistance to carbamate has been observed to date in East Africa whereas pyrethroid and DDT resistance is common (Morgan *et al*. 2010). In West Africa, resistance to pyrethroids, carbamate and DDT is also observed (Okoye *et al*. 2008, Djouaka *et al.* 2011, Riveron *et al.* 2014). Previous studies carried out in southern Africa have suggested a possible cross-resistance between pyrethroids and carbamates, as suggested notably by Brooke *et al.* (2001), with the two resistances having the same underlying P450-based mechanisms after PBO synergist assays. However, this has never been investigated and confirmed. To date, the molecular basis of carbamate resistance in *An. funestus* remained uncharacterised, whereas significant progress has been made in elucidating the molecular basis of pyrethroid resistance (Riveron *et al*. 2013, Wondji *et al.* 2009). In addition, the molecular basis of the possible cross- resistance of carbamate/pyrethroid has not been investigated. It remains to be established whether the resistance genes conferring pyrethroid resistance in *An. funestus* in southern Africa are also involved in carbamate resistance. Elucidation of such cross-resistance is crucial to facilitate the implementation

of successful resistance management strategies, whether methods such as rotation between pyrethroids and carbamates should be recommended or not.

2.1.1 Objective

In this study, genome-wide transcription microarray-based analyses coupled with analysis of the genetic variability of candidate genes were used to characterise the mechanisms of carbamate resistance. Molecular evidences of the existence of carbamate/pyrethroid cross-resistance mechanisms were also investigated in an *An. funestus* population from Chikwawa in Malawi.

2.2 Methods

2.2.1 Field collections

Indoor resting female *An. funestus* were collected in July 2009 and April 2010 (for 2 weeks respectively) using torches and aspirators from the ceilings and walls of houses (Fig. 2.1) by J. Morgan (LSTM) from Chikwawa district ($0°$ 45' N, $34°$ 5'E) in the southern region of Malawi (Fig. 2.2). The climate in Chikwawa is made up of two seasons, the dry and rainy seasons. The rainy season runs from October-April but sees its peak between December-March where there is usually very heavy rains. The dry season is between May-September. The main occupation of the people in Chikwawa is agriculture as they depend on agriculture (maize) for living.

Fig. 2.1: Collection of *An. funestus* using an aspirator (Picture by J. Morgan- LSTM)
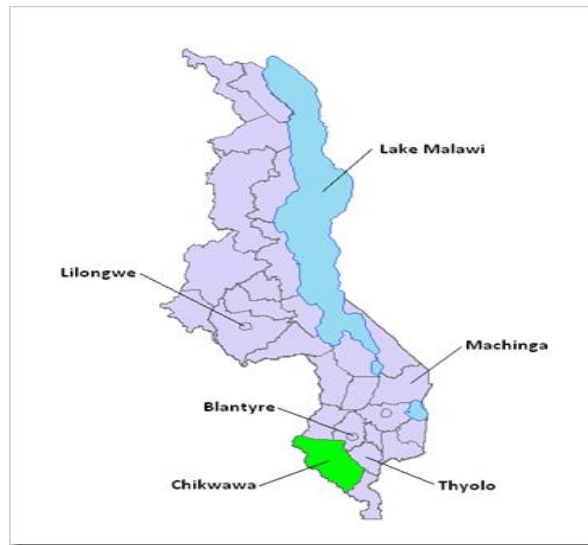
Fig. 2.2: Map of Malawi showing area in green (Chikwawa) where samples were collected.

2.2.2 Mosquito rearing

The female mosquitoes were forced to lay eggs as described in Morgan *et al.* (2010). To attain the gravid stage, the mosquitoes were allowed for 4 to 5 days and later gently introduced into 1.5ml Eppendorf tubes individually. Presence of eggs was checked daily. There was a careful transfer of females that laid eggs into Eppendorf tubes with silica gel. The eggs were then stored at room temperature or at 4ºC for 2 days and brought to the Liverpool School of Tropical Medicine (LSTM). At LSTM, the eggs were allowed to hatch in small cup and then transferred to larvae bowls filled with mineral (bottled) and distilled water for rearing. The larvae were fed with Tetramin (TM) baby fish food daily. To reduce mortality, the water of each larvae bowl was changed every two days.

2.2.3 PCR-Species identification

The female mosquitoes that were used for oviposition were identified morphologically as belonging to the *An. funestus* group following the key of Gillies and Coetzee (1987). For further confirmation that the female that laid eggs were *An. funestus s.s*, a PCR was carried out following the protocol by Koekemoer *et al*. (2002) to confirm that all mosquitoes morphologically identified as belonging to *An. funestus* group were *An. funestus s.s.* Only primers specific to *An. funestus s.s* were used in the first instance to select the *funestus s.s* with the universal primer (UN). For 25µl of PCR product; 2.5µl of reaction Buffer A, 0.2µl of dNTP 25mM, 0.2µl Kapa taq, 1.5µl of 25mM MgCl2, 2.0µl

of universal primer (forward TGT GAA CTG CAG GAC ACA T), 2.0μl of *funestus* reverse GCA TCG ATG GGT TAA TCA TG) 15.6μl of H_2O and 1μl of DNA. (25pmol/μl), were conducted on a DNA Thermal Cycler under the following conditions: 2mins at 94ºC, followed by 35 cycles of denaturing at 94ºC for 30s, annealing at 45ºC for 30s and extension at 72ºC for 40s; finishing with an extension step at 72ºC for 5 min.

2.2.4 Insecticide susceptibility assays (Bioassay)

2-5 days-old F_1 adult mosquitoes were used to assess the susceptibility profile to various insecticides. Around 20-25 mosquitoes per tube were exposed to insecticide-impregnated filter paper (bendiocarb 0.01% and permethrin 0.75%) and in addition, to obtain a more comprehensive susceptibility profile of this population, the following insecticides were also tested: the pyrethroids deltamethrin (0.05%) and lambda-cyhalothrin (0.05%); the organophosphate malathion (5%) and the organochlorines DDT (4%) and dieldrin (4%) for 1h with a control tube that was not exposed to any insecticide. After 1hr, mosquitoes were transferred to a holding tube and allowed to rest for 24hrs. The mortality rate was recorded after 24hrs to get the resistance profile of the mosquitoes. The mosquitoes that were alive including the control mosquitoes were knocked down by freezing them for a couple of minutes and then immediately stored in the -80ºC freezer to prevent RNA degradation.

2.2.5 PBO synergist study

The potential involvement of cytochrome P450 monooxygenases in both carbamate and pyrethroid resistance was assessed in the Chikwawa population using PBO (piperonyl butoxide) which is an inhibitor of P450 activity and also esterases (Scott JG 1999 Gunning *et al.* 1998, young *et al.* 2005, Brooke *et al*. 2001). The effect of the PBO synergist was analysed in combination of 0.01% bendiocarb and 0.75% permethrin. Four replicates of 25 female mosquitoes were pre-exposed to 4% PBO impregnated filter paper for 1hr and immediately exposed to 0.01% bendiocarb or to 0.75% permethrin for 1hr. The control samples were only exposed to PBO for 1hr but not to either of the insecticides. After 24hrs, final mortality of mosquitoes was assessed for both synergized and unsynergized samples.

2.2.6 Genome-wide transcription analysis with microarray to detect candidate genes associated with carbamates and pyrethroid/carbamate cross resistance

In order to detect the genes associated with carbamate resistance or cross-resistance between carbamtes and pyrethroids, a genome-wide transcription microarray-based analysis was performed.

Experimental design

The experiment was designed to compare the resistant (R-alive after exposure to insecticide) against (vs) susceptible (S-FANG laboratory strain), resistant vs control (C-non exposed to insecticide) and control vs susceptible for both bendiocarb and permethrin (Fig. 2.3). This experiment was designed to maximise the chances to identify the main genes that are possibly responsible for bendiocarb and permethrin resistance. The advantage of this design is that, resistance genes could be detected from the list of gene expressed in the R-C comparison with no confounding factors due to difference in genetic background between samples.
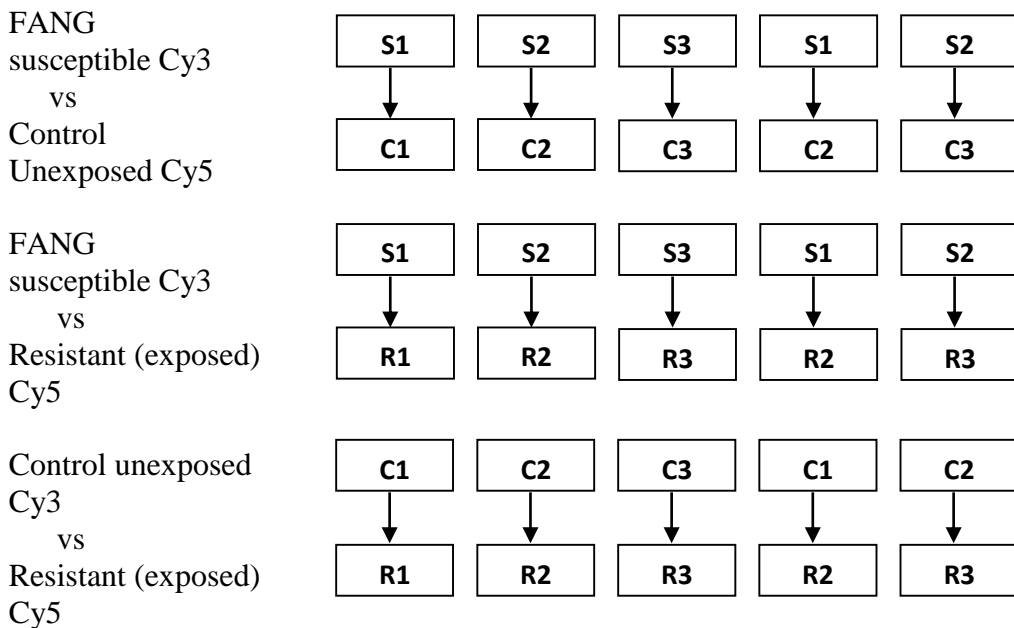
Fig. 2.3: Microarray Experimental Design: FANG S- Susceptible vs Control-C (non exposed to insecticides), S vs R (Resistant- exposed to insecticides) and C vs R for both bendiocarb and permethrin.

In addition, the candidate genes could further be supported by their over-expression in all the comparisons (R-S, C-S and R-C) or in more than one comparison than just R-C. However, because the control samples are a mix of susceptible and resistant mosquitoes

and thus may exhibit a lower fold change when compared to fully resistant samples (R), further attention are given to genes also commonly expressed in R-S and C-S (Fig. 2.3).

2.2.6.1 Microarray 4x44k and 8x60k *An. funestus* chip description

Two different but complementary Agilent microarray chips were used in this study. Firstly, the 4x44k *An. funestus* Agilent chip previously used to characterise pyrethroid resistance (Riveron *et al.* 2013) was initially used to detect potential carbamate and permethrin resistance genes. Secondly, the more comprehensive 8x60k chip was also used to validate and identify further candidate resistance genes. Agilent eArray program was used to design both Agilent microarray chips as described by Riveron *et al.* (2013).

The 4x44k chip is made up of 60 mer probes designed from 8540 ESTs that was generated from the *An. funestus* transcriptom 454 sequencing (Gregory *et al.* 2011) (2 probes for each EST), a set of 2850 *An. funestus* cDNA from Genbank (2 probes for each EST), a set of P450 genes from rp1 and rp2 QTL BAC sequences (Wondji *et al.* 2009, Irving *et al.* 2012) (3 probes for each gene) and the 13,000 transcriptome of the *An. gambiae* genome. All the previous detox genes present on the *An. gambiae* detox were added in this chip with 3 probes so that each gene will take advantage of any likely gene sequence conservation between *An. funestus* and *An. gambiae*. The 8x60k chip was constructed in same way but also by designing probes with an additional 15,000 EST set generated from another RNAseq transcriptome sequencing study (Crawford *et al.* 2010).

2.2.7 RNA Extraction

RNA was extracted from 3 batches of 10 female mosquitoes of *An. funestus* alive after exposure to bendiocarb (0.01%) (Resistant- R), 3 batches of 10 female mosquitoes alive after exposure to permethrin (0.75%), 3 batches of 10 female mosquitoes from wild population, none exposed to insecticide (Control C), and 3 batches of 10 female mosquitoes from the unexposed susceptible laboratory strain (FANG).

RNA was extracted using the Picopure RNA isolation kit (Arcturus). Bench and pipettes were thoroughly cleaned with ethanol and later on sprayed with RNase Zap to prevent any possible contamination. Only sterile and freshly opened tips were used during the process of extraction. Mosquitoes to be extracted were transferred into a 1.5ml Eppendorf tube on ice. A hundred micro liters (100µl) of extraction buffer (XB) was added into the tube and mosquitoes were thoroughly ground using a grinder in the buffer. The mixture

was then incubated at 42ºC for 30min on a heating block. While waiting for incubation to be completed, there was preconditioning of the RNA purification column by pipetting 250µl conditioning buffer (CB) onto the purification column filter membrane and incubated at room temperature for 5min after which column was then centrifuged in the provided collection tube at 16,000 rcf for 1min and sample was again centrifuged at 14,000 rcf for 2min. Supernatant was then transferred into a new Eppendorf tube while avoiding pick up of any pelleted material.

A hundred micro liters (100µl) of 70% ethanol was added to the supernatant and well mixed by pipetting given a final volume of approximately 200µl. The mixture was then transferred into the pre-conditioned purification column and then centrifuged for 2 min at 100 rcf, followed by a centrifugation at 16,000 rcf for 30sec and then transferred into a purification column. Care was made to ensure that the purification column did not have any residual wash buffer. If any remained, purification column was re-centrifuge at 16,000 rcf for 1 min. The flow through waste was discarded and column was re-centrifuged at 16,000 rcf for 1min. The purification column was then transferred to a new 0.5 ml micro-centrifuge tube (Provided in the kit). 30µl of elusion buffer (EB) was pipette directly onto the membrane of the purification column. Care was taken while gently touching the tip of the pipette to the surface of the membrane while dispensing the EB to ensure maximum absorption of EB into the membrane. The purification column was then incubated for 1min at room temperature after which column was centrifuged for 1min at 16,000 rcf to elute RNA. The isolated RNA was ready and immediately stored in -80ºC freezer for further experiments.

2.2.7.1 RNA labelling and amplification

Quality and quantity of RNA were assessed using the NanoDrop (ND1000) spectrophotometer (Thermo-Fisher) and an Agilent Bio-analyzer respectively. Each sample of cRNA was amplified using Agilent Quick Amp labelling Kit (two colours) following the manufacture's guide (Fig. 2.4).

The cRNA from the resistant samples (R) were labelled with Cy5 dye for both bendiocarb and permethrin, while the control cRNA were also labelled with Cy5 and the lab susceptible strain FANG (S) were labelled with Cy3 dyes. However when the resistant samples were to be compared to the control samples, the control samples were labelled with Cy3 using the Agilent Two-color Kit and RNA Spike-in Kit following

manufacturers' guide (Fig. 2.3 and Fig. 2.5). Potential dye-bias was corrected during the normalisation process in Feature Extraction using the Rank consistency probe method (Tseng *et al.* 2001)

Purification of amplified cRNA was done using a column kit (RNeasy Mini Kit, Qiagen). The cRNA was then eluted in 30µl of R-Nase-free water (Invitrogen).

```
┌─────────────────────────────────────┐
│   Template Total RNA with Spike-in    │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│            cDNA Synthesis             │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│            cRNA Synthesis             │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│   cRNA Labelling using CY5 (R&S) and CY3 │
│         (S&C) with Amplification      │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│  Preparation of hybridisation samples (R, C & │
│                    S)                 │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│        17-hour hybridisation (65°C)   │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│               Washing                 │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│      Scanning and Feature Extraction  │
└─────────────────────────────────────┘
```
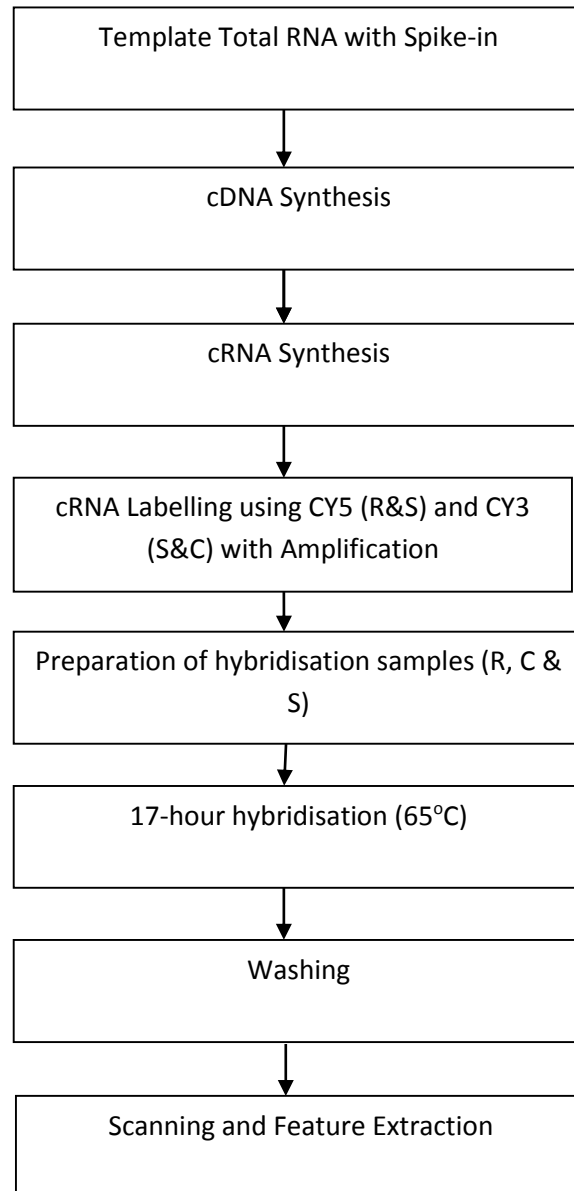
Fig. 2.4: Workflow of amplified cRNA procedure

The NanoDrop™ with the Microarray Measurement setting was used to assess amplification and dye incorporation of samples (Thermo Scientific). The RNA-40 setting was used to measure 1µl of each of the labelled samples and also the cRNA concentration

(ng/µl), concentration of Cy dye (pmol/µl) and ratio of absorbance (260nm/280nm) were recorded.

Yield

cRNA concentration (ng/µl)*30µl (elution volume)/1000=µg of cRNA

Specific activity

Cy3 or Cy5 conc (pmol/µl)/conc. cRNA (ng/µl))*1000=pmol Cy per µg cRNA

A yield >800ng with a specific activity of >8.0pmol/ µg of cRNA was needed to proceed to the hybridisation step of the microarray for each sample.

The Agilent 2100 Bioanalyzer using the RNA 600 Nano Kit was used to analyse an aliquot of each of the cRNA sample for quality following manufacturers guide (Fig. 2.5). RNA with good quality was expected to give a peak between 100-2000 nucleotides.

(a) Cy3 (susceptible) (b) Cy5 (resistant)
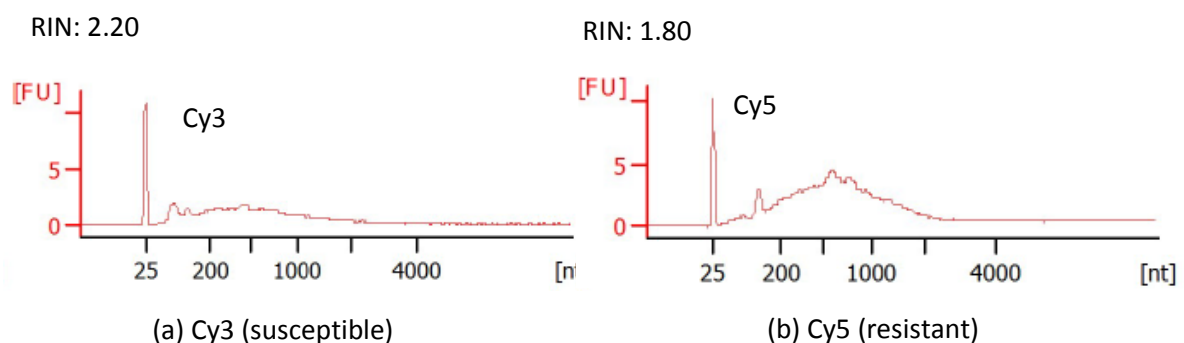
Fig. 2.5: Example Bioanalyzer traces from Agilent 2100 Bioanalyzer showing the quality of *An. funestus* Cy3 (susceptible) and Cy5 (resistant) labelled RNA samples.

2.2.7.2 Microarray hybridisation

For each of the comparisons (R-S, R-C, and C-S), 5 hybridizations were carried out for both bendiocarb and permethrin. The incubation steps were carried in a thermal cycler (Agilent technologies). Eight hunhred and twenty five micro liters (825µg) of each labelled cRNA for comparison on a single array were combined for pre-hybridisation was fragmented for 30 min at 60ºC using Gene Expression Hybridization Kit reagents (Agilent), following manufacturers guide. To end the reaction, 55µl of 2x GEx

Hybridization Buffer HI-RPM (Agilent) was added and mixed with the samples by pipetting. Tubes were briefly spun down and left on ice until hybridization.

A clean 4x44 or 8x60 gasket slide was then placed onto a SureHyb chamber base (Agilent Microarray Hybridization Chamber Kit). 100µl (for 4x44K) and 40µl (for 8x60K) of hybridization sample were then placed at the centre of each of the gaskets by a "drag and dispense" manner followed by full assembly of the chamber. After which the 4x44K or 8x60K array was slowly lowered and were hybridized in the Agilent hybridization oven for 17 hrs at 65ºC following manufactures guide.

Arrays were then washed the following day after hybridization with GE wash buffers 1 and 2 (Agilent) according to manufacturer's guide. They were treated with acetonitrile (Sigma) and later with stabilization and drying solution (Agilent) to protect ozone-induced degradation of the cyanine dyes and also to reduce background noise on the arrays. Arrays were finally stored in light protective boxes until scanned.

2.2.7.3 Scanning and feature extraction of microarray

The arrays were scanned using the G2205B (Agilent technologies) Series Microarray Scanner with the default profile settings following the Agilent Scanner System following the manual.

High Photo-multiplier tube (PMT 100%) and low (PMT 10%) extended dynamic range (XDR) images that have been scanned were extracted and combined using the Feature Extraction (FE) software GE_10.5_Dec08 (Agilent) with the template of the custom array (4x44K or 8x60K) (022094_D_F_20081124.XML) (Fig. 2.6). The default parameters and FULL text output were selected with the resulting QC (Quality Control) reports and result files exported for further analysis. Reports of QC were considered to give an indication about the quality of the array. QC report of 11/11 means the eleven main parameters of the array were passed and a score below 8/11 is an indication of poor quality hybridisation.
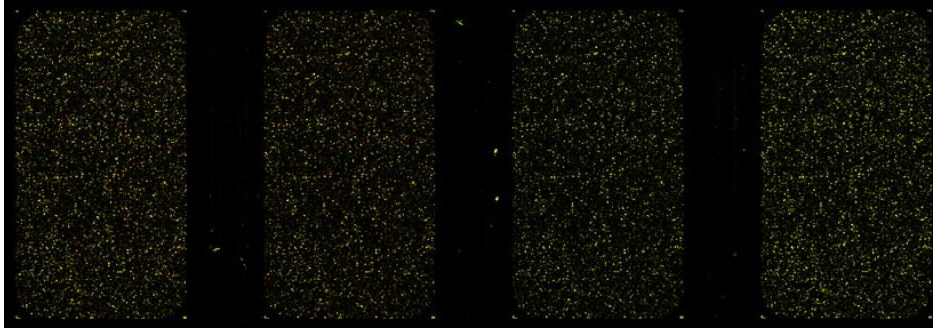
Fig. 2.6: A custom Agilent 4x44K microarray scan image in normal view.

2.2.7.4 Microarray analysis

Microarray analysis was performed using Genespring GX version 12 software (Agilent Technologies, USA). The genes with significant *p-values <0.01* for t-test against zero after Benjamini-Hochberg multiple testing correction, and a fold change (FC) > 2, were considered as significantly differentially expressed between the various comparisons. This was after the submission of the mean expression ratios to a t-test against zero with a multiple testing correction.

2.2.8 Quantitative Real-Time PCR

qRT-PCR experiment was used to validate the expression profile of candidate resistance genes derived from microarray experiments. cDNA was synthesised from 3 replicates of resistant (R) individuals, and same for control (C) and susceptible (S). The genes with the most over-expressed after microarray were selected for analysis including other possible candidate genes. A total of 12 genes were selected and analysed using two housekeeping genes (SP7 and Actin) as control genes (Table 2.1).

Complementary DNA (cDNA) was produced in a 20µl reaction using SuperScript™ III Reverse Transcriptase (Invitrogen) according to manufacturer's guidelines. One micro gram (1µg) of total RNA was added to an initial 13µl reaction containing 2.5µM oligo (dT)$_{20}$ primer (Sigma) and 0.5mM dNTP Mix (ten micro molar (10mM) each of dATP, dGTP, dCTP and dTTP) (Bioline). RNA was then denatured by heating to 65°C for 5mins on a dry heat block and later on placed on ice for 1min. 4ml of 5X First Strand Buffer (250mM Tris-HCL pH 8.2, 375mM KCI, 15mM MgCl$_2$) (SuperScript™ III Kit) was added to RNA along with 5mM DTT (SuperScript™ III Kit), 200 units of SuperScript™ III Reverse Transcriptase (Invitrogen) and 40 units of RNasin (Promega).

Reactions were incubated at 50°C on a dry heat block for one hr. before inactivation by heating to 70°C for 15mins.

First strand cDNA was then treated with 5 units of RNAse H (NEB) in a 50µl reaction according to manufacturer's guide to remove RNA complementary to synthesised cDNA. Complementary DNA was then purified using a column based PCR purification Kit (QIAquick PCR purification kit, Qiagen) and eluted in 30µl of elution buffer (10mM Tris-Cl, pH 8.5) before reading the concentration on the NanoDrop® spectrophotometer (Thermo Scientific), and storing at - 20°C.

2.2.8.1 Performing qRT-PCR

All primer pairs were found to produce a single product when annealed at 60°C so no additional optimisation was required. qRT-PCR for all microarray RNA samples was performed using the universal KAPA™SYBR® FAST kit and the Agilent MX3005 qRT-PCR machine using the (Agilent) Brilliant IIIUltra-Fast SYBR Green QPCR Master Mix (Agilent). Stock cDNA that was synthesised and diluted serially (to a 1ng/µl concentration with nuclease free water (Sigma) after replicate NanoDrop® readings of initial concentration) was used as template for the establishment of standard curve for each gene *and two housekeeping genes RSP7 ribosomal protein SP7(AGAP010592) and Actin 5C (AGAP000651)* (Table 2.1) were diluted. The 1ng/µl stocks were stored at -20°C until required. Serial of each cDNA pool (3 tests, 3 controls for each of the genes for bendiocarb and permethrin and for the laboratory susceptible strain FANG). Each 10µl reaction contained 200nM of forward and reverse primer and 1X master mix containing SYBR® Green I, KAPA™ DNA polymerase and 2.5mM $MgCl_2$. PCR cycling involved 3mins. Incubation at 95°C followed by 40 cycles of 95°C for 3secs, 60°C for 20secs then 72°C for 1sec. Increments with readings every 0.05secs, were performed. qRT-PCR data was analysed after amplification using the MxPro qPCR software (Agilent). According to the $2^{-\Delta\Delta CT}$ method, the fold changes and relative expressions of the candidate genes were calculated with normalisation with the house keeping genes by also combining the efficiency of the PCR (Schmittgen and Livak 2008).

Table 2.1: Genes including their primers used for qRT-PCR

Gene	Forward primer	Reverse primer	Expected size (bp)	Efficiency %/(R²)
Primers used for qRT-PCR				
CYP6P9a	CAGCGCGTACACCAGATTGTGTAA	TCA CAA TTT TTC CAC CTT CAA GTA ATT ACC CGC	92	99(0.99)
CYP6P9b	CAGCGCGTACACCAGATTGTGTAA	TTA CAC CTT TTC TAC CTT CAA GTA ATT ACC CGC	97	100(0.99)
CYP9J11	CAAATTTAAAGAGTGCGCTAGG	GTAGATGGTGCCAAGGATGG	115	99(0.99)
Aldehyde oxidase (Aldoxi)	GACTGGCAGACGATTGGATT	TGTAATCCAGCAACGGTGTC	134	98(0.98)
CYP6AA4	CATCTGGCTGAATGGCACTA	TCAACAATGCCATCAAATCG	109	95(0.97)
CYP6M4	CACTATTCTCTCGCCGAAGG	CAAAGGATCCGCCATTCTAC	119	101(0.99)
CYP6N1	GAAGCATTTCCGTTTTACGC	CGGTGGCTTTATAGCTCGTT	138	93(0.98)
CYP9K1	AGGGCTTCTGGATACGGTTC	CGTACGGTTCGGTTTTGATT	103	910(0.99)
CYP6P2	GCCGACAGAGAAAACACCAG	GAAGGCATGTCCGGTAGTTC	120	100(0.99)
CYP9J13	GATAAAACGCAGGTGCCACT	TTCAAGGTGCTTTGTTCGATT	143	97(0.99)
CYP304b1	GTTTCTGACGTTGGCAGCTT	CCGGTGCGGCTTTATCTC	150	94(0.98)
RSP7*(AGAP010592)*	GTGTTCGGTTCCAAGGTGAT	TCCGAGTTCATTTCCAGCTC	98	100(0.99)
Actin *(AGAP000651)*	TTAAACCCAAAAGCCAATCG	ACCGGATGCATACAGTGACA	111	91(0.99)

2.2.9 Analysis of patterns of polymorphism of candidate resistance genes

To assess a possible correlation between the genetic variability of candidate genes and the resistance phenotype, the full length of genomic DNA of the candidate genes *CYP6P9a* and *CYP6P9b* (including 5'UTR, all exons and the intron) was sequenced for 5 resistant (alive after exposure to bendiocarb or permethrin insecticides) and 5 susceptible (dead after exposure to bendiocarb or permethrin insecticides) mosquitoes with forward and reverse primers for both genes (Table 2.2).

Table 2.2: Primers used for amplification of *CYP6P9a* and *CYP6P9b* for analysis of polymorphism

	Forward	Reverse	sizes
CYP6P9a	ATCCCTAACTATTAA AAGGCAAT	TCA CAA TTT TTC CAC CTT CAA GTA ATT ACC CGC	2.2
CYP6P9b	CATACTCATAATAAC TAGACGCG	TTA CAC CTT TTC TAC CTT CAA GTA ATT ACC CGC	1.8

2.2.9.1 PCR amplification of *CYP6P9a* and *CYP6P9b* genes for bendiocarb and permethrin samples

The following PCR conditions were used for the amplification of the *CYP6P9a* and *CYP6P9b* genes for 5 resistant (alive) and 5 susceptible (dead) samples exposed to bendiocarb or permethrin respectively. Individual PCRs were performed in 30µl reactions containing 3µl of Buffer A, 0.2µl of dNTP 25mM, 1.02µl of forward primer (*CYP6P9a* or *CYP6P9b*), 1.02µl of reverse primer (*CYP6P9a* or *CYP6P9b*), 0.24µl of HiFi Taq, 22.48µl ddH20 and 2µl of DNA; and were conducted on a DNA Thermal Cycler under the following conditions: 5mins at 95ºC, followed by 35 cycles of denaturing at 94ºC for 30s, annealing at 57ºC for 30s and extension at 72ºC for 1min, finishing with an extension step at 72ºC for 10mins.

2.2.9.2 PCR purification

PCR products were purified using the QiaQuick PCR purification kit (Qiagen, Hilden Germany). 8µl of PCR product was added into an Eppendorf tube, followed by an addition of 40µl of Buffer PB. Care was taken to ensure solution was yellow. The entire solution was then transferred into a QiaQuick spin columns with a membrane making

sure membrane was not touched and centrifuged at 13,000 rcf for 1 min. Solution was then discarded. The tube was again centrifuged for 1min to ensure total discard of solution. Column was then transferred on to a new Eppendorf tube and 30μl of purified water was added onto the membrane avoiding contact. Tube was allowed to stand for 1 min and was later centrifuged for 1 min at 13,000 rcf. Column was discarded and the quality and quantity of each sample was checked using NanoDrop™ before being sent off for sequencing to Macrogen (South Korea).

2.2.9.3 Sequence analysis

Sequence quality was checked visually with BioEdit 4.0 (http://www.mbio.ncsu.edu/bioedit/bioedit.html), and alignment was done using CLUSTALW (Thompson *et al*. 1994). The dnaSP 5.10 program (Rozas *et al*. 2010) was used to analyze the sequences after defining the haplotype phase (through the Phase option). It was also used to assess various genetic parameters, such as the number of segregating sites (S), the number of substitutions, the nucleotide diversity π (Nei 1987), number of nucleotide differences (K) (Tajima 1983), θ per nucleotide from the number of mutations (Tajima 1996), haplotype diversity and the D and D^* selection estimates.

The number and location of non-synonymous substitutions were also determined. Changes to amino acid sequence are due to non-synonymous substitutions while only the nucleotide sequence is affected by synonymous substitution. The ratio of synonymous to non-synonymous substitutions can be considered as an indicator of selection pressure (Yang *et al.* 2000). dnaSP 5.10 was used to determine the number of synonymous substitutions per synonymous site (Ks) and non-synonymous substitutions per non-synonymous site (Ka) (Nei and Gojobori, 1986).

2.2.9.4 Test of selection

Various tests of selection were carried out to detect potential signature of selection due to the observed resistance against carbamates or pyrethroids.

2.2.9.4.1 Hudson, Kreitman and Aguadè (HKA) test and the Mcdonald and Kreitman (MK) test (Hudson *et al*. 1987, McDonald and Kreitman, 1991)

The HKA test measures the degree of polymorphism and genetic divergence between species by using the number of variable sites between sequences.

The MK test assesses neutral evolution by using the non-synonymous to synonymous ratio of polymorphisms within species. These statistical tests were performed using dnaSP 5.10 (Rozar *et al.* 2010) with *An. gambiae CYP6P3* genes (AGAP002865-PA) which is the ortholog of the duplicated *CYP6P9a* and *CYP6P9b* genes was used as the out-group.

2.2.9.4.2 The Codon-based Z- test of neutrality, purifying and positive selection was carried out using Mega 5.2 software (Tamura *et al.* 2011, Zang *et al.* 1998, Nei-Gojobori 1986). This test was carried out to further assess signature of selection at each gene. This test uses the Nei-Gojobori method (Nei and Gojobori 1986). This method considers the probability of rejecting the null hypothesis of strict-neutrality (dN=dS) where, dS and dN are the numbers of synonymous and non-synonymous substitutions per site and the variance of difference is computed using the bootstrap method (1000 replicates).

2.2.9.4.3 Tajima's and Fu and Li's test

Tajima's D statistic (Tajima 1998) and Fu and Li's tests D* (Fu and Li 1993) were used to test polymorphism data under the neutral theory of evolution (Kimura 1983). The tests carry out a comparison between the π and θ, which should normally be equal if the equilibrium mutation-drift holds. The theory of neutral evolution assumes that most mutations are neutral and the probability of the fixation of the allele frequency is $1/n$ ($1/2n$ gametes) (Kimura 1983). Values that are negative from the statistical test are an indication of purifying selection while the positive values are for balancing selection. Tajima and Fu and, Li's tests both assume no recombination occurs between sites.

Tajima's test uses two different estimates of θ (π and K) to detect selection (D= (π-K/a)/$\sqrt{V}$ (π-K/a)) (Tajima 1989). The π value is based on nucleotide number that differs per site in two samples from the sequence that are randomly chosen, while K is based on the number of segregating sites that are showing variation among sequences (S) in the sample set. The K value is affected by the existence of low frequency alleles while π is not. When Tajima's test is based on K, the estimate of θ will be greater, and lower when based on π, and D will be negative if samples contain low frequency alleles. However, D will be positive when alleles with intermediate frequency are present within same population and this will not affect K but will increase π. A negative Tajima's D value detects rare mutation and can also indicate earlier events such as hitchhiking, background selection and recent population selection.

2.2.9.5 Phylogenetic tree of *CYP6P9a* and *CYP6P9b* genes for bendiocarb and permethrin samples

A maximum likelihood (ML) phylogenetic tree was constructed using Mega 5.2 (Tamura *et al.* 2011) for the full sequences (coding and non coding) for the *CYP6P9a* and *CYP6P9b* genes for both bendiocarb and permethrin samples. The best model fit for the *CYP6P9a* gene was Jukes-Cantor model (JC) and for the *CYP6P9b* gene, Kimura Nei model was best model fit which the best haplotype dataset out of the candidate models. This model was then used to generate the ML tree as implemented in MEGA 5.2 with 1000 bootstrap replications to assess the trees robustness.

A haplotype network was also built using the TCS program (Clement *et al.* 2000) for both the full length (Coding and non- coding) region (using the 95% connection limit, with gaps been treated as a fifth state) to further assess haplotype diversity and their potential correlation with resistance phenotype.

2.3. Results

2.3.1 Mosquito collection

A total of 300 indoor-resting Anopheles mosquitoes were collected inside houses in Chikwawa over the collection period. Around 240 were morphologically identified as belonging to the *An. funestus* group and 60 were identified as belonging to the *An. gambiae* complex.

2.3.2 Mosquito rearing

Around 130 egg batches were obtained from the 240 oviposition Eppendorf tubes that were set up with gravid *An. funestus* females individually. It was noticed that, mosquitoes that were reared in distilled water grew less quickly and also looked less fit than those reared in mineral (bottled) water. It took 16 days from egg rearing to adult stage and about 4000 F_1 adult mosquitoes were obtained. The PCR results confirmed that eggs were laid by *An. funestus* s.s by running all the samples on 1.5% agarose gel, checking their sizes if it matches the expected size for *An. funestus* at 505bp (Fig. 2.7).
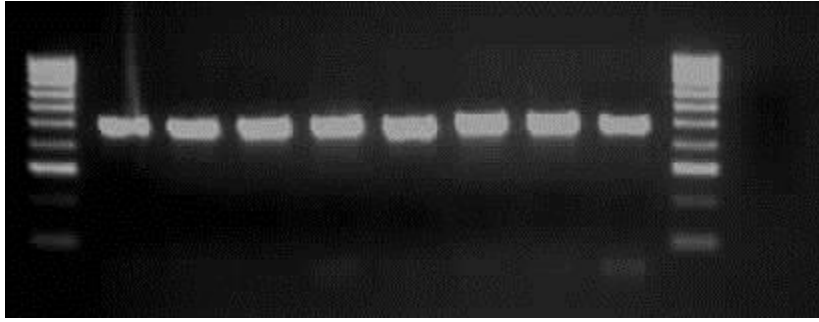
Fig. 2.7: Agarose gel picture showing expected size of *An. funestus* species ID

2.3.3 Insecticide susceptibility tests

WHO Bioassays carried out on F_1 individuals that were generated from the 130 egg batches indicated that, *An. funestus* obtained from Chikwawa was resistant to bendiocarb with a mortality rate of 60% for females and 55.1% for males after 1h exposure to 0.01% bendiocarb. The *An. funestus* population from Chikwawa was also resistant to both type I and type II pyrethroids. A mortality rate of 47.2% was recorded when female *An. funestus* were exposed to 0.75% permethrin, type I pyrethroid, while a mortality rate of 58.1%   was   observed   for   males   (Fig.   2.8).
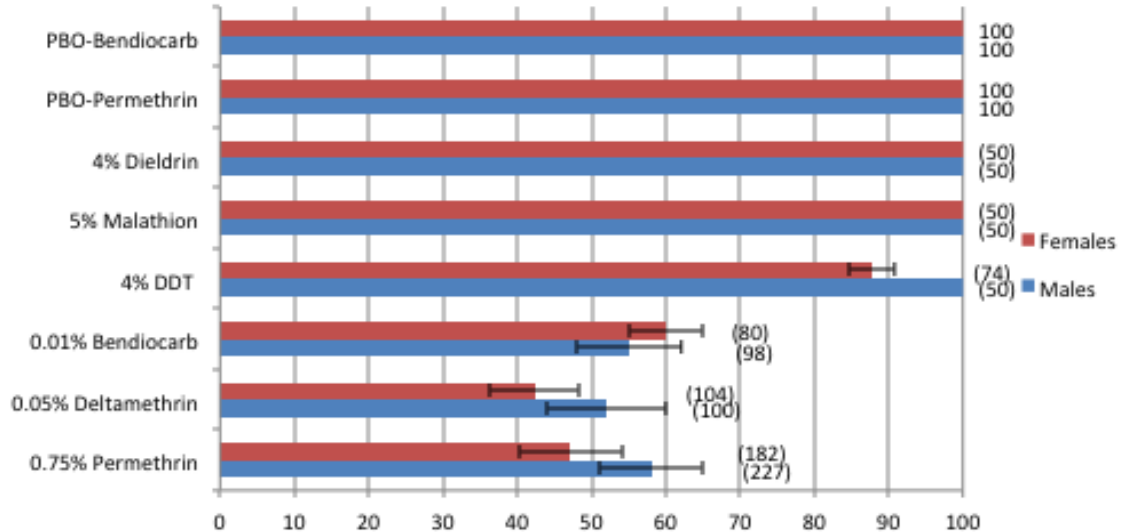
Fig. 2.8: Mortality profile of *An. funestus* exposed to bendiocarb, permethrin, dieldrin, malathion and DDT insecticides and also results of PBO for bendiocarb and permethrin

A similar resistance level was observed after 1h exposure to 0.05% deltamethrin (type II pyrethroid) with a mortality of 42.3% observed for females and 52% for males. A suspected resistance to DDT was observed for females with a mortality rate of 87.8% while a full susceptibility was observed for males after 1h exposure to 4% DDT (Fig.

53

2.8). A full susceptibility was observed for malathion (organophosphate) and dieldrin (organochlorine) with 100% mortality after 1h exposure for respectively 5% malathion and 4% dieldrin (Fig. 2.8). No mortality was observed in the control tubes.

2.3.4 PBO synergist

A full susceptibility recovery was observed for both permethrin and bendiocarb when mosquitoes were pre-exposed to PBO with 100% mortality observed for both insecticides (Fig. 2.8). This suggests that cytochrome P450 monooxygenases could be playing a major role in the resistance mechanisms against carbamates and pyrethroids in the *An. funestus* population from Chikwawa. .

2.3.5 Detection of main genes associated with bendiocarb (carbamate) resistance by microarray

2.3.5.1 RNA extraction and quality assessment

Total RNA was successfully extracted from 3 biological replicates for pools of 10 females for the following samples: Resistant (R- alive after 1hr exposure to bendiocarb or permethrin), Control (C-mosquitoes unexposed to insecticide thus representative of the wild type population), and the Susceptible (S- unexposed mosquitoes from the fully susceptible laboratory strain). The concentrations and quality assessment of these samples are presented in (Table S1) in the appendix.

Complementary RNA (cRNA) were successfully obtained and labelled from these samples as indicated by the quality and quantity estimates in the appendix (Table S2).

2.3.5.2 Quality control of microarray experiments

All eight microarrays were hybridised successfully with good quality images and low background signal. After the feature extraction, the QC reports produced maximum scores of 11 for most arrays with only few with scores of 9 or 10.

Probes differentially expressed were obtained using the set of probes or entities that passed the filtration based on flags present or marginal in at least 1 out of the 5 hybridisations for each comparison.

The number of probes differentially expressed (>2-fold change, *P<0.01*) for each comparison for bendiocarb and between them are presented in Fig. 2.9.
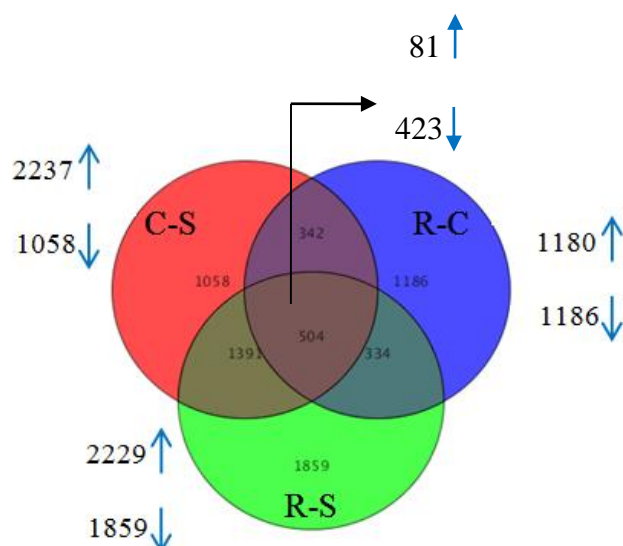
Fig. 2.9: Representation of probes differentially expressed for bendiocarb R-S, R-C and C-S for (>2-fold change, P< *0.01*) 4x44k microarray chip.

The R-S comparison exhibited a total of 3295 differentially expressed at *P<0.01* of which, 2229 probes were up-regulated in the bendiocarb resistant mosquitoes while 1859 probes were down-regulated in comparison to the susceptible FANG strain (S). The R-C comparison had 2366 probes differentially expressed at *P<0.01* of which 1180 were up-regulated and 1186 were down-regulated respectively. The C-S had 3295 probes at *P<0.01* of which 2237 were up-regulated and 1058 were down-regulated respectively. The R-C, C-S and R-S all shared 504 probes in common, of which 81 probes were up-regulated and 423 were down-regulated. The R-S shared 1391 probes with the C-S, and also shared 334 probes with R-C. The R-C shared 342 probes with the C-S. An inspection of the list of probes differentially expressed for each comparison did not detect conflicting results between probes targeting the same gene (where some could be over-expressed while the others are down-regulated. The most common case was that for some genes, some probes were differentially expressed while the others were not. Overall, all these probes were kept individually in the list of over-expressed or down-regulated.

2.3.5.3 Genes up-regulated in bendiocarb resistant mosquitoes

Priority was given to the list of probes significantly up-regulated in the bendiocarb resistant vs control mosquitoes (R-C) (Table 2.3), as this comparison directly compares mosquitoes with same genetic background which only differ for the resistance phenotype. In this study the R-C comparison is likely to detect probes differentially expressed in relation to bendiocarb resistance as the mortality level at 60% provides a relatively

resistance contrast between the alive (R) 100% resistant and the control wild type which is only 40% resistant. In addition, consideration was given to probes that were commonly up-regulated in the R-S and C-S (Table 2.3), and a list of detoxification genes differentially expressed in the various comparisons is presented in the appendix (Table S3, S4 and S5).

2.3.5.3.1 Genes over-expressed in R-C

Analysis of the set of probes over-expressed in the resistant vs control comparison identified several gene families with possible association with bendiocarb resistance. The most highly over-expressed probes were non detoxification genes with unknown names such as combined_c3103 (FC 31.8), combined_c7168 (FC 22.8), the d7- related 1 protein (FC 20). These probes however were not highly over-expressed across the other comparisons. Attention was then paid on possible detoxification genes that might be playing a role in resistance. Among the detoxification genes, probes belonging to cytochrome P450 genes were by far predominant with 22 probes over-expressed (Table 2.3). In contrast, only one or two probes were over-expressed for other detoxification gene families such as glutathione-S transferases, aldehyde oxidases, ABC transporter genes, glucosyl glucuronosyl transferases, argininosuccinate lyase and short chain dehydrogenases. Several probes belonging to cuticle protein genes were also over-expressed suggesting the presence of a reduced penetration mechanism against bendiocarb in this *An. funestus* population. Several probes belonging to salivary gland protein genes were also over-expressed, notably the D7-related 1 protein gene which was highly over-expressed (FC=20). Other genes families detected are commonly associated with insecticide resistance such as peroxidases, heat shock proteins and proteases such as Trypsin. A set of probes belonging to immune response genes were also over-expressed including the TEP1 gene and several CLIP genes.

2.3.5.3.2 Candidate bendiocarb resistance genes

Selection of best possible candidate genes conferring resistance to bendiocarb was done by also considering the probes commonly over-expressed in comparison between resistant and the susceptible FANG strain (R-S) or between the Control and FANG (C-S). Based on this criteria, the cytochrome P450 gene *CYP6P9a* gene appears to be the best candidate detoxification gene associated with bendiocarb resistance in the *An. funestus* from Chikwawa, as two probes of this gene were consistently over-expressed in

the three R-C, R-S and C-S comparisons (Table 2.3). The other gene over-expressed in the 3 comparisons was a cuticle protein gene (CD577515.1), suggesting that a reduced penetration through cuticle thickening could be operating beside a detoxification through elevated expression of P450 genes. The transcript CD577515 is 86% identical to the AFUN004204 gene in the newly released *An. funestus* genome which is in turn 92% identical to the cuticle protein gene AGAP003382-RA in *An. gambiae*.

The list of genes commonly over-expressed in R-C and R-S and is likely associated with bendiocarb resistance, includes several probes that belong to cytochrome P450 genes. Three probes belonging to the *CYP6Z1* gene is consistently over-expressed in both comparisons with FC of 7.3, 6.7 and 5.0 respectively in R-C (Table 2.3). However, one of the *CYP6Z1* probes is down-regulated in C-S while the two others are not significantly expressed. This could suggest an induction of *CYP6Z1* in relation to bendiocarb resistance. Two probes belonging to the *CYP9J11* are also over-expressed in both R-C (FC 6.8 and 5.3) and R-S (FC10.8 and 10.8). This list of cytochrome P450 genes commonly over-expressed in both R-C and R-S also includes two probes of the *CYP6P9b* genes, one probes of *CYP6P4b*, *CYP6P5* and a pseudo P450 gene located between the *CYP6AA2* and *CYP6P9a* in the rp1 P450 cluster (Table 2.3). A set of probes belonging to cuticle protein genes are also consistently over-expressed between R-C and R-S but not in C-S, similar for two probes of a salivary protein gene d7-related protein 1 and a short chain dehydrogenase gene. A set of three probes belonging to trypsin is over-expressed in both R-C and R-S but are all highly down-regulated in the C-S comparison.

Analysis of probes only over-expressed in R-C also detected several cytochrome P450 probes, among which two probes belonging to the *CYP4G16* are the most up-regulated (FC 9.1 and 8.7). However, these two probes are also down-regulated in the C-S comparison. Similarly, many of the other probes up-regulated in the R-C comparison were also down-regulated in the C-S (Table 2.3). The list of probes up-regulated in R-C and not differentially expressed in R-S and C-S is dominated by P450 genes such as *CYP6M4, CYP325D1* and *CYP4H18*. This list also includes other genes such as the glutathione-S transferase GSTO1, an aldehyde oxidase, an arginino-succinate lyase, several peroxidases, odorant receptor genes, salivary proteins or heat shock proteins.

2.3.5.3.3 Probes down-regulated

The most down-regulated probe in all comparison (R-C, R-S and C-S) (Table 2.4) belong to the *Anopheles gambiae* pest agap 01258-pa (FC R-C -10.14, R-S -4.73, C-S -1.07). The zinc finger ccch domain containing protein 15-like protein (FC- R-C -10.1, R-S -5.6 and C-S -7.0). There were other probes like the alanine- glyoxylate aminotransferase, sorbitol dehydrogenase, carboxylpeptidase and the litaf-like protein. There were other down-regulated probes but most of which were either down-regulated in the R-C (*CYP9K1* and carboxylesterase) and up-regulated in the R-S or C-S or mostly down-regulated in the R-S and C-S while up-regulated in the R-C. However, only few detoxification genes were found among the down-regulated probes.

Table 2.3: Probes associated with bendiocarb or cross-resistance to bendiocarb/permethrin: Probes commonly over-expressed in R-C bendiocarb and in other comparisons with P<*0.01* and Fold-change >2 were primarily selected. Significant expression in other comparisons for permethrin are also indicated to detect genes potentially involved in cross resistance.

Probes	Gene names	Bendiocarb FC				Permethrin		Gene function
		R-C	P value	R-S	C-S	R-C	R-S	
Probes commonly over-expressed in all comparisons								
CUST_26_PI406199775	*CYP6P9a*	5.2	0.0026	12.6	5.6	7.9	23.2	cytochrome p450
CUST_25_PI406199775	*CYP6P9a*	2.4	0.0067	17.3	7.8	3.4	37.1	cytochrome p450
Probes commonly expressed at least in R-C and R-S bendiocarb or in R-C and R-S permethrin								
CUST_6134_PI406199769	combined_c3103	31.8	0.0051	-3.4	-20.8	8.1	10.4	---NA---
CUST_13889_PI406199769	combined_c7168	22.8	0.0024		-13.6		29.8	---NA---
CUST_593_PI406199772	EE589616.1	20.0	0.0021	4.7		31.8	9.6	d7-related 1 protein
CUST_1045_PI406199772	EE590131.1	20.6	0.0008	-4.3	-31.6	3.5	11.9	---NA---
CUST_2400_PI406199772	CD578201.1	19.5	0.0027				6.7	---NA---
CUST_2476_PI406199769	combined_c125	19.5	0.0015	-2.0	-27.8	-6.6	7.7	---NA---
CUST_441_PI406199769	combined_c223	17.7	0.0015	-2.0		-2.2	8.6	cytochrome c oxidase subunit viia
CUST_442_PI406199769	combined_c223	16.9	0.0016				7.5	cytochrome c oxidase subunit viia
CUST_359_PI406199772	EE589504.1	16.7	0.0010	4.3	-12.1	26.1		d7-related 1 protein
CUST_1632_PI406199769	combined_c825	16.4	0.0026		-13.0		7.3	---NA---
CUST_717_PI406199772	EE589504.1	16.0	0.0008	4.6	-9.9	28.4	6.9	d7-related 1 protein
CUST_6135_PI406199769	combined_c3103	15.6	0.0090	3.7	-17.9	11.0	7.1	---NA---
CUST_1413_PI406199772	EE589737.1	15.5	0.0019	4.9		31.8		d7-related 1 protein
CUST_4961_PI406199769	combined_c2513	15.3	0.0018	-2.8	-30.3	3.7	7.0	---NA---
CUST_2082_PI406199769	combined_c1050	13.0	0.0017		-16.0		5.2	troponin c

Probes	Gene names	Bendiocarb FC				Permethrin		Gene function
		R-C	P value	R-S	C-S	R-C	R-S	
Probes commonly expressed at least in R-C and R-S bendiocarb or in R-C and R-S permethrin								
CUST_3498_PI406199772	CD577638.1	12.5	0.0022	-2.2	-26.0	6.0	4.9	cytochrome c oxidase subunit viia
CUST_290_PI406199772	EE589911.1	10.3	0.0014			11.4		ge rich salivary gland protein
CUST_9786_PI406199769	combined_c4956 (*CYP4G16*)	9.1	0.0022		-9.5	9.6		cytochrome p450
CUST_2463_PI406199772	CD578169.1	8.7	0.0022			12.7		trypsin
CUST_9787_PI406199769	combined_c4956 (*CYP4G16*)	8.7	0.0016		-9.1	10.1		cytochrome p450
CUST_4508_PI406199772	CD577111.1	7.9	0.0023	-2.7		9.2		atp synthase f0 subunit 6
CUST_45_PI406199775	*CYP6Z1*	7.3	0.0014	2.9		7.8	2.9	cytochrome p450
CUST_13272_PI406199769	combined_c6791 (*CYP9J11*)	6.8	0.0013	10.8			9.3	cytochrome p450
CUST_44_PI406199775	*CYP6Z1*	6.7	0.0021	2.7	-		2.7	cytochrome p450
CUST_3440_PI406199772	CD577671.1	6.6	0.0011		-3.6	7.6		peritrophin a
CUST_9482_PI406199798	AGAP008292-RA	6.5	0.0024	2.7	-2.9		4.3	trypsin
CUST_5118_PI406199772	BU038883	6.1	0.0017		-3.9	7.3		peritrophin a
CUST_5111_PI406199772	BU038886	5.9	0.0039	2.5	-3.3	10.3	3.8	trypsin
CUST_16055_PI406199769	combined_c8512	5.7	0.0035			4.1		serine protease 14
CUST_2288_PI406199772	CD578260.1	5.6	0.0014	2.4	-3.7		3.5	trypsin
CUST_13273_PI406199769	combined_c6791 (*CYP9J11*)	5.3	0.0010	10.8			8.6	cytochrome p450
CUST_43_PI406199775	*CYP6Z1*	5.0	0.0023	2.5	-2.6	6.1	3.1	cytochrome p450
CUST_356_PI406199772	EE589850.1	3.9	0.0087		-7.1	5.5		d7-related 2 protein
CUST_3736_PI406199772	CD577515.1	3.1	0.0066	4.2	3.9			cuticle protein

Probes	Gene names	Bendiocarb				Permethrin		Gene function
		FC						
		R-C	P value	R-S	C-S	R-C	R-S	
Probes commonly expressed at least in R-C and R-S bendiocarb or in R-C and R-S permethrin								
CUST_29_PI406199775	*CYP6P9b*	3.1	0.0054	4.2			5.4	cytochrome p450
CUST_33_PI406199775	Pseudo_P450_between_6AA2_and_6P9a	3.0	0.0030	2.4				cytochrome p450
CUST_3727_PI406199772	CD577519.1	3.0	0.0039	3.3				cuticle protein
CUST_3731_PI406199772	CD577517.1	2.9	0.0039	3.0				cuticle protein
CUST_4100_PI406199772	CD577317.1	2.8	0.0023	2.8				cuticle protein
CUST_9502_PI406199769	combined_c4812	2.8	0.0062	3.9				short-chain dehydrogenase
CUST_3729_PI406199772	CD577518.1	2.6	0.0027	2.6				cuticle protein
CUST_3395_PI406199772	CD577694.1	2.5	0.0030	2.9				cuticle protein
CUST_1977_PI406199772	CD664220.1	2.5	0.0014	3.5				cuticle protein
CUST_28_PI406199775	*CYP6P9b*	2.4	0.0058	3.9			4.8	cytochrome p450
CUST_3734_PI406199772	CD577516.1	2.3	0.0060	3.3				cuticle protein
CUST_4102_PI406199772	CD577316.1	2.2	0.0044	2.8				cuticle protein
CUST_17_PI406199775	*CYP6P4a*	2.1	0.0032	3.1				cytochrome p450
CUST_23_PI406199775	*CYP6P5*	2.1	0.0065	3.5				cytochrome p450
Probes only over-expressed in R-C bendiocarb								
CUST_1719_PI406199772	EE589407.1	12.6	0.0009		-30.2			ge rich salivary gland protein
CUST_1592_PI406199772	EE589544.1	9.2	0.0008					ge rich salivary gland protein
CUST_3256_PI406199772	CD577765.1	7.3	0.0015		-5.1			lethal essential for life l2efl
CUST_4607_PI406199772	CD577060.1	6.4	0.0072	-3.2	-9.9			atp synthase f0 subunit 6

Probes	Gene names	Bendiocarb				Permethrin		Gene function
		FC						
		R-C	P value	R-S	C-S	R-C	R-S	
Probes only over-expressed in R-C bendiocarb								
CUST_4583_PI406199772	CD577072.1	6.4	0.0069	-3.3	-9.8			atp synthase f0 subunit 6
CUST_5095_PI406199772	BU038894	6.1	0.0017		-6.8			lethal essential for life l2efl
CUST_5096_PI406199772	BU038894	5.9	0.0018		-6.3			lethal essential for life l2efl
CUST_1331_PI406199769	combined_c674	5.1	0.0062		-3.1			serine protease 14
CUST_692_PI406199798	AGAP000572-RA	4.8	0.0089					serine protease desc4
CUST_7696_PI406199798	AGAP008141-RA	4.7	0.0042					argininosuccinatelyase
CUST_1736_PI406199772	EE589383.1	4.6	0.0054		-13.9			d7-related 3 protein
CUST_3255_PI406199772	CD577765.1	4.4	0.0066		-5.7			lethal essential for life l2efl
CUST_9088_PI406199798	AGAP006416-RA	4.3	0.0088					serine protease sp24d
CUST_810_PI406199772	EE589416.1	4.1	0.0018		-9.7			d7-related 2 protein
CUST_5673_PI406199769	combined_c2871	3.9	0.0064		-4.9			tep1 protein
CUST_2536_PI406199772	CD578133.1	3.8	0.0034		-5.4			oxidase peroxidase
CUST_633_PI406199788	gb-GSTO1	3.7	0.0015					glutathione s-transferase
CUST_1998_PI406199772	CD664210.1	3.6	0.0064					glutathione peroxidase
CUST_799_PI406199769	combined_c404	3.6	0.0010					serine protease
CUST_5373_PI406199798	AGAP004582-RA	3.6	0.0023		-2.6			heat shock protein 70 b2
CUST_923_PI406199772	EE589310.1	3.3	0.0078		-8.1			gsg7 salivary protein
CUST_759_PI406199772	EE589462.1	3.1	0.0022					sg2a salivary protein

62

Probes	Gene names	Bendiocarb FC				Permethrin		Gene function
		R-C	P value	R-S	C-S	R-C	R-S	
Probes only over-expressed in R-C bendiocarb								
CUST_5662_PI406199798	AGAP004581-RA	3.1	0.0078					heat shock protein 70 b2
CUST_3090_PI406199772	CD577854.1	3.1	0.0058					hsp70 binding protein
CUST_3883_PI406199798	AGAP004247-RA	3.1	0.0056	-2.4	-4.5		2.8	glutathione peroxidase
CUST_5009_PI406199772	BU038937	3.0	0.0065	-3.1	-5.9			cytochrome p450 4g15
CUST_7044_PI406199769	combined_c3564	3.0	0.0063		-4.3			serine protease
CUST_5907_PI406199769	combined_c2988	2.9	0.0044		-2.9			Glucosylglucuronosyl transferases
CUST_8727_PI406199769	combined_c4419	2.9	0.0022		-2.7			abc transporter
CUST_4047_PI406199772	CD577343.1	2.9	0.0075					cuticle protein
CUST_4104_PI406201128	AGAP002418-RA	2.8	0.0034					cytochrome p450
CUST_3438_PI406199772	CD577672.1	2.8	0.0035		-4.9			peritrophin a
CUST_6793_PI406199769	combined_c3435	2.8	0.0024		-2.7			trypsin
CUST_5107_PI406201128	AGAP002204-RA	2.7	0.0030					cytochrome p450 CYP325D1
CUST_157_PI406199798	AGAP000088-RA	2.7	0.0016					cytochrome p450
CUST_5640_PI406199769	combined_c2855	2.7	0.0044					short-chain dehydrogenase
CUST_4033_PI406199769	combined_c2039	2.7	0.0066					ctlma3 protein
CUST_16709_PI406201128	AGAP008212-RA	2.5	0.0053					cytochrome p450
CUST_15404_PI406201128	AGAP006917-RA	2.5	0.0034					gustatory receptor (agap006917-pa)
CUST_15844_PI406199769	combined_c8405	2.5	0.0090					serine protease
CUST_22183_PI406201128	AGAP006226-RA	2.4	0.0022					aldehyde oxidase
CUST_15573_PI406201128	AGAP009033-RA	2.4	0.0056					chorion peroxidase

Probes	Gene names	Bendiocarb				Permethrin		Gene function
		FC						
		R-C	P value	R-S	C-S	R-C	R-S	
Probes only over-expressed in R-C bendiocarb								
CUST_8869_PI406199798	AGAP009393-RA	2.4	0.0063					odorant receptor 13a
CUST_17251_PI406201128	AGAP008392-RA	2.4	0.0091					hsp70 binding protein
CUST_13266_PI406199769	combined_c6787	2.4	0.0093	-2.9	-7.4			clipa6 protein
CUST_14660_PI406199769	combined_c7718	2.3	0.0093					clipb17 protein
CUST_6151_PI406199769	combined_c3111	2.3	0.0053		-4.3			clipb5 protein
CUST_924_PI406199798	AGAP000871-RA	2.2	0.0038					ctlma3 protein
CUST_7029_PI406199769	combined_c3556	2.1	0.0093					cytochrome p450
CUST_1946_PI406199798	AGAP002197-RA	2.1	0.0029					cytochrome p450
CUST_369_PI406199788	gb-CYP4H18	2.0	0.0021					cytochrome p450 4d1
CUST_722_PI406199788	gb-PX5A	2.0	0.0098					oxidase peroxidase

Table 2.3 was built in accordance to the criteria established in section **2.3.5.3** to detect the best resistance genes to bendiocarb or cross resistance bendiocarb/pyrethroid. To avoid a multitude of tables in summarising the results, potential genes over-expressed in the bendiocarb R-C comparison (primarily as this directly compares bendiocarb resistant vs non exposed samples with the same genetic background) but also commonly over-expressed in R-S or C-S were included. This table also indicated whether the bendiocarb R-C over-expressed probes are also over-expressed in the permethrin R-S and R-C comparisons allowing detecting candidate genes associated with cross resistance. The probes in the table are ranked from top to lowest expression in R-C and also according to whether they are commonly over-expressed in all comparisons or in a set of comparisons or only in R-C. Also probes over-expressed only in the R-C are also ranked from top to lowest.

Table 2.4: Probes down-regulated in R-C comparison and other comparisons for bendiocarb and permethrin are also indicated at P<*0.01* and Fold-change >2.

Probes	Gene Name	R-C bend	P Value	R-S bend	C-S	R-S perm	R-C perm	Description
CUST_18430_PI406201128	AGAP012528-RA	-10.14	0.0081	-4.73	-1.07	-1.46	-4.16	anopheles gambiae pest agap012528-pa
CUST_24436_PI406201128	AGAP012528-RA	-10.1	0.0048	-5.6	-7.0		-10.1	zinc finger ccch domain-containing protein 15-like protein
CUST_9504_PI406199798	AGAP008634-RA	-9.6	0.0074				-9.1	candidate odorant receptor
CUST_909_PI406199798	AGAP001012-RA	-7.4	0.0080	15.5				sec61 protein translocation complex beta subunit
CUST_113_PI406199788	gb-COEAE1D	-6	0.0058	4.9				Carboxylesterase
CUST_2754_PI406201128	AGAP000818-RA_Cytoch... CYP9K1	-5.9	0.0063	9.5				Cytochrome P450
CUST_217_PI406199798	AGAP000313-RA___X	-5.8	0.0093	-2.2	-0.02	-0.05	-2.31	alanine-glyoxylate aminotransferase
CUST_650_PI406199769	combined_c327	-5.5	0.0063	0.42	2.36	-0.42	-2.87	sorbitol dehydrogenase
CUST_9986_PI406199798	AGAP006207-RA___2L	-5.2	0.0053	-0.76	-0.17	-0.21	-0.96	carboxypeptidase b
CUST_11440_PI406201128	AGAP004928-RA	-4.92	0.0055	-2.19	-0.25	-0.18	-2.1	litaf-like protein

2.3.6 Investigation of cross resistance between bendiocarb and permethrin

To assess the possible existence of a cross-resistance between carbamate and pyrethroid resistance, the microarray-based transcription profiling of the bendiocarb resistant mosquitoes was compared to that of permethrin resistant mosquitoes from the same location. The data of the microarray experiment for permethrin, notably the R-S and R-C comparison is taken from the analyses published by Riveron *et al.* (2013). In addition, the availability of a recent and more comprehensive 8x60k microarray chip was used to detect possible new genes associated in bendiocarb resistance or cross-resistance to permethrin by comparing the two R-S of permethrin and bendiocarb.

2.3.6.1 Cross-resistance using the 4x44k chip

The number of probes that were differentially expressed (>2-fold change, *P<0.01*) for each comparison for permethrin and between them is indicated in (Fig. 2.10).

The C-S comparison is the same as for the bendiocarb analysis. The R-C had 4158 probes P<*0.01* of which 2657 were down-regulated and 1506 were up-regulated. The R-S had 1974 probes P<*0.01* of which 698 were down-regulated and 1276 were up-regulated in comparison to the susceptible strain. The R-S shared 788 probes with the C-S and also shared 179 probes with R-C. The R-C, C-S and R-S all shared 309 probes in common, of which 108 were up-regulated and 201 were down-regulated. While the R-C shared 1013 probes with the C-S (Fig. 2.10).
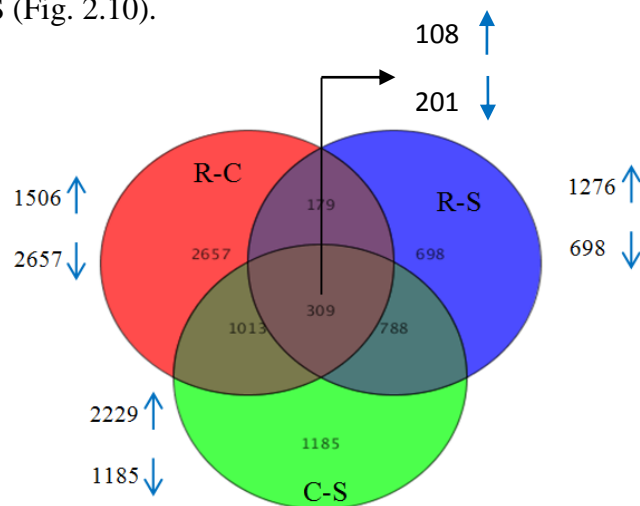
Fig. 2.10: Representation of probes differentially expressed for permethrin R-S, C-S and R-C for the 4x44K microarray chip.

2.3.6.2 Candidate genes associated with cross-resistance for bendiocarb and permethrin

Detection of candidate genes possibly conferring a cross-resistance between bendiocarb and permethrin resistance was done by selecting probes commonly over-expressed in all or most comparisons for both insecticides. With this approach, two probes of *CYP6P9a* were the only probes consistently over-expressed in all the 5 comparisons (Table 2.3) suggesting that *CYP6P9a* over-expression is possibly associated with a cross-resistance between these two insecticide classes. Two probes from *CYP6Z1* were over-expressed in R-C and R-S for both insecticides (Table 2.3), also suggesting a possible role for this other P450 gene in the cross-resistance. Probes for the salivary D7-related 1 protein gene and for trypsin were also up-regulated in both R-C and R-S for bendiocarb and permethrin. Probes for the P450 genes *CYP9J11* and *CYP6P9b* over-expressed in both R-C and R-S comparisons for bendiocarb were also over-expressed in R-S for permethrin. Other probes of these two P450s were also found to be over-expressed in the R-C for permethrin (Riveron *et al.* 2013). This suggests that *CYP6P9b* and *CYP9J11* could also pay a role in the cross-resistance. The Cytochrome P450 *CYP4G16* was the only P450 expressed only in both R-C comparisons

2.3.6.3 New Candidate genes for bendiocarb and cross resistance with permethrin using the new 8x60k chip

When considering the probes that were differentially expressed using the 8x60K chip, the R-S bendiocarb exhibited a total of 2541 probes differentially expressed at *P<0.01* and FC>2, of which 1015 were down-regulated and 1426 were up-regulated in comparison to the FANG strain while R-S permethrin exhibited a total of 1974 probes differentially expressed, of which 317 were down-regulated and 1657 were up-regulated in comparison to the susceptible strain. R-S bendiocarb and R-S permethrin both shared 1588 probes in common, of which 558 were up-regulated and 1030 were down-regulated (Fig. 2.11).
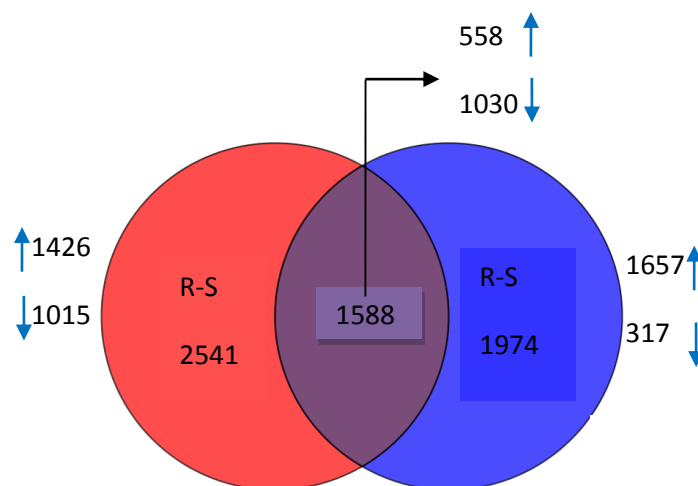
Fig. 2.11: Representation of probes differentially expressed for bendiocarb R-S, permethrin R-S for the 8x60K microarray chip.

Analysis of the list of probes commonly up-regulated between R-S bendiocarb and R-S permethrin with the new 8x60k Agilent chip overall confirmed the results obtained with the 4x44k chip. The main gene families found up-regulated with the 4x44k were again predominant with the 8x60k chip with cytochrome P450s still the most predominant gene families. Among the P450 genes commonly over-expressed in both R-S and possibly associated with cross-resistance, the main candidate genes identified with the 4x44k were all detected for most through several probes including *CYP6P9a*, *CYP6Z1*, *CYP6P9b* (Table 2.5). However, new P450 genes were also detected using the 8x60k such as the *CYP6M7* gene, *CYP9K1* and *CYP6Y2*. The over-expression of the several trypsin probes was also confirmed with the new 8x60k.

The list of probes over-expressed in the R-S bendiocarb but not in the R-S permethrin, detected new genes notably cytochrome P450s such as *CYP6Z3*, *CYP6S2* and others as shown in the (Table S5) in appendix. Other genes from different gene families were also only up-regulated in bendiocarb R-S including an esterase b1, glutathione-S transferases, short chain dehydrogenases and more (Table S6).

2.3.6.4 Down regulated genes for R-S bendiocarb and permethrin

The most down-regulated genes for both bendiocarb and permethrin was the monkey king protein, the Isoform and the cytochrome oxidase subunit 3. They were all down-regulated with high FC in both bendiocarb and permethrin (Table 2.6).

Table 2.5: Probes from detoxification genes or resistance associated genes commonly up-regulated in both R-S comparisons for bendiocarb and permethrin resistance with the new 8x60k *An. funestus* chip.

ProbeName	Systematic name	FC R-S Bendiocarb	P value	FC R-S Permethrin	P value	Description
CUST_9227_PI426302897	Afun009227	31.9	0.0160	66.3	0.0046	argininosuccinatelyase
CUST_27_PI406199775	*CYP6P9a*	24.4	0.0472	39.4	0.0034	cytochrome p450
CUST_30_PI406199775	*CYP6P9b*	17.8	0.0373	24.0	0.0034	cytochrome p450
CUST_7_PI426302915	*CYP6M7*	10.2	0.0224	10.1	0.0054	cytochrome p450
CUST_1096_PI406199769	combined_c557	8.9	0.0352	11.8	0.0044	Trypsin
CUST_26_PI406199775	*CYP6P9a*	8.0	0.0444	9.8	0.0039	cytochrome p450
CUST_1097_PI406199769	combined_c557	7.1	0.0433	6.4	0.0097	Trypsin
CUST_25_PI426302915	*CYP6Y2*	6.1	0.0288	2.9	0.0209	cytochrome p450
CUST_26_PI426302915	*CYP6Y2*	6.0	0.0329	2.8	0.0097	cytochrome p450
CUST_7369_PI426302897	Afun007369 (CYP6P9a)	4.8	0.0165	2.5	0.0136	cytochrome p450
CUST_3394_PI426302897	Afun003394 *(CYP315A1)*	4.8	0.0420	2.4	0.0326	cytochrome p450
CUST_9584_PI426302897	Afun009584 (*CYP6N4*)	4.7	0.0497	3.2	0.0083	cytochrome p450

ProbeName	Systematic name	FC R-S Bendiocarb	P value	FC R-S Permethrin	P value	Description
CUST_8_PI426302915	*CYP6M7*	4.5	0.0052	3.7	0.0031	cytochrome p450
CUST_12197_PI426302897	Afun012197 *(CYP9J11)*	4.4	0.0170	2.8	0.0071	cytochrome p450
CUST_9482_PI406199798	AGAP008292-RA	4.0	0.0135	3.9	0.0102	Trypsin
CUST_8241_PI406199769	combined_c4173	3.8	0.0334	4.8	0.0057	glycoprotein 93
CUST_493_PI426302897	Afun000493	3.7	0.0232	2.2	0.0137	aldehyde oxidase
CUST_8240_PI406199769	combined_c4173	3.7	0.0450	4.6	0.0054	glycoprotein 93
CUST_1458_PI406199769	combined_c738	3.7	0.0245	3.8	0.0160	short Chain dehydrogenase
CUST_7008_PI426302897	Afun007008	3.7	0.0487	2.2	0.0289	ABC transporter
CUST_7469_PI426302897	Afun007469 *(CYP9J5)*	3.6	0.0399	3.1	0.0062	cytochrome p450
CUST_7769_PI426302897	Afun007769 *(CYP9K1)*	3.5	0.0442	2.4	0.0079	cytochrome p450
CUST_28_PI426302915	*CYP6Z1*	3.2	0.0208	2.5	0.0087	cytochrome p450
CUST_5111_PI406199772	BU038886	3.2	0.0358	3.7	0.0091	Trypsin
CUST_899_PI406199772	EE589329.1	3.1	0.0458	7.6	0.0133	d7-related 1 protein
CUST_19_PI426302897	Afun000019	3.1	0.0433	3.5	0.0118	Trypsin
CUST_2287_PI406199772	CD578260.1	3.0	0.0434	3.6	0.0107	Trypsin

ProbeName	Systematic name	FC R-S Bendiocarb	P value	FC R-S Permethrin	P value	Description
CUST_7861_PI426302897	Afun007861 (*CYP6Z1)*	3.0	0.0424	3.1	0.0034	cytochrome p450
CUST_27_PI426302915	*CYP6Z1*	3.0	0.0289	2.5	0.0083	cytochrome p450
CUST_7646_PI426302897	Afun007646	2.7	0.0136	2.0	0.0048	aldehyde oxidase
CUST_29_PI406199775	*CYP6P9b*	2.4	0.0433	2.5	0.0136	cytochrome p450
CUST_7499_PI426302897	Afun007499	2.3	0.0092	2.9	0.0141	glutathionetransferase
CUST_360_PI406199772	EE589855.1	2.0	0.0368	2.7	0.0259	d7-related 1 protein

Table 2.6 Probes down-regulated in R-S comparison for bendiocarb and permethrin indicated at P<*0.01* and Fold-change >2.

Probe Name	Systematic name	FC R-S Bendiocarb	P-value	FC R-S Permethrin	P-value	Description
CUST_3711_PI406199769	combined_c1873	-45.4	0.0054	- 20.0	0.0045	monkey king protein
CUST_13229_PI406199769	combined_c6767	-28.4	0.0098	- 28.1	0.0119	troponin c
CUST_4370_PI406199772	CD577181.1	-23.6	0.0168	- 13.7	0.0028	cytochrome c oxidase subunit ii
CUST_4384_PI406199772	CD577174.1	-22.4	0.0047	- 13.8	0.0027	cytochrome c oxidase subunit iii
CUST_4392_PI406199772	CD577170.1	-15.1	0.0035	- 14.1	0.0182	cytochrome c oxidase subunit ii

2.3.7 Validation of candidate genes with Quantitative Real-Time PCR

Quantitative real-time PCR (qRT-PCR) was used to validate the microarray results for twelve of the most up-regulated detoxification genes in relation to bendiocarb resistance or bendiocarb/permethrin cross resistance. They were 11 P450 genes and one aldehyde oxidase in addition to two housekeeping genes used as control (Fig. 2.12). All the genes had PCR amplification efficiency above 90% which is within the recommended range of 90-110% for Agilent (MX 3005). Results of the qRT-PCR validated the over-expression of the possible candidate genes such as *CYP6P9a* and *CYP6P9b* genes for the R-S, R-C and C-S for both bendiocarb and permethrin. For *CYP6P9a*, R-S and C-S comparison showed higher expression than the R-C for both bendiocarb and permethrin samples with permethrin having higher expression of *CYP6P9a* for R-S (FC 5.82) than R-S for bendiocarb (FC 4.72) and C-S (FC 3.70). The R-C for both bendiocarb and permethrin showed lower expression (FC R-C bendiocarb 1.27 and R-C permethrin 1.57) (Fig. 2.12). In any case, the difference in expression was not highly significant between bendiocarb and permethrin. The *CYP6P9b* genes showed similar expression for R-S for both bendiocarb and permethrin (FC 9.26 and 9.39) respectively. The C-S showed a lower expression (FC 6.92), while the R-C for both comparisons showed similar but much lower expression for both (FC 1.34 for bendiocarb and 1.36 for permethrin). These results indicate that there is a possible cross resistance between bendiocarb and permethrin.

The *CYP9J11* showed a different expression between bendiocarb and permethrin with higher expression of this gene found in the bendiocarb samples (FC: R-S 16.12, R-C 9.33), while there was a much lower expression in the permethrin samples FC: (R-S 3.16, R-C 1.83). The C-S also showed a lower expression (FC 1.73). The possible explanation for the high expression of this gene in the bendiocarb samples could be more involved in bendiocarb resistance than permethrin resistance.

The other genes like aldehyde oxidase (Aldoxi) showed higher expression in permethrin R-S (FC 8.3) and C- S (FC 5.25) and a lower expression in bendiocarb (FC 3.31). The R-C both showed very low expression with almost no differential expression for bendiocarb. The *CYP6AA4* gene was expressed more in C-S and R-S permethrin with very little differential expression in R-C and bendiocarb. The *CYP6M4* showed similar expression for both bendiocarb and permethrin for R-S and R-C. While the *CYP6N1* was over expressed in the C-S, meaning it is only expressed in samples that are not undergoing any induced stress like exposure to insecticide. The other genes like *CYP9K1*, *CYP6P2*,

CYP6Y2, CYP9J11 and *CYP304B1* showed low expression for permethrin, bendiocarb and C-S (Fig. 2.12).The error-bars were plotted using standard-errors. A statistical significance using *P value* was obtained after a Student t-test with the relative expression between samples.
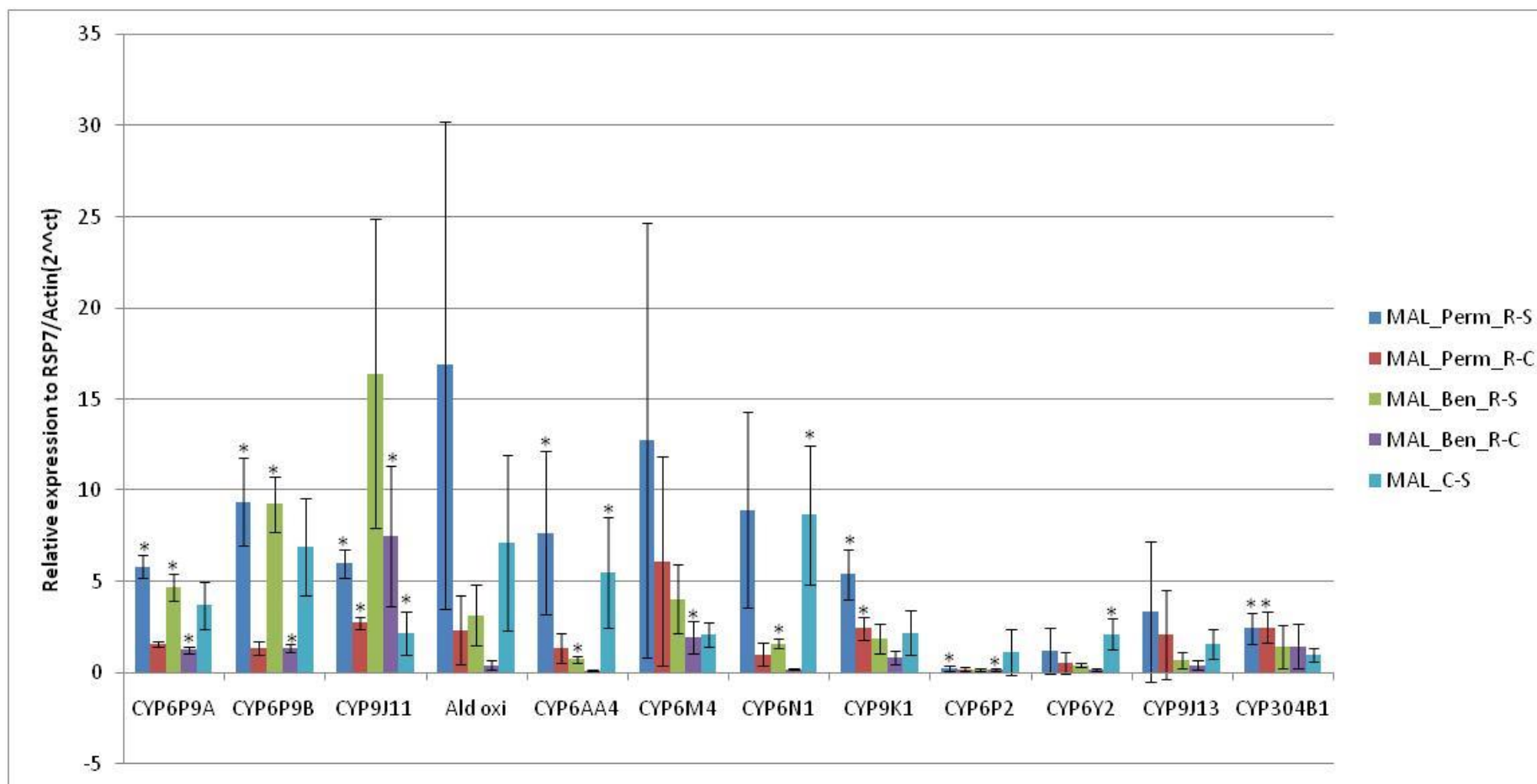
Fig. 2.12: Differential expression by qRT-PCR of 12 genes (*CYP6P9a, CYP6P9b, CYP9J11*, Aldoxi, *CYP6AA4, CYP6M4, CYP6N1, CYP9K1, CYP6P2, CYP6Y2, CYP9J13* and *CYP304b1*), (* represents significant at *P<0.05*).

A trend in support for a correlation between the qRT-PCR and microarray results was observed although not statistically significant probably because of lower expression observed for key genes for qPCR such as *CYP6P9a* and *CYP6P9b* as previously observed in *An. funestus* (Riveron *et al*. 2013) (R^2=0.157; *P*=0.143) as shown in Fig. 2.13).
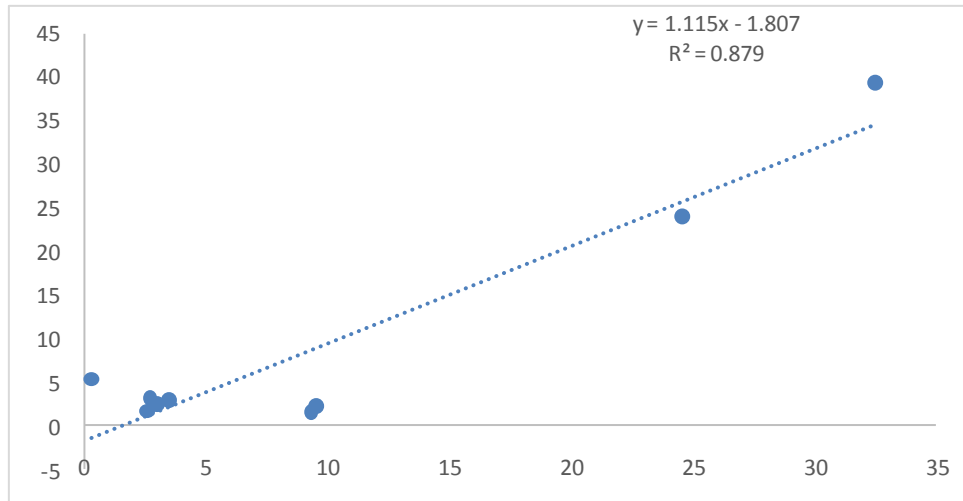
Fig. 2.13: Correlation between microarray fold change and qRT-PCR fold change

2.3.8 Polymorphism analysis of the candidate genes *CYP6P9a* and *CYP6P9b* in relation to cross-resistance

Due to the consistent over-expression of duplicated P450 genes *CYP6P9a* and *CYP6P9b* in the microarray comparisons for both bendiocarb and permethrin resistance, the polymorphism of both genes was analysed to establish their association with both bendiocarb and permethrin resistance. Five resistant (alive after exposure to insecticide) and 5 susceptible (dead after exposure to insecticide) samples for bendiocarb and permethrin were sequenced for *CYP6P9a* and *CYP6P9b* genes so as to establish the possible involvement of both genes in the cross resistance between carbamates and pyrethroids in this *An. funestus* population of Chikwawa in Malawi. By analysing the polymorphism of these two genes, the expectation was to detect: i) possible mutations or haplotypes associated with resistance phenotypes and ii) signatures of selection on each gene that could further confirm its involvement in resistance against one or both classes of insecticides.

2.3.8.1 Polymorphism pattern

The full length plus a 5'UTR fragment of the *CYP6P9a* (2055bp) and *CYP6P9b* (1812) was sequenced for the resistant (alive) and susceptible (dead) mosquitoes for both bendiocarb and permethrin samples.

2.3.8.1.1 Polymorphism pattern between susceptible and resistant bendiocarb samples

For all bendiocarb samples, a total of 33 and 17 substitutions were recorded respectively for *CYP6P9a* and *CYP6P9b*. The resistant samples were more polymorphic than the susceptible samples for *CYP6P9a* genes with 20 polymorphic sites for the resistant (alive) and 13 for the susceptible (dead) samples. However, the high level of polymorphism seen for the resistant samples was as a result of a single mosquito that was very polymorphic unlike for the susceptible samples for which polymorphism was contributed to by almost all of the mosquitoes.

A different pattern was observed for the *CYP6P9b* gene, where the resistant samples had a lower level of polymorphism (7) compared to the susceptible samples (12). The number of haplotypes which are similar between the two phenotypes (4 and 4) respectively for the resistant and susceptible *CYP6P9a* and for the resistant and susceptible samples for the *CYP6P9b* (2 and 4), also does not indicate a strong difference in diversity (Table 2.7).

Considering the nucleotide diversity (π x10^{-3}), the resistant samples were more genetically diverse for the *CYP6P9a* gene (2.44) than the susceptible samples (1.6). This is because of the lone mosquito that was very polymorphic. For the *CYP6P9b* gene, the resistant samples had a much reduced genetic diversity (0.8) than the susceptible (2.58) samples (Table 2.7). This possibly implies that genetic diversity is lower for bendiocarb resistant samples for the *CYP6P9b* genes than the susceptible.

Table 2.7: Summary statistics for polymorphism for resistant (alive) and susceptible (dead) samples for bendiocarb *CYP6P9a* and *CYP6P9b* genes.

	Full length (*CYP6P9*a-2055bp, *CYP6P9b* -1812bp)							Coding region(*CYP6P9a*1437bp, *CYP6P9b* 1437bp)						Non Coding region(*CYP6P9a*943bp, *CYP6P9b* 375bp)					
Samples	N	S	π	Taj D	Fu D	S*	h(hd)	S	π	Taj D	Fu D	S*	h (hd)	S	Π	Taj D	Fu D2	S*	h (hd)
CYP6P9a																			
Alive	10	20	2.44	-1.4205^{n}	-1.2177^{ns}	2	4(0.533)	14	3.9	-2.0091^{ns}	-2.3228^{ns}	0	3(0.378)	8	3.5	-1.6360^{ns}	-1.7327^{ns}	7	4(0.533)
Dead	10	13	1.61	-1.3377^{ns}	-1.4372^{ns}	3	4(0.733)	5	1.7	-1.5860^{ns}	-1.5944^{ns}	0	5(0.667)	3	1.2	-1.5622^{ns}	-1.7844^{ns}	3	2(2.0)
Total	20	33	2.11	-2.1401^{ns}	-2.4239^{ns}	5	8(0.679)	19	2.8	-2.4396^{ns}	-3.4285^{ns}	0	7(0.521)	11	2.4	-2.186^{ns}	-2.966^{ns}	10	5(0.368)
CYP6P9b																			
Alive	10	7	0.88	-1.8733^{ns}	-2.1818^{ns}	2	2(0.200)	7	0.97	-1.8391^{ns}	-2.1369^{ns}	6	2(0.200)	1	0.53	$-1..111^{ns}$	-1.2434^{ns}	2	2(0.200)
Dead	10	12	2.58	-0.9611^{ns}	-0.2611^{ns}	4	4(0.644)	12	2.86	-0.1392^{ns}	1.1713^{ns}	1	3(0.511)	1	1.48	-0.6909^{ns}	-0.2802^{ns}	1	2(0.356)
Total	20	17	17.9	-1.2159^{ns}	-0.5509^{ns}	6	6(0.442)	17	2.9	-1.5219^{ns}	-0.3933^{ns}	7	6(0.363)	3	1.04	-1.4407^{ns}	-1.2549^{ns}	2	4(0.366)

N, number of samples; S, number of segregating sites; $\pi(\times 10^{3})$, nucleotide diversity per site; Taj D Tajima's D; Fu D, Fu and Li's D*; S*, singletons for haplotypes; h, haplotype and hd-frequency of haplotype diversity and ns-not significant.

2.3.8.1.2 Polymorphism pattern between susceptible and resistant permethrin samples

There were a total of 14 polymorphic sites for the resistant and susceptible permethrin samples for *CYP6P9a* and 12 for *CYP6P9b* genes. The susceptible mosquitoes were more polymorphic than the resistant for both *CYP6P9a* and *CYP6P9b* genes. For the *CYP6P9a* gene, 14 polymorphic sites were observed for susceptible mosquitoes whereas only 2 were observed for resistant mosquitoes with similar pattern observed for the coding and non coding regions. A similar pattern was also observed for the *CYP6P9b* gene with susceptible mosquitoes far more polymorphic with 12 substitutions, whereas the resistant mosquitoes had no polymorphism at all (Table 2.8). Similarly, the number of haplotypes observed also supports this difference since the resistant samples have only 3 haplotypes contrary to 6 haplotypes for the susceptible samples for *CYP6P9a. CYP6P9b* exhibited a lower haplotype diversity with the resistant mosquitoes having just 1 haplotype while the susceptible had 2 haplotypes.

Overall, there is a stronger difference observed in the polymorphisms of the susceptible and permethrin resistant mosquitoes for both genes. In addition, the higher number of singleton polymorphic sites in the susceptible samples further confirms a higher polymorphism for the susceptible mosquitoes than the resistant.

Resistant mosquitoes also exhibited a much reduced genetic diversity for both *CYP6P9a* and *CYP6P9b* genes than the susceptible ones. The *CYP6P9a*, resistant mosquitoes showed a much reduced diversity (0.4), whereas the susceptible exhibited a much higher diversity (1.8) (Table 2.8). The pattern was similar for the *CYP6P9b* gene, where there was absolutely no genetic diversity for the resistant samples, while the susceptible samples had a genetic diversity of 1.3 (Fig. 12.14). Overall, this difference in patterns of genetic diversity supports an association between polymorphism at *CYP6P9a* and *CYP6P9b* and permethrin resistance phenotype. In addition, the low diversity observed in the resistant compared to the susceptible samples for permethrin resistance suggests the presence of possible selection acting on *CYP6P9a* and *CYP6P9b* genes for the resistant samples. The positive Tajima results for the resistant samples (Alive) further supports that both genes are under selection in relation to permethrin resistance phenotype. This appears not to be the case for bendiocarb resistance phenotype. This possibly suggests that these genes are less involved in bendiocarb resistance than in

permethrin resistance (Fig. 2.14). Also the bendiocarb samples exhibited more haplotype diversity for both genes than the permethrin samples (Fig. 2.14).

Table 2.8: Summary statistics for polymorphism for Alive and Dead for permethrin *CYP6P9a* and *CYP6P9b* genes.

	Full length (*CYP6P9a*-2055bp, *CYP6P9b* -1812bp)							Coding(*CYP6P9a*1113bp, *CYP6P9b* 1437bp)						Non Coding(*CYP6P9a* 943bp, *CYP6P9b*375bp)					
Samples	N	S	π	Jaj D	FU D	S*	h(hd)	S	π	Taj D	Fu D	S*	h (freq)	S	Π	Taj D	Fu D	S*	h (freq)
										CYP6P9a									
Alive	10	2	0.4	0.1203^{ns}	-0.280ns	1	3(0.644)	1	0.1	-1.1117^{ns}	-1.2434^{ns}	1	2(0.200)	1	1.0	1.3027^{ns}	0.8042^{ns}	0	2(0.533)
Dead	10	14	1.8	-1.2277^{ns}	-1.2366^{ns}	10	6(0.884)	8	1.4	-1.1614^{ns}	-0.8344^{ns}	5	5(0.756)	6	3.0	-1.1028^{ns}	-1.5100^{ns}	5	3(0.644)
Total	20	14	1.1	-1.6394^{ns}	-1.7514^{ns}	9	6(0.763)	8	1.2	-1.6397^{ns}	-0.9198^{ns}	4	5(0.511)	6	1.9	-1.2628^{ns}	-2.2584^{ns}	4	3(0.574)
										CYP6P9b									
Alive	10	0	0	0	0	0	1(0)	0	0	0	0	0	1(0)	0	0	0	0	0	0
Dead	10	12	1.3	-1.9611^{ns}	-2.2985^{ns}	12	2(0.200)	9	1.2	-1.901^{ns}	-2.2187^{ns}	9	2(0.20)	3	0.8	-1.562^{ns}	-1.784^{ns}	3	2(0.200)
Total	20	12	0.6	-2.3161^{ns}	-3.4632^{ns}	13	2(0.200)	9	0.55	-2.219^{ns}	-3.268^{ns}	9	2(0.10)	3	1.0	-1.784ns	-2.386^{ns}	3	2(0.10)

N- number of samples; S- number of segregating sites; $\pi(x10^3)$- nucleotide diversity per site; D1 Tajima's -D; D2- Fu and Li's D*; S*- singletons for haplotypes; h- haplotype and hd-frequency of haplotype diversity and ns-not significant.

2.3.8.1.3 Haplotype distribution for *CYP6P9a* and *CYP6P9b* bendiocarb and permethrin samples

 The most predominant haplotype for *CYP6P9a* with a frequency of 65% in the total bendiocarb sample was haplotype H1 (H1*) (Fig. 2.14a) which is also the predominant haplotype observed for the total permethrin sample (72%) (Fig. 2.14c). For the *CYP6P9b*, the most predominant haplotype was the H2 (H1*) with a 75% frequency for bendiocarb (Fig. 2.14b) and 95% for the permethrin samples. The fact that predominant haplotype in permethrin samples was nearly fixed with 95% frequency (Fig. 2.14d) for the *CYP6P9b* gene further supports that selection was acting on *CYP6P9b* in relation to permethrin resistance.

```
                                  11111111111111111
                 12334444555555566666666677888990012223333444456 6
                 67999056044602466900003334455902301038089155712399 09
Hap#   n         80237084357089469202456825034976671112350250577349 3
H1*    13        ATAAATTAGGATTAACAAGCAACGAACACAAACAATCATTCAAGGTTCGTG
H2R    1         ................................G..............T..
H3R    1         GC.....GCA.....TG.TTTGTT.CTTT..CTGCCTGCCTCGACCCTT..
H4R    1         ....GA............................................
H5S    1         ..............................GG..................
H6S    1         ......C...GCGGGA.G......G....GG...................G.
H7S    1         ...........................................G........
H8S    1         ..............................................C
```

(a) Bendiocarb *CYP6P9a*

```
            11111111111112                                      1
            55666666778890                      11233444555555556661
            99011349162740                      01144355003589991895
Hap# n      89512223165340       H#   n         63358836125381242569
H1*  8      GGTTGCATAAATTC       H1*  15        CCATACCCATAGATCTTCGT
H2*  2      ..............T       H2R  1         .T..G.TT...A....CTA.
H3*  6      .....T........       H3S  2         T..C.AT.GCG.TCTG.T..
H4S  2      .....T..CC....       H4S  1         ...................C
H5S  1      .....TT.......       H5S  1         ..G.................
H6S  1      AACGA.TG..GCC.
```

(b) Permethrin *CYP6P9a*   (c) Bendiocarb *CYP6P9b*

```
                    111111
               112555334444
               234239263478
Hap#   n       525254446992
H1*    19      TCTAATCCAGCG
H2S     1      ATCGGGTTGTTC
```

(d) Permethrin*CYP6P9b*

Fig. 2.14: Haplotype distribution for the full length (a) *CYP6P9a* and (b) *CYP6P9b* bendiocarb samples and (c) *CYP6P9a* and (d) *CYP6P9b* permethrin samples. (Hap# represents haplotype number, n- haplotype frequency, R- resistant samples and S – susceptible samples and H* indicates that haplotype is present in both phenotypes).

2.3.8.1.4 Haplotype network analysis

Analysis of the TCS haplotype network of *CYP6P9a* for bendiocarb sample did not reveal a significant difference between the two phenotypes as a similar haplotype diversity was observed with comparable number of mutational steps between haplotypes within and between phenotypes (Fig. 2.15a). However, between the susceptible samples, more mutational steps were observed (1-9). Haplotype H2R (R-Resistant 2) was an independent haplotype from the rest of the haplotypes as it has more than 20 mutational steps to other haplotypes (Fig. 2.15a). This pattern was not similar for the *CYP6P9b* gene for which both phenotypes exhibited a comparable haplotype diversity with similar number of mutational steps between haplotypes from the predominant haplotype (1-11 for susceptible and 1-7 for resistant mosquitoes) (Fig. 2.15b). In contrast, strong differences were observed between permethrin resistant and susceptible mosquitoes for both *CYP6P9a* and *CYP6P9b*. For *CYP6P9a*, resistant haplotypes did not have any mutational step from the predominant haplotype mutational steps while the susceptible haplotypes have 1-9 mutational steps between them (Fig. 2.15c). For *CYP6P9b*, the single susceptible haplotype has 11 mutational steps from the predominant haplotypes (Fig. 2.15d). These results support that polymorphism of *CYP6P9a* and *CYP6P9b* is not associated with bendiocarb resistance while a strong correlation is observed with permethrin resistance.

(a) Bendiocarb *CYP6P9a*

(b) Bendiocarb *CYP6P9b*
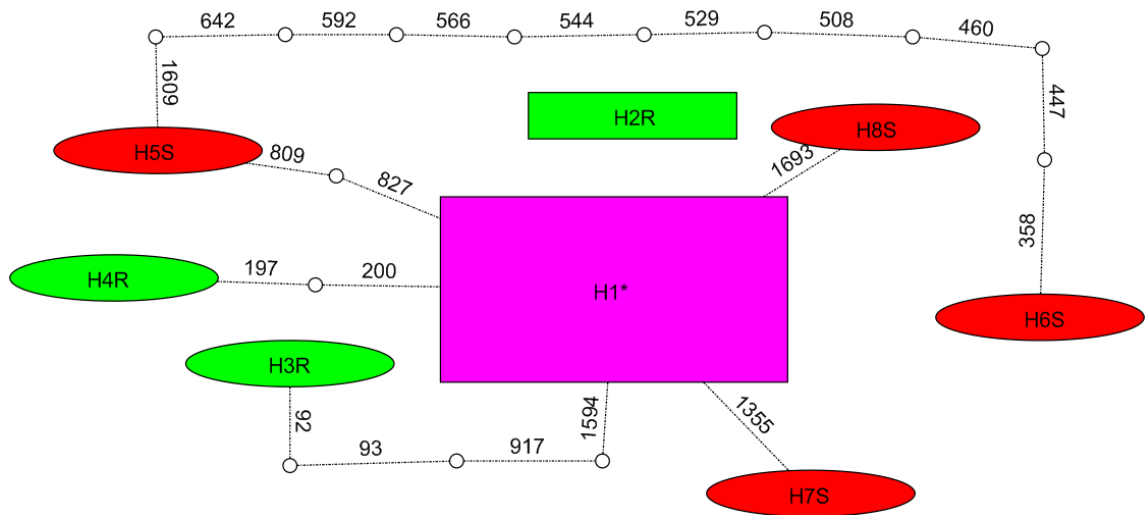
(c) Permethrin *CYP6P9a*
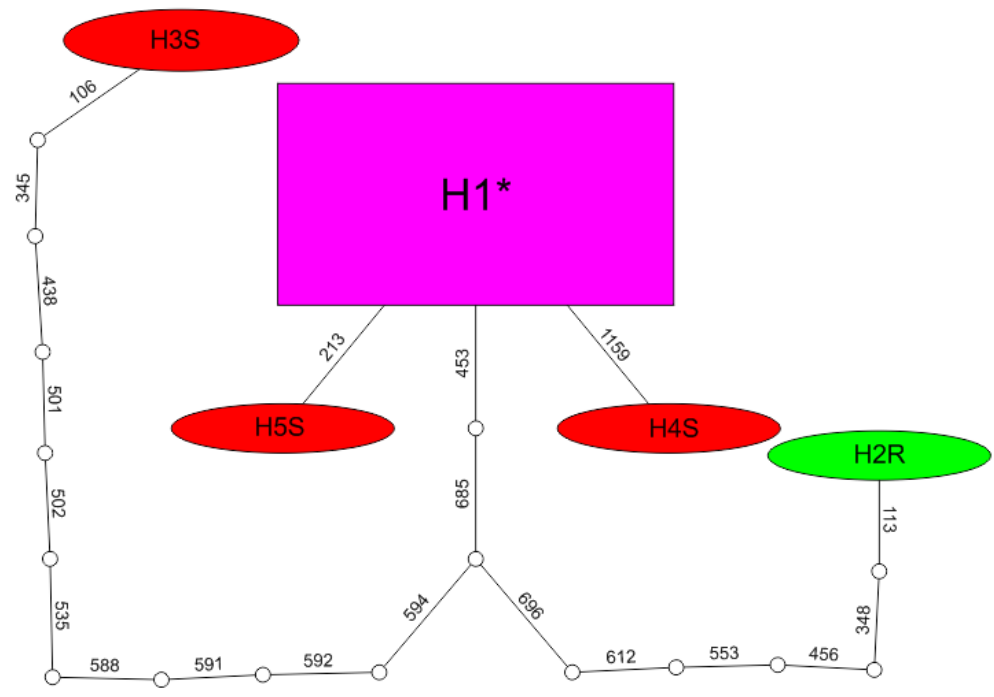
(d) Permethrin *CYP6P9b*
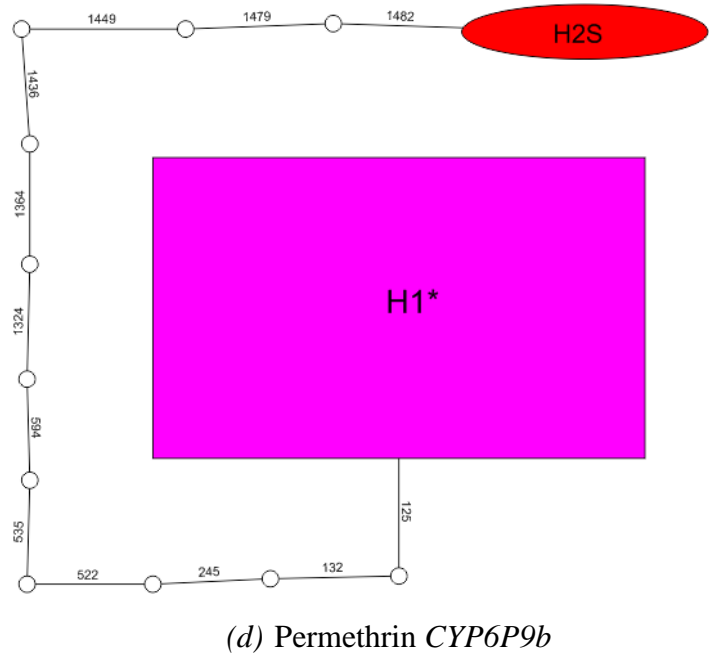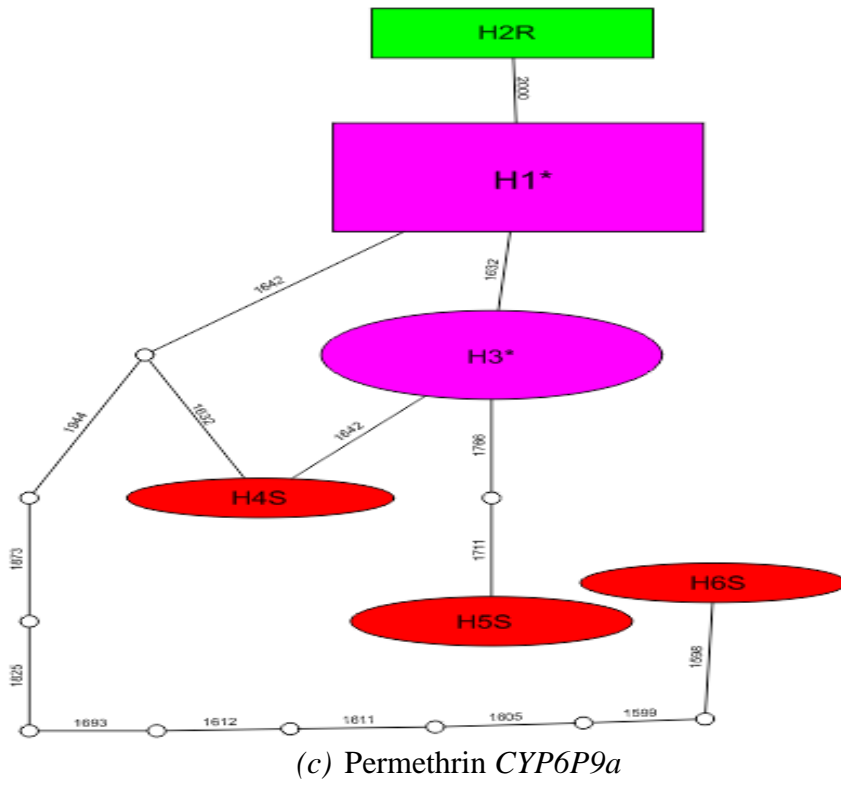
Fig. 2.15: A95% parsimony network of resistant and susceptible samples for *CYP6P9a* (a) and *CYP6P9b* (b) genes for bendiocarb and (c) *CYP6P9a* and (d) *CYP6P9b* permethrin samples. Green colour polygons represent resistant mosquitoes while red represents susceptible mosquitoes and also the pink coloured polygon represents haplotype dominant in both resistant (AL) and susceptible (DE) mosquitoes. The size of the polygon reflects the frequency of the haplotype. Segregating mutation is represented by each node (number) and polymorphic positions are given above the branches.

2.3.8.2 Phylogenetic analysis of *CYP6P9a* and *CYP6P9b* genes

2.3.8.2.1 Maximum likelihood phylogenetic tree for bendiocarb samples

The construction of the maximum likelihood phylogenetic tree for *CYP6P9a* revealed that there was no clustering of haplotypes according to their resistance phenotype. The resistant and susceptible samples both shared common haplotypes with no formation of an independent cluster per phenotype (Fig. 2.16a). This pattern was similar for the *CYP6P9b* gene (Fig. 2.16b).

(a) Bendiocarb *CYP6P9a*
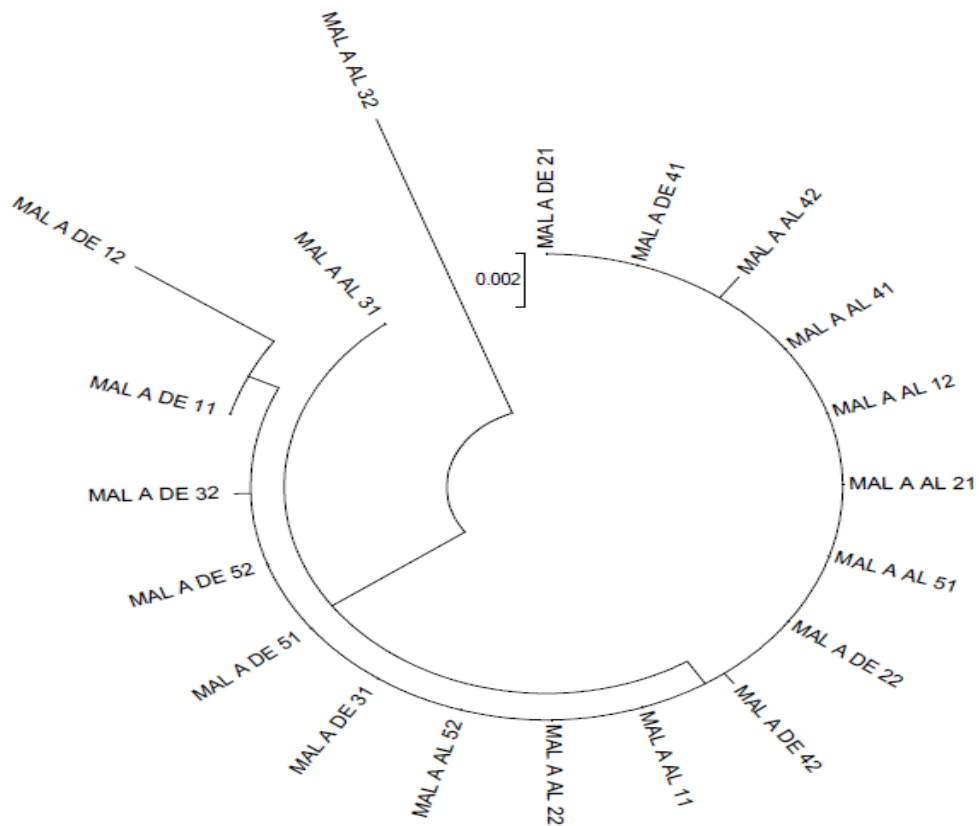
(b) Bendiocarb *CYP6P9b*

Fig. 2.16: Schematic representation of maximum likelihood (ML) tree for bendiocarb samples (a) – *CYP6P9a* and (b) – *CYP6P9b*) "DE" denotes susceptible samples (dead) and "AL" denotes resistant samples (alive).

2.3.8.2.2 Maximum likelihood phylogenetic tree for permethrin samples

In contrast to bendiocarb samples, the maximum likelihood tree (ML) constructed for *CYP6P9a* revealed that haplotypes clustered according to their resistance phenotype (Fig. 2.17a). Similarly for *CYP6P9b*, the ML tree shows that the resistant samples clustered together, whereas the dead samples formed another cluster (Fig. 2.17b). Overall the ML tree further supported a lack of correlation between *CYP6P9a* and *CYP6P9b* polymorphism and bendiocarb resistance phenotype while the opposite is true for permethrin.
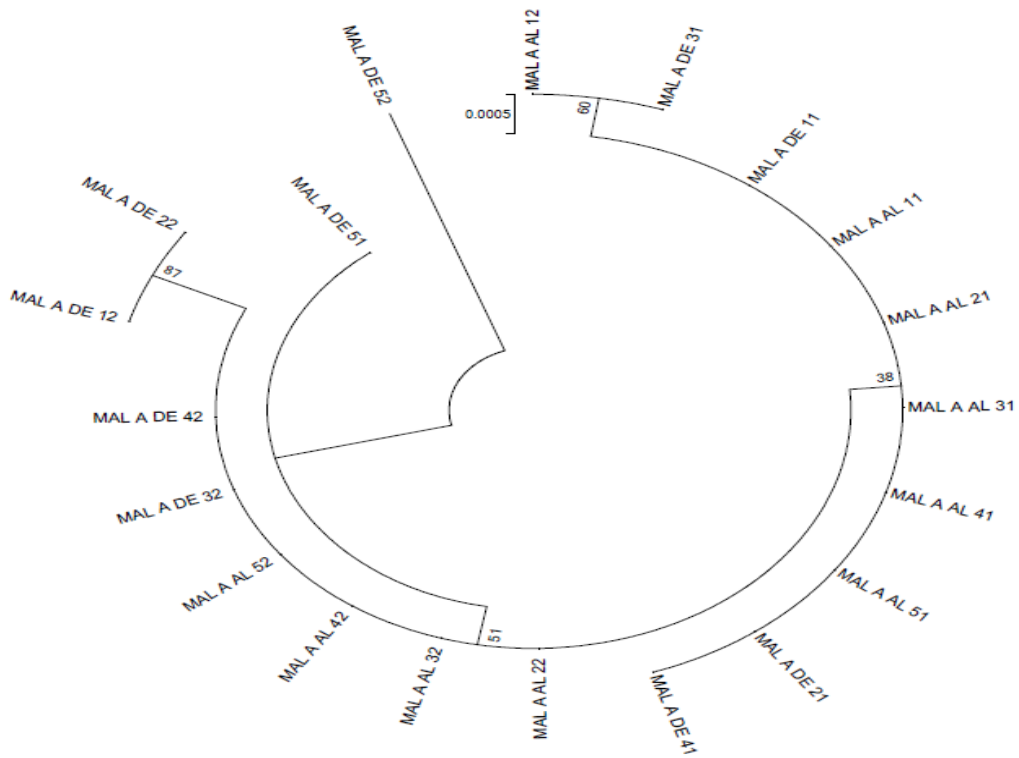
(a) Permethrin *CYP6P9a*
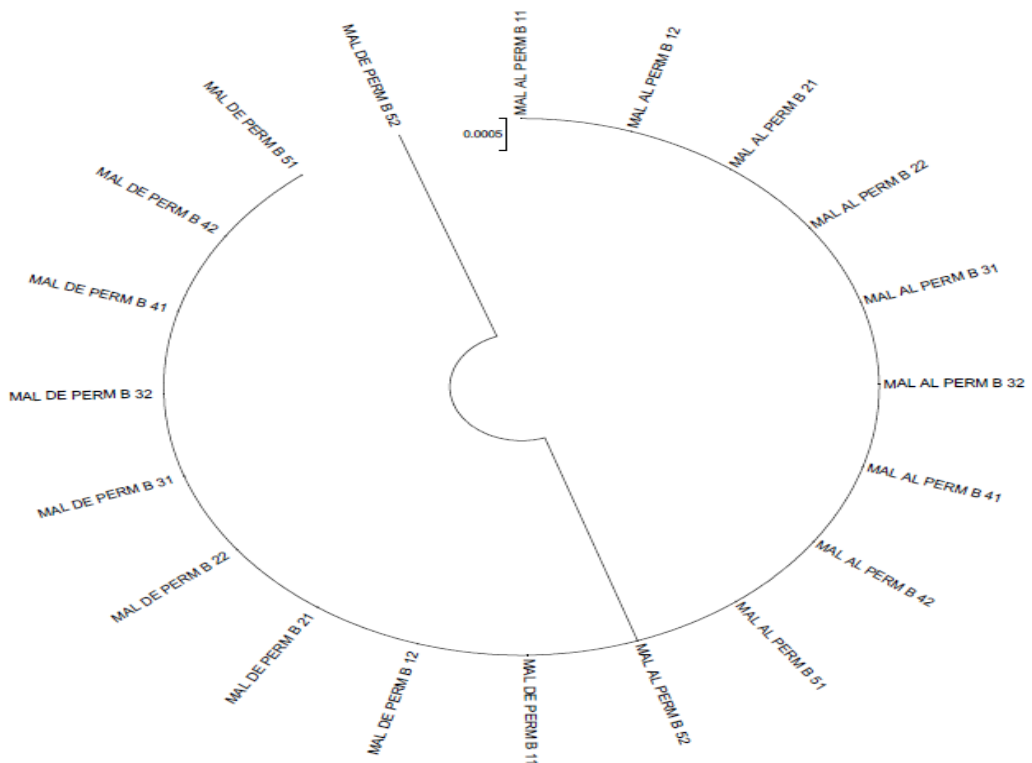
(b) Permethrin *CYP6P9b*

Fig. 2.17: Schematic representation of maximum likelihood (ML) tree for permethrin samples (a) – *CYP6P9a* and (b) – *CYP6P9b*) "DE" denotes susceptible samples (dead) and "AL" denotes resistant samples (alive).

2.3.9 Test for selection on *CYP6P9a* and *CYP6P9b* genes

The test for selection on *CYP6P9a* and *CYP6P9b* resistant samples for bendiocarb using codon-based Z-test as implemented in Mega 5.2, revealed that both *CYP6P9a* and *CYP6P9b* are under purifying selection for both permethrin and bendiocarb samples. This is because the null hypothesis that dN=dS was consistently and significantly rejected in favor of dN<dS (Table 2.9).

Other selection tests such as MK or HKA also showed an indication of purifying selection for the alive samples (Table 2.10). These results confirms the fact that the resistant samples of the *CYP6P9a* and *CYP6P9b* genes is under a stronger selection in the permethrin samples than the bendiocarb samples.

Table 2.9: Test of selection for bendiocarb samples

MK							HKA		dN/dS and Z test				Ka/Ks		
	Fixed Poly Morphic between		Poly Morphic With species		NI	Pvalue	X^2	Pvalue	dN	dS	dN/dS	Pvalue	Ka x10^3	Ks x10^3	Ka/Ks
	syn	Nsyn	Syn	Nsyn											
								CYP6P9a							
Alive	162	67	23	6	0.631	0.3889ns	0.09	0.764ns	0.843	0.185	4.56	0	1.39	19.12	0.31
Dead	171	67	3	7	5.955	0.00892**	3.268	0.0707ns	0.142	0.153	0.93	0.18	1.62	3.66	0.007
Total	161	67	25	13	1.250	0.5692ns	1.055	0.3044ns	0.345	0.077	4.48	0	1.50	11.5	0.18
								CYP6P9b							
Alive	265	281	0	0	0.157	0.064ns	9.4	0.0021**	0.056	0.023	2.44	0.008	0.17	3.38	0.004
Dead	262	280	10	2	0.187	0.0192*	9.9	0.0016**	0.24	0.087	2.76	0.003	0.77	8.98	0.14
Total	262	280	14	3	0.201	0.0061**	15.5	0.0001***	0.135	0.042	3.21	0.001	0.49	6.36	0.09

NI: neutrality index; MK; Mcdonald and Kreitman test; HKA: Hudson, Kreitman and Aguade test; Und: Undetermined, nc: Not calculated, Star (*) denotes significant at P<*0.05*, (**) denotes significant at P<*0.01* and (***) denotes P<*0.0001*, ns-not significant.

Table 2.10: Test of selection for permethrin samples

	MK						HKA		dN/dS and Z test				Ka/Ks		
	Fixed Poly Morphic between		Poly Morphic With species		NI	Pvalue	X^2	Pvalue	dN	dS	dN/dS	Pvalue	Ka x10^3	Ks x10^3	Ka/Ks
	syn	Nsyn	Syn	Nsyn											
								CYP6P9a							
Alive	176	68	0	1	nc	nc	2.0	0.155^{ns}	0.037	0.036	1.03	1.0	0.17	0.0	0.31
Dead	172	67	5	3	1.54	0.69^{ns}	10.2	0.0014**	0.291	0.156	1.87	0.32	0.6	3.75	0.007
Total	172	67	5	3	1.54	0.69^{ns}	11.7	0.0006***	0.666	0.385	1.73	0.04	1.9	0.4	0.18
								CYP6P9b							
Alive	265	281	0	0	nc	nc	nc	Nc	0	0.076	0	1.0	0.0	0.0	0.004
Dead	261	279	6	3	0.47	0.328^{ns}	0.396	0.529^{ns}	0.848	0.412	2.06	0.02	0.51	3.37	0.14
Total	261	279	6	3	047	0.33^{ns}	0.456	0.499^{ns}	0.205	0.098	2.09	0.019	0.25	16.9	0.09

NI: neutrality index; MK; Mcdonald and Kreitman test; HKA: Hudson, Kreitman and Aguade test; Und: Undetermined, nc: Not calculated, Star (*) denotes significant at P<*0.05*, (**) denotes significant at P<*0.01* and (***) denotes P<*0.0001*, ns-not significant.

2.4 Discussion

This study established the resistance profile of the *An. funestus* population in Chikwawa district in Malawi and investigated the molecular basis of carbamate resistance and the possible presence of a carbamate/pyrethroid cross-resistance as previously suggested (Brooke *et al.* 2001).

The *An. funestus* population from Chikwawa is resistant to the carbamate bendiocarb as previously reported in other southern Africa populations of this species notably in Mozambique (Brooke *et al.* 2001, Casimiro *et al.* 2005, Cuamba *et al.* 2010) or in Malawi (Hunt *et al.* 2010). The levels of bendiocarb resistance reported here is similar to that observed in Mozambique (Cuamba *et al.* 2010). In addition, as previously described in Mozambique (Brooke *et al.* 2001), this carbamate resistance seems to be controlled by cytochrome P450 genes as a full recovery of susceptibility to bendiocarb was observed when mosquitoes were exposed to PBO synergist. However, a full susceptibility was observed for the organophosphate malathion indicating that there is no cross-resistance carbamates/organophosphates as also seen in Mozambique (Cuamba *et al.* 2010).

Resistance to both types I and II of pyrethroids was observed in *An. funestus* population of Chikwawa. This study confirms that the pyrethroid resistance front in southern Africa has extended north beyond South Africa and Mozambique where it was originally reported (Hargreaves *et al.* 2000). However, it cannot yet be concluded whether the resistance in Malawi is a direct extension of that seen in southern Mozambique or whether it has arisen independently. Studies of genetic structure of *An. funestus* populations have indicated that there is extensive gene flow between populations of Malawi and Mozambique (Michel *et al.* 2005) supporting the hypothesis of the extension of the resistance front from South Mozambique. However, there are some differences in the resistance profile between the two countries. Firstly, levels of pyrethroid resistance observed in Chokwe (South Mozambique) were significantly higher than in Chikwawa with no mortality observed in Chokwe after 1h30min exposure to both type I and II pyrethroids (Riveron *et al.* 2013) while a mortality around 40% is observed after 1h in Chikwawa. Secondly, a reduced susceptibility to DDT was observed for females in the Malawi population contrary to Mozambique where a full susceptibility was observed for both females and males (Riveron *et al.* 2013). Analysis of patterns of genetic variability of genes conferring this resistance may help to further assess whether this is the same

resistance front or not. The reduced susceptibility to DDT observed in this study in Chikwawa is similar to the case recently reported in Zambia, (Chanda *et al*. 2011), indicating that southern African populations of *An. funestus* may be developing resistance against DDT as seen in East Africa (Uganda) (Morgan *et al*. 2010) and in West Africa (Ghana) (Okoye *et al*. 2008) and Benin (Djouaka *et al.* 2011, Riveron *et al.* 2014).

Control of *An. funestus* in Chikwawa district should take into account the resistance profile reported in this study because organophosphates remain the only insecticide class with full susceptibility in this population, it should be considered for future control interventions such as IRS.

2.4.1   Bendiocarb resistance is driven by cytochrome P450 over-expression

The genome-wide microarray analysis using the custom-made 4x44k and 8x60k Agilent arrays provided a profiling of the transcription associated with bendiocarb resistance. The experimental design chosen in this study allowed the possibility to assess the gene expression changes caused by bendiocarb resistance. Results from the hybridization of the resistant mosquitoes against the control non exposed (R-C) from Malawi allowed a direct assessment of the molecular basis of bendiocarb resistance while minimizing confounding factors associated with difference of genetic background between strains. This positive outcome for the R-C comparison was made possible because the bendiocarb resistance level with 60% mortality rate was appropriate for such R-C comparison. No significant results had in contrast been obtained for such R-C comparison in other studies where resistance levels have been high such as in Mozambique for an *An. funestus* population from Chokwe which was fully resistant to pyrethroid after 2h exposure (Riveron *et al.* 2013). In addition, consideration of consistent over-expression between R-C and R-S or also C-S helped to further select the best candidate bendiocarb resistance genes in this study.

This study revealed that the over-expression of genes from several gene families is associated with bendiocarb resistance. Most of these genes families have been commonly associated with insecticide resistance in previous studies and in various species such as in *An. funestus* (Riveron *et al.* 2013 and 2014), *An. gambiae* (Vontas *et al.* 2005), *Ae. aegypti* (Bariami *et al*. 2012) or *D. melanogaster* (Pedra *et al.* 2004). This include the detoxification gene families such as Cytochrome P450, GSTs, aldehyde oxidase or other resistance associated genes such as ABC transporters, short-chain dehydrogenases and

cuticular protein genes. The involvement of these several families suggests that insecticide resistance induces transcriptional changes involving different molecular pathways than just the few detoxification genes as previously reported (Pedra *et al.* 2004, Vontas *et al.* 2005).

In accordance to the PBO synergist assays suggesting that cytochrome P450 genes were playing a major role in the observed bendiocarb resistance, several probes belonging to this gene families were over-expressed in the bendiocarb resistant mosquitoes. The consistent over-expression of the cytochrome P450 genes *CYP6P9a* and *CYP6P9b* genes in the various comparisons for bendiocarb resistance suggested a possible involvement of these two P450 genes in bendiocarb resistance as also validated by the qRT-PCR experiment. These two duplicated genes are located within the rp1 (resistance to pyrethroid 1) QTL on the 2R chromosome (Wondji *et al.* 2009) and have so far mainly shown to confer resistance to pyrethroids (Amenya *et al.* 2008, Wondji *et al.* 2009, Riveron *et al.* 2013). In general, members of the cytochrome P450 family have been mainly and repeatedly associated to insecticide resistance to pyrethroids in mosquito species (Müller *et al.* 2008b, Wondji *et al.* 2009, Stevenson *et al.* 2012) or in other insects such as the *CYP6BQ9* P450 gene of the red flour beetle *Tribolium castaneum* (Zhu *et al.* 2010). HoweverP450s have also been shown to metabolise other insecticide classes such as DDT in *An. gambiae* by *CYP6Z1* (Chui *et al.* 2008) and *CYP6M2* (Mitchell *et al.* 2012) or *CYP6G1* in *D. melanogaster* (Daborn *et al.* 2002). Interestingly, over-transcription of a P450 gene *CYP6M2* was recently shown to confer carbamate resistance in *An. gambiae* (Edi *et al.* 2014) further supporting that over-expression of cytochrome P450s such as *CYP6P9a* and *CYP6P9b* could well be responsible for the bendiocarb resistance in the Chikwawa *An. funestus* population.

The *CYP6Z1* gene was also over-expressed in the various comparisons and for the R-C comparison it even exhibited a higher FC than *CYP6P9a* and *CYP6P9b,* although not so for the R-S and C-S comparisons suggesting that *CYP6Z1* is equally a good candidate gene for bendiocarb resistance. However, further functional analyses are needed to confirm such a role. The over-expression of other P450 genes in the R-C such as *CYP9J11*, *CYP6P4a* and *CYP6P5* suggests that bendiocarb resistance in this population is the result of the contribution of several genes than just a single major one as it is the case for the GSTe2 gene conferring DDT resistance in *An. funestus* (Riveron *et al.* 2014).

A significant observation in the microarray profiling of the bendiocarb resistance is the consistent over-expression of several probes belonging to cuticle protein genes, notably in the R-C comparison. Such consistent over-expression could suggest that besides the up-regulation of cytochrome P450 genes, a reduced penetration of insecticides through a thickening of the mosquito cuticle could be another resistance mechanism associated with the bendiocarb resistance in the *An. funestus* population of Chikwawa. The presence of this reduced penetration mechanism could be supported by a previous report that cuticle thickening was playing a role in the pyrethroid resistance observed in the laboratory resistant strain FUMOZ originally from Mozambique in southern Africa (Wood *et al.* 2011). Such mechanisms have also been reported in a resistant strain of aphid *Myzus persicae* (Puinean *et al.* 2010) with *in vivo* radio labelled insecticide penetration assay that provided evidence of the resistant strain showing reduced penetration. Similar experiment will also help to validate the role of the reduced penetration in the bendiocarb resistance in *An. funestus*.

2.4.2 Cytochrome P450 genes possibly confer a bendiocarb/permethrin cross resistance

The list of genes commonly over-expressed in bendiocarb and permethrin resistant mosquitoes revealed that Cytochrome P450 genes are the main gene families possibly conferring a cross-resistance between these two insecticide classes.

The consistent over-expression of the duplicated P450 genes *CYP6P9a* and *CYP6P9b* in all comparisons for both permethrin and bendiocarb resistance suggested that these genes are the main genes possibly involved in the carbamate/pyrethroid cross resistance in the Chikwawa populations. These genes were both highly over expressed in the R-S of bendiocarb and permethrin with moderate over-expression in the R-C. *CYP6P9a* and *CYP6P9b* have already been conclusively shown to confer pyrethroid resistance in the same population, especially using transgenic expression in *D. melanogaster* (Riveron *et al.* 2013). Similar experiments will be needed in the future to confirm that these genes could also do the same for carbamates. *CYP6Z1* could also play a role in the carbamate/pyrethroid cross resistance as this gene was up-regulated in both R-C/R-S for both insecticides. *CYP6Z1* involvement in such cross-resistance will not be too surprising as its ortholog in *An. gambiae* has previously been shown to confer cross-resistance to pyrethroid/DDT/carbaryl (Chiu *et al.* 2008). However, functional characterisation such

as transgenic expression in Drosophila flies and heterologous expression of recombinant enzymes in *Escherichia coli* (Riveron *et al.* 2013) is needed to confirm this role in *An. funestus*. Equally the possible involvement of other P450s such *as CYP9J11, CYP6P4a* and *CYP6P5* will need to be validated.

No esterase gene was significantly highly over-expressed in this population and cytochrome P450 genes were by far the most over-expressed in bendiocarb or permethrin resistant mosquitoes suggesting that the main resistance mechanisms were captured in this study. However, future work taking advantage of the recently published full genome of *An. funestus* should be performed to ensure that the full mechanistic pathway of the carbamate or pyrethroid/carbamate resistance is elucidated.

2.4.3   Genetic variability patterns

Contrary to permethrin resistance, genetic variability patterns of *CYP6P9a* and *CYP6P9b* genes do not support their involvement in bendiocarb resistance since the analysis of patterns of genetic polymorphisms of *CYP6P9a* and *CYP6P9b* revealed a significant difference between bendiocarb resistant and permethrin resistant mosquitoes. While a reduced polymorphism is clearly observed for both genes in permethrin resistant mosquitoes in comparison to susceptible ones, this is not the case for bendiocarb resistant mosquitoes. This absence of correlation between the polymorphisms of both genes and bendiocarb resistance phenotype could suggest that although *CYP6P9a* and *CYP6P9b* even if involved in bendiocarb resistance, they may not be playing a major role as for permethrin resistance or the bendiocarb resistance is not associated with the directional selection observed for permethrin resistance. However, the sample size used in this study could be further increased to confirm this difference in polymorphism for bendiocarb and permethrin samples. Nevertheless, it is well possible that such difference of polymorphism be observed between permethrin and bendiocarb samples for the candidate genes *CYP6P9a* and *CYP6P9b* if these genes have different role in both resistances. This is because the permethrin and bendiocarb samples are two different phenotypes, and may well present different polymorphism profiles. In this regard, the difference observed between alive and dead permethrin here suggests that these two genes are associated with pyrethroid resistance whereas, the absence of difference between alive and dead bendiocarb suggests that both genes are not associated with this resistance. Additionally, the same sample size had been shown to be sufficient to reveal association between

polymorphism and association with insecticide resistance. For example Riveron *et al.* (2014) used 5 DDT resistant and 5 susceptible to demonstrate that variation of polymorphism at GSTe2 is associated with DDT resistance (Riveron *et al.* 2014). Similarly, same sample size allowed detecting the association of *CYP6P9a* and *CYP6P9b* with permethrin resistance in Malawi, Zambia and Mozambique consistently (Riveron *et al.* 2014). Further study of the polymorphism of both genes and the genomic regions adjacent to these genes with a higher sample size could provide more insights into the association between these duplicated P450 genes and carbamate resistance. Future work should also investigate the polymorphism of other genes such as *CYP6Z1* or *CYP9J11* in relation to bendiocarb resistance. The reduced polymorphism observed in relation to permethrin resistance with highly predominant haplotypes for both genes is similar to that seen in *D. melanogaster* for the P450 gene *CYP6G1*, where the *CYP6G1*/ Accord haplotypes in DDT-resistant for field samples is highly selected (McCart and Ffrench-Constant 2008). The implementation of IRS and LLINs and use of other agricultural pesticides might be responsible for the predominance of the H1 Haplotype for *CYP6P9a* and H2 for *CYP6P9b* genes for both bendiocarb and permethrin. A similar dominant haplotype (H2-*CYP6P9a* and H1-*CYP6P9b*) were observed in Mozambique as in Malawi suggesting that these haplotypes might stem from a single origin for pyrethroid resistance in southern Africa (Riveron *et al.* 2013).

Conclusion

The possible cross-resistance revealed in his study could inform vector control programmes in southern Africa in future resistance management strategies as rotation of pyrethroids and carbamates should be avoided because this might actually further increase the selection of resistance. The best option recommendable is the use of organophosphates for control interventions such as IRS but the challenge is the high cost of using this insecticide class as it does not last as long as pyrethroid or DDT (Ware *et al.* 2010).

Chapter Three

Assessing the genetic structure of *An. funestus* population in Africa

3.1 Introduction

Genetic differentiation between *An. funestuss.s* populations has been associated to polymorphic chromosomal inversions playing a role in their adaptation to heterogeneous environments (Coluzzi 1982, Coluzzi *et al.* 1985, Rieseberg 2001). Different parts of Africa revealed significant differences in frequency between locations and also the uncovering of inverted arrangements that were endemic to particular regions through the characterization of chromosomal inversions (Green and Hunt 1980, Dia *et al.* 2000). No compelling evidence of population substructure has been found in study carried out using cytogenetic studies on *An. funestus* samples from Kenya, Nigeria, Senegal and southern Africa (Green and Hunt 1980, Dia *et al.* 2000b, Kamau *et al.* 2002).Molecular evidence based on microsatellite study also supported the view that at a local scale, *An. funestus* populations are panmictic (Braginets *et al.* 2003, Cohuet *et al.* 2004, Temu *et al.* 2004). Contradicting this general picture, studies carried out with samples of *An. funestus* from Burkina Faso revealed evidence of chromosomally incipient species parallel to the situation seen in *An. gambiae* (Costantini *et al.* 1999, Guelbeogo *et al.* 2005). Based on chromosomal inversion data, two chromosomal forms named Kiribina and Folonzo in Burkina Faso were described (Michel *et al.* 2005). The Folonzo has characteristic high frequencies of inversions on the 3Ra, 3Rb and 2Ra while the Kiribina has high frequencies in corresponding un-inverted arrangements (Michel *et al.* 2005).

The existence of at least one main division between *An. funestus* population from West-Central and East Africa has also been suggested by using sequence differences of ITS2 and D3 regions of ribosomal DNA molecular markers (Giros *et al.* 2004). Michel *et al.* (2005), using microsatellite markers found three subdivisions namely: Eastern (Coastal Tanzania, Malawi, Mozambique and Madagascar), Western (Nigeria, Mali, Burkina Faso and Western Kenya) and Central (Gabon, Coastal Angola) in *An. funestus.* This revealed 3 clusters that were specific to geographical origins suggesting the presence of barriers to gene flow. They also showed the presence of two clades using mtDNA in southern Africa where clade I represented other samples across Africa and clade II was assigned only to individuals sampled from Mozambique and Madagascar. There have been few studies on population genetics of *An. funestus*, and most of the studies have been limited to

geographic regions limiting the extent to which their conclusions can be applied Africa-wide (Michel *et al.* 2005). Moreover, there are increasing reports of resistance to insecticide classes in *An. funestus* populations as reported in chapter one. Wondji *et al.* (2011) detected *Rdl^R* allele associated to dieldrin resistance in four out of the six countries with highest frequencies observed in West (Burkina Faso) and Central Africa (Cameroon), respectively with moderate frequency observed in Benin. In East Africa (Uganda), the *Rdl^R* was also detected although at a very low frequency and only as heterozygote. This allele was not observed in the two southern African countries of Mozambique and Malawi indicating that this allele has probably not yet spread to *An. funestus* populations from these regions. Understanding variation of the resistance profiles between *An. funestus* from different regions in Africa will help in finding out if the resistance profile is associated to the genetic structure of this vector which will provide an insight to possible barriers to gene flow between *An. funestus* populations in order to predict the spread of resistance genes in this species.

3.1.1 Objective

The objective of this chapter is to: Assess the genetic structure of *An. funestus* population in Africa using microsatellite loci genotyping in order to estimate the risk of spread of insecticide resistance genes across Africa.

By establishing patterns of the genetic structure of *An. funestus* and the level of gene flow across Africa. This study will help to predict the direction and speed of insecticide resistance in this major malaria vector across Africa. This will be achieved by using 12 microsatellite markers genome-widely distributed (Chromosome X, 2R, 2L 3R and 3L) (Fig. 3.3)

3.2 Materials and methods

3.2.1   General mosquito collection

Randomly blood fed adult female *An. funestus* mosquitoes resting indoors were collected as already described in chapter two.

3.2.1.1 Sample site description

Tororo (Uganda)

Indoor resting *An. funestus* females were collected in Tororo in eastern Uganda (0° 45'N, 34° 5'E), part of the District of Tororo near the border with Kenya. It is also about 230km (140miles) from the capital city by road east of Kampala. Mosquitoes were collected for two weeks in April and November 2009 by J. Morgan (LSTM). The mosquitoes were collected and immediately transported to the laboratory of the National Livestock Resources Research Institute based in Tororo which is an area of high malaria transmission with 1586 infectious bites per person per year in 2001 (Okello *et al.* 2006). The climate is characterised by two main seasons, the rainy and the dry seasons. The rainy season is between March-May and a light rainy season between October-November. The dry season is from December-February and June-August. Annual rain fall ranges between 500mm to 2500mm with relative humidity between 70%- 100% with an average temperature of around 26℃. The inhabitants of Tororo primarily rely on mixed-crop (e.g millet, cassava and maize) agriculture as a main source of living. In Tororo, there is 62% mortality to permethrin after 1hr exposure (Morgan *et al.* 2010)

Lagdo (Cameroon)

Indoor resting female *An. funestus* mosquitoes were collected in November 2006 from Lagdo by C. Wondji (LSTM). Lagdo is situated in the Northern part of Cameroon (9℃ 05'N, 13^0 40' E). It is about 50km from the main Northern city of Garoua and about 911km from the national capital Yaoundé. Lagdo experiences two major climate types, the rainy and the dry seasons that alternate all through the year. November-March experiences very dry seasons accompanied by hash harmattan winds. The heat during this period is very intense until April when there is torrential rainfall and lower temperatures. Rain closes up in June even though during this period temperatures still remain low with relatively high humidity. Lagdo receives between 500-1000mm per year of annual rainfall. Temperatures can be as high as 38℃. Lagdo has a dam that helps supply electricity to the Northern part of the country. The people of Lagdo practice mostly postural farming and depend on the animals for living. Only reduced susceptibility to permethrin is observed (90%) mortality in Lagdo (Wondji *et al.* unpublished data).

Pahou (Benin)

Indoor resting female *An. funestus* were collected in March 2010 by R. Djouaka (Benin) from Pahou (6° 23' 0" N, 2° 13' 0" E) located in the region of Ouidah, Atlantique in South-Benin. Pahou has two climatic seasons, the rainy and the dry that alternates throughout the year. There is a long dry season from December-March and a long rainy season from April-July, a short dry season in August and September and a short rainy season from October-November. Average temperatures are approximately 25°C with average relative humidity throughout the year ranging between 70-90%. Average rain is about 1200mm. The main occupation of the people is subsistent farming with fishing being part of it.
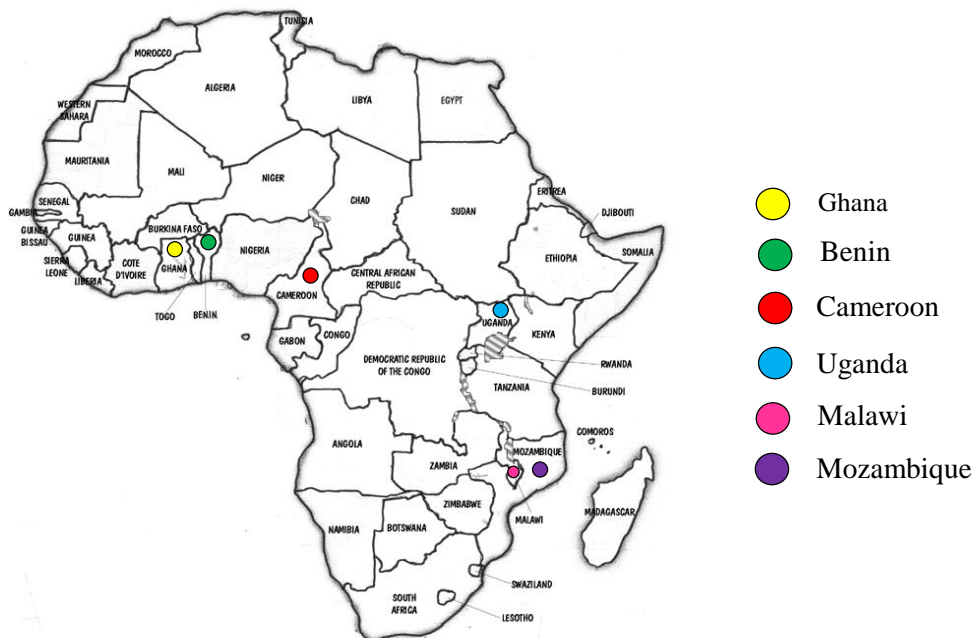
Fig. 3.1: Map of Africa showing location of the six sample collection countries (Map taken from www.mongabay.com).

Obuasi (Ghana)

Samples from Obuasi Ghana were collected in January 2010 by C. Wondji. Obuasi is a town in Ghana situated (6° 12' 0" N, 1° 40' 0" W). It is the Municipal capital of Obuasi, lying south of Kumasi the Ashanti capital city. It is approximately 50miles North of Kumasi. Obuasi features a tropical wet and dry season. The wet season is long and runs from March-July and a shorter rainy season from September-November with relatively low precipitation in August. The dry season is accompanied by the harmattan and it is

between December-November. Gold mining is the main source of income for the people of Obuasi. Okoye *et al.* (2008) observed permethrin resistance in Obuasi.

Chokwe (Mozambique)

Samples were collected in February 2009 by C. Wondji from Chokwe which is the capital of the Chokwe district in the province of Gaza in southern Mozambique (24° 33' S, 33° 01' E). It is situated about 230km from the capital city of Maputo. Chokwe has two climatic conditions, the rainy and dry seasons. The dry season is very long from April-October with high temperatures and short rainy season from November-March. The precipitation is about 761mm per year. The people primarily practice mixed-crop and live-stock agriculture for living. Chokwe samples have high pyrethroid resistance with no mortality to permethrin after 3hrs (Riveron *et al.* 2013)

Malawi (Chikwawa)

An. funestus samples from Chikwawa were collected as described already in chapter two.

3.2.1.2 DNA Extraction

Genomic DNA was extracted from whole mosquitoes using Livak extraction protocol (Livak KJ 1984) from forty eight samples for each country. Each sample was digested in 100µl of extraction buffer (1.6ml 5M NaCl, 5.48g Sucrose, 1.57g Tris, 10.16ml 0.5M EDTA, 2.5ml 20% SDS). Mosquitoes were ground using a pestle and a rotor and immediately transferred in to water bath at 65°C and incubated for 30mins. Samples were briefly microfuged to collect condensation. Potassium acetate 14µl was added and mixed well. Mixture was then incubated in ice for 30mins. Samples were removed from ice and centrifuged at 14,000rpm at 4°C for 20mins. Supernatant was transferred to fresh 1.5ml Eppendorf tubes avoiding debris. 200µl of absolute ethanol was added and flicked gently a few times. Samples were spin for 15min at 4°C at 14,000 rpm. Supernatant was removed and discarded and pellet was rinsed with approximately 100µl ice cold 70% ethanol avoiding dislodging pellet. Pellet was dried by leaving the opened tubes at room temperature for approximately an hour. Pellet was then re-suspended in 100µl of PCR grade water and stored in -20° C for further use.

3.2.1.3 Microsatellite Genotyping

Individual PCR was carried out for each sample and then pulled into multiplex reactions according to their dye colour and sizes as indicated below (Table 3.1). Loci (Fig. 3.2) with differing product size ranges and labelled with different dyes (D2, D3 or D4) were multiplexed in order to increase the throughput of the genotyping and also to save time and cost (Table 3.1).

PCR reagents and conditions for microsatellite genotyping were as follows, 15μl of PCR product, 1.5μl of reaction Buffer, 0.2ul of dNTP 25ml, 0.325μl for both primers, and 9.8 μl of H_2O, 0.2μl of Hotstartaq, 1.0μl of $MgCL_2$ and 1μl of DNA. (15pmol/μl), were conducted on a DNA Thermal Cycler under the following conditions: 15mins at 95ºC, followed by 35 cycles of denaturing at 94ºC for 45s, annealing at 57ºC for 30secs and extension at 72ºC for 45secs finishing with an extension step at 72ºC for 10mins.

Species were ID as described in chapter two and for each sample 1μl of PCR product was added to 30μl CEQ Sample Loading Solution (Beckman Coulter) and 0.5μl CEQ DNA Size Standard 400. Products were sized using the Beckman CEQ8000 fragment analysis software.

Table 3.1: Primers used for microsatellite genotyping including size ranges (Sharakhov *et.al. 2002, Cohuet et al. 2002, Sinkins et al.* 2000)

Primer name	References	Chromosome	Left primer	Right Primer	Label	Size Range	Multiplex with
AFND12		X 3B	GTAGTCGACAATCCGTACGGTTCTCCATCGCTGTTCTACTC	TATAACGTTTCGTACACACGCC	D3	87-107	FUNR, AFUB10
FUNQ		X 5D	ATCGGACTCGAGCTAAGCGGCAAACTGCTAGTAAATGTTTCC	ACATTTCCACAATTTGCGC	D4	84-98	AUB6
FUNO		2R 18A	ATCGGACTCGAGCTAAGCGGCACACATTTCAGGCAGC	GCCCACATTCTGCACCTT	D4	110-132	AFUB11, AFND7
FUNR		2R 12A	CACGACGTTGTAGAACGACGTAGTCGATGGTGCCGTGTG	ACCGTCCCTTCCATCTGTGA	D2	132-148	AFND12, AFUB10
AFUB6		2R 11A	GTAGTCGACAATCCGTACGCCAGCAGGTGTGGAGGAC	GTCGTACAAAAGCACCACCA	D3	141-151	ALONE
FUNL		2L 24C	GTAGTCGACAATCCGTACGAACAGTGGAAGGCAAATTGC	GCACGGTTACCACTGCTCA	D3	140-166	AFN19
AFUB10		2L 26CD	ATCGGACTCGAGCTAAGCGTGTCCATGTACAACCGCAAC TGTCCATGTACAACCGCAAC	TTCTCCAGCATCATCAGCAC	D4	195-210	AFND12, FUNR
AFUB11		2L20D	GTAGTCGACAATCCGTACGCAGTTTCTGCGTGGAGGAAT	AGCAGCTGATGAGCCATCTC	D3	188-191	AFND7, FUNO
AFND19		3R 23A	CAGAACCACTTCGATTCAAC	CCTGCACTCAGAAACACAC	D4	172-205	FUNL
AFND7		3R 33A	CACGACGTTGTAGAACGACTGCATCATTCGACTCGGAAG	AACGGCACTACCGTTCACTG	D2	70-84	FUNO, AFUB11
FUNF		3L 43A	GTAGTCGACAATCCGTACGGCCTTCAGTTTCGATTGGCG CCTTCAGTTTCGATTGGCG	AATAAGATGCGACCGTGGC	D3	104-118	AFUB11
AFUB12		3L 46C	GTAGTCGACAATCCGTACGTGGGGAACTGGTCGTTAGAG TGGGGAACTGGTCGTTAGAG	CTGGTGATGGGATTGAGGAT	D3	152-158	FUNF
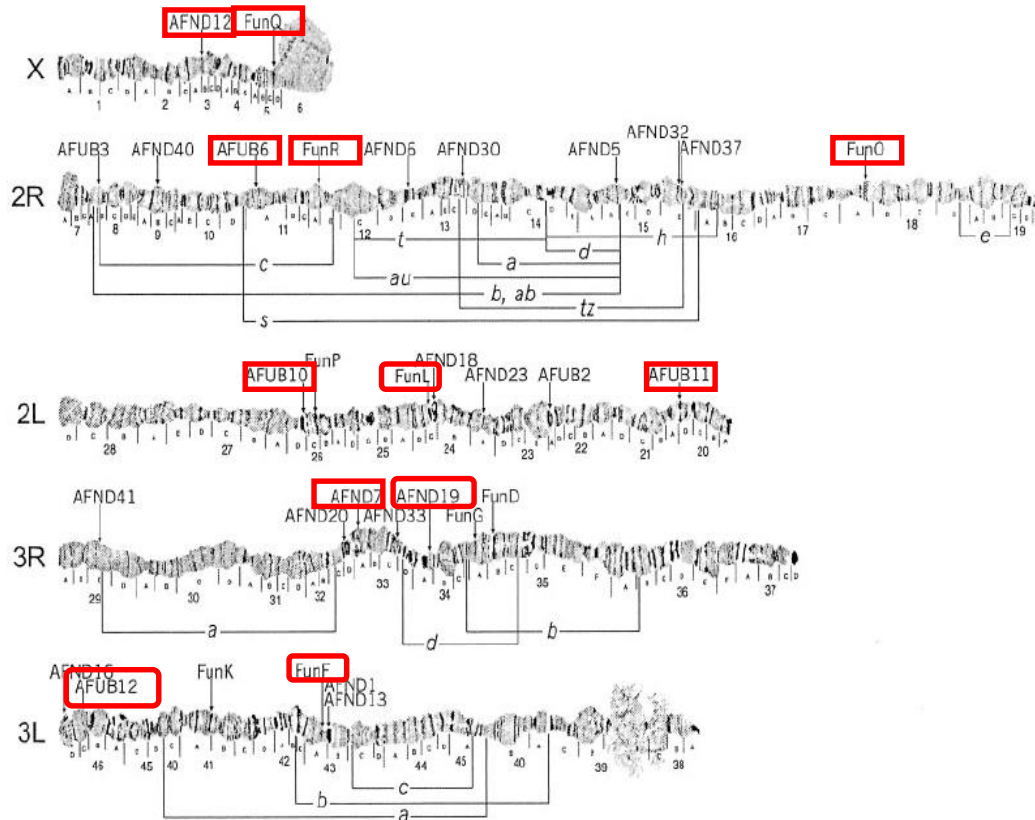
Fig. 3.2: Distribution of microsatellite loci (red boxes) and inversions (brackets) across *An. funestus* polytene chromosomes (from Sharakhov *et al.* 2002).

3.2.1.4 Quality control of allele calling

A visual inspection of each marker was carried out to score the allele sizes following the expected primer size range, and Micro Checker software (version 2.2.3) was used to check for null alleles and scoring errors (Oosterhout *et al*. 2005). Microchecker uses a Monte Carlo simulation (bootstrap) method to generate expected homozygote and heterozygote allele size difference frequencies. The application used the Hardy-Weinberg (HW) theory of equilibrium to calculate expected allele frequencies and the frequency of any null alleles detected (Oosterhout *et al.* 2005).

3.2.2 Data Analysis

Microsatellite data analysis was mainly carried out using Genepop 4.0.10 software (Raymond and Rousset 1995, http://www.genepop.curtin.edu.au).

3.2.2.1 Test for Hardy-Weinberg Equilibrium

Test for deviation from Hardy-Weinberg Equilibrium (HWE) of each individual locus were carried out to determine if each sample was a panmictic population as well as free from locus specific constraints such as null alleles or preferential amplification. Heterozygosity, number of alleles per locus and allele sizes observed in each population was used to measure genetic diversity. Genepop option 1.3 analyses were carried out assuming heterozygote deficit (Raymond and Rousset 1985), and Bonferroni correction (Holm 1979) made at ($P<0.01$) significance level. F_{is} which measures inbreeding within a population with high F_{is} values indicates a possible level of inbreeding. Negative F_{is} values represent excess heterozygosity while positive values represent a deficit. These figures are only of interest if they are calculated to be significant at $P<0.01$ level after Bonferroni correction. Bonferroni correction was carried out so as to consider the effect of multiple testing while determining the statistical significance of the test.

3.2.2.2 Linkage Disequilibrium

Genepop option 2.1 was used to carry outLinkage Disequilibrium (LD) which is the non-random association between alleles at two or more loci on a chromosome and will increase around regions under selection since there is inheritance of alleles close to each other. LD test was carried out to determine if there is no physical linkage between the markers used since the presence of such linkage will make the genetic information redundant and also to confirm whether there is Wahlund effect or not in the population. The test is also described as the composite linkage disequilibrium test (Weir, 1996). LD can be used to confirm deviation from HWE because a Wahlund effect always induces an increase of LD between markers. The default test statistic is the log likelihood ratio statistics (G-test) and tables are created using Markov chain algorithm of Raymond and Rousset (1995a). LD analysis can also be used to determine the likelihood that a sample set is a single population. Lower linkage figures are expected and LD above 5% means that there is the possibility that samples have come from multiple populations and true departure from HW is seen.

3.2.2.3 Genetic differentiation

Genepop option 6.2 was used to determine the differentiation between population pairs using F_{st} estimates. F_{st} is estimated by a "weighted" analysis of variance (Weir and Cockerham 1984). F_{st} is the measure of genetic divergence between populations and can be used to estimate the level of gene flow between individuals from different populations. The greater the F_{st} values (closer to 1), the greater the divergence (less similarity) between populations.

Option 3.4 was used to analysis the significance of differentiation of genotype frequencies. The test is concerned with the diploid genotypes distribution among the various populations. An unbiased estimate of the *P-value* of a log-likelihood (G), based on the exact test is performed (Goudet *et al*. 1996) to assess the genotypic differentiation between populations. Significance of differences between populations was corrected using Bonferroni method.

The F_{st} values were used to construct a neighbor joining tree in MEGA5.2 (Tamura *et al*. 2011).

The spatial genetic structure and the relationship between the pair-wise genetic distance $F_{st}/(1-F_{st})$ and the logarithm of genetic distance was analyzed using the Mantel's test (Mantel 1967). The IBDWS (Isolation by distance web service (http//www.ibdws.sdsu.edu/) program version 3.23 was used for the determination of the relationship between the genetic differentiation and the geographical distance. The geographical distance in kilometers (km) was plotted to determine the relationship between F_{st} and distance. To determine the strength of association distance and genetic similarity (F_{st}), 'r' was calculated. The mantel test was used to calculate the relationship between F_{st} (genetic distance) and geographical distance (km) because it properly accounts for populations as a unit of replication and not as population pairs. By using Fisher's combination probabilities test using *P* values from each locus and/or population, multilocus and/or multi-sample tests of deviation from HWE were carried out as implemented in GENEPOP. In a similar way, multi-locus test of population differentiation was also carried out. In order to evaluate the significance of locus-specific departures with populations, sequential Bonferroni procedure (HOLM 1979) was used since it can detect single test-specific departure when multiple tests are performed.

3.2.2.4 Analysis of genetic population structure using Bayesian approaches

3.2.2.4.1STRUCTURE Model

The population structure was estimated using STRUCTURE version 2 (Pritchard *et al.* 2000). The program calculates the estimate of the posterior probability of the data for a given K, Pr(X/K) (Pritchard *et al.* 2000). STRUCTURE performs a Bayesian analysis that can identify hidden population structure by clustering individuals to groups that are genetically distinguishable on the basis of their LD and allele frequencies (Schug *et al.* 2007). Individually based admixture models were used to identify the ancestral allele source observed in the different individuals, where the ancestral source population is unknown (Pritchard *et al.* 2000).

For this analysis, 20 runs were performed to quantify the likelihood variation for each K-value that ranged from 1-9, where K was the potential number of genetic clusters that might exist in the sampled individuals. The number of known populations in this analysis were 6 (true number of populations) plus 3 (Evanno *et al.* 2005), because varying K-values will provide a guide to deciding the most appropriate value or values. The program performed a Bayesian analysis that assigned individuals to a number of predefined clusters based on the probabilistic analysis of multilocus genotypes. In STRUCTURE, K is a fixed value and that is why different levels of K were used for both the individual and the admixture analysis with a burning period of 50,000 generations and MCMC (Markov Chain Monte Carlo) simulations of 100,000 iterations. The MCMC are useful in visualizing samples from the distribution probability π (θ) which cannot be directly simulated (Pritchard *et al.* 2000).The admixture model was used for the analysis because it allows individuals to have mixed ancestry where a fraction q_k of the genome of an individual comes from subpopulation K (where $\sum_k q_k = 1$) (Pritchard *et al.* 2002). The no-admixture model was not used for this analysis because individuals are assumed to come from single K populations and individuals might have experienced a recent ancestor in more than one population (Pritchard *et al.* 2002).

Estimating K (number of populations) should be treated with care as advised by the authors because, it is computationally difficult to obtain an accurate estimates of Pr (X/K) and their methods only provides an ad hoc approximation; and it may not be straightforward to biologically interpret K (Pritchard *et al.* 2000a). The number of individuals used for this analysis was 288 and number of loci was 12 and 10 (excluding

2 markers FUNR and AFUB6) for six samples from six African countries (Benin, Cameroon, Ghana, Malawi, Mozambique and Uganda).

Structure Harvester (http://taylor0.biology.ucla.edu/structureHarvester/) was used to assess and visualize likelihood values across multiple values of K and hundreds of iterations for easier detection of the number of genetic groups that best fit the data. In addition, it reformats data for use that was used in CLUMPP (Cluster Matching and Permutation Program).

CLUMPP (http://www.resenberglab.bioinformatics.med.umich.edu/clumpp.html) was used for aligning multiple replicate analyses of the same data set by assigning individuals and fraction of individuals to clusters. CLUMPP takes as input the estimated cluster membership coefficient matrices of multiple runs of STRUCTURE program and outputs the same matrices permutated so that all replicates have as close a match as possible (Rosenberg and Jakobsson 2007). CLUMPP also provides finer control of the graphical output.

3.2.2.5 Assignment Test

Assignment test was carried out using Gene class 12.0 software (http//www.genepop.curtin.edu.au). This was to identify the origin of each individual sample within perceived populations. This method using multiple loci, assigns individuals probabilistically to populations (Pritchard *et al.* 2000). Each sample set is assessed as a whole and a representative sample created and each individual sample is then assigned to a population on the basis of their genotype at all the different loci.

3.3 Results

3.3.1 Species ID

All samples used for this study were all confirmed to be *An. funestus* s.s by running all the samples on 1.5% agarose gel checking their sizes if it matches the expected size for *An. funestus* at 505bp as seen in chapter two. Fifty samples were each tested (48 were used) for each location to be used for microsatellite genotyping (Fig. 3.3).
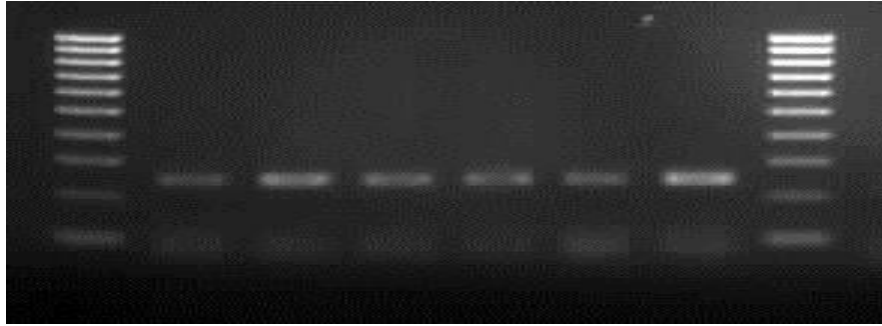
Fig. 3.3: Example of the gel picture of individual PCR of one microsatellite marker (FUNL)

3.3.2 Genotyping and quality control

3.3.2.1 Visual Scoring

PCR amplification for most of the markers was easy with the exception of AFUB6 which needed further optimization. Multiplex amplification was not easy despite various multiplex combinations. Finally individual PCR was carried out (Fig. 3.4) and multiplexing was done during genotyping on the Beckman Coulter machine as shown in Table 3.1. All 12 markers were successfully genotyped and visually scored as shown in Fig. 3.4.
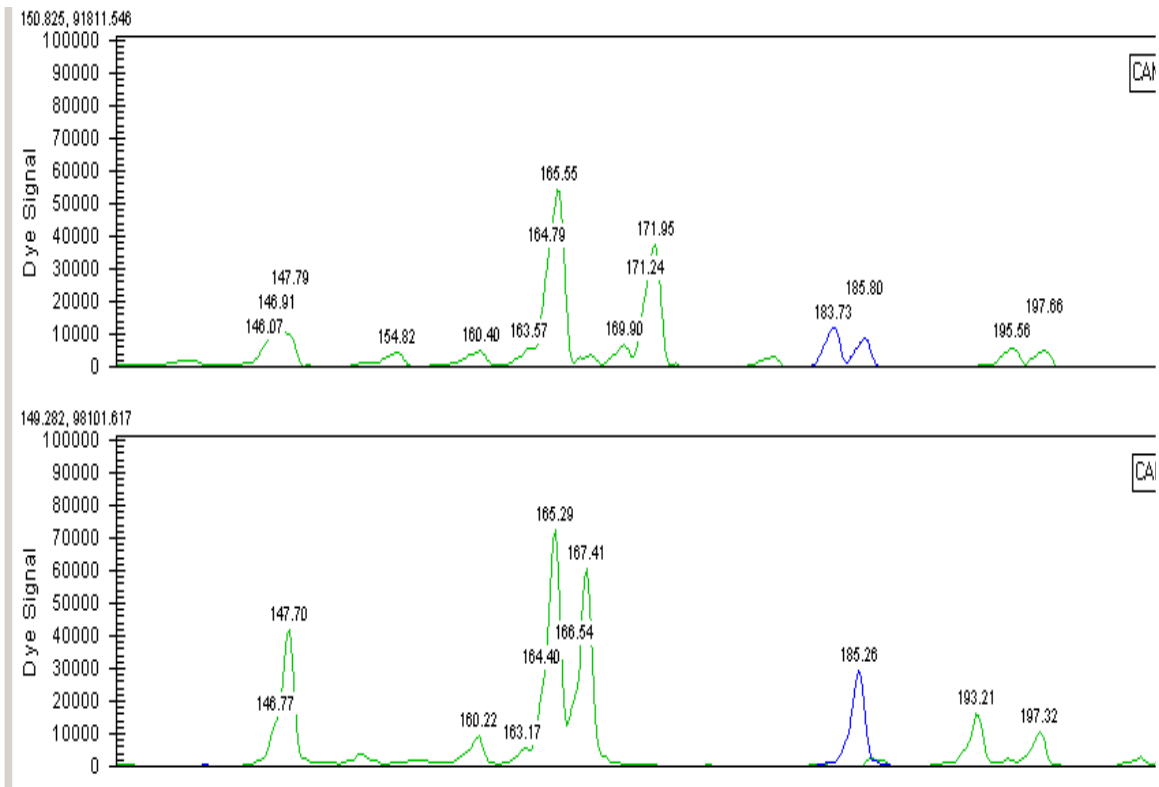
Fig. 3.4: Beckman genotype profile showing peaks for homozygotes and heterozygotes for multiplex primers (FUNL 166-172 green colour, AFND19 184-186 blue.

3.3.2.2 Micro-checker application

Test carried out by micro-checker application (Fig. 3.5) confirmed the absence of scoring errors for all 12 markers, implying that the visual scoring was good for markers across all populations. The Micro-checker detected the probability of null alleles for six markers (AFUB6, AFND12, AFUB10, FUNL, FUNF and FUNQ). This was mostly single markers for a single population across the locations. Some markers deviated significantly from HWE at $P<0.01$ but not across all the locations (4 markers FUNL, AFND12, AFUB10 and AFUB12), at $P<0.05$ (AFND19), and at $P<0.001$ (AFUB10, FUNL and FUNQ) considering the 12 markers across the locations. Two samples showed large allele drop out for FUNF. Large allele drop out can be as a result of preferential amplification of shorter allele (Wattier *et al.* 1998).
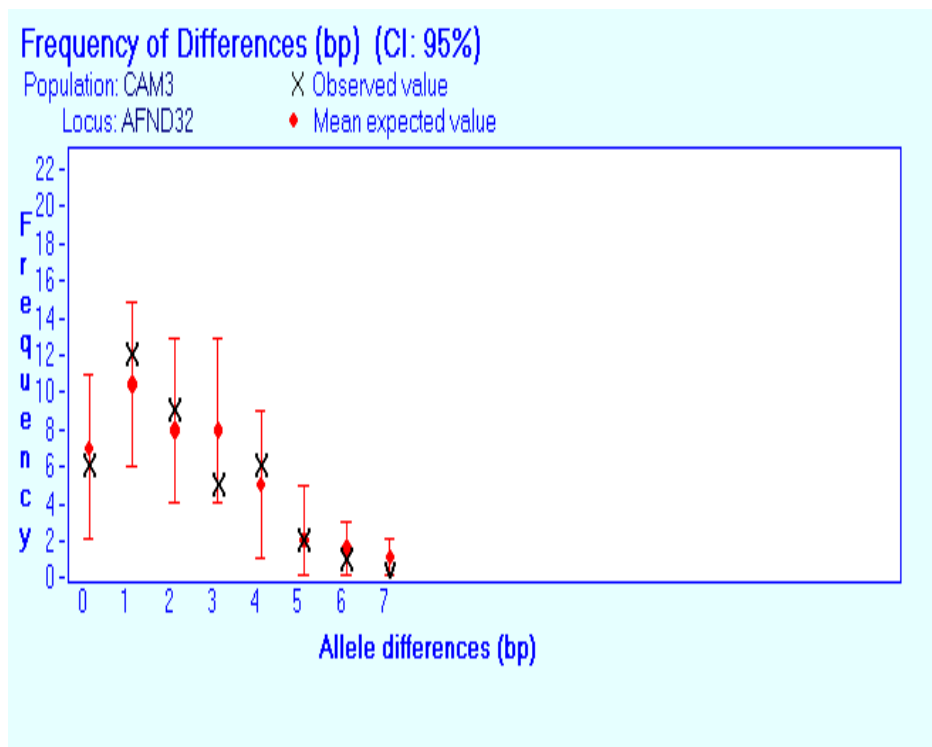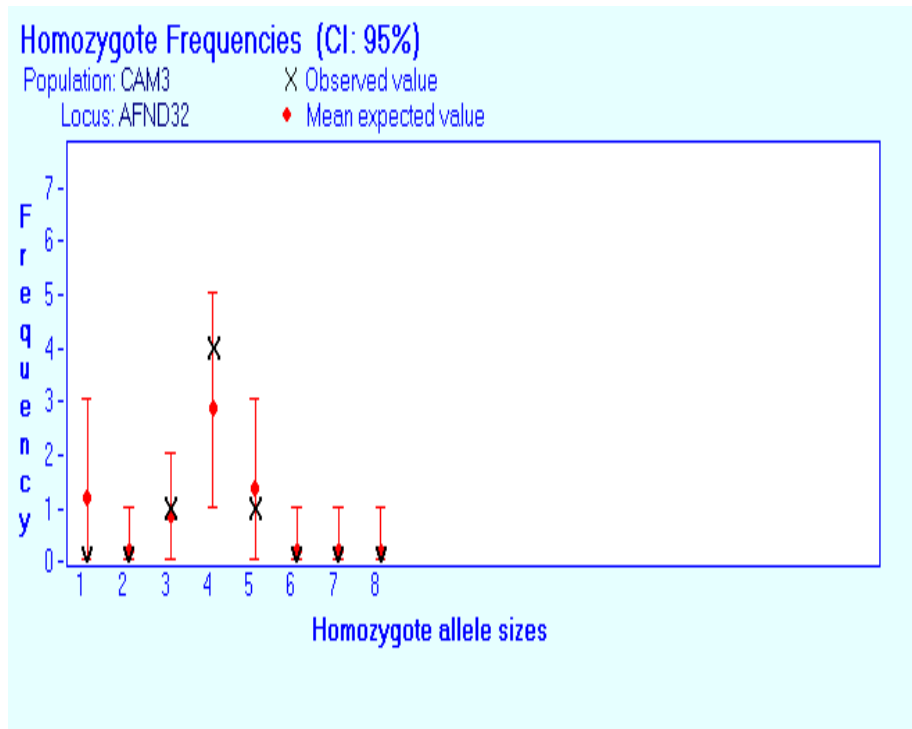
Fig. 3.5: Micro-checker Graphs showing allele frequency differences and homozygote allele sizes.

Distribution of the allele frequency in each sample for each marker was achieved (Fig. 3.6)

Distribution of Allele Frequencies
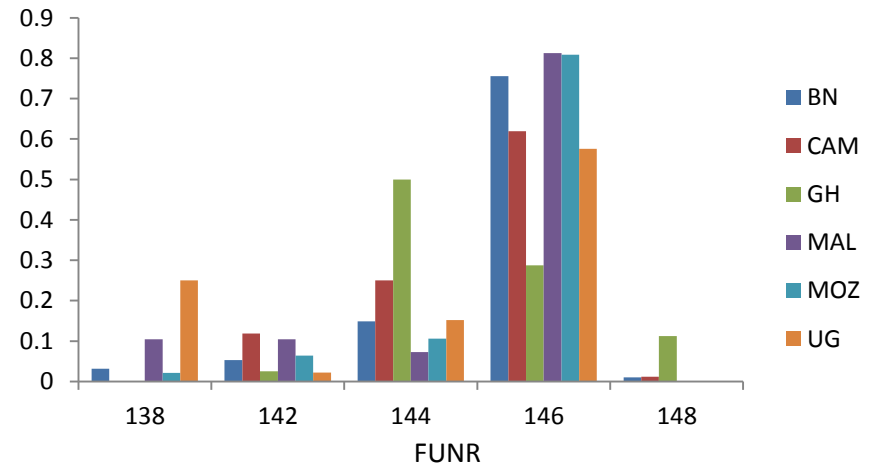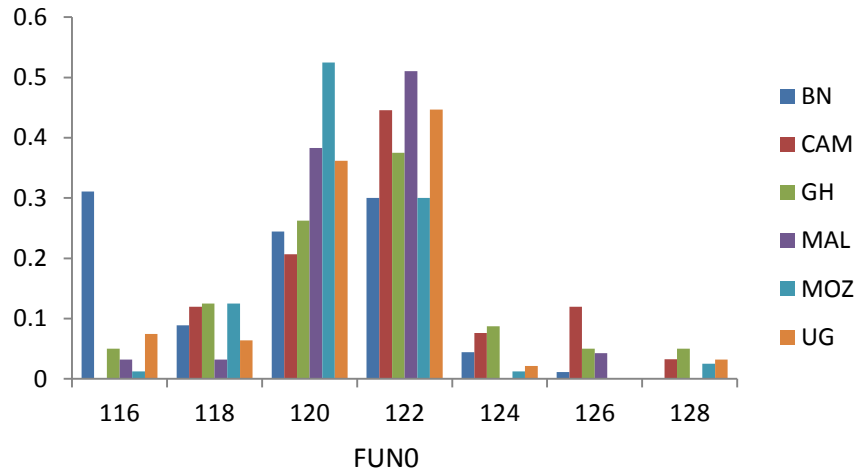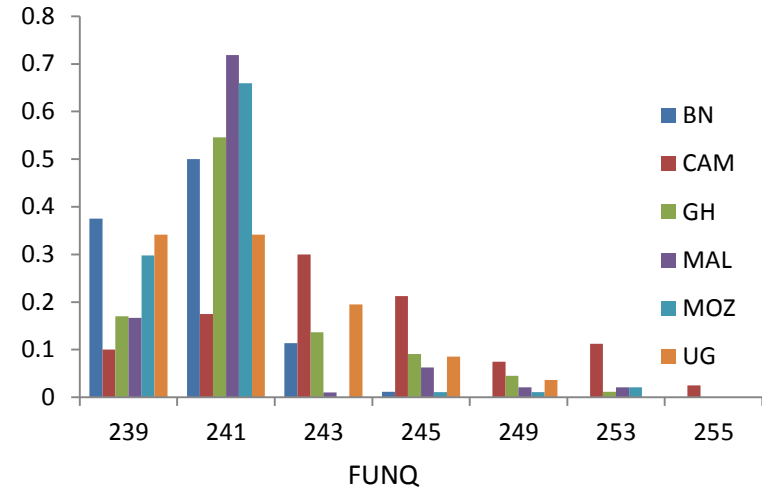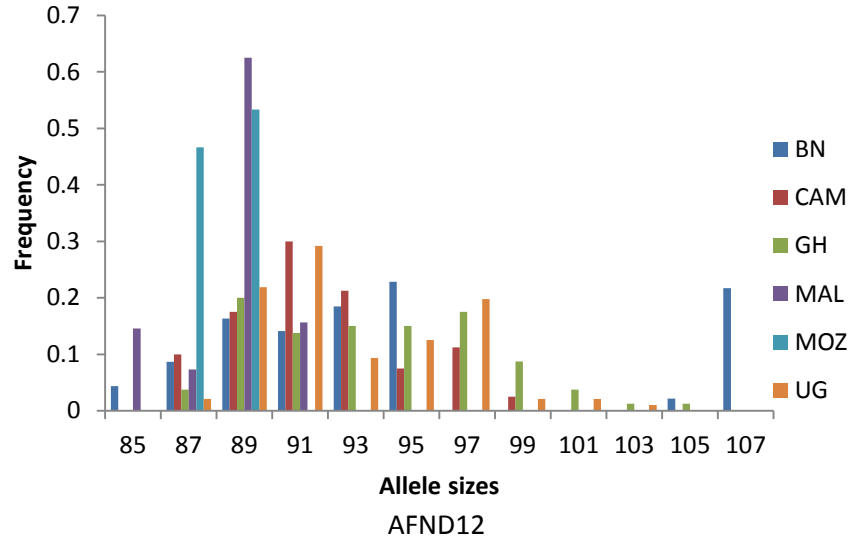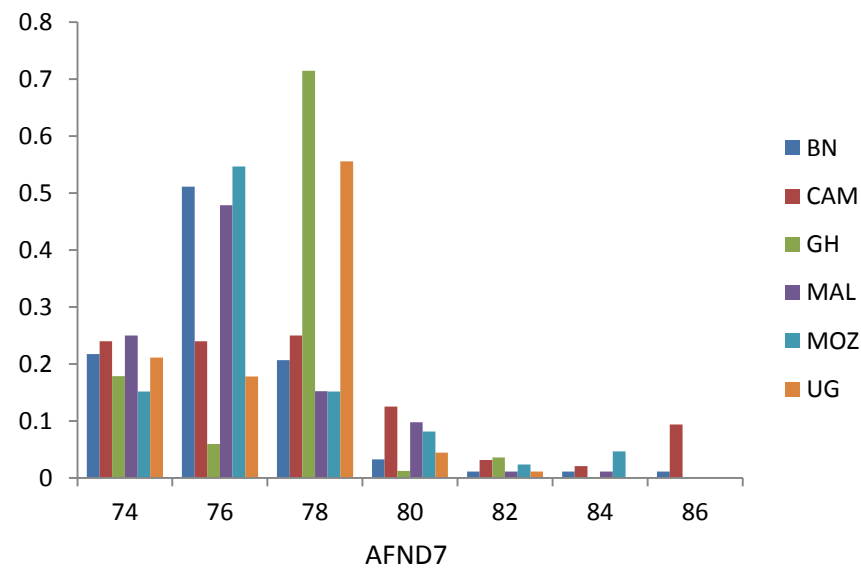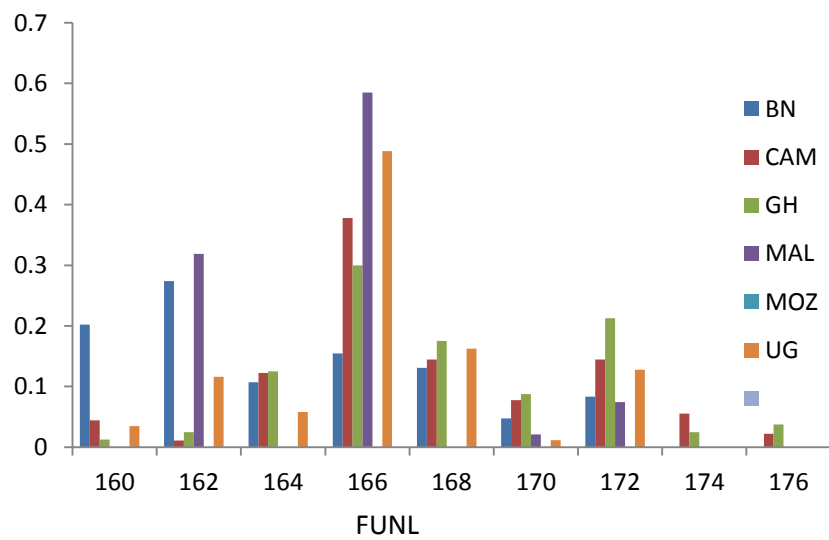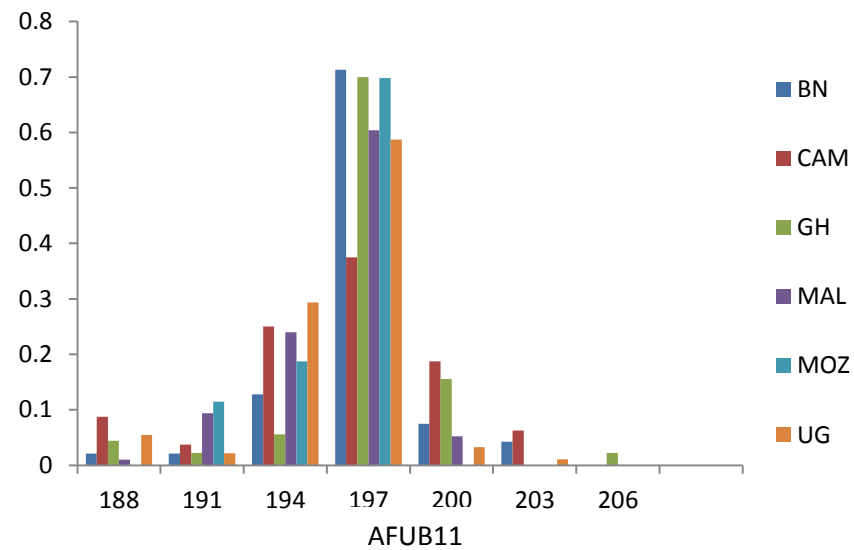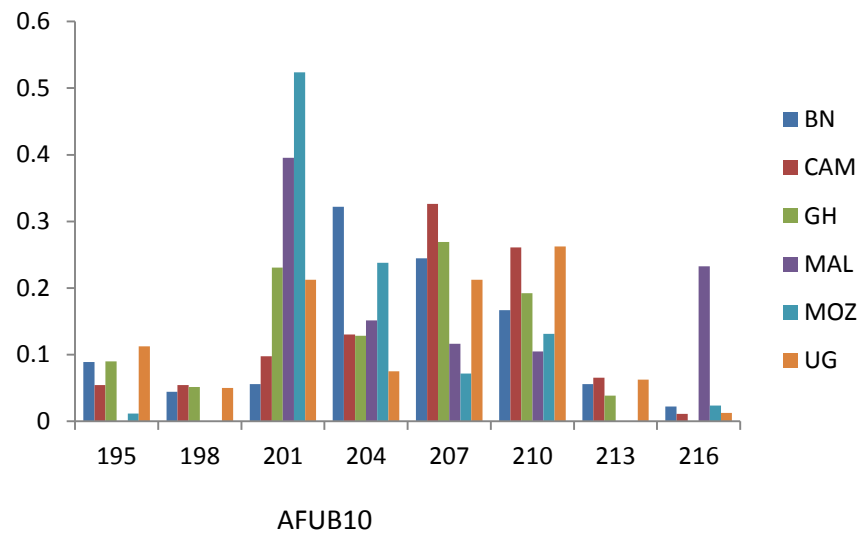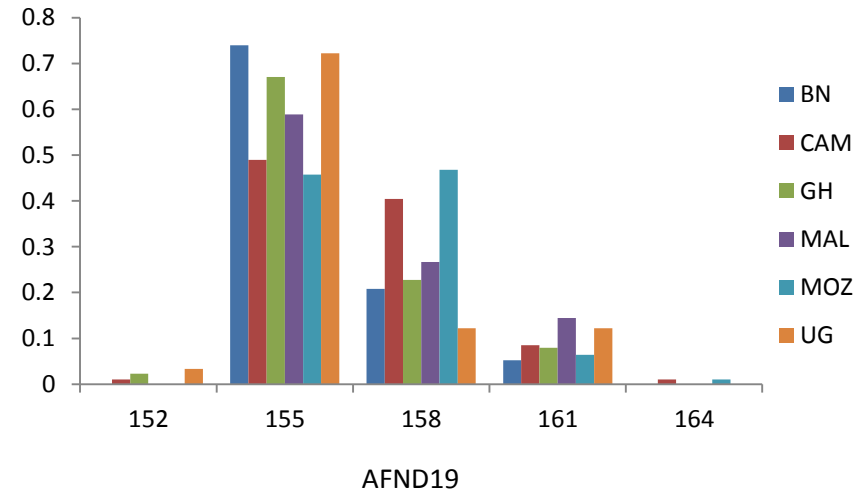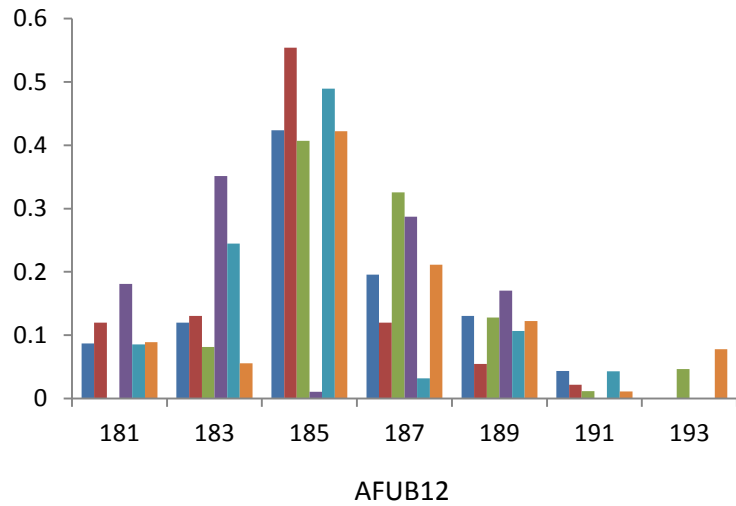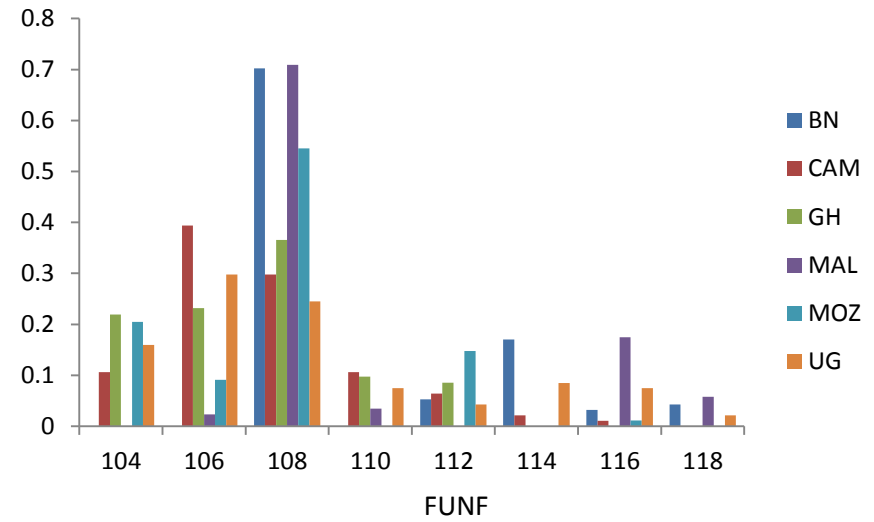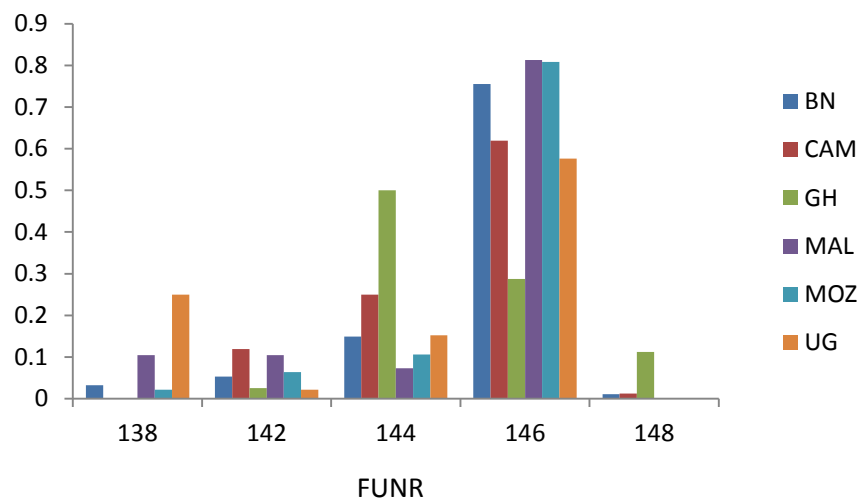
Fig. 3.6: Allele frequency distribution in each sample of *An. funestus* for each microsatellite locus. Alleles are denoted by their estimated sizes (bp). (BN (Benin), CAM (Cameroon), GH (Ghana), MAL (Malawi), MOZ (Mozambique, UG (Uganda).
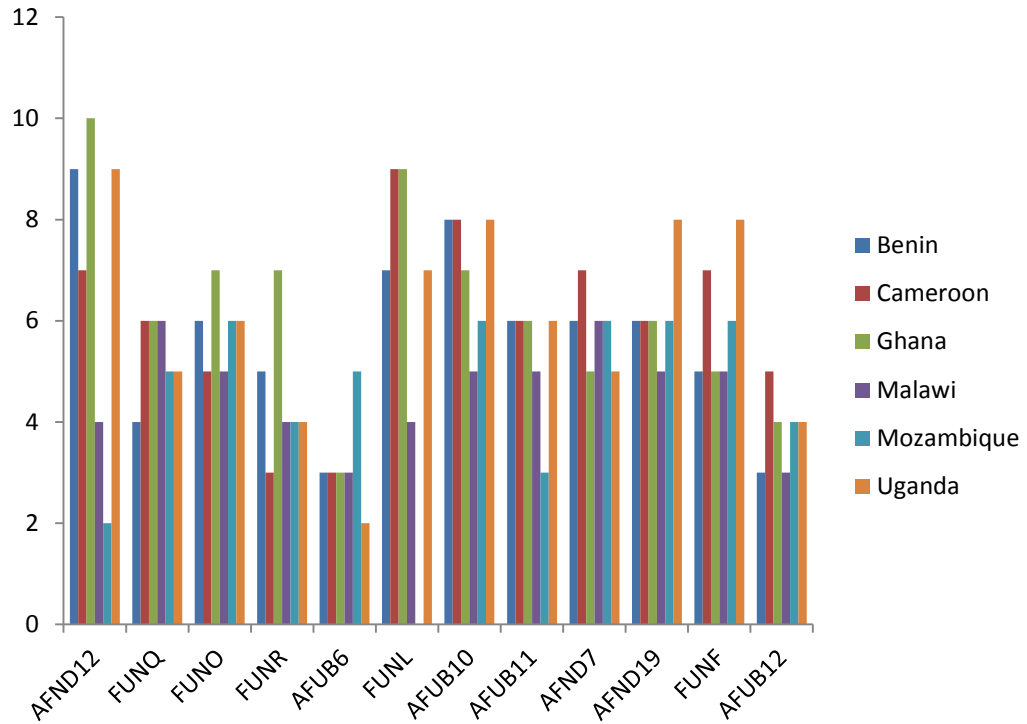
Fig. 3.7: Allele Distribution frequency across the 12 microsatellite loci for each sample of *An. funestus*. Alleles are denoted by their estimated sizes (bp) for the six populations

3.3.3 Test for Hardy-Weinberg

Table 3.2: Genetic diversity and Test for Hardy-Weinberg in *An. funestus*

Makers		BENIN = 48	CAM = 48	GHANA = 48	MAL = 48	MOZ = 48	UGANDA = 48	Mean of All (2N = 288)
AFND12	Nall	9	7	10	4	2	9	6.8
	Hobs	0.84	0.80	0.78	0.46	0.93	0.80	0.77
	Hexp	0.85	0.81	0.87	0.56	0.50	0.85	0.74
	F_{is}	0.0076	0.0193	-0.1058	**<u>0.1893</u>**	-0.8793	**0.0495**	**<u>0.2085</u>**
FUNQ	Nall	4	6	6	6	5	5	5.3
	Hobs	0.32	0.43	0.43	0.23	0.13	0.46	0.33
	Hexp	0.60	0.72	0.65	0.46	0.48	0.73	0.61
	F_{is}	**<u>0.4754</u>**	**<u>0.3997</u>**	**<u>0.3401</u>**	**<u>0.4995</u>**	**<u>0.7365</u>**	**<u>0.3672</u>**	**<u>0.6156</u>**
FUNO	Nall	6	5	7	5	6	6	5.8
	Hobs	0.77	0.85	0.73	0.51	0.53	0.87	0.71
	Hexp	0.75	0.73	0.77	0.60	0.63	0.67	0.69
	F_{is}	-0.0349	-0.1615	**0.0583**	**0.1432**	**0.1626**	-0.315	**<u>0.0525</u>**
FUNR	Nall	5	3	7	4	4	4	4.5
	Hobs	0.38	0.52	0.33	0.17	0.17	0.76	0.39
	Hexp	0.41	0.55	0.67	0.33	0.33	0.59	0.48
	F_{is}	0.0312	0.0420	**<u>0.5104</u>**	**<u>0.4929</u>**	**<u>0.4931</u>**	-0.2974	**<u>0.0496</u>**

AFUB6	Nall	3	3	3	3	5	2	3.2
	Hobs	0.65	0.73	0.18	0.12	0.16	0.75	0.48
	Hexp	0.64	0.32	0.21	0.20	0.59	0.47	0.41
	F_{is}	-0.0150	0.1429	0.1373	**0.0406**	**0.0213**	-0.5932	**_0.0987_**
FUNL	Nall	7	9	9	4	No Data	7	7
	Hobs	0.74	0.60	0.58	0.36		0.58	0.57
	Hexp	0.83	0.80	0.82	0.56		0.71	0.74
	F_{is}	0.1143	**0.2500**	**_0.3000_**	0.3516		**0.1813**	**_0.2394_**
AFUB10	Nall	8	8	7	5	6	8	7
	Hobs	0.71	0.50	0.52	0.63	0.45	0.60	0.67
	Hexp	0.80	0.80	0.82	0.75	0.65	0.83	0.78
	F_{is}	**0.1134**	**_0.3756_**	**_0.3770_**	-0.1656	**_0.3106_**	**0.2764**	**_0.2698_**
AFUB11	Nall	6	6	6	5	3	6	4.8
	Hobs	0.46	0.73	0.33	0.71	0.52	0.74	0.58
	Hexp	0.48	0.76	0.49	0.57	0.47	0.57	0.56
	F_{is}	**0.0548**	**0.0444**	**_0.3154_**	-0.2416	-0.1106	-0.2988	**0.0498**
AFND7	Nall	6	7	5	6	6	5	5.8
	Hobs	0.72	0.77	0.31	0.59	0.56	0.67	0.60
	Hexp	0.65	0.81	0.46	0.68	0.65	0.62	0.65
	Fis	-0.0968	0.0432	**0.3274**	-0.1423	-0.1479	-0.0762	**_0.1263_**
AFND19	Nall	6	6	6	5	6	8	6.2
	Hobs	0.65	0.54	0.67	0.76	0.64	0.71	0.66
	Hexp	0.75	0.65	0.71	0.74	0.69	0.75	0.72
	F_{is}	**_0.1310_**	**0.1664**	0.0525	-0.0350	0.0710	**0.0566**	0.0271
FUNF	Nall	5	7	5	5	6	8	6
	Hobs	0.47	0.72	0.80	0.44	0.66	0.43	0.59
	Hexp	0.48	0.74	0.75	0.47	0.64	0.81	0.65
	F_{is}	0.0199	0.0185	-0.0645	0.0539	-0.0340	**_0.4799_**	**_0.0857_**
AFUB12	Nall	3	5	4	3	4	4	3.8
	Hobs	0.44	0.55	0.45	0.67	0.68	0.38	0.53
	Hexp	0.41	0.60	0.50	0.57	0.57	0.45	0.52
	F_{is}	-0.0647	-0.0725	0.0414	-0.1770	-0.0189	**0.1666**	**0.0310**
Combined Data for All Loci	Nall	5.7	6	6.3	4.9	4.8	6	5.5
	Hobs	0.60	0.75	0.51	0.47	0.52	0.61	0.54
	Hexp	0.67	0.69	0.64	0.54	0.56	0.67	0.63
	F_{is}	0.0054	0.1098	0.0088	0.0148	0.0017	-0.0649	**_0.1320_**

Polymorphism at 12 microsatellite loci in *An. funestus* collected from six African countries (Nall–Number of alleles, Hobs -Observed heterozygosity, Hexp - Expected heterozygosity (Nei 1978). Significant after Bonferroni correction is represented by bold (P<*0.05*), bold and underlined (P<*0.01*) after accounting for multiple testing (Holm 1979). Deviation from Hardy-Weinberg is shown by F_{is} calculated according to Weir and Cockerham (1984). Mean of all F_{is} values is calculated by the sum of values for each locus divided by the total number of populations (6) Cam = Cameroon, Mal = Malawi, Moz = Mozambique.

Table 3.3: Deviation from HWE for each marker in all populations

Markers	Before Bonferroni correction	After Bonferroni correction
AFND12	2	1
FUNQ	6	6
FUNO	3	0
FUNR	3	3
AFUB6	3	0
FUNL	4	1
AFUB10	5	3
AFUB11	3	1
AFND7	2	0
AFND19	3	1
FUNF	2	1
AFUB12	1	0
Total	35	17

The number of sample sets with significant deviation from HWE for each marker before and after Bonferroni correction

Table 3.4: Deviation from HWE for each sample set

	Before Bonferroni correction	After Bonferroni correction at (P<0.05)	After Bonferroni correction at (P<0.01)
Benin	4	2	2
Cameroon	5	3	2
Ghana	7	5	5
Malawi	6	3	3
Mozambique	6	5	3
Uganda	7	5	2

The number of markers (out of 12) with significant deviation from HWE for each sample set before and after Bonferroni correction at (P<0.05) and (P<0.01)

3.3.4 Genetic diversity

Moderate to high levels of polymorphism were observed in all the samples from the six African populations of *An. funestus* (Table 3.2). In total, the average number of alleles per locus was 5.5 (ranging from 2 for AFUB6 to 12 for AFND12).The average heterozygosity was 0.54 (range 0.12-0.93), with AFUB6 having lowest heterozygosity (0.12). Samples from Benin and Ghana showed an average of 6 (range 3-10) alleles per locus with average heterozygosity of 0.55 (range 0.31-0.84). The Malawi and Mozambique samples had an average of 4.9 alleles (range 2-6), an average heterozygosity of 0.49 (range 0.12-0.71). The Cameroon samples showed an average of 6 alleles (range 3-9) with average heterozygosity of 0.75 (range 0.50-0.85), and the Uganda samples had

an average of 6 alleles (2-9) with average heterozygosity of 0.61 (range 0.38-0.87). There was low polymorphism at 2 loci; FUNR and AFUB6 for Malawi and Mozambique (average Hobs 0.17 for FUNR and 0.14 for AFUB6) compared with samples from Benin and Ghana with (average Hobs 0.36 for FUNR and 0.42 for AFUB6). The presence of null alleles observed for AFUB6 was not seen across the six populations and possibly does not explain the low allelic richness since the FUNR also had low allelic richness. Presence of selection acting on these loci could be a possible reason why they exhibit a low allelic richness since they are located within the rp1 QTL on the 2R chromosome (Wondji *et al.* 2007).

3.3.5 Test of Hardy-Weinberg Equilibrium (HWE)

Significant heterozygosity deficit was observed in 35 out of 71 tests across the markers before Bonferroni correction at ($P<0.05$), and 17 out of 71 tests were observed after Bonferroni correction at ($P< 0.01$) across the markers. FUNQ which is located outside chromosomal inversion showed significant deviation across all the samples for both before and after Bonferroni corrections at $P<0.05$ (all six samples) and $P<0.01$(6 samples) (Table 3.3). AFND7, FUNF and AFUB12 showed the least significant deviation from HWE for all samples (2 samples) after Bonferroni correction at $P<0.05$ (Table 3.3). Across the samples, Ghana, Malawi and Mozambique 5 and 6 markers deviating before Bonferroni correction at $P<0.05$ and 5, 3 and 2 markers respectively at $P<0.01$ after Bonferroni correction (Table 3.4). Benin, Cameroon and Uganda showed the lowest number (2) of markers showing significant deviation from HWE after Bonferroni correction at $P<0.05$. 7 out of 24 ($P< 0.01$) tests (FUNR and AFUB6) showed significant heterozygote deficit.

LD analysis showed an absence of LD between markers in all the samples with proportion of LD below 5% in all populations. Overall, the lack of significant LD in the samples reveals that the different alleles of each microsatellite marker are evenly distributed between the different populations and therefore the observed deviations from HWE in this study are less likely to be caused by genetic substructure within each sample. In the presence of inbreeding in a population, it is expected that different gene pools will coexist inducing a lack of LD which was not the case in this study therefore supporting the view that the departure from HWE is less likely to have been caused by the presence of a Wahlund effect or inbreeding. However, for each locus pair across all populations, the

following five different loci out of the 66 pairwise comparisons were significantly deviating FUNR-FUNL, FUNQ-AFND19, AFUB11-FUNF AFND7-FUNF and FUNF-AFUB12.

3.3.6 Genetic differentiation between populations

Despite some markers showing deviation from HWE, this is more likely the results of locus-specific constraints such as null alleles (Callen *et al.* 1992), limited allelic range (Epplen *et al.* 1994a) or preferential amplification of one allele in heterozygotes (Wattier *et al*. 1998) rather than population substructure or inbreeding within the different samples. It is very common to observe deviations from HWE in population genetics studies as performed in several *Anopheles* species using microsatellite markers either in *An. funestus* (Cohuet *et al.* 2005, Michel *et al.* 2005), in *An. gambiae* (Wondji *et al.* 2002, Lehman *et al.* 2000) or even in *An. arabiensis* (Simard *et al.* 2000). Many observations point to the fact that, the observed departures from HWE are more locus-specific and therefore could have less impact in the patterns of genetic differentiation assessed using F_{st} estimates. Firstly, it could be noticed that most of the markers deviating from HWE in this study are doing so in many populations showing that it is not due to Wahlund effect or inbreeeding. This is the case of FUNQ which exhibits a positive F_{is} in all populations possibly due to presence of null alleles. Similarly AFUB10 departs from HWE for 5 out of 6 populations again supporting that such deviations are due to locus-specific constraints not Wahlund effect. Secondly, these loci are found to be in linkage equilibrium at every stage of the analysis suggesting that the different alleles of each marker are evenly distributed between the different populations and therefore that the observed deviations from HWE are not caused by genetic substructure within each samples. Based on these observations, markers showing a deviation from HWE were also included to assess the genetic differentiation levels based on F_{st} estimates. Nevertheless, this was done with caution and effect of this inclusion was assessed by estimating F_{st} with and without these markers. No significant impact was found.

Levels of genetic differentiation between populations were estimated using F_{st} (Weir and Cockerham, 1984) and significance of F_{st} estimates were carried out using Bonferroni correction at F_{st} (P < 0.05) and (P <0.001). Overall F_{st} were significant for most of the loci across the populations (Table 3.5). Average F_{st} ranged from 0.0342 for Cameroon-Uganda to 0.2060 for Ghana-Malawi. Populations sharing the same geographical region are expected to show reduced genetic differentiation between samples. While Malawi and

Mozambique showed greatest similarity for the 12 markers (0.0600), Ghana and Benin showed a higher genetic differentiation (0.1745). The population from Cameroon showed greater genetic homogeneity with the population from Uganda for the 12 markers (0.0342). Samples from Malawi and Mozambique showed high genetic differentiation to the samples from Ghana and Benin e.g. (Benin-Mozambique, 0.0922, Ghana-Malawi, 0.2060, Ghana-Mozambique, 0.1820). Overall, the samples from Malawi and Mozambique showed much reduced genetic similarity when compared to the rest of the populations. Genetic differentiation revealed a significant differentiation between the Ghana and Benin samples and samples from Malawi and Mozambique as well as the Central and East African samples. The F_{st} values also showed very high genetic differentiation across the markers for the pair wise comparisons. In situations where genetic differentiation is negative, this implies that the variance of the within-samples in allele size is bigger than the variance of among samples (Goodman 1997). F_{st} for markers outside chromosomal inversion was considered and a much reduced genetic differentiation was observed between the West African population (Benin and Ghana range 0.0381) and there was still a higher genetic differentiation between the West African populations and southern African populations (Malawi and Mozambique e.g. Benin-Mozambique 0.1198, Ghana-Mozambique 0.0939). Overall the F_{st} for markers outside chromosomal inversions showed that the samples from Benin and Ghana were distant from Malawi and Mozambique samples. When pair wise analysis was carried out for 10 markers excluding the two markers (FUNR and AFUB), the samples from Benin showed reduced genetic differentiation from samples from Ghana (F_{st} 0.0674) when compared with the F_{st} values for the 12 markers (F_{st} 0.1745). This possibly suggests that the two markers (FUNR and AFUB6 located on the 2R) were impacting on the results since they are found around the rp1 QTL where the resistance genes CYP6P9a and CYP6P9b are located (Wondji *et al.* 2007) but not FUNQ that is located on the X chromosome. No significant difference was observed with analysis carried out with markers seen to be deviating from neutrality like FUNQ and AFUB10.

Table 3.5: Genetic differentiation per locus of populations from Benin, Cameroon, Ghana, Malawi, Mozambique and Uganda

	BN-CM	BN-GH	BN-ML	BN-MZ	BN-UG	CM-GH	CM-ML	CM-MZ	CM-UG	GH-ML	GH-MZ	GH-UG	ML-MZ	ML-UG	MZ-UG
AFND12	0.0196	0.0035	0.1796	0.2343	0.0249	0.0162	0.1008	0.1744	0.0082	0.1859	0.2528	0.0077	0.2015	0.1257	0.2141
FUNQ	0.0479	0.0254	0.0770	0.0232	0.0129	0.0976	0.2110	0.1515	0.0008	0.0269	0.0314	0.0365	0.0080	0.1395	0.0957
Average F_{st} X	0.0588	0.0449	0.1518	0.1224	0.0204	0.0100	0.0976	0.0726	0.0111	0.1412	0.0881	0.0078	0.0846	0.0868	0.0932
FUNO	0.0747	0.0399	0.0897	0.0999	0.0541	-0.0004	0.0314	0.0844	0.0270	0.0252	0.0447	0.0085	0.0468	-0.0011	0.0318
FUNR	0.0236	0.2451	0.0087	-0.0067	0.0674	0.1267	0.0848	0.0524	0.0600	0.3274	0.3050	0.1762	0.0006	0.0752	0.0984
AFUB6	0.0712	0.8653	0.0305	0.2041	0.3129	0.7390	0.0067	0.0549	0.1198	0.7924	0.5528	0.6530	0.1087	0.1813	0.0695
Average-2R	0.0754	0.2602	0.0467	0.0360	0.0901	0.0997	0.0673	0.0330	0.0328	0.2764	0.1567	0.1532	0.0793	0.1089	0.0285
FUNL	0.0738	0.0662	0.1479	-	0.0887	-0.0064	0.1107	-	0.0099	0.1388	-	0.0261	-	0.0504	-
AFUB10	0.0206	0.0278	0.1150	0.1451	0.0433	0.0008	0.1111	0.1516	0.0044	0.0601	0.0794	-0.0093	0.0343	0.0693	0.0942
AFUB11	0.0987	0.0025	0.0208	0.0112	0.0348	0.0984	0.0509	0.1093	0.0445	0.0442	0.0415	0.0653	0.0056	0.0017	0.0260
Average F_{st} -2L	0.0627	0.0329	0.0901	0.1331	0.1168	0.0355	0.0776	0.2163	0.1389	0.0840	0.2071	0.1376	0.1971	0.1645	0.1797
AFND7	0.0498	0.2864	-0.0041	-0.0018	0.1475	0.1664	0.0383	0.0671	0.0646	0.2973	0.3283	0.0254	-0.0009	0.1568	0.1864
AFND19	0.0090	0.0077	0.0549	0.0207	-0.0043	0.0499	0.0880	0.0098	0.0183	0.1092	0.0754	0.0037	0.0789	0.0622	0.0420

Average F_{st} -3R	0.1788	0.1232	0.0407	0.1266	0.0257	0.1793	0.1230	0.2338	0.1678	0.0841	0.1805	0.0904	0.1126	0.0602	0.1309
FUNF	**0.2249**	**0.1672**	**0.0438**	**0.0847**	**0.1983**	0.0185	**0.2203**	**0.1077**	**0.0026**	**0.1650**	0.0349	0.0100	**0.0930**	**0.1971**	0.0893
AFUB12	**0.0835**	-0.0037	**0.0256**	**0.1236**	**0.0046**	0.0449	0.0173	-0.0052	**0.1046**	0.0018	**0.0801**	**0.0041**	0.0455	**0.0278**	**0.1515**
Average F_{st} -3L	0.0426	0.0138	0.0471	0.0765	0.0403	0.0313	0.0285	0.0108	0.1037	0.0152	0.0515	0.0256	0.0326	0.0592	0.1047
Average F_{st}12 loci	**0.0677**	**0.1745**	**0.0784**	**0.0922**	**0.0780**	**0.1317**	**0.0969**	**0.0921**	0.0342	**0.2060**	**0.1820**	**0.1072**	**0.0600**	**0.0928**	**0.1045**
Average F_{st} 10 loci	**0.0704**	**0.0674**	**0.0832**	**0.0916**	**0.0654**	0.0483	**0.1007**	**0.0974**	0.0268	**0.1150**	**0.1184**	0.0168	**0.0602**	**0.0888**	**0.1081**

The F_{st} estimates were calculated using genepop. Pairwise Significance after standard Bonferroni corrections is represented by bold (P<*0.05*) and underlined (P<*0.01*). (on top) Genetic differentiation per locus and across all Loci for all sample pair comparisons (BN-Benin, CM-Cameroon, GH-Ghana, ML-Malawi, MZ-Mozambique & UG-Uganda) and average F_{st} across the various chromosomes (X,2R,2L,3R&3L) and loci(12&10).

3.3.7 Structure Model

In estimating the best K in this study, results from structure were loaded into the genetic structure harvester software, where Pr (K) is small for K < the real value (effective zero), and more plateaus for larger K. For this model, the highest likelihood was observed at K=3 for both 12 and 10 markers (Fig. 3.8a and 3.8b) since it is advised to aim for the smallest value of K that captures the major structure in the data set (Pritchard *et al.* 2000a). When structure analysis was carried out for the 12 markers, the individuals from Benin (West Africa) where represented mostly in Cluster 1 with Uganda (East Africa) (0.865 for Uganda and 0.775 for Benin), the individuals from Cameroon (Central Africa) where represented mostly in Cluster 2 with Ghana (West Africa) (0.896 for Cameroon and 0.722 for Ghana), and the samples from Malawi and Mozambique both from southern Africa were represented most in cluster 3 (0.745 for Malawi and 0.935 for Mozambique) (Fig. 3.10a and Table 3.6a). The analysis for the 10 markers showed similar distribution like the 12 markers but Ghana had almost 50% distribution between cluster 1 and cluster 2 (0.447) and (0.419) respectively (Fig. 3.10(b) and Table 3.6(b)).
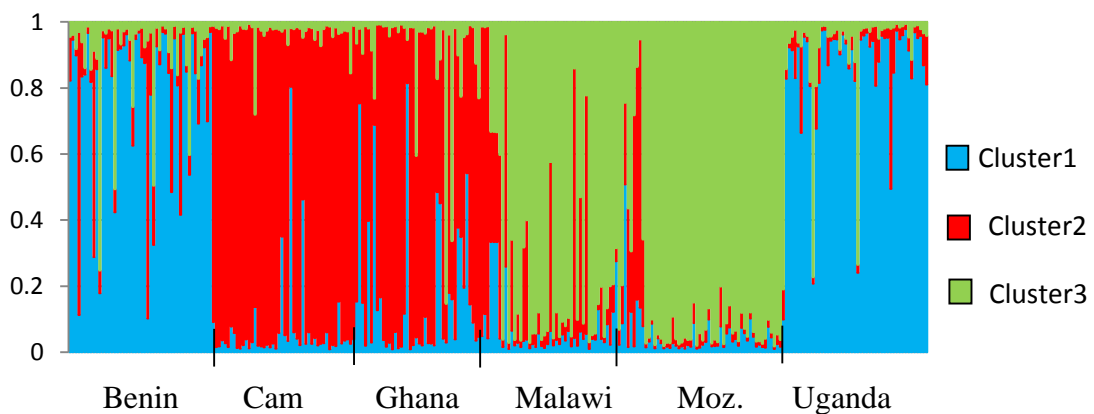
(a) 12 Markers
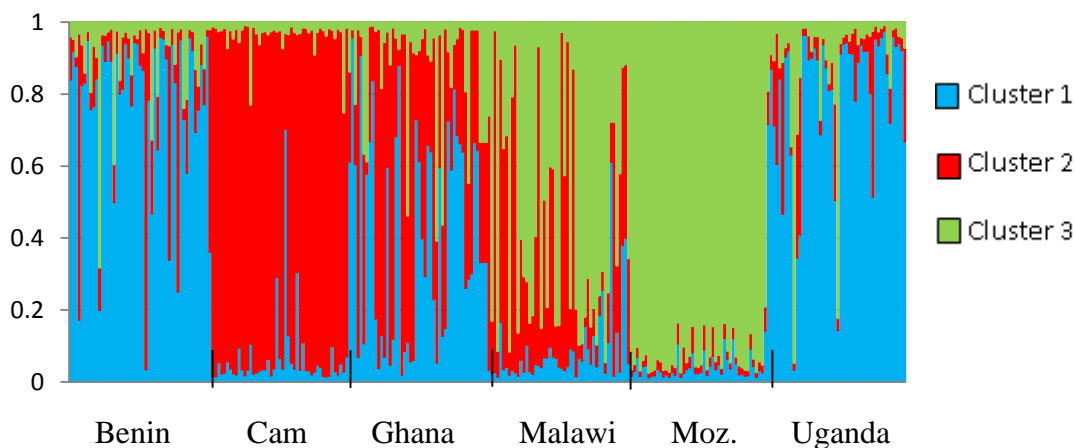
(b) 10 Markers

Fig. 3.8(a-b): Graphical representation of the data set for the most likely K (K=3) for Bayesian cluster analysis using STRUCTURE, where each color represents a suggested cluster, Blue color corresponds to (Cluster 1 = Benin + Uganda), Red, (Cluster 2 = Cam-Cameroon + Ghana) and Green (Cluster 3 = Malawi + Moz-Mozambique) for 12 and 10 makers respectively.

Table 3.6(a): Proportion of membership of each predefined population in each of the three clusters inferred by STRUCTURE (12 Markers)

Population	Cluster 1	Cluster 2	Cluster 3
Benin	0.774	0.130	0.095
Cameroon	0.073	0.888	0.038
Ghana	0.419	0.447	0.132
Malawi	0.089	0.349	0.561
Mozambique	0.042	0.033	0.924
Uganda	0.790	0.093	0.115

Table 3.6(b): Proportion of membership of each predefined population in each of the three clusters inferred by STRUCTURE (10 Markers)

Population	Cluster 1	Cluster 2	Cluster 3
Benin	0.775	0.121	0.102
Cameroon	0.067	0.896	0.036
Ghana	0.172	0.722	0.105
Malawi	0.063	0.191	0.745
Mozambique	0.035	0.029	0.935
Uganda	0.865	0.057	0.077

3.3.8 Assignment test

Population structure can be further studied by using multilocus Bayesian method by assigning individual samples to different populations. Assignment tests were used to assign the potential origin of individual samples used for this study. For the 12 markers, about 85-100% of individuals were correctly assigned to their original populations. To investigate the impact of the two markers, the markers (FUNR and AFUB6) were removed from the analysis and about 46-92% of individuals were correctly assigned to their original populations. In Uganda, 45.8% of the samples were correctly assigned to their origin and 23% were assigned to Ghana, 20% to Cameroon and 10.4% to Benin. In Malawi, samples that were allocated outside their origin were allocated mostly to Mozambique for both 12 (Table 3.7a and Fig. 3.9a) markers as well as the 10 markers (Table 3.7b and Fig. 3.9b). This correlates with the F_{st} estimates seen between Malawi and Mozambique. Ghana had a 100% allocation to itself for the 10 markers. While for the 12 markers, most of its samples were not allocated to Benin which is found in same geographical origin but to Uganda and Cameroon. This result also correlates with the high genetic differentiation seen between Ghana and Benin for the 12 markers, which is similar to results observed in STRUCTURE and BAPS.

Table 3.7(a): Assignment of individuals to different population groups (12 microsatellite)

COUNTRY	BN	CAM	GH	MAL	MOZ	UG
BN	83.3	0	0	4.2	0	2.1
CAM	2.1	85.4	0	0	6.7	10.4
GH	0	2	100	0	0	0
MAL	2.1	4.2	0	83.3	0	2.1
MOZ	4.2	4.2	0	8.3	91.7	0
UG	8.3	4.2	0	0	2	85.4
Total %	100	100	100	100	100	100

Table 3.7(b): Assignment of individuals to different population groups (10 microsatellite)

COUNTRY	BN	CAM	GH	MAL	MOZ	UG
BN	81.3	0	2.2	2.1	0	10.4
CAM	2.1	77.1	11.1	2.1	4.2	20.8
GH	0	8.3	64.4	0	0	23
MAL	4.2	6.3	2.2	87.5	0	0
MOZ	2	2	0	6.3	91.6	0
UG	10.4	6.3	20	2	4.2	45.8
Total %	100	100	100	100	100	100

(a) 12 microsatellite markers (b) 10 microsatellite markers
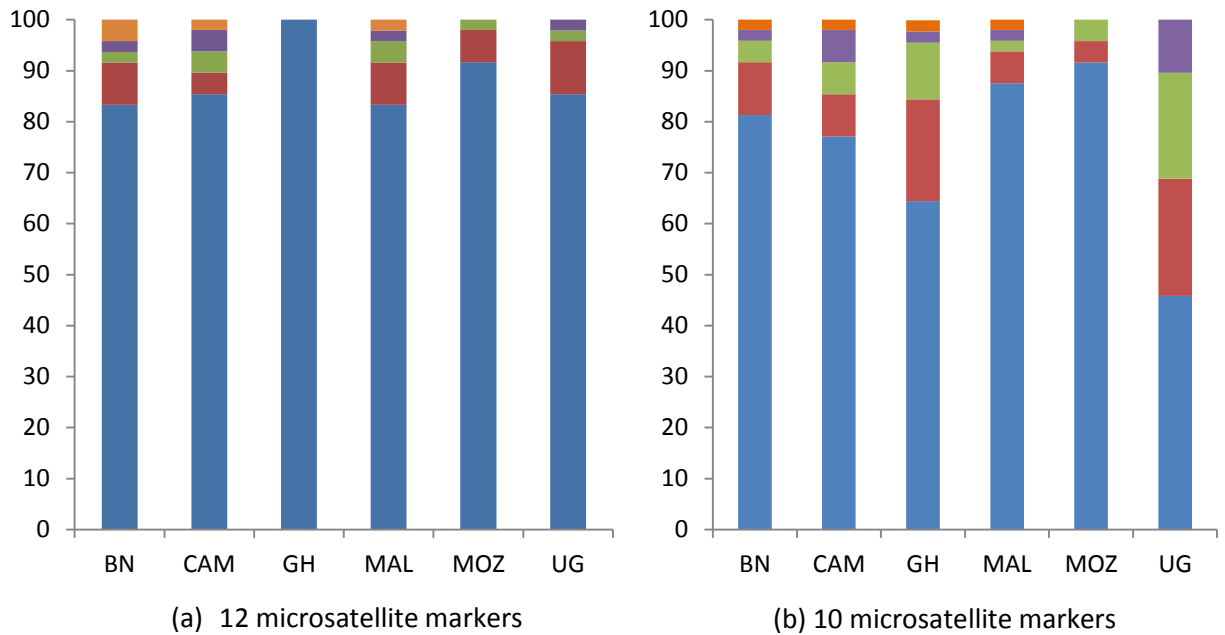
Fig. 3.9: Assignment of individuals from different sample sets to populations of other sample sets. Blue indicates the proportion of the correctly assigned sample to the location of origin and the other colours indicate other allocation of samples to the other locations (Red-second highest, green-third highest, purple-fourth and orange-fifth as seen in table 3.7a and 3.7b).

3.3.9 Neighbour joining tree

Neighbour joining tree (Fig. 3.10a) was constructed using the F_{st} values in order to determine the evolutionary distance between the different populations. The samples from Malawi and Mozambique showed a group of their own for the 12 markers, which is similar to results from STRUCTURE, while the samples from Benin and Ghana despite being from same geographical region did not form a common group. The other samples (Benin, Ghana, Cameroon and Uganda) break down. Similar pattern was also seen (Fig. 3.10b) when the two markers FUNR and AFUB6 were pulled out from the analysis allowing 10markers, the other populations still break down.
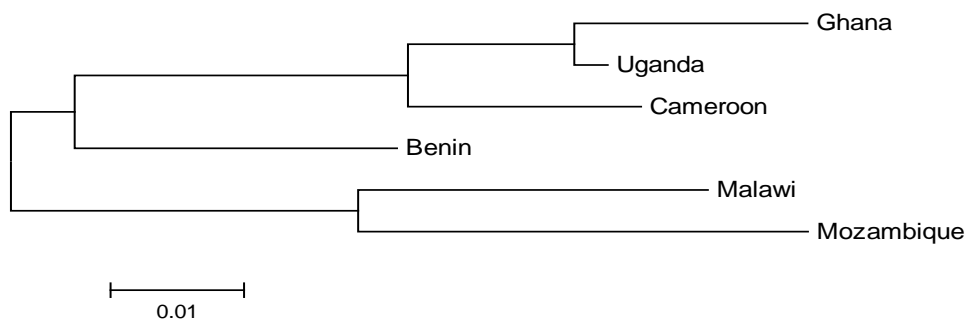
Fig. 3.10(a): Neighbour joining phylogenetic distance tree of F_{st} values of the samples (12 markers)

125
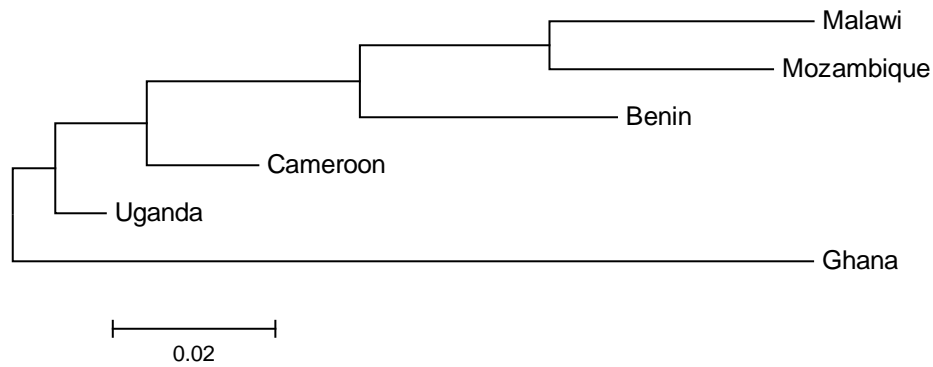
Fig. 3.10(b): Neighbour joining phylogenetic distance tree of F_{st} values of the samples (10 markers)

3.3.10 Isolation by distance

There was no significant correlation between the genetic (pairwise F_{st} /(1- F_{st})) and geographical distance (km) (3.8) when considering all 12 markers r=0.0318 and P=0.4270, Likewise for 10 markers (r=0.3545 and p=0.1150) with the exclusion of FUNR and AFUB6, there was also no significant correlation (at P<*0.05*). However, because the limited number of pairwise comparisons (N=15) these analyses have a limited power and further assessment conducted using higher number will be more informative.

Table 3.8: Genetic differentiation (*Fst*) and geographical distance between pair wise populations

	Benin	Cameroon	Ghana	Malawi	Mozambique	Uganda
Benin	-	1297km	431.5km	4368km	4789km	3602km
Cameroon	0.0677	-	1722km	3636km	4286km	2455Km
Ghana	0.1745	0.1317	-	4716km	5078km	4026km
Malawi	0.0784	0.0969	0.206	-	961km	1861km
Mozambique	0.0922	0.0921	0.182	0.06	-	2806km
Uganda	0.078	0.0342	0.1072	0.0928	0.1045	-

F_{st} is below the diagonal, distance in kilometres (km) is above the diagonal.
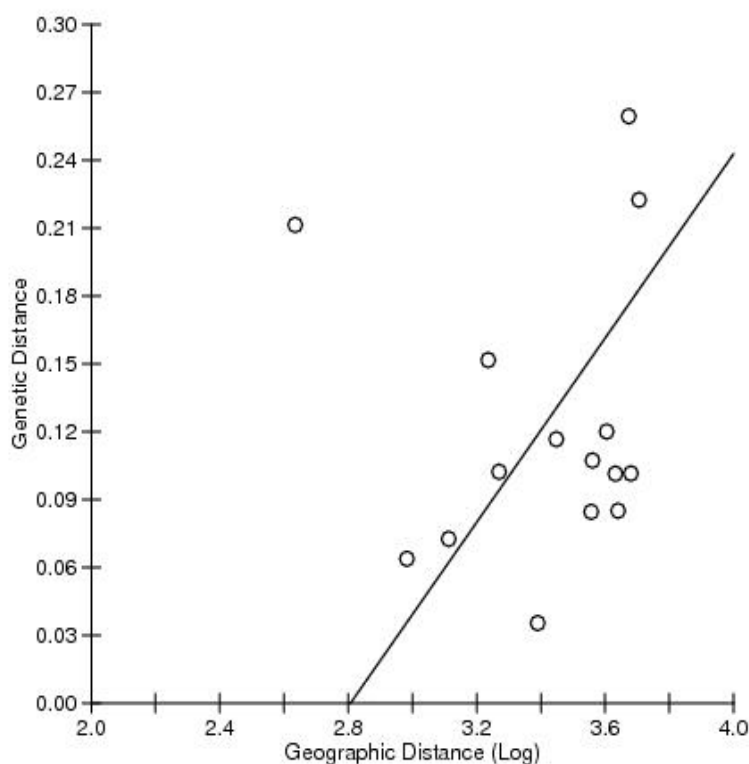
Fig. 3.11: Correlation between the $F_{st}/(1-F_{st})$ and logarithm of distance (in kilometers) for pair wise comparison of 6 African countries at 12 analyzed microsatellite loci. (X=$F_{st}/(1-F_{st})$) and Y-Log distance (km)

3.4 Discussion

The population structure of *An. funestus* in Africa was studied using 12 microsatellite loci spread throughout the genome (Fig. 3.3). Eleven of the 12 microsatellite markers amplified for all the 285 individual mosquitoes with an indication of selective amplification for FUNL for Mozambique which lead to difficulty in scoring and as a result was not included in the analysis.

3.4.1 *An. funestus* populations exhibit some genetic diversity differences

The markers were not physically linked to each other and each of the markers was polymorphic. There have been consistent reports of the presence of strong differences in the profiles of insecticide resistance between *An. funestus* populations across Africa particularly between different geographical regions of East, southern and West/Central Africa (Wondji *et al.* 2011, Dia *et al.* 2013). This observation could reflect the presence

of barriers to gene flow between populations of this species Africa-wide leading us to investigate the genetic structure of populations of this major malaria vector Across Africa. There was an overall high level of polymorphism in all the samples (Table 3.2). Study carried out by Michel *et al.* (2005) using most of these makers also showed a high level of polymorphism. Average allelic richness was seen to be higher for the samples from Benin, Ghana, Cameroon and Uganda than for samples from Malawi and Mozambique. Even though some of the markers were located close to chromosomal inversions as seen in Fig. 2.3, these inversions are not known to have any impact on genetic structure patterns as shown by Ayala *et al.* (2011) who did not observe any significant difference between microsatellite loci lying within and outside polymorphic chromosomal inversions. Additionally, the decreased allelic richness across the populations on the 2R is unlikely to have been influenced by inversions as the markers used are outside inversions that have been reported to have any impacts on the population like the 2Rh (Dia *et al.* 2013). Impact of chromosomal inversions on patterns of genetic structure has extensively been shown in *An. gambiae* between the former M (now *An. coluzzii*) and S (now *An. gambiae*) molecular forms with and increased genetic differentiation levels generally observed in markers located within inversion than in those outside (Lehman *et al.* 2000). In *D. melanogaster*, it has also been shown that chromosomal inversion can impact patterns of genetic differentiation due to reduced recombination within these regions (Russell *et al.* 2013). However, because no chromosomal karyotyping was performed on the samples used in the present study in *An. funestus*, one cannot completely rule out that contrasting chromosomal inversions arrangements between the different populations is not one of the reason why there is a reduced allele richness in southern Africa although the proximity of the major pyrethroid resistance rp1 suggests that pyrethroid resistance could be the major underlying cause.

Observed heterozygosity values were relatively high for some of the loci (Table 3.2) than the expected. Using 10 out of the 12 microsatellite markers that were used in this study, Michel *et al.* (2005) also reported higher observed heterozygosity at most loci. Some of the microsatellite loci showed heterozygote deficiency which was very pronounced for two markers FUNR and AFUB6 for samples from Malawi and Mozambique. Such Heterozygotes deficits are rather more common for chromosomal inversion markers. For example, significant heterozygote deficits were observed in *An. funestus* population from Burkina Faso with significant LD among some chromosomal rearrangements leading to

the identification of two chromosomal forms By Costantini *et al.* (1999) named Kiribina and Folonzo, which was in parallelism with the *An. gambiae* chromosomal forms from Mali (Coluzzi *et al.* 1985, Coluzzi *et al.* 2002). Heterozygote deficits have also been detected in some areas of Cameroon and Senegal and were detected mostly in the inversions of the 3R and 3L arm (Dia *et al.* 2000, Cohuet *et al.* 2005, Ayala *et al.* 2011, Dia *et al.* 2000). The heterozygote deficit detected in this study for microstellite markers is possibly not as a result of presence of inversions on the 2R but possibly as a result of selection acting on this chromosomal arm since the RP1 QTL that has been linked to insecticide resistance is located on this chromosome (Wondji *et al.* 2007). This can also be supported by the fact that the heterozygote deficit detected in Cameroon and Senegal based on inversions was not observed when using microsatellite markers (Ayala *et al.* 2011, Cohuet *et al.* 2004) showing that the extensive difference at FUNR and AFUB6 is not influenced by difference in chromosomal inversion arrangements. This deficiency in heterozygote is consistent with other factors like null alleles that can cause true heterozygotes to be scored as homozygotes (Michel *et al.* 2005). However in this study, it might not be null-alleles since it was not observed throughout the populations but more so for populations with very high reports of insecticide resistance like Malawi and Mozambique. Average heterozygosity (Hobs) and allelic richness was lower for samples from Malawi and Mozambique from southern Africa than for samples from the four other countries. The striking reduction in diversity for FUNR and AFUB6, for the southern populations for both loci may reflect selection acting on these loci because both loci are found around the RP1 QTL, where the genes (*CYP6P9a* and *CYP6P9b*) associated to insecticide resistance are found (Wondji *et al.* 2007). Michel *et al.* (2005) carried out a similar study using the same microsatellite markers, where they found reduced heterozygosity and allelic richness were lower for the Eastern cluster than for West and Central Africa. Their results were in line with previous studies carried out by Braginets *et al.* (2003) from corresponding geographic region. Deviation from HWE measured by inbreeding coefficient F_{is} were detected across most of the markers except for FUNQ, FUNL and AFUB10. Departure from HWE through heterozygote deficit may be caused by a Wahlund effect (pooling of separate gene pools) (Wahlund 1928), presence of null alleles at specific loci (Dakin and Avise 2004, Jarne and Lagoda 1996), inbreeding or selection (Jimenez *et al.* 1994, Chen 1993). Deviations from HWE are common in population genetics studies using microsatellite markers in *An. funestus* (Michel *et al.* 2005, Cohuet *et al.* 2004) and other mosquitoes like *An. gambiae* (Lanzaro *et al.* 1998).

There was no evidence of LD (Table 3.5) which suggests that the samples were drawn from homogeneous and randomly mating populations as there is an even distribution of alleles for each marker within each sample which is contrary to a structured sample made of mixed populations with different gene pools. No evidence of LD between samples from East, Central and southern African countries were also found by Temu *et al.* (2004). It is therefore likely that the deviation from HWE observed in this study is mainly due to locus specific constraints such as selection or a result of null alleles and not due to a Wahlund effect or inbreeding.

3.4.2 Gene flow among *An. funestus* populations is restricted between geographical African regions

Pairwise genetic differentiation between samples showed that samples from Malawi and Mozambique showed an overall reduced genetic differentiation while a higher genetic differentiation is observed between the samples from Benin and Ghana. Higher levels of genetic differentiation were observed when comparing populations from different regions such as southern and western Africa while the population from Uganda in East Africa is closer to West/Central Africa although also close to southern populations suggesting it is intermediate. Another study carried out by Temu *et al.* (2004) showed results of discontinuous population structure between southern African and East African populations with *An. funestus* samples from Uganda, Kenya, Malawi, Mozambique and South Africa.

The results shown by STRUCTURE also distinguished the southern populations from the rest of the other samples where they consistently formed a cluster together for both the 12 and 10 markers. The clustering seen in the Bayesian results bears a close similarity to the population substructure of *An. gambiae* where a continental split has been found between populations of the East and the Northwest (Lehmann *et al*. 2003). This can explain why samples from East Africa (Uganda) were more closely related to samples from West Africa (Benin). A similar population structuring was observed in Garros *et al.* (2004) using Restriction Fragment Length Polymorphisms (RFLP) typing of *An. funestus* domain III, where two genotypes were observed with one occurring in West African samples and the other in East African and Malagasy samples, and both were found to occur in the southern African samples.

Overall, the above study showed restriction to gene flow across African *An. funestus* population since the southern population where unique in their genetic diversity to the other populations, which may be an indication that the dispersal of *An. funestus* is not as wide spread as earlier studies reported (Gillies and De Meillon 1968). Based on chromosomal inversion polymorphisms, there is evidence of lack of gene flow between sympatric populations of *An. funestus* (Costantini *et al*. 1999, Michel *et al*. 2005). Genetic distance was reported to limit gene flow among populations by Temu *et al*. (2004) and also seen to promote genetic differentiation among populations.

3.4.3 Selection due to insecticide resistance on 2R could be influencing patterns of genetic diversity among other factors

Insecticide resistance pattern in *An. funestus* across Africa mirrors the population genetic structure that is seen in the microsatellite and STRUCTURE analysis, where the southern population turns to form a separate group. The southern populations of *An. funestus* present the same resistance pattern with higher reports of resistance to various insecticides that is different from the other African populations. Previous reports have shown that *An. funestus* in Malawi and Mozambique were resistant to pyrethroids with evidence of carbamate resistance (Hargreaves *et al*. 2000, Brooke *et al*. 2001, Casimiro *et al*. 2006, Wondji *et al*. 2012). There are also high reports of DDT resistance in West/Central and East Africa for *An. funestus* but until present, there are no reports of DDT and organophosphate resistance in southern Africa (Djouaka *et al*. 2011, Cuamba *et al*. 2010, Hunt *et al*. 2010). Thus selection for resistance correlates with the fact that the southern populations being genetically different to the other populations as also indicated by study carried out by Temu *et al*. (2004). However, difference in profiles of resistance to insecticide could actually reflect an already existing barrier to gene flow between these regions and could have taken advantage of such restriction to further act on certain regions of the genome to increase differentiation such as on rp1on the 2R chromosome. Geographical proximity might be another reason why the southern population cluster together and forming one large population even though this does not apply for the West African populations of Benin and Ghana that are also geographical close to each other. Although no karyotype information exist for Benin and Ghana, it cannot be ruled out that the higher differentiation observed between them in contrast to southern populations is impacted by chromosomal inversions. This should be further assessed in future studies.

The Rift Valley has also been previously shown to be acting as a barrier to gene flow for *An. gambiae* (Lehmann *et al.* 1999, 2000) and a possible contributor to its population subdivision (Lehmann *et al.* 2003). Braginet *et al*. (2003) also suggested the same effect of the Rift Valley in *An. funestus* populations across Kenya where an important sub-division due to the Rift valley was detected. This study suggests the possibility that the rift valley could be acting as a barrier to gene flow between *An. funestus* in Africa as results from STRUCTURE indicate that the Uganda (western side of Rift Valley in East Africa) samples had very few of its individuals allocated to the southern populations of Malawi and Mozambique (located on the eastern side of the Rift Valley) but more of its individual to Benin (West Africa) and Cameroon (Central Africa). However the possible effect of inversions cannot be ruled out. Similarly, assignment tests also show that more samples from Uganda are assigned to Central Africa (Cameroon) than to West Africa samples (Benin and Ghana). However, the impact of the rift valley could not be properly tested in this study since no clear correlation could be established since no direct comparison between samples from the East and the West side of the rift valley was performed in the same country to minimise other confounding factors. Such study should be performed in the future in a larger scale to assess the influence of the rift Valley and inversions among other factors in shaping the Africa-wide patterns of gene flow among *An. funestus* populations.

Test of isolation by distance showed a significant correlation between differentiation by genetic distance (pairwise F_{st}) and geographic distance (Km) for the 10 markers and only differentiation by geographic distance (km) and no significant correlation by genetic distance (pairwise F_{st}) for the 12 markers. This can possibly be as a result of the fact that, isolation by distance is due to the whole loci set (Cohuet *et al.* 2004). However, the IBDs results is not very informative since there were only very few pairwise comparisons and future studies on this topic should increase this number of comparisons so as to obtain a more informative result.

Conclusion

The findings in this study requires further investigation with deeper sampling and more analysis of genetic loci to thoroughly elucidate the forces that shape and maintain the population structure of *An. funestus* across Africa. More studies are required to characterize the three clusters seen from the STRUCTURE results regarding their

ecological and malaria susceptibility and also detail cytogenetic study of the populations should be investigated.

In this study, the results have suggested that, the population structure of *An. funestus* is structured such that the spread of genes responsible for insecticide resistance would be easier within than between genetic units. The effects of FUNR and AFUB6 located on the 2R chromosome turn to skew up the results. This led to further microsatellite genotyping of makers on the 2R chromosome in order to assess the signature of selective sweep around the rp1 QTL on the 2R chromosome in chapter four.

CHAPTER FOUR

Assessing signature of selective sweep around the pyrethroid resistance loci on the 2R Chromosome

4.1 Introduction

Insecticide resistance offers one of the best opportunities to study microevolution processes notably the impact of selection on natural populations (Silva *et al.* 2012, ffrench-Constant *et al.* 2004). The amount of divergence between species is proportional to DNA variations which are as a result of neutral mutation within that species. Differentiation of intra-specific variation is as a result of non-neutral mutations. Under negative or positive selection pressure, non neutral mutations either increase in frequency within populations or are eliminated from a gene pool. This process is known as selective sweeps or background selections (Smith and Haigh1974).

Mutations that confer insecticide resistance spread quickly and have a selective advantage over those that do not confer resistance. There is reduced genetic variability at locus under selection and in the flanking regions surrounding it (Hitchhike with it) as this beneficial mutation spreads (Smith and Haigh 1974). It is possible to identify regions that have undergone selective sweep by detecting the regions with reduced variability (Kim and Stephan, 2002).

In *An. funestus*, the identification of the pyrethroid resistance QTL (rp1) on the 2R chromosome in FUMOZ-R laboratory strain (Wondji *et al.* 2009) has been confirmed in field populations in Mozambique and Malawi (chapter 3; Riveron *et al.* 2013). The main pyrethroid resistance genes *CYP6P9a* and *CYP6P9b* are found within this rp1QTL (Wondji *et al.* 2009) and these genes have been the most associated to insecticide resistance in Malawi (Wondji *et al.* 2012, Riveron *et al.* 2013). Additionally, two microsatellite loci, the FUNR and AFUB6, showed much reduced genetic variability in chapter three, which is a possible indication that these loci might be under selection due to LD with resistance loci. A typical characteristic of selective sweep is the high levels of LD that decreases with distance away from the selected site. A narrow selective sweep is produced when a weak selection pressure acts on a region with high recombination and vice versa (Kim and Stephan 2002).

Selective sweep is linked with reduced genetic variability at a locus under selection. However, diversity can be affected by both selective and demographic events.

Demographic events reduce whole genome wide diversity with events such as bottlenecks while selection leads to allele fixation at particular regions in the genome.

Genetic markers such as Single Nucleotide Polymorphisms (SNP), microsatellite and sequence data can all be used for the detection of selective sweep. Microsatellites occur less frequently than the SNPs but their high rates of mutations make them more informative than the SNPs, while the higher density of SNP offers the possibility of a fine-scale analysis.

The rate and strength of a mutation can affect the size of the genomic region which is under selection and using such variation in pattern can be a very good example in identifying genomic regions under selection (Sabeti *et al*. 2002, Vigouroux *et al*. 2002, Schlötterer 2003). Change in phenotype has been previously linked to the change in the genomic region of the candidate gene. Some of these genes were for example insecticide resistance genes in mosquitoes (Yan *et al.* 1998) and in the Malaria parasite *Plasmodium falciparum* (Wootton *et al.* 2002, Nair *et al*. 2003). There is occurrence of sequential mutations in *Drosophila simulans* for DDT resistance with *CYP6G1* where strong selective sweep was associated to a transposon insertion (Schlenke and Begun 2004).

Different tests are required to analyse microsatellite data since different alleles exist at one locus, and determining the microsatellite ancestral state is very difficult. It has been shown by Wiehe *et al.* (1998) that hitchhiking effect affects microsatellite variation even though the effect is time dependent on the fixation of the favourable linked mutation since hitchhiking is quickly obscured by high mutation rates. Thus microsatellites show only the effects of selective sweep that falls within a certain time frame dependent on the recombination and mutation rate. In a population affected by selective sweep event, the effect seen on the microsatellite locus is that there is reduced variability. To account for the neutral variability on the locus, demographic factors should be taken into account since there is strong difference in mutation rates between loci (Nielsen 2001), although it is difficult to infer neutral assumption for each microsatellite. A statistical test that is independent of locus specific or demographic effects was developed by Schlötterer (2002b). This test relies solely on comparing the gene diversities of different populations that might have been affected by different selection pressures. LnRV and LnRH statistics compares pairwise variance in repeat units (V) or heterozygosities (H) at multiple loci. There is the assumption that, loci that are found at same genomic location, are presumed

to have evolved under same constraints and mechanisms. Pairwise comparison tends to eliminate locus specific effects such as variable mutation rates. Loci of one population can be affected by inclusion of multiple loci that controls demographic effects.

There are other tests such as the Lewontin-Krakauer test which uses genetic distance estimators like the F_{st} values to identify loci that are deviating extremely from the other loci (Baer 1999). Many false positive sweeps might be caused when there is substantially very high variance of F_{st} values per locus that is under neutrality if this method is solely used (Schlötterer 2003). Deviation from neutrality can be inferred from the observed and expected allele frequency distribution within populations as seen in the case of the study using microsatellite allele distribution on human by Payseur *et al*. (2002).

The results from chapter three indicated a significant reduction of heterozygosity for FUNR and AFUB6 markers, which are located on the 2R chromosome close to the *CYP6P9a* and *CYP6P9b* resistance genes. Since reduced heterozygosity is an indication of selection, further analysis was needed to assess if there is the presence of signature of selective sweep around this region. To investigate the possible presence of selective sweep on the 2R chromosome, five additional microsatellite markers (Table 4.1 and Fig. 4.1) on the 2R chromosome were genotyped and analysed in addition to the 3 markers (AFUB6, FUNR and FUNO) of the 2R that were used previously in chapter three.

Overall the hypothesis of this study is that in areas of high insecticide resistance, resistant genes are most likely subjected to strong selection pressures resulting in reduced genetic diversity in the genomic regions around them.

4.1.1 Objective

The objective of this chapter is to investigate the presence of signature of selective sweep around the pyrethroid resistance loci on the2R chromosome using microsatellite loci on the 2R markers.

4.2 Materials and method

Samples used for microsatellite genotyping of 2R chromosome markers for assessing the signature of selective sweep around the rp1 QTL are the same samples as those used in chapter three for the genetic structure analysis.

The collection of samples, general laboratory techniques such as DNA extraction and Species ID are the same as already described in chapter two for molecular characterisation and chapter three for microsatellite genotyping. All the analysis except statistical analysis is same as already described in chapter 3.

Table 4.1: Primers used for microsatellite genotyping including size ranges (Sharakhov *et al.* 2002, Cohuet *et al. 2002,* Sinkins *et al.* 2000)

Primer name	Chromosome	Left primer	Right Primer	Label	Size Range	Multiplex with
AFND40	2R 9B	GTAGTCGACAATCCGT ACGATTCATCCTGTGAT GCTTTG	AGGCTCTTCT TTGCACTGT	D3	188-200	AFND32, AFND30
AFUB3	2R 8B	CACGACGTTGTAGAAC GACGGGAAGGATTCGA CCTTAGC	GCCGCCATTT AGTAGCAGTT	D2	164-182	AFND6
AFND32	2R 15E	GTAGTCGACAATCCGT ACGGTTCTCCATCGCTG TTCTACTC	TTTTCGTACG GAGAAAAAT G	D4	103-121	AFND30, AFND40
AFND6	2R 12E	GCTTCTTCTCCCCTAAT CTG	TCCTGCTTTT TAGTTTGTCG	D3	99-101	AFUB3
AFND30	2R 13C	GTAGTCGACAATCCGT ACGGTTCAGCTGTTGGT GTGTTAG	TTTTCGTACG GAGAAAAAT G	D3	81-107	AFND32, AFND40

CYP6P9a and *CYP6P9b* genes
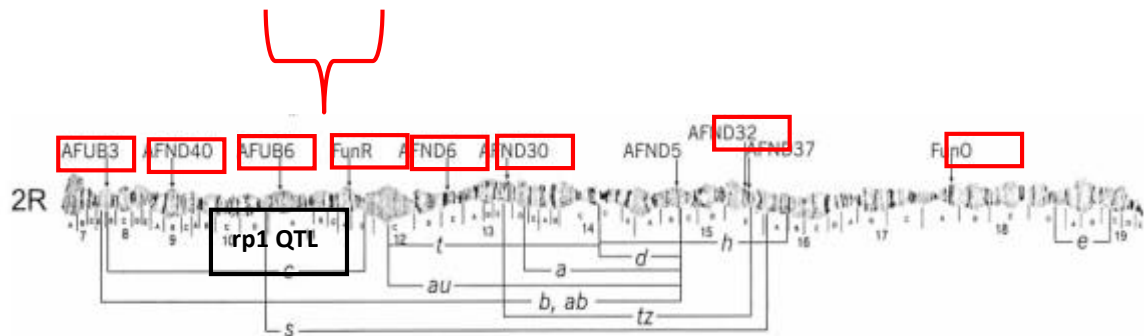
Fig. 4.1: Distribution of microsatellite loci (red boxes) across 2R and inversions (brackets) with respect to *An. funestus* polytene chromosomes photomap, the black box indicates the position of rp1 QTL and red inverted braces indicating location of *CYP6P9a* and *CYP6P9b* genes (from Sharakhov *et al.* 2002, Wondji *et al.* 2007).

4.2.1 Statistical analysis of the selective sweep

To identify loci under selection pressure, the LnRH statistics according to Schlötterer (2002b) and Kauer *et al.* (2003) were estimated. The LnRV was not calculated to also identify loci under selection pressure, because previous studies (Schlötterer 2002, Kauer

137

et al. 2003, Tescheke *et al*. 2008), have shown that there is no significance difference in results seen between LnRV and LnRH. The LnRH approach is based on the following equation:

$$LnRH = Ln \frac{\left(\frac{1}{1-H(loc1,pop1)}\right)^2 - 1}{\left(\frac{1}{1-H(loc1,pop2)}\right)^2 - 1}$$

$H = heterozygosity$

Based on the estimator:

$H = 1 - (1/(1 + 2\theta)^{1/2})$

(Ohta & Kimura 1973)

Kauer *et al.* (2003) was used to calculate LnRH statistics where there was comparison of variability levels between the various pairwise samples (e.g Benin vs Malawi) as well as population wise comparison West Africa population (Benin and Ghana) vs southern Africa (Malawi and Mozambique) population. In some cases, the significance was calculated by estimating the normalized LnRH values with reference sample which was the more susceptible samples (Cameroon and Uganda) and also to the West African samples.

The LnRH values were calculated by comparing all the single samples against each other which resulted in 15 different comparisons (pairwise comparison). In addition to that, there was comparison by geographical group of populations West African samples vs Central Africa (Cameroon), West vs southern and southern vs Central. The group population comparison was done so as to establish if the selection is more geographical or not.

4.2.2 Microsatellite data analysis

Microsatellite loci were analyzed as previously described in chapter three. The 5 markers on the 2R chromosome (Table 4.1) were genotyped and pooled together with three 2R markers (AFUB6, FUNR and AFUB6) that were used previously in chapter three. Analysis was then carried out with the eight markers found on the 2R. As a genomic control, further analysis were carried out with the 9 markers (AFND12, FUNQ, FUNL, AFUB10, AFUB11, AFND7, AFND19, FUNF AND AFUB12) these are markers found outside the 2R chromosome in chapter three. Furthermore, analysis was carried out with all the 17 markers (AFND40, AFUB3, AFND32, AFND6, AFND30, AFND12, FUNQ, FUNO, FUNR, AFUB6, FUNL, AFUB10, AFUB11, AFND7 AND AFND19, FUNF AND AFUB12) consisting of 5 new 2R markers and the 12 initial markers used in chapter

three. These comparisons were done to further confirm the presence of signature of selective sweep around the 2R chromosome.

4.3 Results

4.3.1 Quality control and Micro-checker application

The six samples were all genotyped and scored as described in chapter three. Micro-checker application confirmed no scoring errors due to stuttering for all five microsatellite loci, implying that the visual scoring was good for markers across all samples.

4.3.2 Genetic diversity

Moderate to high levels of polymorphism were observed in all the samples from the six African populations of *An. funestus* (Table 4.2). Samples from all six countries were successfully genotyped for all the 5 microsatellite loci and analysis was carried out for the 8 2R microsatellite loci. All 8 loci were polymorphic across all the samples with allele sizes ranging from 2-12 alleles (Fig. 4.2). Most of the alleles were distributed across all the samples and there was an indication of an allele belonging to a single population (private allele) for AFND30 found in Benin and AFND6 found in Uganda. Overall, the observed heterozygosity (Hobs with average mean=0.53) was lower than expected heterozygosity (Hexp with average mean=0.69). For most of the samples and markers, predominant alleles were almost the same. When all the alleles were pooled together (Fig. 4.3a), results from the 8 markers showed much reduced heterozygosity and also reduced allele numbers for three markers (AFND40 mean average allele size of 5.5, AFUB6 mean average allele size of 3.2 and FUNR mean average allele size of 4.5) which are markers closer to the *CYP6P9a* and *CYP6P9b* candidate genes located on the rp1 QTL (Wondji *et al.* 2009) for all the samples even though much more reduced for Malawi and Mozambique as already mentioned in chapter three. The other markers showed higher allele sizes and higher genetic diversity across the populations.

In total, the average number of alleles per locus was 6.4 ranging from 2 (FUNR) to 12 (AFUB3) and the average heterozygosity of 0.52 (range 0.12-0.93). The West African populations showed an average of 6.8 (range 3(AFUB6)-9(AFND32)) alleles per locus with average heterozygosity of 0.55 (range 0.18(AFUB6)-0.70(AFND40)). The southern populations had an average of 5.8 alleles (range 3(AFUB6)-9(AFND32)), an average heterozygosity of 0.39 (range 0.12(AFUB)-0.65(AFND32)). The Central African

samples showed an average of 6.4 alleles (range 6(AFUB6)-10(AFUB3)) with average heterozygosity 0.64 (range 0.42 (AFND6) 0.73 (FUNO)), and the East African samples had an average of 6.8 alleles (2-(AFUB6)-12(AFUB3)) with average heterozygosity of 0.63 (range 0.54 (AFND6)-0.85 (FUNO)). The alleles were distributed across the various samples for the different markers as seen in Fig. 4.2
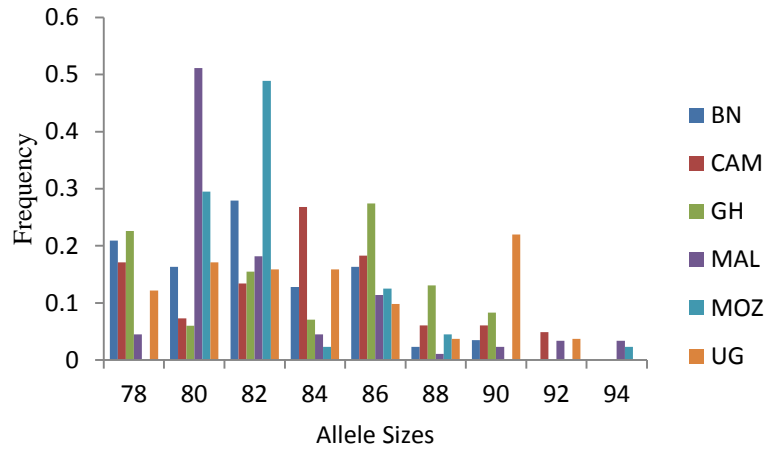
(a)   AFND40
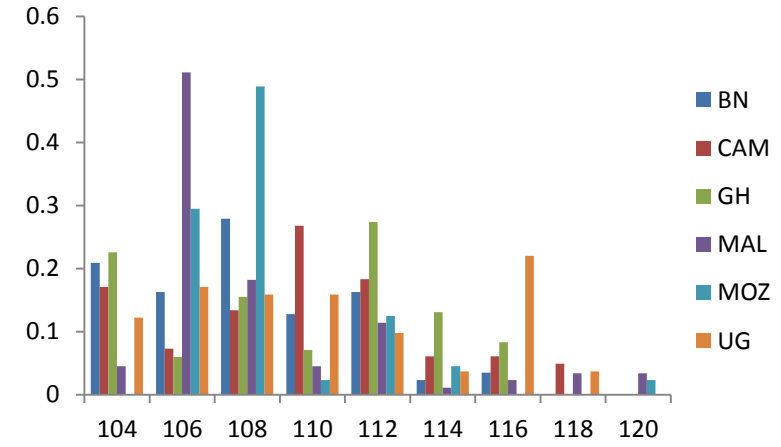
(b) AFUB3
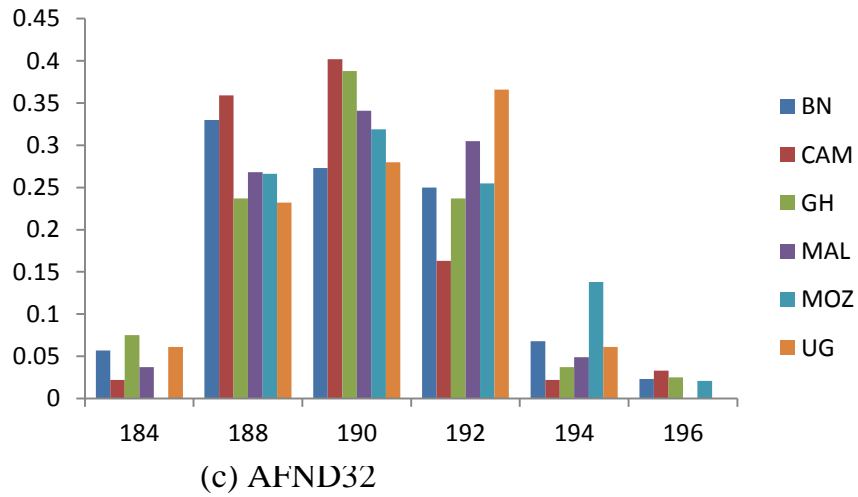
(c) AFND32
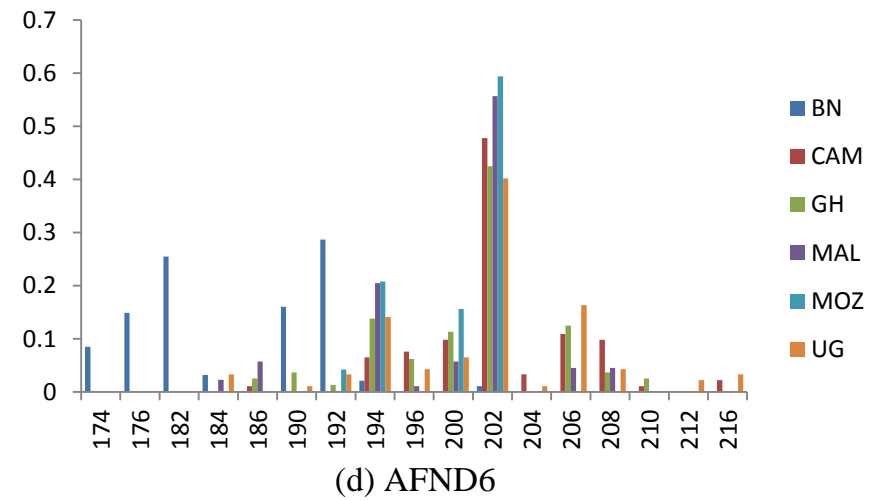
(d) AFND6
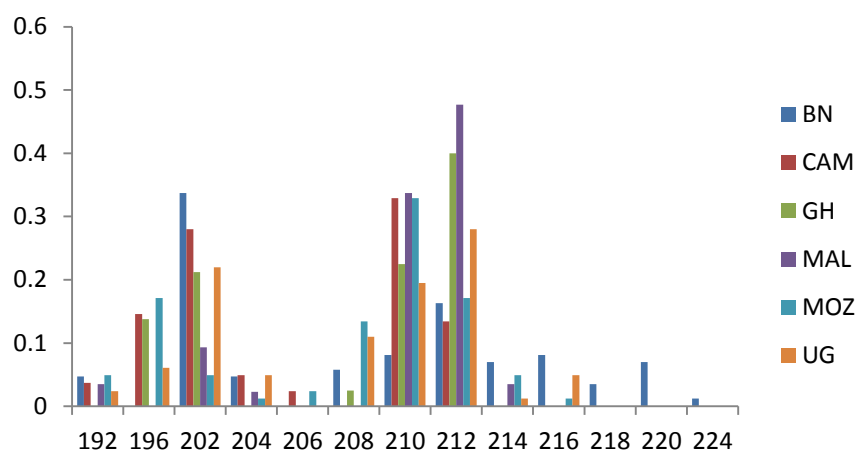
(e) AFND 30

Fig. 4.2: Allele frequency distribution in each sample of *An. funestus* for each microsatellite locus. Alleles are denoted by their estimated sizes (bp). (BN (Benin), CAM (Cameroon), GH (Ghana), MAL (Malawi), MOZ (Mozambique, UG (Uganda) – (a-e)

Table 4.2: Test for Hardy-Weinberg Equilibrium

		BN N=48	CAM N=48	GH N=45	MAL N=48	MOZ N=48	UG N=48	Mean of All
AFND40	All Hobs Hexp *Fis*	6 0.70 0.75 -0.0242	6 0.58 0.67 0.0117	6 0.6 0.66 0.0871	5 0.38 0.62 **0.3962**	5 0.40 0.74 **0.4641**	5 0.60 0.63 0.0385	5.5 0.54 0.68 0.1798
AFUB3	All Hobs Hexp *Fis*	8 0.93 0.79 -0.1927	10 0.67 0.71 0.0560	10 0.58 0.69 0.0964	8 0.46 0.59 **0.2259**	4 0.40 0.58 **0.3245**	12 0.58 0.76 **0.1492**	8.7 0.60 0.69 0.1233
AFND32	All Hobs Hexp *Fis*	7 0.58 0.73 0.0204	8 0.73 0.72 -0.0134	7 0.47 0.77 **0.3983**	9 0.60 0.64 0.0499	6 0.62 0.61 -0.0291	8 0.62 0.73 -0.1465	7.5 0.60 0.7 0.1261
AFND6	All Hobs Hexp *-Fis*	11 0.58 0.75 -0.2245	7 0.42 0.67 **0.3765**	5 0.36 0.65 **0.4581**	6 0.33 0.59 **-0.4348**	10 0.25 0.70 **-0.6447**	9 0.54 0.70 -0.2325	8 0.41 0.68 0.3952
AFND30	All Hobs Hexp *Fis*	9 0.4 0.77 **0.4916**	8 0.48 0.74 **0.3578**	8 0.67 0.75 -0.1076	8 0.65 0.72 0.0100	4 0.50 0.62 -0.1950	8 0.31 0.71 **0.5651**	7.5 0.50 0.72 0.3029
Combined for all loci	All Hobs Hexp Fis	8.2 0.63 0.76 0.1408	7.8 0.58 0.70 0.1787	7.2 0.54 0.70 0.2295	7.2 0.48 0.63 0.2414	5.8 0.43 0.65 0.3198	8.4 0.53 0.70 0.2264	7.44 0.53 0.69 0.2255

Polymorphism at 5 microsatellite loci in *An. funestus* collected from six African countries (Nall–Number of alleles, Hobs -Observed heterozygosity, Hexp - Expected heterozygosity (Nei 1978) Significant after Bonferroni correction is represented by bold (P<0.05), bold and underlined (P<0.01) after accounting for multiple testing (Holm 1979). Deviation from Hardy-Weinberg is shown by F_{is} calculated according to Weir and Cockerham (1984). Mean of all F_{is} values is calculated by the sum of values for each locus divided by the total number of populations, Benin, Cam = Cameroon, Ghana, Mal = Malawi, Moz = Mozambique and Uganda.

Selective sweep around rp1 QTL
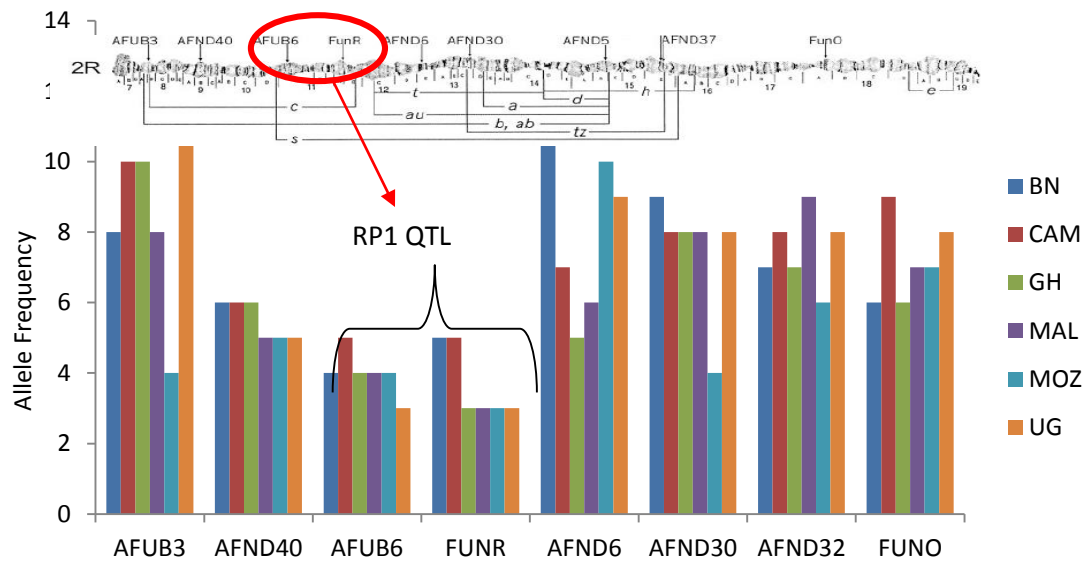
Fig. 4.3: Allele distribution for 2R markers showing reduced genetic diversity with low allele numbers around the rp1 QTL.
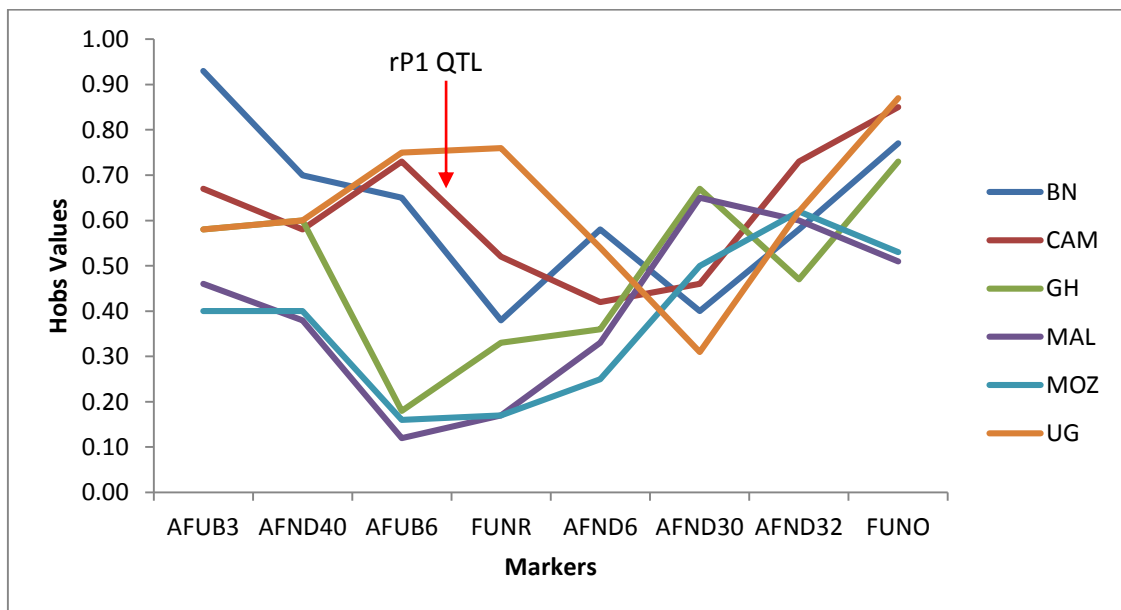
(a)
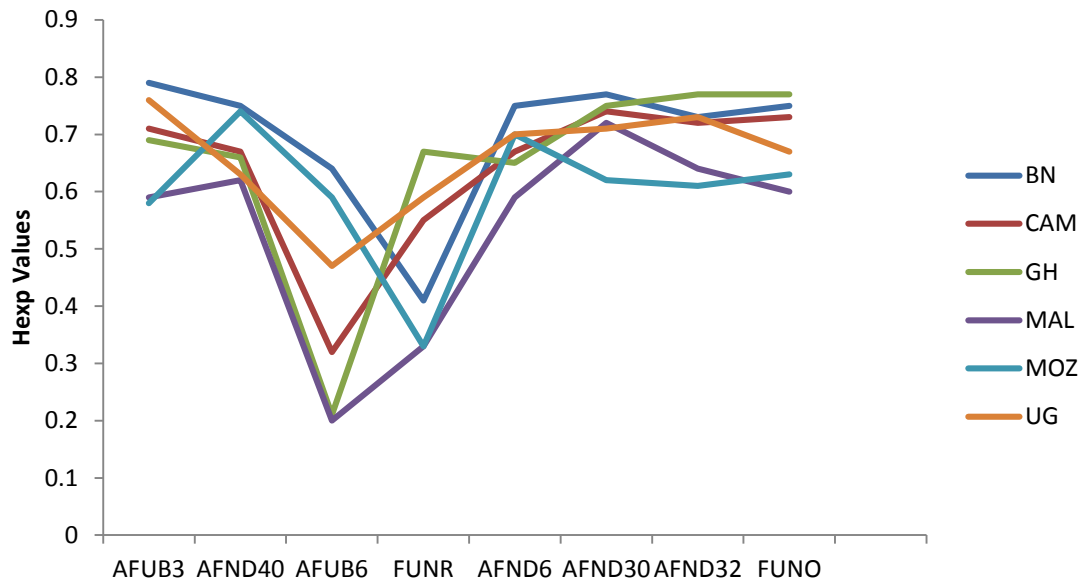
(b)

Fig. 4.4: Observed reduced heterozygosity (Hobs) (a) and expected heterozygosity (b) around the rp1 QTL for Malawi, Mozambique and Ghana and less for Uganda with lower resistance and a recovery away from the QTL

4.3.3 LnRH analysis

The LnRH calculations were done on different levels. There was a simple pairwise comparison which was comparing each sample against all others and there was also pooling of the *An. funestus* populations from Central, West and southern Africa and compared against each other. Table 4.3a and Table 4.3b show the summary of all the sample pairwise comparisons and also the pairwise population comparisons that were performed. The resulting distributions and characteristic such as the mean, standard deviation and the 95% and 5% percentiles were calculated (Table 4.3a and 4.3b).

The LnRH distribution resulted in a normal distribution except for the fact that there were extreme values at the edges of the distribution, which is shown in Fig. 4.5a and 4.5b. These extreme LnRH values correspond to the high differences in variability seen in two loci (AFUB6 (0.8376) and FUNR (0) due to equal expected heterozygosity values of 0.33), for most of the samples and especially for the southern populations. If these characteristics are applied, the AFUB6 and FUNR loci can be considered as potential sweep loci since they exhibit extreme values in Benin comparisons, with Malawi (-0.88285 and -0.16797) and Mozambique (-0.4524 and -0.16797), likewise, Cameroon with Malawi, and Ghana with Mozambique. This is also confirmed from their allele

distribution as seen in Fig. (4.3). AFND40 and AFND6 loci that are also located closer to AFUB6 and FUNR loci also showed extreme values but the values were not common across most of the comparison as seen for the AFUB6 and FUNR loci. The other markers showed similar variability for example AFUB3 (-0.08301) and FUNO (-0.08604).

The loci below the 95% and 5% threshold of the normal distribution can be considered as different because they all exceed the threshold which is a characteristic of a normal distribution. There was a different distribution of threshold for each pairwise comparison and population wise comparison, depending on which samples and populations that were compared. The population wise comparison did not show any significant difference between the populations (Table 4.3b)

The extreme differences observed between the two loci (FUNR and AFUB6) found around the rp1 QTL is probably the only reason for the extreme values seen around the edges. In any case, it is very difficult to set an overall threshold to identify a selective sweep locus as the general characteristic of the samples or populations influences the threshold values. In this study, the numbers of markers were few which make it very difficult to be conclusive from these results alone. This is because LnRH test normally should involve more markers (Schlötterer *et al.* 2002) for effect result interpretation. However only few microsatellite markers have been identified for *An. funestus* and more so on the 2R. A total of 10 markers were tested but only the 8 markers used in this study did work across all the samples. This can be illustrated by analysing the pairwise comparisons against the southern populations. The southern populations generally exhibited lower genetic diversity around the rp1 QTL. As a result, almost all the comparisons with the southern population (ML and MZ), are slightly skewed to more negative values. Therefore, the lower genetic diversity of the southern population should be taken into account. In a very conservative assumption, loci can be considered as potential sweep loci if they exhibit extreme values in at least three comparisons of the same population (Teschke *et al*. 2008).

Analysis carried out with the control (markers outside the 2R chromosome) showed a more narrow distribution (Fig. 4.5c). This is expected since these microsatellite loci have not been linked to insecticide resistance which is the suspected selection force. This distribution reveals an extreme difference between the 2R microsatellite loci and the loci outside the 2R in term of their distribution patterns. However, only 9 markers were used

as control (markers outside the 2R) which also is not large enough to provide a conclusive and definitive absence of a sweep on markers away from the 2R chromosome.

The LnRH calculation is less dependent on the distribution of the alleles but takes into consideration the alleles frequencies which will result in extreme values of LnRH for AFUB6 and FUNR (Ihle 2004). Other loci exhibit very high polymorphism but AFND40 which is closer to AFUB6, exhibits a lower polymorphism across all the samples even though not as obvious as for AFUB6 and FUNR. Overall, there is an exhibition of extreme LnRH in loci that exhibits lower number of alleles because higher heterozygosity differences is as a result of differences in allele distributions. This thus implies that, loci FUNR and AFUB6 and AFND40 exhibit the expected allele distribution for a possible selective sweep event with an extreme reduction of diversity in one population (southern) with substantial variations in West and Central populations.

The distribution of allele for a potential sweep in the southern population follows a possible selective sweep event expectation. AFUB6 and FUNR most especially was recognised by the LnRH statistics with extreme reduction of polymorphism within all populations but more so for the southern population and samples from Ghana.

Although the first indicator of a possible sweep event is extreme values of LnRH, there are other further methods of analysis of the allele distribution that could be carried out to confirm the selective sweep event (such as Kolmogorov LnRV). Since the sample sizes and number of markers used in this study were not large, as such other possible statistical tests could not be carried out.

In summary, analysis and inspections of the allele frequency distributions performed all point to the fact that the AFUB6 and FUNR loci are very likely bear the signature of a sweep event across all the populations (Fig. 4.5a and 4.5b). There is visible reduction in polymorphism in the region flanking the rp1QTL for AFUB6 and FUNR locus for all samples and more so for the southern populations for all the comparisons where there was a strong reduction in heterozygosity (Fig. 4.4a).The expected heterozygosity is much higher than the observed for the two loci (Fig. 4.4b). However, similar reduced polymorphism is also observed in the other populations although to a lesser degree than in southern Africa probably an indication that the rp1 also plays a role in the pyrethroid resistance which is observed in all regions at different degree. The use of more markers and large sample sizes to confirm these results will be helpful.

146

Table 4.3 (a) Pairwise sample comparison

	BN-CM	BN-GH	BN-ML	BN-MZ	BN-UG	CM-GH	CM-ML	CM-MZ	CM-UG	GH-ML	GH-MZ	GH-UG	ML-MZ	ML-UG	MZ-UG
N	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48
Mean of LnRH	-0.05	-0.08	-0.19	-0.09	-0.01	-0.03	-0.14	-0.31	0.04	-0.01	-0.01	0.07	0.1	0.18	0.07
Stdev	0.19	0.33	0.28	0.06	0.13	0.14	0.15	0.23	0.11	0.15	0.36	0.24	0.30	0.24	0.17
Lower 5% border	-0.42	-0.72	-0.74	-0.21	-0.26	-0.30	-0.43	-0.76	-0.18	-0.30	-0.71	-0.40	-0.49	-0.29	-0.26
Lower 95% border	0.32	0.57	0.36	0.03	0.24	0.24	0.15	0.14	0.26	0.28	0.69	0.54	0.69	0.65	0.40

Table 4.3(b) Combined pairwise population comparison

	West vs Central Africa	West vs southern Africa	Central vs southern Africa
N	96	96	96
Mean of LnRH	-0.02	0.01	0.07
Stdev	0.08	0.13	0.04
Lower 5% border	-0.18	-0.24	-0.008
Lower 95% border	0.14	0.26	0.15

Characteristics of all pairwise LnRH distributions (Table 4.3a) shows the result of comparison of the pairwise samples (Table 4.3b) shows the summaries of all combined pairwise population comparisons and the results of the pooled analysis for microsatellite loci.
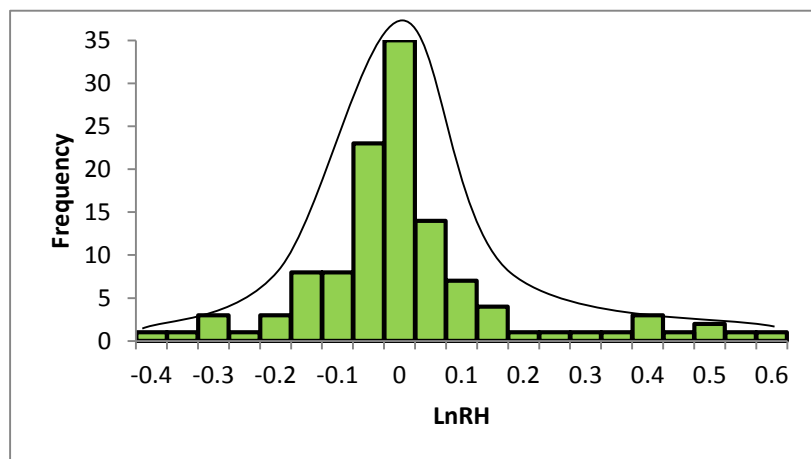
(a)
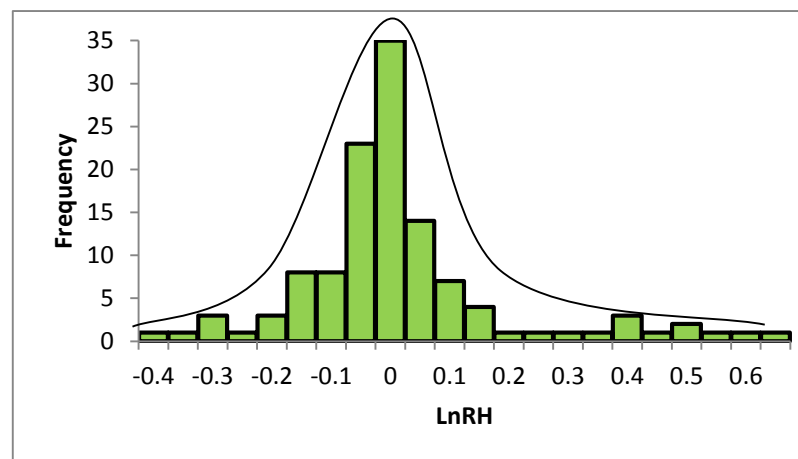
(b)
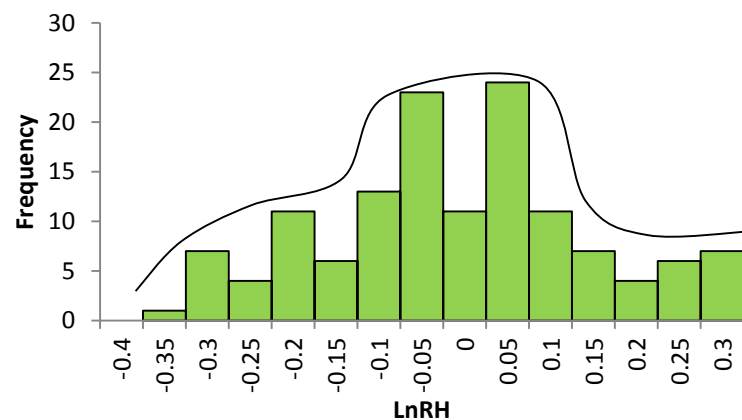
(c)

Fig. 4.5a -c: Distribution of LnRH values calculated for microsatellite markers. The distributions represent the summary of all pairwise (a) and population wise (b) – Central Africa vs southern Africa, southern Africa vs West Africa and West Africa vs Central Africa) and (c) control which are markers outside the 2R chromosomes.

Table 4.4a: Deviation from HWE for each marker at P<*0.05* and P<*0.01*

5 Markers	Before Bonferroni Correction	After Bonferroni Correction at P<*0.05*	After Bonferroni Correction at P<*0.01*
AFND40	3	2	2
AFUB3	4	3	3
AFND32	2	2	1
AFND6	5	4	4
AFND30	4	3	3

The number of markers sample sets with significant deviation from HWE for each marker before and after Bonferroni correction for 5 markers.

Table 4.4b: Deviation from HWE for each sample set

	Before Boferroni Correction	After Bonferroni Correction at P<*0.05*	After Bonferroni Correction at P<*0.01*
Benin	6	4	2
Cameroon	5	2	2
Ghana	4	3	2
Malawi	4	3	3
Mozambique	4	3	3
Uganda	2	2	1

The number of markers (out of 5) with significant deviation from HWE for each sample set before and after Bonferroni correction at (P<*0.05*) and (P<*0.01*).

4.3.4 Test of Hardy-Weinberg Equilibrium

The results here are reported including all the 8 markers used on 2R (i.e. including the 3 (2R) markers used previously in chapter 3). Most loci were in HWE for most of the samples (Table 4.4). Significant heterozygosity deficit was observed in 10 out of 48 tests (P < 0.001) across the markers. After Bonferroni correction was carried out at P<*0.05* and

P<*0.01*. The highest number of deviation was seen at 4 loci (AFUB3, AFND6, and FUNO) where 4 of the samples showed deviation from HWE. Samples from Malawi and Mozambique showed the highest number of loci deviating from HWE (4 loci) both before and after (3) Bonferroni corrections. Uganda showed the lowest number of loci deviating before (2) and after (1) Bonferroni correction.

LD analysis was carried out to further assess HWE patterns and the results showed that there is no significant LD for pairwise comparisons for each population of *An. funestus*. Also only two marker pairs (AFND6-FUNR for Benin and AFND40-AFND32 for Cameroon) showed LD. The observed heterozygote deficit is possibly locus specific unlike the Wahlund effect and inbreeding which affects the whole genome, since the LD across all the samples was below 5%. High HW disequilibrium is possibly due to selection acting on the 2R as selection is known to lead to deviation from HWE. More so no marker pairs where observed to show LD when considering all locus pairs across all the populations.

Despite the fact that genetic differentiation has been done in chapter three, it was analyzed again in this chapter as an alternative approach to detect signature of selection. It has been shown that signature of selective sweep could also be detected through the observation of locus-specific high population differentiation detected by statistics such as F_{st} (Kryukov *et al.* 2007).

So if differentiation is abnormally higher in the 2R than in the others, then this may further support the potential presence of selective sweep (Nielsen *et al.* 2007). Several other methods have been proposed for detecting selection based on this idea, for example, Akey *et al*. (2002) identifies areas of increased F_{st}, the traditional population genetic measure of population subdivision.

Study between population differences in allele frequencies is another approach for detecting ongoing selection. Among populations, levels of genetic differentiation may increase as a result of positive selection for two reasons; firstly, selection might be local and thus related to the local environmental adaptations with an example seen in humans where genes related to skin pigmentation selection is probably related to local environmental adaptations to climate. The second is selection acting on mutations that come from specific geographic regions that might cause increased levels of population

subdivision at a period that the mutation keeps increasing in frequency (Charlesworth *et al.* 1997, Slatkin *et al.* 1998).

4.3.5 Genetic differentiation between populations

Based on *Fst*

Seven of the 8 markers on the 2R were found inside polymorphic inversions (a,c,d) (Green and Hunt 1980), tz (Lochouarn), s,t,ab and au (Boccolini*et al.* 1998) and only one (FUNO) was found outside (Fig. 4.1). In any case, the position of the markers is not known to have any effect in this study since they are not common anywhere and are also not known to be polymorphic.

When considering estimates of pairwise genetic differentiation between samples for the 8, 9 and 17 markers based on F_{st} Benin-Cameroon (8 markers (0.0764), 9 (0.0699) and 17 (0.0708) markers), Cameroon-Uganda (8 markers-0.0231, 9-0.0268 and 17-0.0292), Benin and Malawi, Benin and Mozambique likewise Ghana and Malawi, Ghana and Mozambique and Malawi and Mozambique also showed no significant difference between all 3 comparisons. Other pairwise comparisons had higher F_{st} value for the 8 markers than the 9 markers like Benin-Ghana (0.1423 for 8 markers and 0.0707 for 9-markers) and Cameroon and Malawi. This is a sign that there could be difference of role of rp1 in the resistance between these samples. The F_{st} values where higher for the 8 markers than the 9 for some comparisons such as Ghana and Uganda. This might also be an indication of the role of the rp1 even though this is not definitive. Most of the comparisons showed similar values for the 9 and 17 markers (Table 2.5).

Cameroon and Uganda samples showed the lowest genetic differentiation for all the markers (F_{st} 0.0231) 9 (0.0268) and 17 (0.0292). When analysis was carried out with microsatellite loci completely outside the 2R chromosome (9 markers), Benin and Ghana showed reduced genetic differentiation (0.0707). This can only be explained by the fact that the 2R markers located around the rp1 QTL where the resistant Cytochrome P450 genes *CYP6P9a* and *CYP6P9b* are located (Wondji *et al*. 2007) might be impacting the patterns observed with 2R markers since the southern Africa samples are more resistant. It was observed for comparisons involving Uganda that the F_{st} values were not unusually high like the case seen between the southern Africa samples and the West Africa (Benin and Ghana) samples (Table 4.5). The 2R markers turn to have a great impact on the results.

Table 4.5: Genetic differentiation per locus of populations from Benin, Cameroon, Ghana, Malawi, Mozambique and Uganda

	BN-CM	BN-GH	BN-ML	BN-MZ	BN-UG	CM-GH	CM-ML	CM-MZ	CM-UG	GH-ML	GH-MZ	GH-UG	ML-MZ	ML-UG	MZ-UG
AFND40	0.0082	0.0032	-0.0052	-0.004	0.004	0.0037	0.0095	0.0123	**0.0395**	-0.0087	-0.0005	**0.0067**	-0.0078	-0.0084	**0.0023**
AFUB3	**0.2251**	**0.1976**	**0.267**	**0.2898**	**0.0231**	-0.0021	**0.0178**	**0.0375**	0.0027	0.0127	**0.0298**	-0.0075	**0.0032**	**0.02**	**0.0447**
AFND32	**0.0197**	0.0182	**0.0903**	**0.0641**	0.0231	0.0202	**0.1411**	**0.1466**	0.0205	**0.1471**	**0.1371**	0.037	**0.088**	**0.0917**	**0.1114**
AFND6	**0.049**	**0.0645**	**0.1274**	**0.09**	0.0198	**0.0407**	**0.0959**	**0.029**	**0.0244**	0.0217	**0.0732**	**0.0732**	**0.0732**	**0.0917**	**0.0283**
AFND30	**0.0354**	**0.0685**	**0.0673**	**0.1263**	**0.0283**	-0.0026	**0.0079**	**0.0528**	0.0051	0.0013	**0.0488**	0.0089	**0.0328**	**0.0275**	**0.0617**
FUNO	**0.0343**	0.0219	**0.0689**	0.0229	**0.0244**	0.006	**0.035**	0.0146	0.0067	**0.0441**	-0.0074	0.0109	**0.0381**	**0.0111**	0.0111
FUNR	0.0066	**0.1895**	0.0294	0.0238	**0.0309**	**0.159**	0.0448	0.0102	**0.0206**	**0.339**	0.0789	**0.1752**	**0.1149**	**0.0777**	**0.0435**
AFUB6	**0.1699**	**0.5049**	**0.0283**	**0.06623**	**0.198**	**0.1487**	**0.1116**	**0.0678**	**0.0665**	**0.4293**	**0.3584**	**0.2749**	**0.096**	**0.2118**	0.0342
Average 8 LOCI	**0.0764**	**0.1423**	**0.0996**	**0.0983**	**0.0702**	0.0431	**0.0603**	0.0494	0.0231	**0.1239**	**0.0905**	**0.0621**	**0.0516**	**0.0603**	0.0435
Average 9 loci	**0.0699**	**0.0707**	**0.0824**	**0.0904**	**0.0667**	**0.0537**	**0.1079**	**0.0990**	**0.0268**	0.1244	0.1271	**0.0177**	**0.0618**	**0.0980**	**0.1165**
Average 17 loci	**0.0708**	**0.1432**	**0.0931**	**0.1044**	**0.0722**	**0.0965**	**0.0844**	**0.0804**	**0.0292**	**0.1572**	**0.1407**	**0.0778**	**0.0526**	**0.0752**	**0.0864**

The F_{st} estimates were calculated using Genepop. Pairwise Significance after standard Bonferroni corrections is represented by bold ($P<0.05$) and underlined ($P<0.01$). Genetic differentiation per locus and across all loci for all sample pair comparisons. (BN-Benin, CM-Cameroon, GH-Ghana, ML-Malawi, MZ-Mozambique, UG -Uganda).

The average F_{st} values (Table 4.5) were used to construct the neighbor joining tree (Figs. 4.6a, 4.6b and 4.6c) to determine the evolutionary distance between the populations. The samples from Malawi and Mozambique from southern Africa formed a common group for all the markers (8, 9 and 17), while the samples from Benin and Ghana even though from West Africa did not form a common group. The samples from Cameroon in Central Africa and also the samples from Uganda in East Africa did not form a common group. Looking at the neighbor joining tree, Benin turn to be closer to Uganda that comes from East Africa than to Ghana that comes from West Africa for the 8 markers (located on the 2R). The samples from the other African countries show another pattern for the 9 and 17 markers while Malawi and Mozambique still got closer to each other. Thus the other populations formed no pattern while the southern African populations formed a clear cut pattern across all the comparisons. (Fig. 4.6a, b and c).
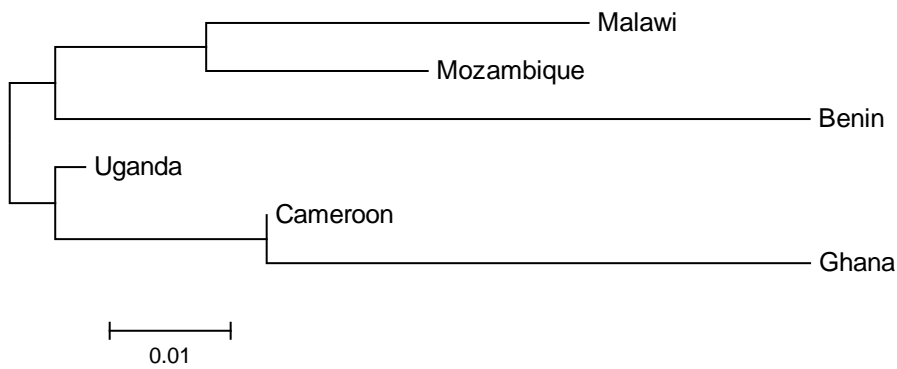
(a) 8 Markers on the 2R
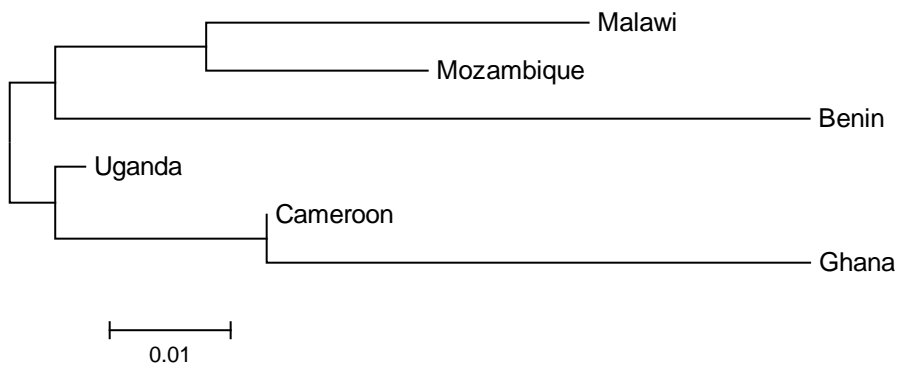
(b) 9 Markers outside the 2R

153
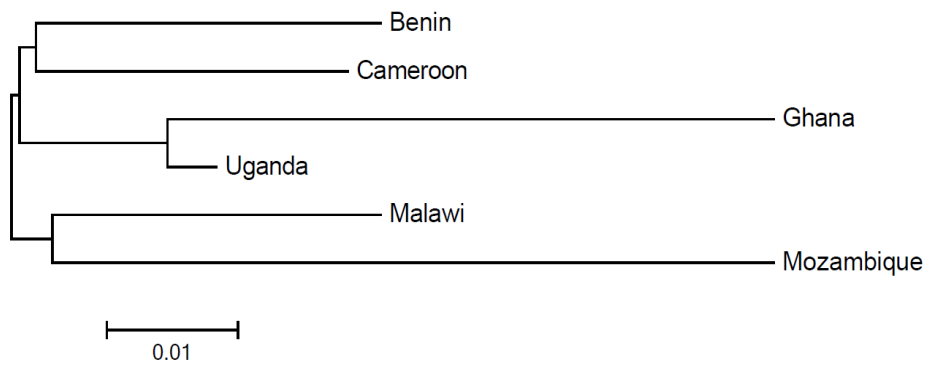
(c) 17 Markers (all markers)

Fig. 4.6: Neighbour joining phylogenetic distance trees of F_{st} values of the samples

Based on STRUCTURE model

The Bayesian STRUCTURE population analysis showed that K=3 was the best K that suited the data and the samples were clustered same as in chapter three. The samples from Benin and Uganda were represented most in cluster 1 for the 8 markers (assignment 0.904 and 0.407 respectively), while the samples from Cameroon and Ghana were represented most in cluster 2 (assignment 0.880 and 0.468 respectively). Malawi and Mozambique samples were highly represented in cluster 3 (assignment 0.705 and 0.736 respectively) (Fig. 4.7a, Table 4.6a and Fig. 4.8a). In any case, the Ugandan samples had a great number of its samples assigned to Cameroon in cluster 2 (0.301) which is what is expected since Cameroon is closer to Uganda in East Africa (Table 4.7a) than to cluster 3 that belongs to samples from southern Africa.

This pattern was similar for the 9 markers (markers outside the 2R chromosome) even though there was a slight twist with the samples from Ghana that were almost evenly distributed between cluster 2 (0.468) and cluster 1 (0.414) (Fig. 4.7b, Table 4.6b and Fig. 4.8b). This can only imply that the 2R markers do impact the population structure of the samples and in their absence, the Ghana samples turn to show close similarity to Benin from West Africa The resistance profile of Ghana is unknown since no bioassay was done but the results here suggests it is probably resistant with the involvement of the *CYP6P9a* and *CYP6P9b* genes. Also in cluster 3, the samples from Malawi shared more of its individuals with the other clusters like cluster 2 (0.425) than for the 8 markers (0.179). This possibly means that the absence of the 2R markers located around the rp1 QTL induces a higher genetic differentiation between Malawi and Mozambique.

The 17 markers showed similar pattern of assignment like the other markers but Ghana had more of its samples assigned to cluster 2 (0.782) than was the case for the 9 markers. Malawi and Mozambique also had most of their samples assigned to their cluster (3). Although Uganda formed a cluster with West Africa (Benin), it still shared more samples with central Africa than Southern Africa (Fig. 4.7c, Table 4.6c and Fig. 4.8c).
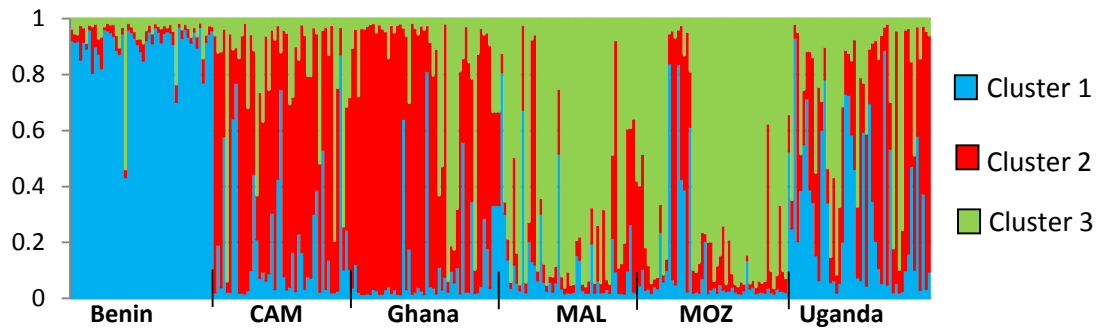
(a) 8 Markers on 2R
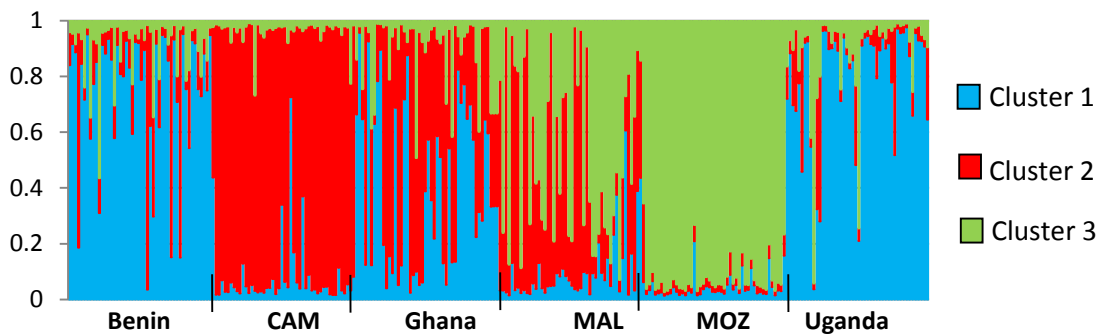
(b) 9 Markers on 2L, 3R and 3L
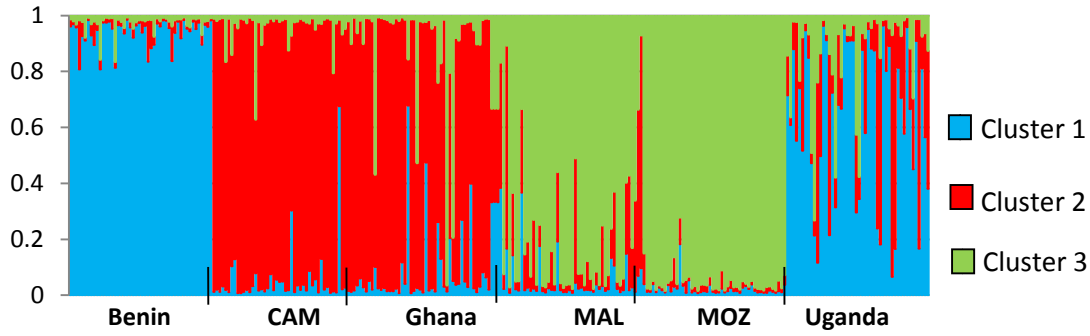
(c)   Total 17 Markers

Fig. 4.7(a-c): Bayesian cluster analysis using STRUCTURE. Graphical representation of the data set for the most likely K (K=3), where each color represents a suggested cluster, Blue color corresponds to (West Africa = Benin + Uganda), Red, (Central Africa = Cam-Cameroon + Ghana) and Green-(Southern Africa = Mal- Malawi + Moz-Mozambique) for 8, 9 and 17 makers respectively.

Table 4.6(a): Proportion of assignment of each predefined population in each of the three clusters inferred by STRUCTURE (8 Markers)

Population	Cluster 1	Cluster 2	Cluster 3
Benin	0.904	0.038	0.056
Cameroon	0.184	0.630	0.185
Ghana	0.110	0.697	0.192
Malawi	0.115	0.179	0.705
Mozambique	0.104	0.159	0.736
Uganda	0.407	0.310	0.282

Table 4.6(b): Proportion of assignment of each predefined population in each of the three clusters inferred by STRUCTURE (9 Markers)

Population	Cluster 1	Cluster 2	Cluster 3
Benin	0.744	0.153	0.101
Cameroon	0.080	0.880	0.039
Ghana	0.414	0.468	0.116
Malawi	0.100	0.425	0.474
Mozambique	0.043	0.031	0.924
Uganda	0.787	0.096	0.116

Table 4.6(c): Proportion of membership of each predefined population in each of the three clusters inferred by STRUCTURE (17 Markers)

Population	Cluster 1	Cluster 2	Cluster 3
Benin	0.939	0.031	0.029
Cameroon	0.050	0.903	0.045
Ghana	0.096	0.782	0.120
Malawi	0.056	0.149	0.794
Mozambique	0.022	0.022	0.955
Uganda	0.632	0.232	0.135

(a) 8 markers

(b) 9 markers
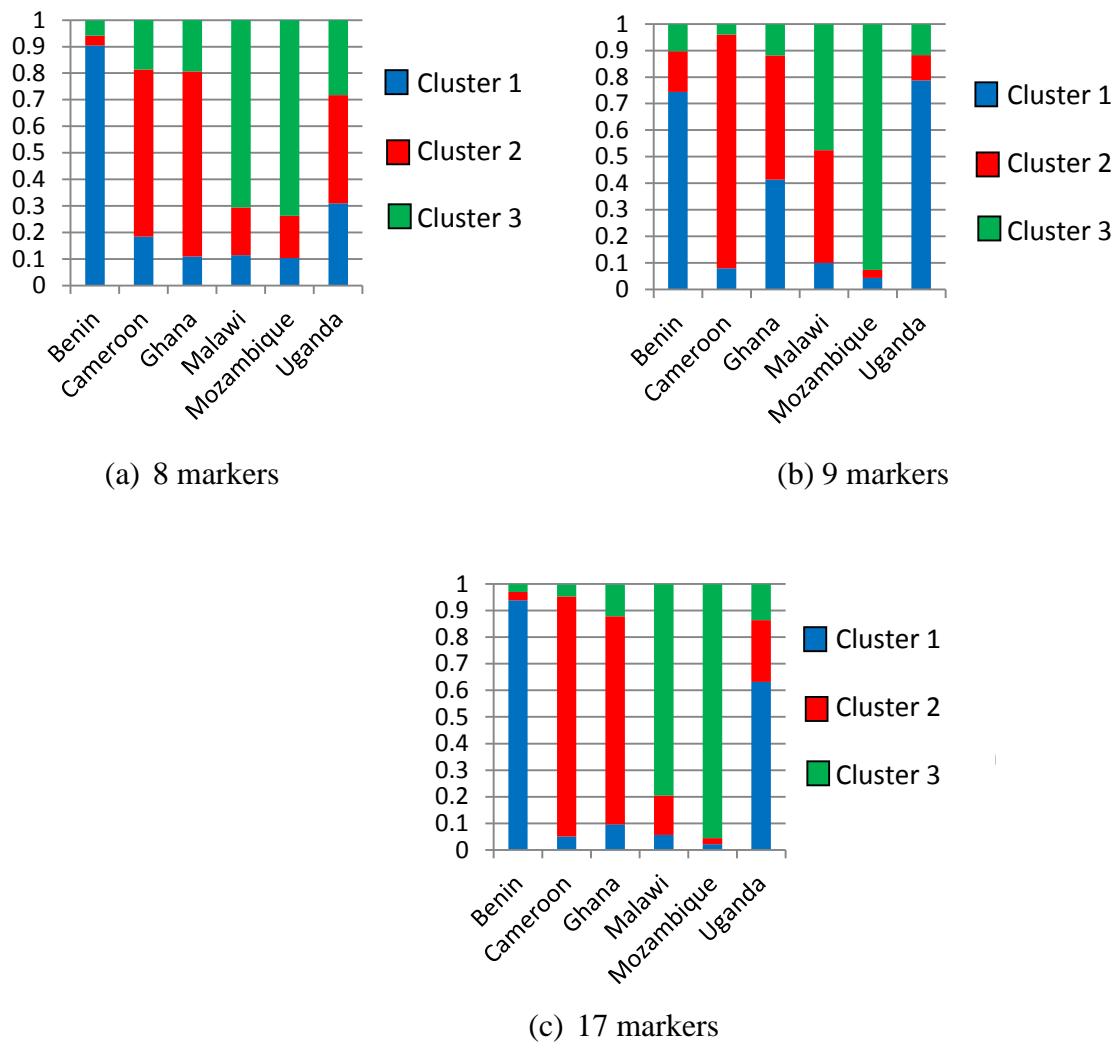
(c) 17 markers

Fig. 4.8(a-c): Graphical representation of membership of each predefined population in each of the three clusters inferred by STRUCTURE (8, 9 and 17 Markers respectively). The blue colour represents the sample allocation in cluster 1, red in cluster 2 and green in cluster 3.

4.4 Discussion

Detection of signature of selection is important to assess the genetic signature of important adaptive traits such as insecticide resistance. Microsatellite loci genotyping provides an important tool for the identification of genomic regions that are affected by such selection (Schlötterer *et al*. 1997, Payseur *et al*. 2002, Schlötterer 2002b). This study has investigated the presence of selective sweep around a major pyrethroid resistance rp1 QTL, on 2R chromosome using microsatellite markers notably two loci (FUNR and AFUB6) that are located at the close vicinity of the *CYP6P9a* and *CYP6P9b* pyrethroid resistance genes. This study identified potential signatures of selective sweep based on (1) a reduced diversity of microsatellite markers around the rp1 pyrethroid resistance

QTL, (2) on patterns of LnRH and also based on (3) patterns of genetic differentiation and structure around the 2R markers in comparison to those located in other chromosomes. However, these potential signatures of selection will need further confirmation in future studies.

4.4.1 Potential signatures of selective sweep from reduced genetic diversity of markers around rp1

Overall, there was reduced allelic diversity on the 2R chromosome particularly around the AFND40, FUNR and AFUB6. The AFND40 is closer to the FUNR and AFUB6 which are found around the rp1 QTL a major pyrethroid resistance locus (Wondji *et al*. 2009).A sharp increase in genetic diversity is seen immediately when moving away from the FUNR and AFUB6. This can be explained by the presence of a selective sweep as it is commonly associated with a reduced genetic diversity (Schlötterer 2002b). The reduced variation seen for those two markers that are around same position as the resistance genes *CYP6P9a* and *CYP6P9b* on the 2R, could only be linked to selection as seen in the study carried out by Kane and Rieseberg (2007) associating strong selection to reduce genetic variation observed around some locus that contained candidate genes in common sun flower. When looking at the allele distribution of the 9 markers outside the 2R, there is no defined pattern seen across the samples as that seen with the 2R and more so around FUNR and AFUB6 which has a clearly defined reduced heterozygosity across all the samples. The AFUB6 and FUNR markers also showed lowest genetic diversity across all the samples suggesting that selection could be acting around this region. However, this reduced diversity is observed in all six countries although different resistance profiles are observed in these countries. This general reduced diversity could indicate that it might not be caused by the presence of a selective sweep but by other factors such as a chromosomal inversion or a historical population event such as bottleneck or it could also be the nature of the markers for example, some of the markers were di-nucleotide, tri-nucleotide and some had mixed mortifs. The nature of the markers can also influence the results since tri-nucleotide markers are known to be less diverse than the di-nucleotides (Luro 2000). It is nevertheless unlikely that a single chromosomal inversion will induce such reduced diversity across the continent as most inversions are very localised as shown previously in *An. funestus* populations (Sharakhov *et al*. 2004, Ayala *et al*. 2009, Dia *et al*. 2013). Similarly, it is unlikely that this reduced diversity on the 2R is caused by a historical population event as these tend to affect the genome uniformly and not in a

localised way as seen on 2R. In addition, the observation of this reduced diversity in all 6 populations could be as a result of pyrethroid resistance at different degree in all these populations under the contribution of the rp1 at various extents. It has been shown that pyrethroid resistance is present in all these regions (Cuamba *et al* 2010, Morgan *et al.* 2010, Okoye *et al.* 2008) explaining the common reduced diversity in all these populations.

The different levels of resistance previously reported with very high levels in southern Africa could explain the difference in the extent of the reduction of this genetic diversity. The West, Central and East Africa samples showed higher heterozygosity across all the markers respectively than the southern Africa (Malawi and Mozambique) population. The samples from Malawi and Mozambique (southern Africa) with high reports of insecticide resistance and also a different resistance profile from the rest of the samples showed the lowest diversity for 2R markers. Nevertheless, further work is needed with more samples and markers to assess the consistency of the reduced genetic diversity and to confirm whether or not it is conclusively as a result of selective sweep caused by pyrethroid resistance from rp1 locus.

4.4.2 Evidences of selective sweep from LnRH statistics of markers around rp1

By using the LnRH statistics of Kayser *et al.* (2003), this study further investigated the presence of selective sweep around rp1. A normal distribution was seen with analysis carried out using the LnRH and also extreme values were seen in the summary of the pairwise population and sample comparisons. This result is similar to study carried out by Ihle (2004) using house mouse where extreme values for pairwise comparison at some loci was observed and loci were considered to be under selective sweep. Where the LnRH statistics showed a deviation from the normal distribution which might have been caused by the high amount of the extreme values seen at the edges of the distribution. These extreme values at the edges could not be considered as normal statistical deviations as reported by Schlötterer (2002b) but as a possible result of loci that are possibly under selective sweep. This is because real sweep locus must exhibit significant results (Schlötterer 2002b). Selective sweep detection based on pairwise population comparisons was that at least three pairwise comparisons should show reduction in polymorphism in the same population (Schlötterer *et al.* 2002b and 1997). In this study two possible selective sweeps were detected mostly for the southern populations because they

exhibited reduction in polymorphism. Other studies have also used LnRV, Rst and LnRH to identify selective loci (kayser *et al*. 2003, Teschke *et al*. 2008). However, further test are required to confirm the presence of selection on both loci since only a single statistical test was carried out and thus interpretation could be biased.

The comparisons between the populations resulted in a narrower normal distribution with a lower standard deviation than the lineage comparisons between the populations. However, this result did not bring much in confirming the presence of a selective sweep since there was no practical evidence of presence of selection acting on the various population pair comparisons. Compared to other studies that used markers that are distributed throughout the genome, the identification of a potential sweep in 2 out of 8 microsatellite loci is rather high but, the number of markers used in this study is quite small which could cause a bias in the interpretation of the results. Using multi-locus candidate screen in maize, 151 loci out of 501 loci were exhibiting evidence of selective sweep (Vigouroux *et al*. 2002). This situation is similar to the findings in this chapter where two loci are indicating a possible sweep since results from markers outside the 2R that excludes the 2 loci did not show any indication of selection. Also as seen in chapter three, the 2 loci were skewing up the results. Eleven out of 332 microsatellites were identified as potential sweep candidates in human populations by Kayser *et al*. (2003). Similar to the data presented in this study, Schlötterer *et al*. (1997) identified out of 10 microsatellite loci, only one locus in random screen in *D. melanogaster*. The approach applied in this study was not pre chosen but because of the reduced heterozygosity, allele frequency and their position around the rp1 QTL of the 2 loci (FUNR and AFUB6), their reduced diversity might be as a result of selection. This is possible since they are very evident across all the samples even though possibly higher in the more resistant southern populations of Malawi and Mozambique that have very high reports of insecticide resistance (Charlwood *et al*. 1998, Brooke *et al*. 2002, Wondji *et al*. 2012). In addition, the fact that the loci closest to these loci have a sharp increase in heterozygosity and allele frequencies is another indication that these loci are located away from the sweep window with the advantageous mutation. However, further tests have to be carried out to confirm this by increasing the number of markers and possibly sample sizes. Microsatellite loci genotyping is a suitable tool to identify the size of the sweep region (Harr *et al*. 2002, Kayser *et al*. 2003). Such analysis requires the use of microsatellite loci with same characteristics such as same length, repeat structure of the repetitive region so as to ensure

they evolve under same rate of mutation (Ellergren 2000). In *P. falciparum*, a minimal variation of microsatellite allele was found surrounding the dhfr for 12kb with continuous reduced variation 100kb flanking the locus (Nair *et al.* 2003). In the pfrct locus, a larger selective sweep was detected at approximately 200kb (Wootton *et al.* 2002). In this study, the exact size of the possible selection is not known but could be larger than the originally sequenced 120kb of the BAC clone spanning rp1 QTL (Wondji *et al.* 2009).

4.4.3 Suggestions of selective sweep from patterns of genetic differentiation of 2R markers

From the F_{st} values, there was a clear suggestion that a selective sweep was acting on the 2R loci than on the other loci. This selection seems to be stronger in the southern African populations of Malawi and Mozambique than the other samples. A reason why the selection is stronger in southern Africa might be due to geographic proximity since even very low migration rates is sufficient to spread an advantageous allele (Morjan and Rieseberg 2004). The selection is not across all the loci on the 2R but around AFUB6 and FUNR because that is where the resistance genes *CYP6P9a* and *CYP6P9b* are located. Rieserberg and Burke (2001) proposed that, selective sweep linked to a favourable mutation maygenetically homogenise populations in some loci while differentiating in others. The results presented in this study is similar to the proposition since very high differentiation is seen with analysis with markers outside the 2R for the southern populations that have high reports of insecticide resistance while the genetic differentiation for the other samples was not much of a difference or even lower for the markers outside the 2R.

Conclusion

The two microsatellite loci identified in this chapter as potentially under a selective sweep are located at the close vicinity of the two pyrethroid resistance genes *CYP6P9a* and *CYP6P9b* on the 2R. Further assessment of the polymorphism patterns of these genes or around the rp1 BAC clone of 120kb could allow to further confirm the presence of the selective sweep in these populations of *An. funestus*. This led to fine scale analysis of selective sweep around the pyrethroid resistance QTL (rp1) genomic region in chapter 5.

CHAPTER FIVE

Fine scale analysis of selective sweep around the pyrethroid resistance QTL (rp1) genomic region

5.1 Introduction

Genetic changes provide the basis for adaptation in domesticated and natural populations (Schlötterer 2001). There has been long-standing interest in biology in identifying those genetic changes associated with adaptation. Insecticide resistance offers one of the best examples of adaptation to a changing environment. In chapter 4, a likely signature of selective sweep was detected on the 2R chromosome and more so around two microsatellite loci (AFUB6 and FUNR) located around the rp1 QTL associated with pyrethroid resistance in *An. funestus*. This possible signature of selective sweep was detected through a reduced genetic diversity, LnRH statistical analysis and difference in genetic differentiation patterns between microsatellite markers and those in other chromosomes. Signature of this selection event was stronger in the more resistant southern populations (Malawi and Mozambique) than those in other regions. The southern population have high reports of insecticide resistance which might explain the existence of the signature of selective sweep found in this region. The location of the two loci coincides to the location of the main pyrethroid QTL explaining around 85% of genetic variance to resistance (Wondji *et al*. 2009). Furthermore, analysis of the BAC sequence spanning rp1 has revealed that the FUNR locus is located in the 5'UTR region of *CYP6P9a* gene shown to be a major pyrethroid resistance gene with *CYP6P9b* both in laboratory or field pyrethroid resistant strains (Wondji *et al.* 2009, Riveron *et al.* 2013).

However, one could not rule out completely that the reduced diversity observed around rp1 was not caused by other factors such as presence of a chromosomal inversion or possible historical population events. To address this concern, it could be helpful to perform of fine-scale analysis of the genetic diversity around the genomic region spanning the rp1 pyrethroid resistance QTL. The assumption is that, if this region has experienced a selective sweep event, a significant decrease of genetic diversity should be observed around the *CYP6P9a* resistance gene followed by a gradual recovery as one moves away in both direction of this gene along the rp1 genomic region. This U-shape genetic diversity pattern is a common characteristic of selective sweep as observed previously around *CYP6G1* in *Drosophila* (Schlenke and Begun, 2004) or in *Plasmodium*

falciparum (Nair *et al.* 2003). Additionally, if the presence of such selection on the 2R chromosome was to be confirmed, the scale and extent of this selection will also need to be established in order to help inform control programs of the impact of ongoing control strategies in selecting resistance in local populations of *An. funestus*. A comparative analysis of this polymorphism pattern between samples from highly resistant countries and more susceptible ones will help to validate the presence of this selective sweep. Furthermore, microarray analysis have revealed that rp1 role in pyrethroid resistance is less important away from southern Africa (Mulamba *et al.* 2014; Djouka *et al.* 2011, Riveron *et al.* 2014) suggesting that even if pyrethroid resistance is observed in other countries, a significant difference in polymorphism around rp1 could still be observed when compared to southern Africa.

Therefore, in this chapter, analysis of the genetic diversity of rp1 pyrethroid resistance genomic region through DNA sequencing is expected to provide a fine-scale assessment of the selection associated with pyrethroid resistance than possible with microsatellite loci used in chapter 4.

The previously sequenced 120kb BAC clone spanning the rp1 pyrethroid resistance QTL will offer a great opportunity for fine-scale analysis of the extent and strength of the selective sweep caused by pyrethroid resistance in field populations more than previously obtained using microsatellite loci.

5.1.1 Objective

The objectives of this chapter are

1. Validate the signature of selection around the main pyrethroid resistance gene *CYP6P9a* and establish its strength by analysing its polymorphism pattern in several populations with different resistance profiles across Africa.
2. To carry out fine-scale analysis of the extent of selective sweep around the rp1 QTL genomic region by assessing the polymorphism of 5 different fragments of BAC across the 120kb region between the highly resistant southern African population and the more susceptible Cameroon sample in Central Africa.

To achieve these objectives, the full-length of the resistance gene *CYP6P9a* was sequenced for the same six countries as in previous chapters from different geographical regions of Africa showing varied profiles of resistance to pyrethroids. Additionally, 5

genomic fragments (Table 5.1) were sequenced to span the 120kb rp1 BAC clone by comparing the more resistant population from Malawi and Mozambique to the more susceptible samples from Cameroon in order to determine if there will be differences in selection between the resistant and susceptible samples.

5.2 Materials and methods

The genomic DNA used for this study is same as those used in chapter three.

Table 5.1: BAC primers used for BAC fragments sequencing including size ranges (Wondji*et al.* 2009)

Primers	Forward Sequence	Reverse Sequence	Size (bp)
0BAC	TGGTAGCTGCTTGAGG AGAAA	TCACACTAGCTGCCAA ATCG	716
25BAC	GTGACGGATCTGGACC TTGT	TCATCTCCTTGAGTGCA TCG	823
70BAC	ATATTCATGGGCGATTG TGC	AAGAGTTGGCAAAGGA AGGA	702
95BAC	CCTCTTTACTGGCCACC GTA	CTCCTCCGTAACGTGAT CGT	755
120BAC	TCTTCCGCCATTGTGTA TCA	AGTTCGAGCACCAGCT CAAG	711

Five BAC clones (Table 5.1) that span the entire 120kb of the *CYP6P9a* gene of the *An. funestus* used for this study was from a BAC library of *An. funestus* that was constructed by the Institute for Genomic Research (TIGR) using genomic DNA fragments from the FUMOZ colony (Wondji *et al.* 2009 in pers. Comm. with N.Lobo)

5.2.1 Primer design

For the full-length amplification of *CYP6P9a*, the primers were designed to cover the 5' flanking region and to the 3' end of the coding region of *CYP6P9a* gene (Fig. 5.1). The BAC primers were also selected to cover the entire 120kb of the rp1 BAC clone (Fig. 5.2).
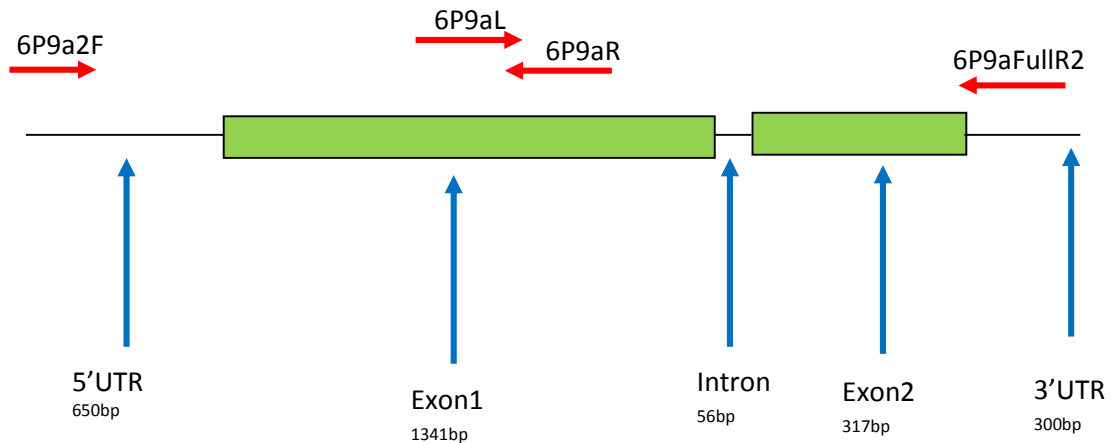
Fig. 5.1: Schematic representation of *CYP6P9a* gene with the different primers used in this study.
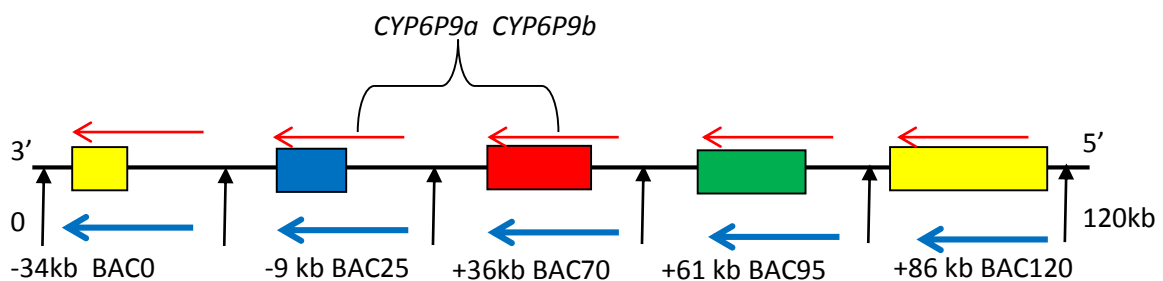
Fig. 5.2: Schematic representation of BAC across the *rp1*. Blue arrows represent BAC fragments while red arrows show the 5'-3' direction.

5.2.2 PCR amplification of *CYP6P9a* gene and BAC fragments

The following PCR conditions were used for the amplification of the *CYP6P9a* gene for an average of 10 mosquitoes per country is as described in chapter two (samples were same genomic DNA used in chapter 3) and an average of 8 mosquitoes per country for BAC fragments. Individual PCRs were performed in 30µl reactions containing Buffer A 3µl, dNTP 25mM 0.2µl, forward BAC primer 1.02µl, BAC Reverse primer (*CYP6P9a* or BAC), KapaTaq 0.24µl (BAC), ddH20 22.48µl and DNA 2µl; and were conducted on a DNA Thermal Cycler under the following conditions: For *CYP6P9a* gene, the conditions are same as previously described in chapter two and for BAC fragments, 5mins at 95°c, followed by 35 cycles of denaturing at 94°c for 30secs and extension at 57°c for 30secs; finishing with an extension step at 72°c for 1min and 72°c for 10mins.

5.2.3 PCR purification

PCR products were purified using the QiaQuick (Qiagen, Hilden Germany) as previously described in chapter two

5.2.4 Sequence analysis, test for selection and phylogenetic tree analysis

The analyses are same as previously described in chapter 2. In any case, for the MKA test, the *CYP6AA2* (AGAP002862-PA) of *An. gambiae* was used as the out-group for the BAC25 fragment which spans the *CYP6AA2* gene in *An. funestus.*

5.3 Results

5.3.1 Genetic diversity of *CYP6P9a* across Africa

5.3.1.1 Polymorphism patterns of *CYP6P9a*

The full fragment of 1965bp of *CYP6P9a* gene was sequenced and aligned for a total of 59 samples for all the six countries (Benin, Cameroon, Ghana, Malawi, Mozambique and Uganda). The coding region was 1470bp while the non coding region was 496bp. There were a total of 113 polymorphic sites (Table 5.2). There was an indel in the 5' UTR mainly in southern Africa and not in the other countries. The number of polymorphic sites varied between the populations (Benin- 25, Cameroon-34, Ghana- 14, Malawi-5, Mozambique-61 and Uganda-16). The observed high level of polymorphism in Mozambique is a result of 3 mosquitoes that were very close to the susceptible mosquitoes as similarly described in Riveron *et al.* (2013). The polymorphism shown by these three mosquitoes was closer to that observed in East or West Africa. Also Mozambique had the highest number of sequences (28) which might also be a contributing factor to the high level of polymorphism seen. In the coding region, 83 polymorphic sites were found which also varied across the various populations. The coding region was less genetically diverse than the non coding region when considering the nucleotide diversity (Table 5.2). There was an overall 59 haplotypes with Malawi and Ghana exhibiting the lowest number of haplotypes (8 and 9) respectively. Mozambique showed high number of haplotypes which is as a result of three mosquitoes as earlier described. There was also an overall 21 haplotypes that were singletons. The southern population (Malawi and Mozambique) had the lowest number of singletons (1 and 2). Benin (16) and Cameroon (17) had very high

number of singletons. Ghana (4) and Uganda (8) had an average number of singletons but were still high when compared to the southern population (Table 5.2).

The Tajima (D), test showed all negative values across all the samples except for Malawi (Table 5.2). Negative Tajima value is an indication of purifying selection while positive values indicate balancing selection with the presence of alleles with average frequency (Tajima 1986). The Fu and Li (D) values were negative for all the samples except for the southern population. This might also be indicating purifying selection. Mozambique samples were significant at P<*0.02* for the Fu and Li test for the coding region and significant at P<*0.05* for the non coding region. Uganda was significant at P<*0.05* for Tajima test for the non coding region.

Table 5.2: Genetic parameters for *CYP6P9a*

	Complete sequenced fragment (1967bp)							Coding region (1470bp)						Non-Coding region (496bp)					
S	N	S	π	Taj D	Fu D	S^*	h (hd)	S	π	Taj D	Fu D	S^*	h (hd)	S	π	Taj D	Fu D	S^*	h (hd)
BN	16	25	2.2	-1.72^{ns}	-1.58^{ns}	16	11(0.908)	12	1.2	-1.89^{ns}	-1.95^{ns}	9	6(0.617)	13	5.1	-1.35^{ns}	-0.99^{ns}	7	10(0.825)
CAM	20	34	2.9	-1.57^{ns}	-1.15^{ns}	17	10(0.832)	25	2.9	-1.49^{ns}	-0.79^{ns}	11	9(0.789)	9	2.9	-1.47^{ns}	-1.72^{ns}	6	5(0.568)
GH	16	14	2.0	-0.44^{ns}	-0.06^{ns}	4	9(0.917)	11	2.0	-0.69^{ns}	-0.05^{ns}	3	8(0.883)	3	2.1	-0.47^{ns}	-0.04^{ns}	1	5(0.758)
MAL	20	5	0.8	0.31^{ns}	0.39^{ns}	1	8(0.837)	3	0.5	-0.23^{ns}	-0.12^{ns}	1	5(0.621)	2	1.5	0.895^{ns}	0.866^{ns}	0	3(0.653)
MOZ	28	61	6.1	-0.95^{ns}	1.615^{ns}	2	13(0.786)	48	6.7	-0.79^{ns}	1.53^{**}	2	11(0.677)	13	4.0	-1.33^{ns}	1.494^{*}	0	6(0.648)
UG	18	16	1.5	-1.34^{ns}	-0.95^{ns}	8	10(0.908)	7	1.2	-0.38^{ns}	0.70^{ns}	1	8(0.837)	9	2.4	-1.924^{*}	-2.11^{ns}	7	4(0.399)

N = Number of sequences; S = Number of polymorphic sites; h = Number of haplotypes; hd = Haplotype diversity; π = Nucleotide diversity (π is raised to 10^{-3}); D = Tajima's statistics; Taj D = Fu and Li's statistics; S* = singleton, star (*) on top of figure means significant at P<*0.01* P<*0.05* and (**) denotes significant at P<*0.02*, Not significant (ns)

5.3.1.2 Genetic diversity

The genetic diversity across the countries varied. Malawi showed much reduced nucleotide diversity ($\pi \times 10^{-3}$) (0.8), while higher levels of nucleotide diversity was observed for the other populations; Benin (2.2), Cameroon (2.9), Ghana (2.0), Uganda (1.5). Surprisingly, Mozambique even though found in southern Africa, showed the highest nucleotide diversity (6.1 x 10^{-3}), (Table 5.2 and Fig. 5.3). The high genetic diversity in Mozambique was contributed by the presence in this population of very few mosquitoes (3) which were very different from the remaining sample and were more similar to mosquitoes from other geographical regions. When these 3 mosquitoes are removed, the Mozambique sample also shows a reduced diversity ($\pi \times 10^{-3}$) (0.6), while other populations were more homogeneous in the polymorphisms observed between mosquitoes.
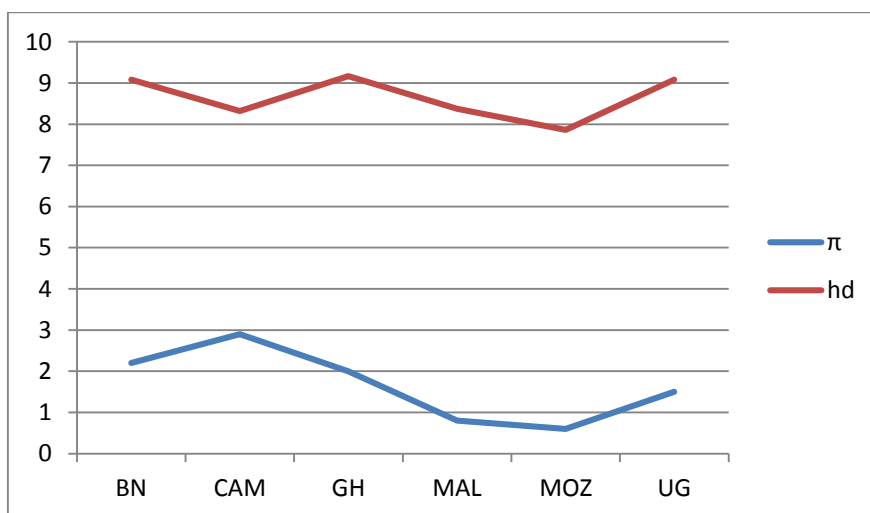
Fig. 5.3: Distribution of nucleotide diversity (π x 10^{-3}) and haplotide diversity (hd x 10) across samples for CYP6P9a gene (BN-Benin, CAM-Cameroon, GH-Ghana, MAL-Malawi, MOZ-Mozambique, UG-Uganda).

5.3.1.3 Haplotype distribution

A total of 59 haplotypes were observed for the full *CYP6P9a* gene (Fig. 5.4a) and 46 for the coding region (Fig. 5.4b). When considering the full gene, the resistant haplotype MAL/MOZ33 was predominant accounting for 16.9% of the total sample mainly because of its high frequency in Malawi (35%) and Mozambique (65%). This same haplotype corresponds to the resistant haplotype that have already been seen in the previous chapter of resistance mechanisms when comparing mosquitoes alive after permethrin exposure to

those dead. There was also haplotype BN/CAM/GH/UG3 that accounted for 5.6% of the entire samples. This haplotype was found in the four countries but not in any of the southern countries. In addition, there were many singletons among the populations (BN-16, CAM-11, GH-4). In any case, there were very low (1) singleton in Malawi and Uganda where one haplotype is predominant and also to some extent in Mozambique.

When considering the coding region only, the most predominant haplotype was MAL/MOZ28 which accounted for 23.7% of the entire sample with Malawi (42.9%) and Mozambique (57.1%) (Fig. 5.4b and Fig. 5.5). There were no other coding haplotype predominant to other regions but to individual countries like CAM 15 (6.8%).

There were 16 non-synonymous equivalent of amino acid changes (Fig. 5.6a) and also 16protein variants (Fig. 5.6b) on the coding region for this gene. The first amino acid change was found on 633bp and the last at 1349bp. The amino acid changes were clustered in some regions of the protein sequence. The most predominant protein (MAL/MOZ9 with a frequency of 38 (32.2% of entire samples)) variant was mainly in southern Africa with Malawi (55.3%) and Mozambique (44.7%). There were other predominant protein variant like BN/CAM/GH1 with a frequency of 34 (27.1%) was found in West and Central Africa but not in southern Africa and East Africa (Uganda). There was also GH/UG7 with a frequency of 7%.

```
Hap#   Freq. Sequences
BN1:     1   BN32-1
BN2:     1   BN32-2
BN3:     5   BN33-1 BN34-1 BN34-2 BN36-1 BN44-1
BN4:     1   BN33-2
BN5:     1   BN35-1
BN6:     1   BN35-2
BN7:     1   BN36-2
BN8:     1   BN44-2
BN9:     2   BN45-1 BN45-2
BN10:    1   BN46-1
BN11:    1   BN46-2
CAM12:   3   CAM2A-1 CAM2A-2 CAM10A-1
CAM13:   1   CAM2D-1
CAM14:   1   CAM2D-2
CAM15:   8   CAM3D-1 CAM5D-1 CAM6D-1 CAM6D-2 CAM9D-1 CAM9D-2 CAM11A-1 CAM11A-2
CAM16:   1   CAM3D-2
CAM17:   2   CAM5D-2 CAM6A-1
CAM18:   1   CAM6A-2
CAM19:   1   CAM9A-1
CAM20:   1   CAM9A-2
CAM21:   1   CAM10A-2
GH22:    2   GH1-1 GH1-2
GH23:    1   GH4-1
GH24:    1   GH4-2
GH25:    2   GH6-1 GH6-2
GH26:    1   GH7-1
GH27:    1   GH7-2
GH28:    4   GH8-1 GH8-2 GH9-1 GH9-2
GH29:    2   GH10-1 GH10-2
GH30:    2   GH5-1 GH5-2
```

```
GH31:      2  ML1-1 ML14-1
GH32:      1  ML1-2
MAL/MOZ33: 20   ML2-1 ML2-2 ML3-1 ML4-1 ML9-1 ML9-2 ML13-1 MOZ6-1 MOZ23-1 MOZ4D-1 MOZ7A-
1 MOZ7A-2 MOZ2D-1 MOZ2D-2 MOZ3D-1 MOZ3D-2 MOZ6A-1 MOZ6A-2 MOZ3A-1 MOZ3A-2
MAL/MOZ34: 5  ML3-2 ML4-2 ML7-1 ML12-1 MOZ15-1
MAL35:     1  ML7-2
MAL36:     3  ML8-1 ML12-2 ML13-2
MAL37:     1  ML8-2
MAL38:     1  ML14-2
MOZ39:     1  MOZ6-2
MOZ40:     1  MOZ15-2
MOZ41:     1  MOZ23-2
MOZ42:     1  MOZ4D-2
MOZ43:     2  MOZ8A-1 MOZ8A-2
MOZ44:     2  MOZ9A-1 MOZ9A-2
MOZ45:     2  MOZ5D-1 MOZ5D-2
MOZ46:     1  MOZ21-1
MOZ47:     1  MOZ21-2
MOZ49:     1  MOZ22-2
UG50:      2  UG2-1 UG2-2
UG51:      2  UG3-1 UG5-1
UG52:      2  UG3-2 UG8-1
UG53:      2  UG4-1 UG4-2
UG54:      1  UG5-2
UG55:      5  UG6-1 UG7-1 UG9-1 UG9-2 UG10-2UG56:1 UG6-2
UG57:      1  UG7-2
UG58:      1  UG8-2
UG59:      1  UG10-1
```

(a) Full length of CYP6P9a

```
Hap# Freq Sequences]
BN1:  10 BN32-1 BN33-1 BN34-1 BN34-2 BN35-1 BN35-2 'BN36-1 BN36-2 BN44-1 BN46-1
BN2:   1  BN32-2
BN3:   1  BN33-2
BN4:   1  BN44-2
BN5:   2  BN45-1 'BN45-2
BN6:   1  BN46-2
CAM7:  3  CAM2A-1 CAM2A-2 CAM10A-1
CAM8:  1  CAM2D-1
CAM9:  1  CAM2D-2
CAM10: 9  CAM3D-1 CAM5D-1 CAM6A-2 CAM6D-1 CAM6D-2 CAM9D-1 CAM9D-2 CAM11A-1 CAM11A-2
CAM11: 1  CAM3D-2
CAM12: 2  CAM5D-2 'CAM6A-1
CAM13: 1  CAM9A-1
CAM14: 1  CAM9A-2
CAM15: 1  CAM10A-2
GH16:  2  GH1-1GH1-2
GH17:  1  GH4-1
GH18:  1  GH4-2
GH19:  4  GH6-1 GH6-2 GH10-1 GH10-2
GH20:  1  GH7-1
GH21:  1  GH7-2
GH22:  4  [GH8-1 GH8-2 GH9-1 GH9-2
GH23:  2  GH5-1 GH5-2
GH24:  3  ML1-1 ML1-2 ML14-1
MAL/MOZ25: 28   ML2-1 ML2-2 ML3-1 ML3-2 ML4-1 ML4-2 ML7-1 ML7-2 ML9-1 ML9-2 ML12-1 ML13-1
MOZ6-1 MOZ15-1 MOZ23-1 MOZ4D-1 MOZ7A-1 MOZ7A-2 MOZ9A-1 MOZ9A-2 MOZ2D-1 MOZ2D-2 MOZ3D-1
MOZ3D-2 MOZ6A-1 MOZ6A-2 MOZ3A-1 MOZ3A-2
MOZ26: 3  ML8-1 ML12-2 ML13-2
MOZ27: 1  ML8-2
MOZ28: 1  ML14-2
MOZ29: 1  MOZ6-2
MOZ30: 1  MOZ15-2
MOZ31: 1  MOZ23-2
MOZ32: 1  MOZ4D-2
MOZ33: 2  MOZ8A-1 MOZ8A-2
MOZ34: 2  MOZ5D-1 MOZ5D-2
MOZ35: 1  MOZ21-1
MOZ36: 1  MOZ21-2
MOZ37: 1  MOZ22-1
MOZ38: 1  MOZ22-2
UG39: 2  UG2-1 UG2-2
UG40: 2  UG3-1 UG5-1
```

172

```
UG41:  2  UG3-2 UG8-1
UG42:  7  UG4-1 UG4-2 UG6-1 UG7-1 UG9-1 UG9-2 UG10-2
UG43:  1  UG5-2
UG44:  1  UG6-2
UG45:  1  UG7-2
UG46:  2  UG8-2 UG10-1
```

(b) Coding region only

Fig. 5.4: Number of haplotype distributed across all the six countries for the full length - (a) (coding and non coding) and for the coding region – (b) only. Hap# = haplotype number, Freq = frequency of haplotype and Sequence is country where haplotype is mostly found.

```
Hap#  Freq. Sequences]
BN1     10CAAGGTTTCGTTCTCGTTGATCCGCGATGCATACGGTACGCTGCCAAGTGCTGTCCCTTCGATGCTCGAGGCCGCACCCTGAA
BN2      1.......................................................................T.C..............
BN3      1..................................................................AT........GG.T....
BN4      1.....................................................................T........
BN5      2...........................................................A........................
BN6      1T.......G...............................................................TT..G.
CAM7     3T...............................A...............................................
CAM8     1TG.T....G........A.................C.GA.........................................
CAM9     1....A...................A.........CCGA.C...........................T....ATC.....C..G
CAM10    9...............................A................................................
CAM11    1...............................A................................A...............
CAM12    2...............................A...............................T................
CAM13    1........................T....A......................................C...........
CAM14    1.......G....C...........C.........C....C...........................T....ATC.....C..G
CAM15    1T.............................A...............A.............T.C.................
GH16     2T..T....G........C......................TA......................T.............A..
GH17     1T..T....G........C......................TAT.....C.............................A..
GH18     1T..T....G........C......................TAT.....C.............T...............A..
GH19     4TG.T....G........A......................TAT.................T.................A..
GH20     1T.......................................TAT.................T...................
GH21     1T......G........C.....C.................TAT.......G........T...................
GH22     4T..T....G........C.....C................TAT.................T.................A..
GH23     2T..T....G........C.....C................TAT................TT.................A..
ML24     3.....GCA.ACGAC....................TC.GA.CATAT....ACA..TTC.A.G..T.....TATC........
ML/MZ25 28.....GCA.ACGAC....................TC.GA.CATAT....ACA..TTC.AAG..T.....TATC........
MAL26    3.....GCA.ACGAC....................TC.GA.CATAT....ACA..TTC.AAG..T.T...TATC........
MAL27    1.....GCA.ACGAC....................TC.GA.CATAT....ACA..TTC.AAGC.T.....TATC........
MAL28    1.....GCA.ACGAC....................TC.GA.CATAT....ACA..TTC.A.G..T.T...TATC........
MOZ29    1T....GCA.ACGACGTCCCTGTT.T...A.GCCTATC.GA.CATAT....ACA..TTC.AAG..T.....TATC........
MOZ30    1T....GCA.ACGACGTCCCTGTT.T...A.GCCTATC.GA.CATAT....ACA..TTC.AAG..T...C..ATC.......G.
MOZ31    1T....GCA.ACGACGTCCCTGTT.....A.GCCTATC.GA.CATAT....ACA..TTC.AAG..T...CA.ATC.......G.
MOZ32    1.....GCA.ACGAC.............C.......TC.GA.CATAT....ACA..TTC.AAG..T.....TATC........
MOZ33    2TG.T............A.................TC.GAT..T.TTACAAC...TTC.A.G.................C...
MOZ34    2..T..GCA.ACGAC....................TC.GA.CATAT....ACA..TTC.AAG..T.....TATC........
MOZ35    1.....GCA.ACGAC....................TC.GA.CATAT....ACA..TTC.AAG..T.....T............
MOZ36    1T....GCA.ACGACGTCCCTGTT.....A.GCCTATC.GA.CATAT....ACA..TTC.AAG..T.....TATC........
MOZ37    1.....GCA.ACGAC....................TC.GA.CATAT....ACA..TTC.AAG..T......ATC.........
MOZ38    1T....GCA.ACGACGTCCCTGTT.T...A.GCCTATC.GA.CATAT....ACA..TTC.AAG..T...................
```

174

```
UG39    2T..T....G........C.....C..C.........C....C.T.TTAC..C...TTC.AAG..T......A...........
UG40    2T..T....G........C.....C..C.........C....C.T.TTAC..C...TT...AG........A...........
UG41    2T..T....G........C.....C..C.........C....C.T.TTAC..C...TTC.A....T......A...........
UG42    7T..T....G........C.....C..C.........C....C.T.TTAC..C...TT..AAG..T......A...........
UG43    1T..T....G........C.....C..C.........C....C.T.TTAC..C...TTC.AG..T......A...........
UG44    1T..T....G........C.....C..C.........C....C.T.TTAC..C...TT..AAG..T..C...A.........G.
UG45    1T..T....G........C.....C..C.........C....C.T.TTAC..C...TT..AAG..T......A.........G.
UG46    2T..T....G........C.....C..C.........C....C.T.TTAC..C...TT..AAG........A...........
```

Fig. 5.5: Schematic representation of the haplotypes of *CYP6P9a* gene across Africa (BN- Benin, CAM-Cameroon, GH-Ghana, MAL-Malawi, MOZ-Mozambique and UG-Uganda) for the coding region. The polymorphic positions are indicated with numbers besides the haplotype (haplotypes are labelled by the country initials where the haplotype is predominant), and the second numbers indicates the haplotype frequency.

```
Hap#    Freq.  Sequences
BN/CAM/GH1   34    ATTTCCGGTCGGAAACTCTGTGGAC
BN2           1    ....................A..GG
BN3           2    ................A........
CAM4          1    ....................A....
CAM5          7    .............T...........
GH6           1    ..........C...TG.........
GH7           7    ..........C...T..........
GH8           2    ..........C...T....T.....
MAL/MOZ9     38    .GAGA.........T..A.......
MAL10         1    .GAGA.........T..AC......
MOZ11         4    .GAGAGTCGT.AGCT..A.......
MOZ12         1    .GAGAGTCGT.AGCT..A....A..
MOZ13         2    ..............T..A.......
MOZ14         2    TGAGA.........T..A.......
UG15         14    ..........C...T..A.......
UG16          1    ..........C...T..A...C...
```

(a)

```
                  000000000000011112233333344
                  45566778889001879011157833
                  1123612156137174912715489
BN/CAM/GH1   34    HSLFQREMLLEVHNNLMHFXNRSKR
BN2           1    ....................K..EG
BN3           2    ................K........
CAM4          1    ....................K....
CAM5          7    ..............I..........
GH6           1    ..........D...IV.........
GH/UG7        7    ..........D...I..........
GH8           2    ..........D...I..........
MAL/MOZ9     38    .AQLK.........I..Q.......
MAL10         1    .AQLK.........I..QL......
MOZ11         4    .AQLKGDIVF.IRTI..Q.......
MOZ12         1    .AQLKGDIVF.IRTI..Q....N..
MOZ13         2    ..............I..Q.......
MOZ14         2    LAQLK.........I..Q.......
UG15         14    ..........D...I..Q.......
UG16          1    ..........D...I..Q...P...
```

(b)

Fig. 5.6: Haplotype distribution of non-synonymous equivalent of amino acid change – (a) and protein variants – (b)

5.3.1.4 Genetic differentiation based on *CYP6P9a* gene

The results of the genetic differentiation test using K_{ST} showed that the two southern African populations of Malawi and Mozambique are two genetically close populations with only limited genetic differentiation with a K_{ST} estimate of 0.045 (*0.01<P<0.05*). However these two pyrethroid resistant populations of southern Africa are highly genetically differentiated to all other 4 populations with K_{ST} estimates from 0.34696 for Benin and Malawi, 0.57404 for Uganda and Mozambique with P<*0.001*. The other four populations exhibit higher levels of genetic differentiation between them than seen between the two southern African populations, although in accordance to geographic proximity, the Benin samples is closer to Cameroon compared to the rest (Table 5.3). However despite been geographically close to Benin in West Africa, the Ghana sample

is more differentiated to Benin (K_{ST}=0.50) than the Cameroon populations (K_{ST}=0.13) in line with microsatellite results.

A Neighbour joining tree that was drawn from the K_{ST} values (Table 5.3) further supports that the southern population form a separate group from the other four populations with the Ugandan population positioned as an intermediate between southern and west/central Africa similar to microsatellite genetic differentiation pattern.

Table 5.3 K_{ST} Values

Countries	Benin	Cameroon	Ghana	Malawi	Mozambique
Cameroon	0.14494***				
Ghana	0.53047***	0.47588***			
Malawi	0.34696***	0.64195***	0.49400***		
Mozambique	0.56865***	0.58241***	0.57271***	0.04823*	
Uganda	0.52071***	0.62288***	0.43902***	0.48077***	0.57465***

PERMTEST calculates Hudson's K_{ST} statistic of genetic differentiation. K_{ST} is equal to 12 K_S/K_T, where K_S is a weighted mean of K1 and K2 (mean number of differences between sequences in subpopulations 1 and 2, respectively) and K_T represents the mean number of differences between two sequences regardless of their subpopulation. The null hypothesis of no genetic differentiation will be rejected (P<*0.05*) when K_S is small and K_{ST} is close to 1.PM test; Probability obtained by the permutation test with 1000 replicates); ns, not significant; *, *0.01<P<0.05*; **, *0.001<P<0.01*; ***, P<*0.001*
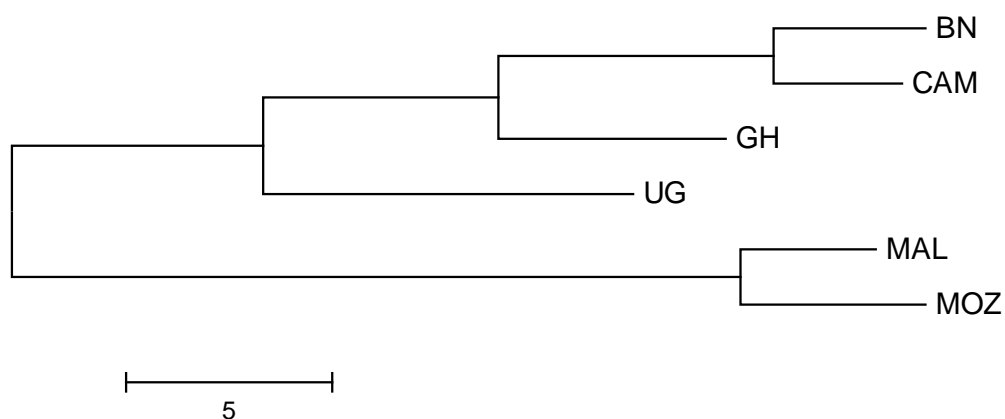
Fig. 5.7: Neighbour joining tree based on genetic distances from K_{ST} estimates of pairwise population comparison.

5.3.2 Phylogenetic analysis

5.3.2.1 Maximum likelihood tree

In the more pyrethroid resistant populations from southern Africa (Malawi and Mozambique), several individuals belonged to the same haplotype (ML/MOZ33) when analysing the full-length or only the coding region (ML/MOZ25). But for the other four countries, no haplotype is shared between countries as it is the case in southern Africa. This further indicates the genetic homogeneity of the southern African populations potentially due to the selection from insecticide or perhaps geographic proximity. Consequently, the maximum likelihood phylogenetic tree of *CYP6P9a* haplotypes clearly showed that the southern population (Malawi and Mozambique) formed a unique and defined cluster from the other four populations. Benin and Ghana despite being from West Africa did not form a common cluster despite their geographical proximity. This is in line with microsatellite results possibly suggesting some barriers to gene flow between the southern and the West African countries. Benin formed a cluster with Cameroon which is more in Central Africa, while Ghana was seen closer to Uganda although they did not form a defined cluster (Fig. 5.8).

All haplotypes from Uganda formed a cluster which is intermediate between southern, West and Central African countries which correlate well with the genetic structure patterns obtained with microsatellite loci (Fig. 5.9). This also correlates with the geographical position of Uganda in East Africa which is between southern and West/Central Africa.

Maximum likelihood trees constructed for each country confirmed the low genetic diversity in Malawi with most individuals belonging to a predominant haplotype corresponding to the resistant haplotype ML/MOZ33. The same haplotype is also predominant in Mozambique but other individuals belong to more diverse haplotypes than observed in Malawi. This is in line with the overall genetic diversity observed for both countries. One mosquito from Mozambique rather clustered with mosquitoes from Uganda.

This phylogenetic relationship between the mosquitoes analysed is further confirmed when only analysing the non-synonymous sites or the protein variants. The ML tree obtained for these non-synonymous sites shows that a highly predominant *CYP6P9a* protein variant is present in southern African mosquitoes at a frequency of 32.2% of the entire sample and similarly for Benin and Cameroon which shares a major allele, only present in two

mosquitoes from Ghana. Ghana still remains different from Benin and Cameroon with different protein variants found while Uganda remains intermediate between the regions.

(a) Coding region
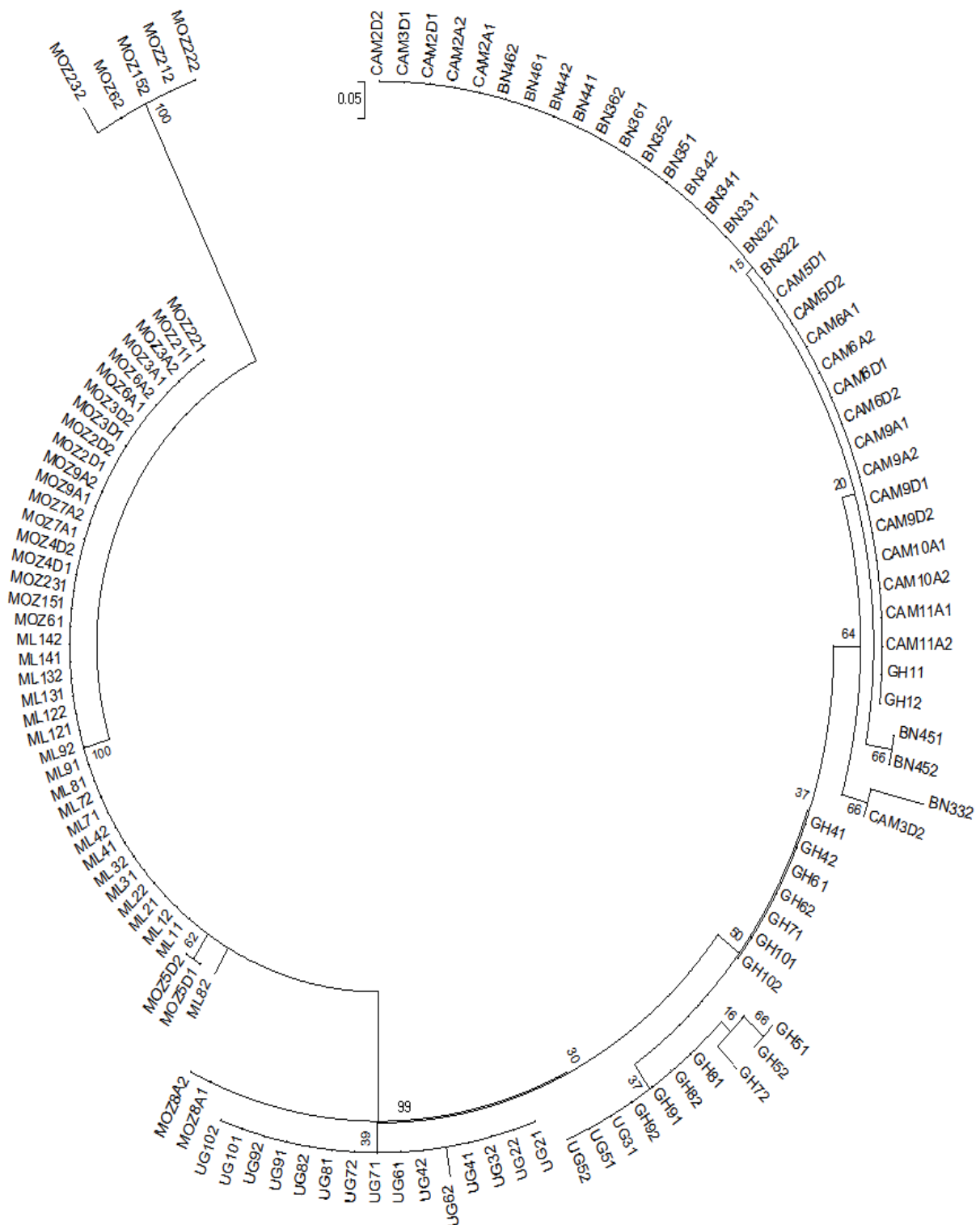
(b) Non synonymous protein variants

Fig. 5.8: Maximum likelihood tree of the *CYP6P9a* gene amplified for six samples (BN – Benin, CAM – Cameroon, GH- Ghana, MAL – Malawi, MOZ – Mozambique and UG - Uganda).

Fig. 5.9: Coding maximum likelihood tree of the CYP6P9a gene for the six individual countries – (a)Benin, (b)Cameroon, (c)Ghana, (d)Malawi, (e)Mozambique and (f) Uganda.

181

5.3.2.2 Haplotype network with TCS

As shown by the ML phylogenetic tree, the haplotype network for *CYP6P9a* using TCS shows that haplotypes from the southern populations are less diverse as separated by few mutational steps between them. For example, 7 out of 13 southern African haplotypes are only separated by 1 mutational step from the predominant MOZ/MAL33 resistant haplotypes. A similar pattern is observed with each country as within Malawi, there is just one mutational step between most haplotypes. Similarly, most Mozambique haplotypes have few mutational steps between them apart from the 3 mosquitoes already shown to be different from the rest. This reduced haplotype diversity in Malawi and Mozambique supports that *CYP6P9a* is under directional selection in these southern African countries. However, southern African haplotypes are highly different from those from other countries with high number of mutational steps observed leading to southern African haplotypes forming a separate group due to the presence of more than 20 mutational steps between these haplotypes. For example, more than 20 mutational steps difference are observed between MOZ/MAL33 and BN1 (Fig. 5.10). Analysis of haplotypes from other regions revealed that they are more diverse with more mutational steps between haplotypes within each country (between 1 to 13 mutational steps) than in southern Africa, suggesting that selection if present is less marked in these regions. A similar pattern was for the coding region only (Fig. 5.10) and for the haplotypes from nonsynonymous substitutions only (Fig. 5.11). For the haplotype network for the non-synonymous equivalent of amino acid change, the most predominant haplotype was the pyrethroid resistant MAL/MOZ9 haplotype with a frequency of 38% present only in southern Africa where the reduced diversity is further revealed. In other region, the BN/CAM/GH1 haplotype is predominant in West/Central Africa (34%) but still with more diversity than in southern Africa with Uganda located at an intermediary position (Fig. 5.11).
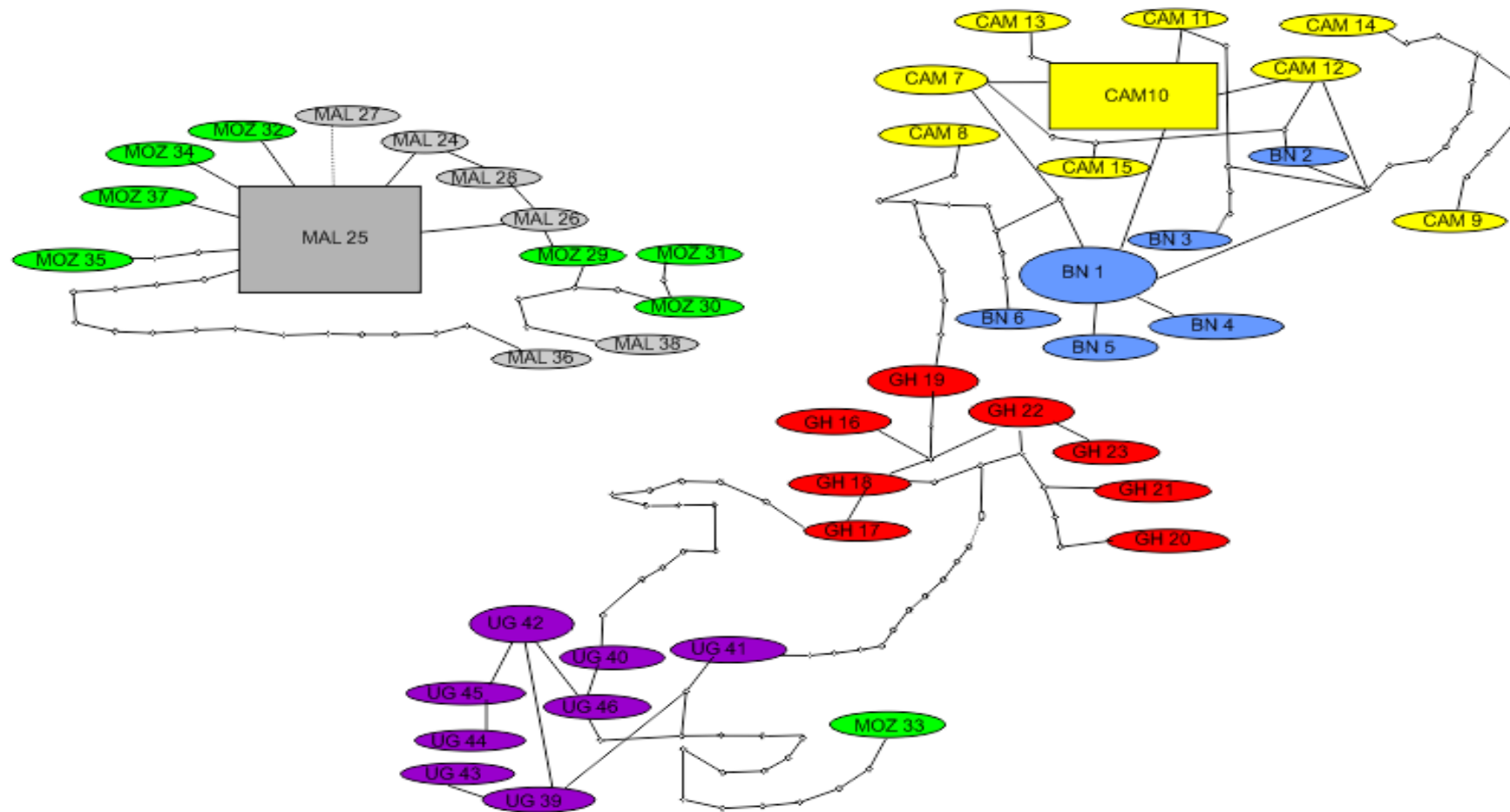
Fig. 5.10: Haplotype network of *CYP6P9a* for the entire six countries for coding region. The size of the polygon reflects the frequency of the haplotype and colour represents the countries (BN (Benin) –Blue, CAM (Cameroon) –Yellow, GH (Ghana) –Red, MAL (Malawi) -Grey, MOZ (Mozambique) –Green and UG (Uganda) –Purple). Segregating mutation is represented by each node and rectangular boxes represent major haplotype.
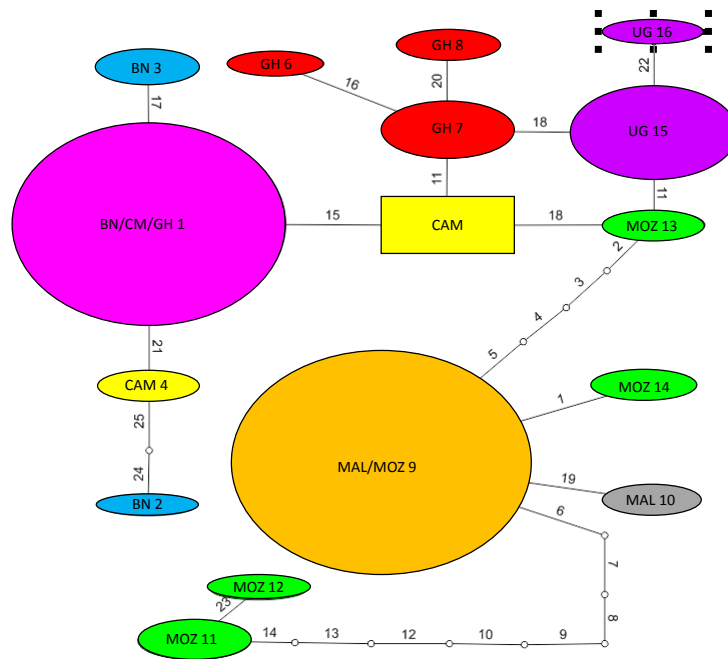
Fig. 5.11: Haplotype network for non-synonymous equivalent of amino acid change for *CYP6P9a*. Orange colour represents haplotype shared between MAL (Malawi) and MOZ (Mozambique), Pink represent haplotype shared between BN (Benin), CAM (Cameroon), and GH (Ghana), Blue represents BN (Benin), Red represents GH (Ghana), Yellow represents CAM (Cameroon), Green represents MOZ (Mozambique), Grey represents MAL (Malawi) and Purple represents UG (Uganda).
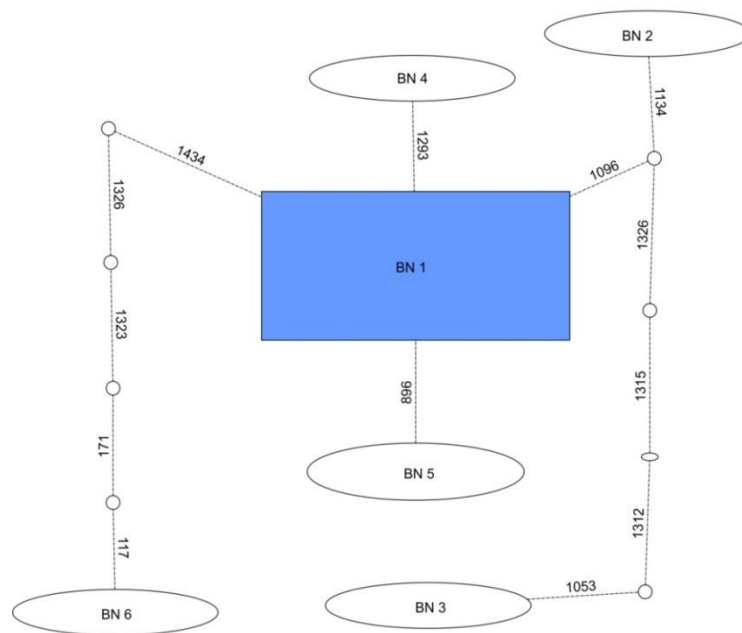
(a) Benin
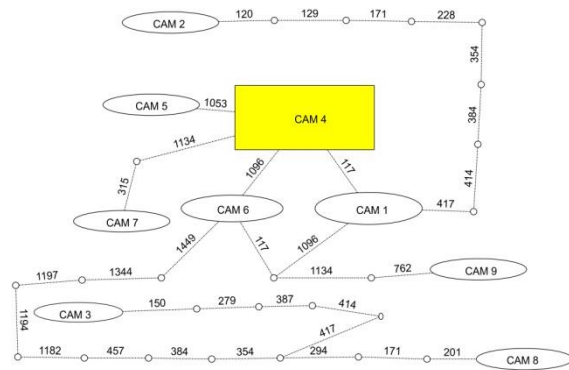
(b) Cameroon
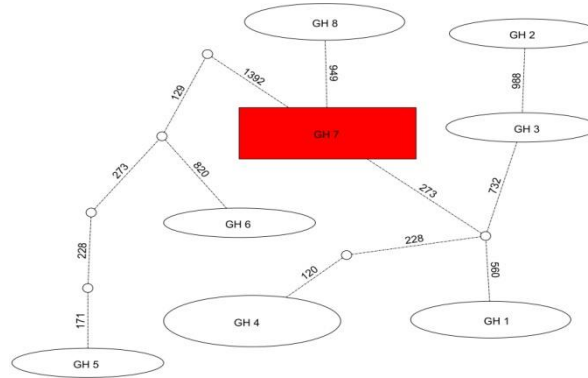
(c) Ghana
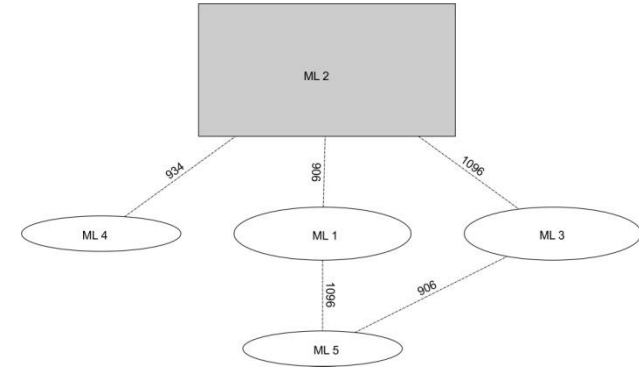
(d) Malawi
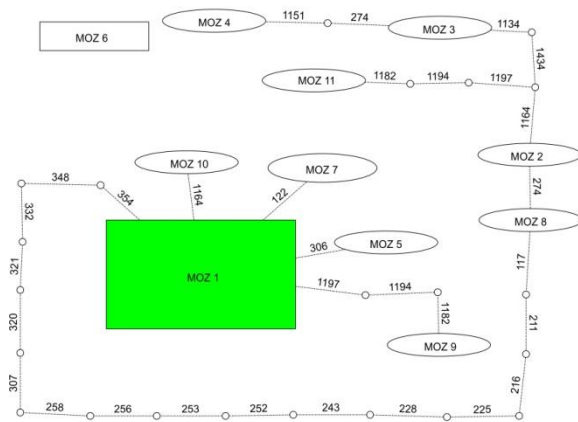
(e) Mozambique
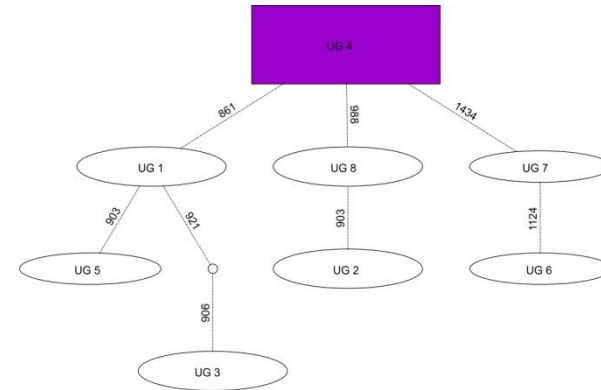
(f) Uganda

Fig. 5.12: Haplotype network of *CYP6P9a* for individual countries for coding regions. The size of the polygon reflects the frequency of the haplotype. Segregating mutation is represented by each node and polymorphic positions are given above the branches. Rectangular boxes represent major haplotype.

5.3.3 Test for selection for *CYP6P9a* gene

To further assess the selection observed on *CYP6P9a,* other tests of selection such as the HKA, MK, ka/ks and dN/dS ratios were performed but these did not show any signature of selection on this gene even in Malawi (Table 5.4). This could be due to the fact that the selective sweep might already be extensive in southern African and about to reach near fixation. In a case of near fixation, it has been shown that signature of selective sweep is better revealed rather by reduced levels of genetic variations as observed in this study than by through the additional tests performed here (Biswas and Akey 2006). For the MKA test, Cameroon and Mozambique were significant at P<*0.001* for the *CYP6P9a* gene (Table 5.4).

Table 5.4: Selection parameter of *CYP6P9a*

	MK					HKA				dN/dS and Z test			Ka/Ks		
	Fixed Poly Morphic between		Poly Morphic With species		NI	Pvalue	X^2	Pvalue	dN	dS	dN/dS	Pvalue	Ka x10^3	Ks x10^3	Ka/Ks
	syn	Nsyn	Syn	Nsyn											
Benin	251	247	8	4	0.51	0.38	0.82	0.36	0.186	0.840	0.221	1.0	1.1	3.6	0.31
Cam	248	247	24	1	0.04	0.00***	4.6	0.03*	0.418	0.00	Und	0.34	0.009	12.49	0.007
Ghana	249	246	8	4	0.51	0.38	2.86	0.09	0.097	0.023	4.22	0.47	0.99	5.46	0.18
Mal	257	242	2	1	0.53	1.0	1.2	0.27	0.951	0.00	Und	1.0	0.009	1.99	0.004
Moz	246	238	35	14	0.41	0.007	14.1	0.000***	0.620	1.724	0.359	0.01	2.82	19.98	0.14
Uganda	251	246	5	2	0.41	0.45	2.4	0.12	0.887	0.150	3.913	0.23	0.36	4.19	0.09
Total	239	238	59	25	0.43	0.001***	11.1	0.001***	1.85	1.69	1.094	1.0	3.61	49.81	0.07

NI: neutrality index; MK; Mcdonald and Kreitman test; HKA: Hudson, Kreitman and Aguade test; Und: Undetermined, nc: Not calculated, Star (*) denotes significant at P<*0.05* and (***) denotes P<*0.0001*.

The test result of codon-based test with the null hypothesis that dN<dS (so that there is purifying selection acting on the *CYP6P9a* gene) showed that there is a significant probability at P<*0.05* for all countries except for Malawi. The lack of significant P-value in Malawi is an indication that this population with the predominant allele is moving to positive selection (Table 5.5). The presence of significant value in Mozambique is probably due to the presence of other genetically polymorphic mosquitoes despite the presence of the predominant haplotype from Malawi. Thus the presence of positive (directional) selection is masked by the presence of some susceptible mosquitoes. Sequencing of a higher number of mosquitoes in Mozambique may help to better assess the selection in Mozambique.

Table 5.5: Codon based Z-test on the hypothesis that dN<dS

	P value	dS-dN
BN	0.012	2.300
CAM	1E-06	4.964
GH	0.21	2.056
MAL	0.088	1.361
MOZ	8E-07	5.063
UG	0.023	2.012

5.3.4 Sequencing of BAC fragments around rp1 QTL

To further assess the extent and scale of the selection around *CYP6P9a* in pyrethroid resistant populations of southern Africa, the polymorphism of five loci within a genomic region around *CYP6P9a* across the previously sequenced120kb rp1 BAC clone was analysed for the resistant Malawi and Mozambique population in comparison to the more susceptible Cameroon population.

5.3.4.1 Polymorphism patterns

Across the five loci analysed around *CYP6P9a*, the Malawi population consistently exhibited a lower polymorphism level to that of Cameroon from -9 to +86kb. A similar pattern was observed in Mozambique although a lower polymorphism to Cameroon was observed only from -9 to +61kb. The extent of this low polymorphism is characterised by the complete absence of polymorphic site at +61kb from *CYP6P9a* (BAC95) or only 1 polymorphic site at +86kb or 2 sites at +36kb from *CYP6P9a*. This low polymorphism is also noted in Mozambique but less marked and mainly focused around *CYP6P9a* with

only 1 polymorphic site observed at -9kb from *CYP6P9a*. At +86kb from *CYP6P9a*, the polymorphism in Mozambique has already completely recovered and is similar to that of Cameroon.

The more susceptible sample from Cameroon showed higher nucleotide diversity (π) than the more resistant southern populations of Mozambique and Malawi (Table 5.6). In correlation with the polymorphism level, nucleotide diversity was significantly lower in Malawi over a distance of more than 95kb from -9kb of *CYP6P9a* to +86kb (Fig. 5.13, Fig. 5.14 and Fig. 5.15). This is confirmed by a striking reduction of the relative diversity in Malawi over the same distance as shown by the lower ratio of the genetic diversity (π) between Malawi and Cameroon (Fig. 5.16). A similar reduced genetic diversity is observed in Mozambique but over a shorter distance from -9kb of *CYP6P9a* to +61kb (Fig. 5.16). Overall the signal of this reduced diversity is stronger and more extensive in Malawi than in Mozambique where the selection appears to be more focal around *CYP6P9a*. Analysis of the population from Cameroon shows that the relative diversity remains high across the *rp1* genomic fragment although a slight reduction is observed at *CYP6P9a* compared to other loci (Fig. 5.13).
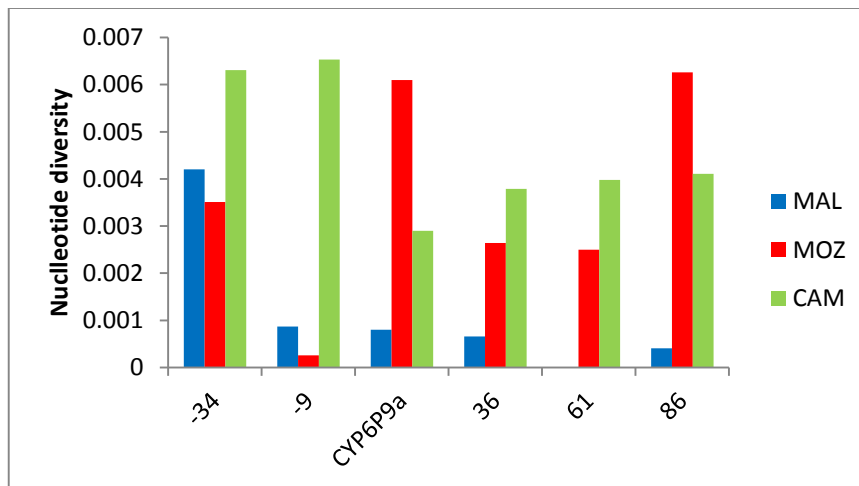
Fig. 5.13: Distribution of nucleotide diversity between samples (Cameroon- Central Africa) and southern population (Malawi and Mozambique) across the BACs with *CYP6P9a* gene.
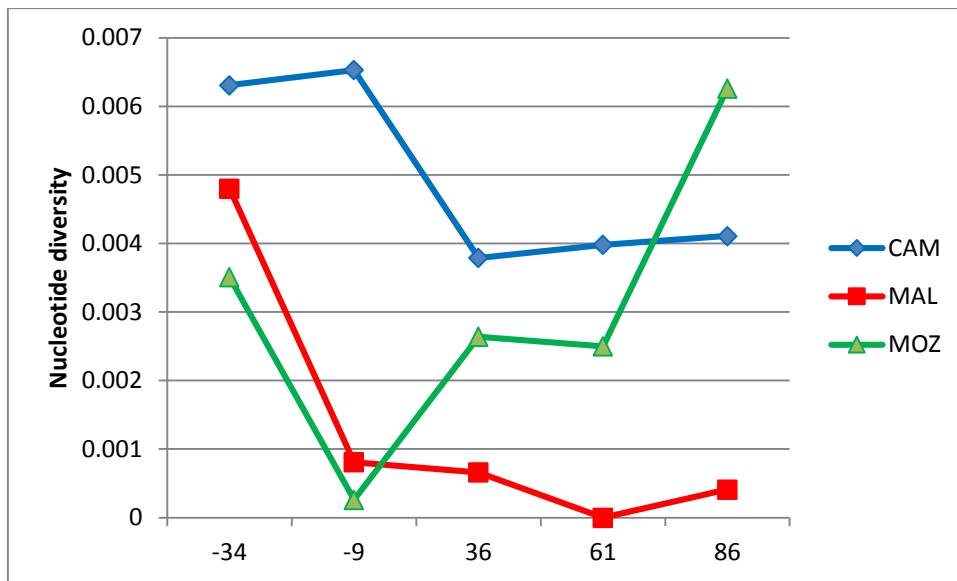
Fig. 5.14: Distribution of nucleotide diversity between samples (Cameroon- Central Africa) and southern population (Malawi and Mozambique) across the BACs without *CYP6P9a* gene.

The more resistant populations showed lower nucleotide diversity for all the BACs, however, the lowest nucleotide diversity was observed for Malawi. This can be clearly seen in (Fig. 5.15), where the graphical representation indicates the resistant population exhibiting overall reduced nucleotide diversity.
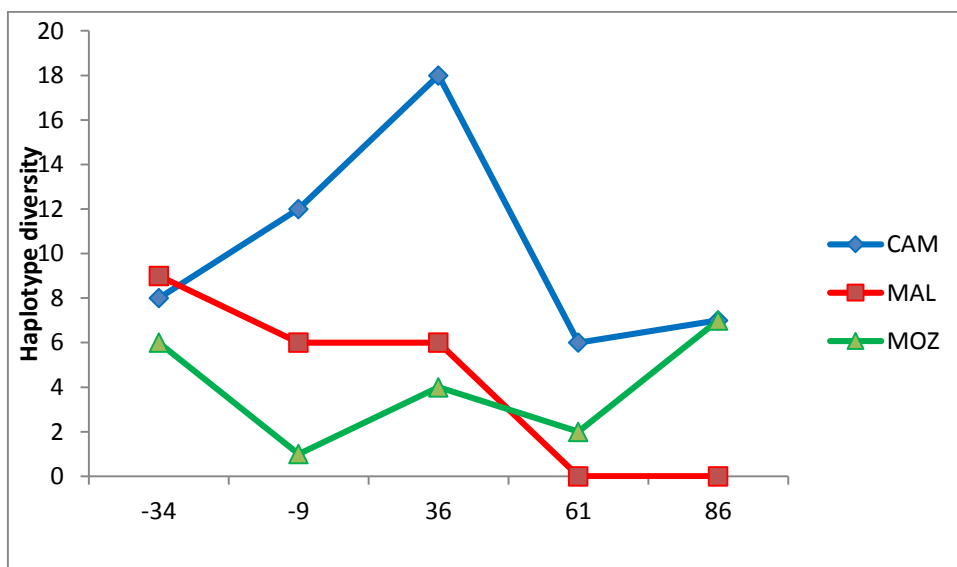
Fig. 5.15: Haplotype Diversity (h) across the BACs CAM (Cameroon), MAL (Malawi), MOZ (Mozambique).

Malawi showed a much reduced haplotype diversity from -9 to +86 and almost no diversity at +61 and +86. The pattern was also similar for Mozambique even though

Mozambique showed a much reduced haplotype diversity at -34 to Cameroon than Malawi. In any case, at +86, Mozambique and Cameroon had a similar haplotype diversity.

Table 5.6: Summary statistics for polymorphism for BACs

Entire Sequence Set								Coding Region						
Samples	N	S	π	Taj D	Fu D	S*	h	N	S	π	Taj D	Fu D	S*	h
BAC 0 CAM	14	16	0.00631	-0.8857^{ns}	-1.1266^{ns}	10	8 (0.868)	-	-	-	-	-	-	-
BAC 0 MAL	14	12	0.0048	-0.6075^{ns}	0.7426^{ns}	2	9 (0.934)	-	-	-	-	-	-	-
BAC 0 MOZ	10	8	0.00351	-0.7545^{ns}	-0.3853^{ns}	4	6 (0.844)	-	-	-	-	-	-	-
BAC 0 ALL	38	36	0.01458	0.2499^{ns}	0.05708^{ns}	9	23 (0.964)	-	-	-	-	-	-	-
BAC 25 CAM	16	22	0.00653	-1.0014^{ns}	-1.0999^{ns}	12	10 (0.867)	16	22	0.00653	1.13789^{ns}	-1.2023^{ns}		10 (0.867)
BAC 25 MAL	18	6	0.00087	-2.034^{ns}	-2.8307^{ns}	6	2 (0.111)	18	6	0.00087	-2.0342^{ns}	-2.83076	6	2 (0.111)
BAC 25 MOZ	10	1	0.00026	-1.1117^{ns}	-1.2434^{ns}	1	2 (0.200)	10	1	0.00026	-1.11173^{ns}	-1.2434^{ns}	1	2 (0.200)
BAC 25 ALL	44	34	0.00728	-1.0570^{ns}	-2.6449^{ns}	18	13 (0.640)	44	34	0.00728	-1.05708^{ns}	-2.6449^{ns}	19	13 (0.640)
BAC 25 MAL	18	6	0.00087	-2.0342^{ns}	-2.8307^{ns}	6	2 (0.111)	-	-	-	-	-	-	-
BAC 70 CAM	14	10	0.00379	-0.9001^{ns}	-0.2576^{ns}	4	6 (0.802)	-	-	-	-	-	-	-
BAC 70 MAL	18	2	0.00066	-0.6848^{ns}	0.8846^{ns}	0	2 (0.209)	-	-	-	-	-	-	-
BAC 70 MOZ	12	6	0.00264	-0.5854^{ns}	0.70614^{ns}	0	4 (0.561)	-	-	-	-	-	-	-
BAC 70 ALL	44	29	0.01136	0.29613^{ns}	0.74291^{ns}	4	11 (0.687)	-	-	-	-	-	-	-
BAC 95 CAM	12	13	0.00398	-1.6589^{ns}	-1.759^{ns}	10	6 (0.682)	-	-	-	-	-	-	-
BAC 95 MAL	12	0	0	0	0	1	0	-	-	-	-	-	-	-
BAC 95 MOZ	10	5	0.0025	0.02348^{ns}	1.30011^{ns}	0	2 (0.356)	-	-	-	-	-	-	-
BAC 95 ALL	34	25	0.00807	-0.3420^{ns}	-0.6422^{ns}	8	9 (0.656)	-	-	-	-	-	-	-
BAC 120 CAM	10	10	0.00411	-1.4324^{ns}	-1.2537^{ns}	7	4 (0.644)	-	-	-	-	-	-	-
BAC 120 MAL	14	1	0.00041	-0.3414^{ns}	0.71557^{ns}	0	2 (0.264)	-	-	-	-	-	-	-
BAC 120 MOZ	12	11	0.00626	0.05987^{ns}	0.11352^{ns}	7	9 (0.909)	-	-	-	-	-	-	-
BAC 120 ALL	38	22	0.00431	-1.7480^{ns}	-1.4996^{ns}	9	14 (0.606)	-	-	-	-	-	-	-

N, number of samples; S, number of segregating sites; π, nucleotide diversity per site; Taj D Tajima's D; Fu D, Fu and Li's D*; S*, singletons; h (hd)haplotype diversity; ns denotes not significant.

5.3.4.2 Phylogenetic Analysis of BAC fragments

To visualize the phylogenetic relationships between the haplotypes and to gain additional insights into the evolutionary history of the three samples, a maximum likelihood neighbor joining tree was constructed.

The BAC sequences showed an overall maximum likelihood network with few reticulations for Malawi and Mozambique population than Cameroon. The Malawi and Mozambique population formed a clade of theirs while Cameroon formed its own clade (Fig. 5.16a, b, c and d). This was the case from -34 to +61 except for +86, where some of Mozambique samples shares a clade with Cameroon (Fig. 5.16e). The Malawi and Mozambique populations were also very distant from that of Cameroon. This might also be linked to their natural geographic locations in Africa and their resistant profiles.

Across all the BACs, the Malawi and Mozambique population shared a common haplotype while the Cameroon population had its own haplotypes (Fig. 5.16a, b, c and d) except for +86 where both the more susceptible Cameroon and the more resistant southern populations shared some haplotypes although Cameroon exhibits a higher haplotype diversity (Fig. 5.16e).
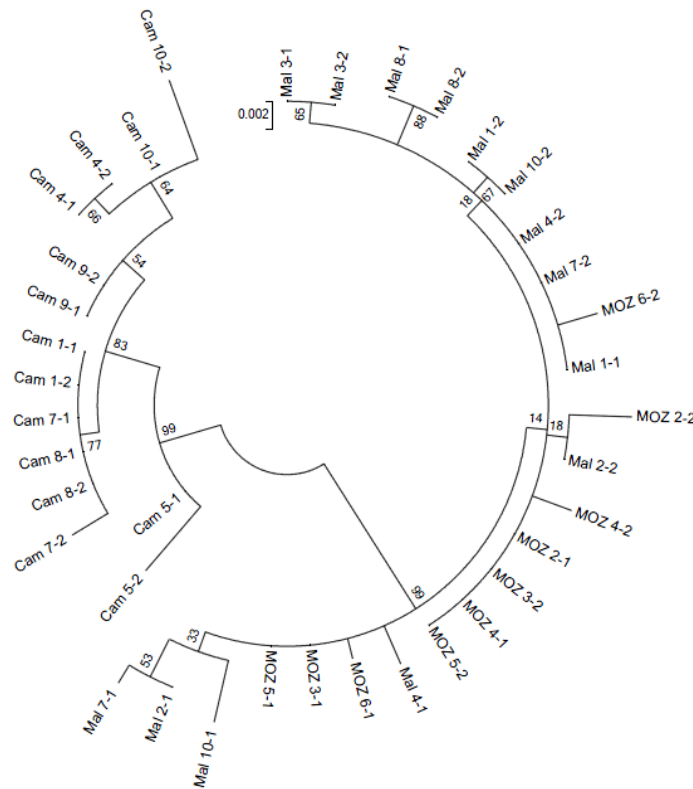
(a) -34kb (BAC 0)
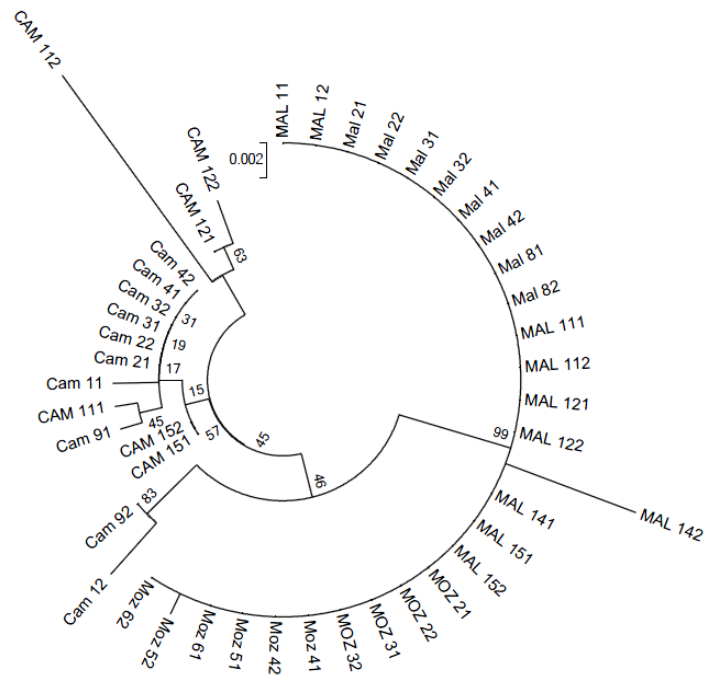
(b) -9 kb (BAC 25)
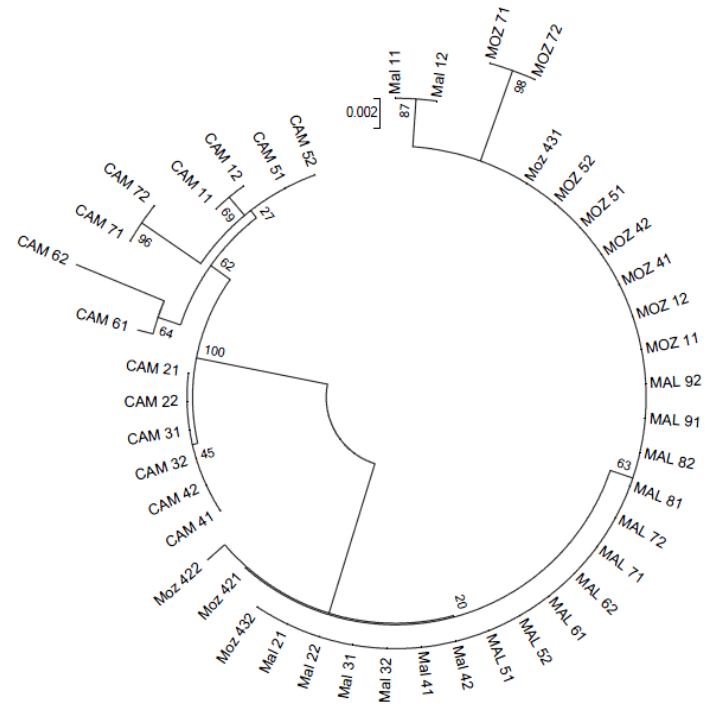
(c) +36 kb (BAC 70)
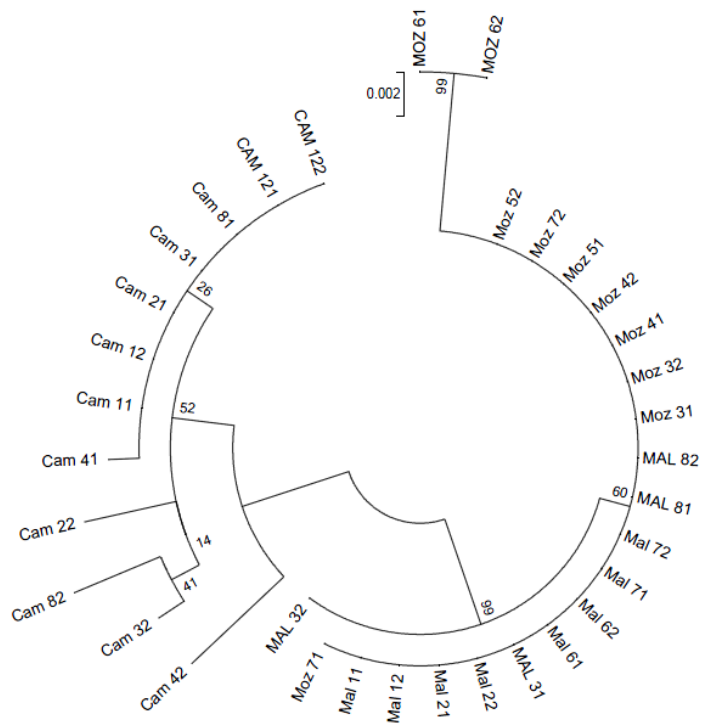
(d) +61kb (BAC 95)
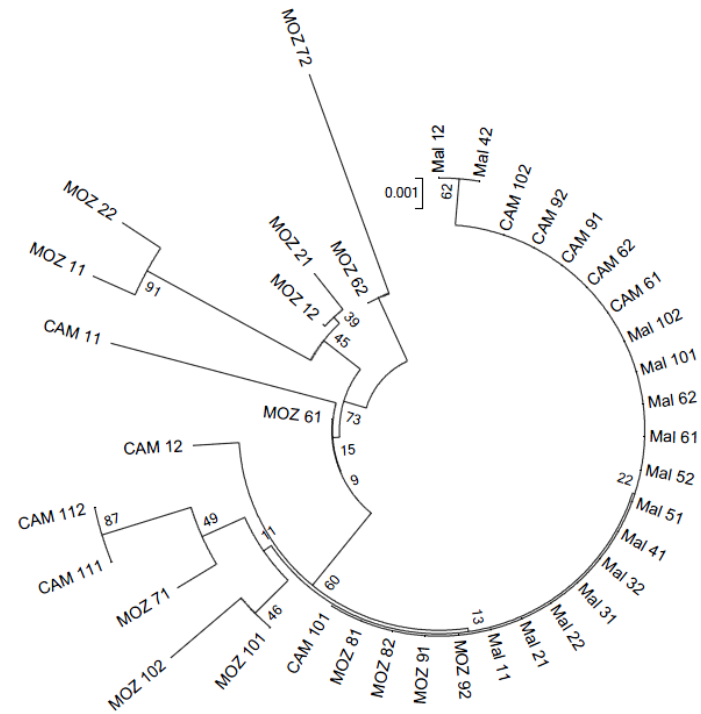
(e) +86kb (BAC 120)

Fig. 5.16: Maximum likelihood tree of the five BAC loci amplified for resistant samples (MAL – Malawi and MOZ – Mozambique) and susceptible sample (CAM – Cameroon)

```
                   1111222222422222223333334444445555 6
                   4975933569011224466689922248000126234 7 4
 Hap#  n1456791178201201207956716806178034724 0 9
CAM1   5GGATTTTACCA-CAACCAGGACGTCATGGGT-GGGGGAT
CAM2   2..T........-T.....A..........AG-.......
CAM3   1.AT.......-.................A-.......
CAM4   1.ATC.......-.......A.......A..A-..A....
CAM5   1...........-..G..G.............-.......
CAM6   2...........-.................AG-.......
CAM7   1...........-T.....A..........AG-.......
CAM8   1......GG...-T...G.A..........AG-......A
MOZ9   4.AT.-A..TA.ATG.T...AG..C.G....A-....AG.
MOZ10  1.AT.-A..TAGATG.T...AG..C.G....A-..AA.G.
MOZ11  2.AT.-A..TA.ATG.T....G..C.G....A-....AG.
MOZ12  1.AT.-A..TAGATG.T...AG..C.G.A..A-....AG.
MOZ13  1AAT.-A..TA.ATG.T....G..C.G....A-....AG.
MOZ14  1AAT.-A..TA.ATG.T...AG..C.G....A-A.A.AG.
MAL15  3.AT.-A..TA.ATG.T...AG..C.G....AC..A.AG.
MAL16  2.AT.-A..TA.ATG.T...AG..C.G...A.AC..A.AG.
MAL17  1.AT.-A..TA.ATG.T....G..C.G.A..ACA.A.AG.
MAL18  1.AT.-A..TA.ATG.T...AG..C.G....AC...AAG.
MAL19  1.AT.-A..TA.ATG.T...AG..CAG....AC..A.AG.
MAL20  1.AT.-A..TA.ATG.T....G..C.G....AC.AA.AG.
MAL21  1.AT.-A..TA.ATG.G....G..C.G.A..ACA.A.AG.
MAL22  1.AT.-A..TA.ATG.T...AGGCC.G....AC..A.AG.
MAL23  1.AT.-A..TA.ATG.G....G..C.GG...ACA..AG.
```

(a)   -34kb from CYP6P9a (BAC0)

```
                       1111111111112222344446667
                   34677778881134455668000 15823681263
 Hap#   n          91716780174605609120247 68638508432
MAL/MOZ1  26       TGAGAAAAGCGATTCTTCCGCATCGGTCGCAAAA
MAL2       1       ..........C...G.GGGA..............
CAM3       1       ...A.CC........C........A....GGT.T
CAM4       1       GC.ACGC..A....................GGT.T
CAM5       6       ...A..CC.......C........A....GGT.T
CAM6       1       ......CCC......C........A....GGT.T
CAM7       1       GC....C..A....................GGT.T
CAM8       1       ......CC.A.....C........A....GGT.T
CAM9       1       ..T...CC.A.TA.......TGCA.AC.AGGTGT
CAM10      1       ...A..CC.................ACT.GGT.T
CAM11      1       ...A..CC.......C.......AACT.GGT.T
CAM12      2       ...A..CC.......C............GGT.T
MOZ13      1       ............C.....................
```

(b)   -9kb from CYP6P9a (BAC 25)
```

```
                 1111112223333334444445555556
                 8891144490782668900023524578 3
Hap#      n      0902705623144467115648220246 7
CAM1      2      ACAGACCACTGAGGAACGGCGTATGTATC
CAM2      6      ......G........C............
CAM3      2      ...............C............
CAM4      1      ...............C......G......
CAM5      1      CT..CT.........C......G......
CAM6      2      ..........A..A.C........A....
MAL/MOZ8  24     ....C.G.AA.G...CAATAAC.A.ATCG
MOZ9      2      ...AC.GCAA.GC.GC.AT.AC.A.ATCG
MOZ10     1      ......G.AA.G...C.AT.AC.A.ATCG
MOZ11     1      ..T...G.AA.G...C.AT.AC.A.ATCG
```

(c)+36kb from CYP6P9a (BAC 70)

```
                 1111222222333334445566 6
                 7811452224673367804502478
 Hap#     n      3206883672584658016569783
CAM1      7      GAATCTCCGAATTTACACATCTCTA
CAM2      1      .G.C.......CC............
CAM3      1      .G.......G.....G.........
CAM4      1      ....................A....
CAM5      1      .G.......G....G.G.G.T.T..
CAM6      1      ....G........G.G.........
MAL/MOZ7  19     A.T...........G..TGC.A.AG
ML8       1      AGT...........G..TGC.A.AG
MOZ9      2      A.T..AAAA.T...G..TGC.A.AG
```

(d) +61kb (BAC 95)

```
                     1122222223373 4444444
                     7773922456770 0361133559
Hap#              n  4690069418792 8353438595
MOZ1             1   CCTCGGAGAGTAATAGCGCACGT
MOZ2             1   ...T.AG.G.C............
MOZ3             1   ..AT.AG.G.C............
MOZ4             1   ..GT..................
MOZ5             1   ...T.AG.G.C.......G....
MOZ6             1   ...T..G.G.C.C.....G....
MOZ7             1   ...T.AG.G.C...C.TAG..A.
MOZ8             1   A..T..G.G.C.CC....GGTA.
MAL/MOZ/CAM9    22   ...T.AG.G.C.....TAG....
MOZ10            1   ...T.AG.G.C.....TAG...G
MOZ11            1   ...T..G...C.....TAG...G
MAL12            2   .T.T.AG.G.C.....TAG....
CAM13            1   .....AGAGACG...C..G....
CAM14            1   ...T.AG.G.C....TTAG....
CAM15            1   ...TA.G.G.C...C.TAG....
```

(e) +86kb (BAC 120)

Fig. 5.17: Schematic representation of the haplotypes of BAC loci across the more resistant population (MAL – Malawi and MOZ – Mozambique) and the more susceptible samples (CAM – Cameroon). The polymorphic positions are indicated with numbers besides the haplotype (haplotypes are labelled by the country initials where the haplotype is predominant (Hap)), and the second numbers (n) indicates the haplotype frequency.

5.3.4.3 Haplotype network at the *rp1* loci

Analysis of the haplotype network using TCS (Clement *et al.* 2000) confirms profiles obtained with the maximum likelihood phylogenetic trees for the 5 loci from -34kb to +86kb from *CYP6P9a*. Across all the loci, the haplotypes were distributed according to profiles of pyrethroid resistance. In general, the more resistant populations of Malawi and Mozambique formed their own clusters across the rp1 genomic region with evidence of significant reduced genetic diversity shown by the fact that haplotypes from both countries had only significantly less mutational steps between them (1-3 steps), while the more susceptible samples formed its own clusters with higher mutational steps (4-16 steps) for almost all of the BACs except for +86 (BAC 120). Where the more resistant population shared haplotypes with the more susceptible samples, there were higher mutational steps between Malawi and Cameroon for -34 to+61 than between Malawi and Mozambique except for +86 (BAC 120). However, for +86, Mozambique (more resistant) had already recovered its diversity as it has many mutational steps than even the more susceptible samples with both populations sharing some common haplotypes (Fig. 5.18a,

b, c, d and e). This suggests that the selective sweep on the rp1 does not go beyond +86kb from *CYP6P9a*.

Individual analysis of these loci shows that no signature of selection is observed at -34kb from *CYP6P9a* (BAC0) as a high aplotypic diversity is noted at this locus. No predominant haplotype is observed with the most common haplotype CAM1 having only a low frequency of 13.1% (5/38) and found only in the more susceptible sample from Cameroon. A high proportion of haplotypes were singletons (16 out of 23) even in Malawi (7) and Mozambique (4). Additionally, the haplotype diversity was high with high number of mutational steps observed between haplotypes from each (>12) or between (>20) country (Fig. 5.17).

A drastic change is observed at -9kb from *CYP6P9a* where a strong signature of selection probably from pyrethroid resistance is observed in both Malawi and Mozambique. This reduced haplotype diversity is shown by the presence of a highly predominant haplotype MAL/MOZ1 only found in Malawi and Mozambique with a frequency of 59.1% (26/44) for the entire sample and 92.8% in Malawi and Mozambique. In contrast the most predominant haplotype in Cameroon only has a frequency of 13.6% (6/44) with a high proportion of singletons observed in Cameroon (8 out of the 10 haplotypes). Additionally, only one haplotype exhibits a difference of 6 mutational steps to the other two haplotypes in Malawi and Mozambique (Fig. 5.18) which further supports the directional selection acting on this locus probably because of the proximity with the *CYP6P9a* resistance gene. To contrast in Cameroon, the ten haplotypes are highly diverse with more than 20 mutational steps difference between some of them supporting an absence of selection in this sample.

A similar reduced haplotype diversity is observed at +36kb and +61kb in Malawi and Mozambique with for example a difference of only 1 mutational step between the haplotypes of Malawi and Mozambique at the +36kb loci in contrast to those from Cameroon (>6 steps). For the +61kb loci a predominant haplotype is still present in Malawi and Mozambique with only a single haplotype from Mozambique (MOZ9) having more than one mutational step difference to other haplotypes.

However at the +86kb loci, a reduced haplotype diversity is only observed in Malawi not in Mozambique which is highly diverse as in the susceptible Cameroon. If the Malawi population remains less diverse with only two haplotypes observed, a high proportion of

singleton haplotypes is found in Mozambique with more than 13 mutational steps difference between them. This diversity is even higher than in Cameroon where only 5 haplotypes are noted with only 1 haplotype having a difference of 10 mutational steps to others.

Overall, these analyses show that the pyrethroid resistance selection acting on the rp1 QTL genomic region locus is more extensive in Malawi than in Mozambique. In Malawi, the region under selection spans at least from -9kb from *CYP6P9a* and beyond +86kb which reasonably could be above a region of 100kb. In Mozambique the region under selection spans from -9kb to +61kb which is around 70kb. In addition to the larger region under selection in Malawi, it is also observed that the selection is stronger than in Mozambique because at loci +36 and +61kb, the reduction of haplotype diversity is higher in Malawi than Mozambique.

(a) -34kb (BAC0)

(b) 9kb (BAC25)

(c) BAC 70

(d) +36 (BAC 95)

(e) +61 (BAC 120)

Fig. 5.18: Haplotype network of rp1 loci for more resistant southern population (MAL – Malawi (Gray colour) and MOZ – Mozambique (Green colour)) and the more susceptible samples from CAM (Cameroon (Yellow colour)). Pink coloured polygon represents haplotype dominant in both Malawi and Mozambique. The size of the polygon reflects the frequency of the haplotype. Segregating mutation is represented by each node and polymorphic positions are given above the branches. All the networks were generated with the same scale from TSC and the scaling for frequency was area of the shapes.

5.3.5 Test for selection for BAC fragments

The reduced genetic diversity seen for the resistant samples and not for the susceptible samples suggest that there is possible selection acting around the BAC fragments for the resistant populations from southern Africa (Malawi and Mozambique). Negative Tajima results observed from -34 to +86 further confirms that there is an indication of purifying selection since negative Tajima values implies a possible purifying selection (Tajima 1986).

Further test carried out using K_{ST} also confirms that Malawi and Mozambique are not significantly genetically differentiated from -34 to +61, while both countries are significantly genetically differentiated to Cameroon from -34 to +61. However, for +86, Malawi and Mozambique are both genetically significant to Cameroon at *0.01<P<0.05*, while Malawi is highly significant to Mozambique at *P<0.001* (Table 5.7) The K_{ST} values thus confirm the signature of selective sweep in Malawi and Mozambique since reduced genetic differentiation is linked to a common selection factors homogenising the genetic profiles at these position with selected predominant haplotypes.

Table 5.7: K_{ST} values

	CAM/MAL	CAM/MOZ	MAL/MOZ
BAC 0	0.38234***	0.35197***	0.06986^{ns}
BAC 25	0.29986***	0.26060***	0.59670^{ns}
BAC 70	0.32634***	0.19697***	0.0910^{ns}
BAC 95	0.37273***	0.29118***	0.06033^{ns}
BAC 120	0.03462*	0.05193*	0.13133***

PERMTEST calculates Hudson's K_{ST} statistic of genetic differentiation. KST is equal to 12 K_S/K_T where K_{ST} is a weighted mean of K1 and K2 (mean number of differences between sequences in subpopulations 1 and 2, respectively) and K_T represents the mean number of differences between two sequences regardless of their subpopulation. The null hypothesis of no genetic differentiation will be rejected (P<*0.05*) when K_S is small and K_{ST} is close to 1.PM test; Probability obtained by the permutation test with 1000 replicates); ns, not significant; *, *0.01<P<0.05*; **, *0.001<P<0.01*; ***, P<*0.001*

However, further tests of selection such as the HKA, MK, ka/ks and dN/dS ratios on the only locus with a coding region [-9kb loci (BAC25) spanning the *CYP6AA2* P450 gene], did not detect any significant signature of selection (Table 5.8) probably because of the near fixation of the selective sweep at this locus.

Table 5.8: Selection parameters on the -9kb locus (BAC 25)

	MK				NI	P value	X^2	P value	dN	dS	dN/dS	P value	Ka x10^3	Ks x10^3	Ka/Ks
	Fixed polymorphic between species		Polymorphic with species		HKA					dN/dS and Z test			Ka/Ks		
Cameroon	13	10	0	0	0.35	0.02	2.45	0.12	0.003	0.002	1.5	0.11	0.003	0.017	0.176
Malawi	2	40	0	0.90	1.0	4.2	0.04	0.001	0.001	0.000	Und	0.09	0.001	0.001	1
Mozambique	0	1	1	0	0.0	1.0	0.89	0.34	0.00	0.00	0	1.0	0.0	0.0	0
Total	19	16	0	0	0.38	0.01	2.992	0.083	0.02	0.00	Und	0.002	0.002	0.021	0.095

NI: neutrality index; MK; Mcdonald and Kreitman test; HKA: Hudson, Kreitman and Aguade test; Und: Undetermined, nc: Not calculated

5.4 Discussion

This study has used analysis of polymorphism of the pyrethroid resistance gene *CYP6P9a* and surrounding loci to validate the presence of a signature of selective sweep around the rp1 pyrethroid resistance QTL and also assess the extent and strength of this selection across Africa. These analyses have further strongly supported the observation in previous chapters using microsatellite loci that pyrethroid resistance in southern Africa has a selection footprint in the genomic region harbouring the main resistance genes.

5.4.1 Directional selection acting on *CYP6P9a* gene

The presence of a signature of selective sweep around the rp1 QTL has been supported in this chapter by the analysis of the polymorphism patterns of the *CYP6P9a* pyrethroid resistance gene across Africa which showed that this gene is under strong directional selection in southern Africa populations of Mozambique and Malawi. There was an overall significant reduced polymorphism in Malawi and Mozambique known to be highly resistance to pyrethroid insecticides (Brooke *et al.* 2001, Cuamba *et al.* 2010, Wondji *et al.* 2012, Riveron *et al.* 2013). This reduced polymorphism was consistent in southern Africa and observed at various levels such as number of substitutions, genetic and haplotype diversities, patterns of genetic differentiation or phylogenetic analysis. Such reduced genetic diversity could only be linked to a selection pressure exerted by pyrethroids on these *An. funestus* populations possibly through vector-based control intervention such as LLINs and IRS which are largely implemented in these countries (Casimiro *et al.* 2006, Wondji *et al.* 2012). Even though Mozambique exhibited higher levels of polymorphism than Malawi, this was as a result of very few mosquitoes contributing to the polymorphism pattern since the polymorphisms of these mosquitoes were similar to that of susceptible mosquitoes. Analysis carried out without these mosquitoes saw a reduced polymorphism for Mozambique as that seen for Malawi. However, the close proximity of these two countries might also be the reason why they are exhibiting similar polymorphism pattern although the BAC sequencing showed that such similarity disappeared when moving away from *CYP6P9a*.

The presence of the highly predominant haplotype in Malawi and Mozambique is exactly the same as the pyrethroid resistant haplotype recently found in both countries when comparing the resistant and susceptible mosquitoes directly (Riveron *et al.* 2013). This supports the fact that the high frequency of this haplotype in both countries is linked to

selection of a favourable allele which is a signature of directional selection. The strength of this directional selection in southern Africa was also shown by the reduced haplotype diversity in this region with only few haplotypes clustering together and with only few mutational steps between them in contrast to the four countries (Benin, Cameroon, Ghana and Uganda) showing a higher diversity and higher genetic differentiation levels to southern populations.

A similar signature of directional selection was recently detected on the glutathione S-transferase GSTe2, another detoxification gene conferring resistance to DDT in *An. funestus* populations in Africa (Riveron *et al.* 2014), and particularly in resistant samples from Benin where a nearly fixed haplotype was detected in contrast to DDT susceptible populations where a high genetic diversity was observed. The findings in this chapter are in line with the microsatellite results presented previously in chapter three, where two loci (AFUB6 and FUNR) showed very reduced genetic diversity for the southern population (Malawi and Mozambique) and at a lower extent for the other populations. It was also observed that when these two loci were removed from the analysis, the overall genetic diversity for the southern population increased and was similar to other populations pointing to possible selection acting on these loci which are known to be located on the rp1 QTL on the 2R chromosome, with the FUNR located just in the 5' upstream region of *CYP6P9a* gene. In chapter 4, a reduced diversity was seen in all the populations around the rp1 with the microsatellite markers located only on the 2R. However, selection is stronger in southern on CYP6P9a and only limited in other countries.

Additionally the reduced diversity of *CYP6P9a* in southern Africa further confirms the detection of a selective sweep also observed in chapter 4 whereby, there was much reduced diversity around markers situated in same position as the *CYP6P9a* gene. The strong correlation between all these results is a robust indication that pyrethroid resistance has very likely induced a selective sweep around the resistance genes.

 The F_{st} obtained from microsatellite in the previous chapters where Malawi and Mozambique where less genetically differentiated than the other four populations is also supported by the K_{ST} result where Malawi and Mozambique show less genetic differentiation to the other populations. This similarity is possibly due to the fact that both countries genetically are close due to geographical proximity and gene flow. The overall

pattern of differentiation obtained with the *CYP6P9a* gene in this study is identical to that obtained in the GSTe2 study by Riveron *et al.* (2014) where Malawi and Mozambique is still close, likewise for Ghana and Uganda, while Benin and Cameroon were intermediate. This shows that the sequencing of these two genes located in two independent chromosomal positions (2R for *CYP6P9a* and 2L for GSTe2) combines with microsatellite analysis to support a restriction of gene flow in *An. funestus* populations across Africa as previously suggested by Michel *et al.* (2005). Directional selection has previously been detected on insecticide resistance genes in other insect species such as for the case reported for *CYP6G1* gene that confers resistance to DDT in *D. melanogaster.* This gene has also been shown to be under directional selection with a single haplotype whose resistance contains an Accord transposable element in the 5' UTR region (Daborn *et al.* 2002). In *D. simulans*, the same *CYP6G1* gene has also been observed to present an extensive reduced heterozygosity with a 100-kb due to a selective sweep around this gene (Schlenke *et al*. 2004).

5.4.2 The haplotype distribution of the *CYP6P9a* gene correlates with pyrethroid resistance profiles

Similar to the GSTe2 gene, the haplotype distribution of the *CYP6P9a* appeared to be more influenced by the pyrethroid resistance status than geographical proximity. The predominant haplotype in the resistant countries of southern Africa was completely absent in the other susceptible countries indicating a low level of gene flow between southern African populations and the others as shown already with the distribution of the same resistance mutations such as the L119F GSTe2 and the A296S RDL mutation which are completely absent in southern Africa (Wondji *et al*. 2011). Furthermore, the selection of a favourable haplotype in the resistant population for *CYP6P9a* was also observed for GSTe2 for which a single haplotype was nearly fixed in Benin with a frequency of 95% (Riveron *et al*. 2014). This shows that, haplotype diversity of insecticide resistance genes can provide a good indication of the level of selection acting on these genes but also provide insight into contemporary gene flow between populations.

It was observed in the phylogenetic analysis that the southern population formed a cluster of theirs which was independent of the other populations. This finding is similar to that observed in Michel *et al*. (2005), where the southern population formed an independent cluster from the other African populations. These results are in line with the

STRUCTURE results in chapter three and chapter four where the southern population formed a group of theirs. This is possibly due to the resistance profile and pattern in the southern population which is different from the other populations. Geographical proximity might also be playing a role. Although this is not consistent as Ghana and Benin which are geographically close to each other, did not share similar haplotypes or show close genetic differentiation as observed for the southern population. The other population did not show a defined pattern. However, the exact resistance profile of the samples used in this study is not known since no bioassay was carried out for these field collected mosquitoes. Thus, the full extent of pyrethroid resistance on these genetic diversity patterns of *CYP6P9a* will need further work to be fully established.

5.4.3 The reduced genetic diversity across rp1 supports the presence of selective sweep in southern Africa

The presence of the selective sweep around the rp1 pyrethroid resistance QTL was further confirmed by the analysis of the polymorphism of 5 DNA fragments from a BAC clone spanning 120kb of this locus including the *CYP6P9a* gene.

There was much reduced polymorphism across the rp1 region for the more resistant Malawi and Mozambique populations while very high levels of polymorphism were observed for the more susceptible samples from Cameroon. This result is similar to the finding using VSGC in Jones *et al*. (2012) showing extensive reduction of heterozygosity around the kdr mutation for resistant samples.

Selection in the southern population is confirmed by the continuous reduced genetic diversity for Malawi and Mozambique and also the reduced haplotype diversity in both countries across the rp1 away from *CYP6P9a*. This reduced diversity could be caused by other factors such as chromosomal inversion, a historical population event, local adaptation to local environmental factors. However, it is unlikely that this is down to chromosomal inversion as the size of the region is small for the known chromosomal inversion which is bigger than 120kb and also the lowest diversity coincides with *6p9a* and recovers away from it showing that it is not inversion but very likely selection on *6p9*. It is also not possibly due to a historical event as such event tends to impact the whole genome which is contrary to what is observed in this study where reduction of diversity is well centred on *CYP6P9a*.

The size of the region under selection is around 70kb, which is similar to the extent of the sweep observed in Drosophila for *CYP6G1* where a region of 100kb was under selective sweep in resistant populations (Schlenke and Begun 2004).It seems that the scale and extent of the selective sweep is smaller with the BAC sequencing than with microsatellites. This could be explained by the different mode of evolution and variability patterns between the two markers as DNA sequences evolve rapidly than the microsatellite which might explain the longer signature seen with microsatellite. However, the two microsatellite markers (FUNR and AFUB6) are located within the 100kb showing that rp1 is the center of this selective sweep.

Sequencing of more loci and the newly release genome of *An. funestus* will help to determine the real extent of this selective sweep, since such selective sweep will show the impact of control interventions on malaria vectors. This is one of the first research work to show that malaria metabolic resistance can also leave a significant footprint on the genetic profile of malaria vectors. So far, mainly target site mutations such as Kdr (Lynd *et al.* 2010, Jones *et al.* 2012) have been associated with selective sweep. Here, this study shows that metabolic resistance could do the same. The strong signature observed in Malawi or Mozambique is also an indication that resistance is being highly selected in these populations either by control interventions such as LLINS or IRS or from agriculture. This calls for urgent implementation of resistance management strategies in these countries to prevent the resistant alleles to become fixed. Unfortunately, the predominant resistant haplotype is getting close to fixation in Malawi and Mozambique which could make it very difficult to remove it from the population. This is because, the frequency is so high and the mosquitoes may have already developed an alternative strategy to mitigate any potential fitness cost associated with this resistance.

As seen for Rdl and GSTe2, it seems resistance allele distribution in *An. funestus* will be mainly limited to the geographical region of origin and will only spread to other regions over time, which in some way is good as these barriers of gene flow will prevent the rapid spread of these resistance mutations as seen for kdr in *An. gambiae*. However it could also be a disadvantage if in the future, one wanted to drive in the population of the continent with other advantageous genes such as for *Plasmodium* refractoriness.

Conclusion

The DNA sequence results confirms the fact that there is selection acting on the more resistant samples, since reduced nucleotide and haplotype diversity is linked to selection. The results also indicate that, the *CYP6P9a* is under a stronger selection in the southern Africa countries than in the other countries as seen from the nucleotide diversity. The sequence based test of BAC also clearly supports the fact that there is selection acting around the rp1 QTL and *CYP6P9a* gene from the reduced nucleotide diversity, haplotype diversity and the reduced polymorphism seen for the resistant samples than the susceptible.

Overall, this study has investigated the previous detection in chapter 4 for a possible selective sweep on 2R chromosome using Microsatellite loci. It has shown that this selection focused on the *CYP6P9a* resistance gene, extends beyond it to neighbouring genomic region on a distance of around 100kb as seen with the BAC clone analysis. This suggests that insecticide-based control interventions could induce significant genetic changes in mosquitoes populations in response to the selection exert. Although other factors could be behind this selection, these factors are very unlikely and the selection is mainly caused by pyrethroid and the strength seen in southern Africa calls for action to reduce the selection.

Selection on resistance gene has also been observed by Nair *et al.* (2003) around the *dhfr* (dihydrofolate reductase) gene which is resistant to the anti-malaria drug pyrimethamine, where markers surrounding this gene showed reduced diversity around 100Kb. This result is similar to the findings seen in this study where reduced diversity was seen around 70kb of the *CYP6P9a* gene which is an indication of selection acting on this gene. In rice (*Oryza sativa* L), reduced nucleotide and haplotype diversity, low polymorphism and high LD was seen around 100Kb of the entire 150Kb of *OsAMT1;1* gene which is a gene responsible for ammonium uptake at low nitrogen concentration but it fell on either side of the gene indicating selective sweep in the *OsAMT1;1* gene (Ding *et al.* 2011). The result is similar with the case of the *CYP6P9a* gene seen in this study where there is low diversity and away from the 70Kb and towards the 120Kb of the gene, while there is increased diversity for the more resistant samples. The *CYP6P9a* gene that is associated with insecticide resistance in *An. funestus* (Amenya *et al.* 2008, Wondji *et al.* 2009 and Riveron *et al.* 2013) is possibly under selection.

Chapter Six

Conclusion and recommendations for future work

Successful control of the malaria vector *An. funestus* as well as other malaria vectors requires an effective control program that relies on the suitable application, management and use of available insecticides either in the form of IRS or LLINs. Use of insecticide-based control interventions, the main means of controlling the malaria vectors, are facing the serious challenge of the resistance developed by the main malaria vectors such as *An. funestus* across Africa. Increasing reports of resistance to different classes of insecticides such as carbamates, pyrethroid and DDT have now been made in *An. funestus* populations from various regions of Africa such as in Cameroon (Central Africa (Wondji *et al*. 2011), Benin (Djouaka *et al*. 2011) and Ghana (West Africa (Okoye *et al*. 2008)), Uganda (East Africa (Morgan *et al.* 2010)), Malawi (Hunt *et al*. 2010) and Mozambique (southern Africa)(Hargreaves et al. 2000, Casimiro *et al*. 2006, Cuamba *et al*. 2010)).

There is fear that such resistance could negatively impact on the success of ongoing and future insecticide-based control interventions if suitable resistance management strategies are not implemented. However, in order to ensure the continuous effectiveness of these insecticides, it is without doubt that a good knowledge into the mechanisms that causes resistance is required. In addition, a proper understanding of the genetic structure and patterns of gene flow between populations of vectors is also crucial in order to predict the spread of resistance and to advise on the appropriate alternative control strategies such as insecticide rotation or mixtures. Knowledge of cross-resistance patterns between pyrethroids and carbamates is essential to design such alternative control strategies. The present study has investigated and tried to address these questions in the malaria vector *An. funestus*.

Although relatively less studied than *An. gambiae, An. funestus* has received more attention recently notably for insecticide resistance. This is shown by increasing efforts to characterise mechanisms of insecticide resistance in this species, which has all stemmed from the colonisation of a resistant strain from Mozambique (FUMOZ-R) that is resistant to pyrethroid and a susceptible strain (FANG) from Angola (Hunt *et al*. 2005). Genetic studies of this species has also been facilitated by studies such as the construction of a physical map for *An. funestus* by Sharakhov *et al.* (2004), and integrated genetic and

physical map by Wondji *et al.* (2005), with the identification of SNPs (Wondji *et al.* 2007a) and also *An. funestus* transcriptome sequencing (Crawford *et al.* 2010, Gregory *et al.* (2011). Recently, the first draft of the genome of *An. funestus* was publicly made available (www.Vectorbase.org), further facilitating genetic and genomic studies in this species.

6.1 Contrasting resistance in *An. funestus* across Africa

Analysis of pattern of susceptibility of an *An. funestus* population from Chikwawa in South Malawi further supported the contrasting resistance profile between populations of this species from geographical regions of Africa. The bendiocarb resistance observed in Chikwawa (Malawi) contrast with the full susceptibility to this insecticide class in East Africa (Morgan *et al.* 2010), whereas the relatively susceptibility to DDT in Chikwawa contrast to DDT resistance in East and West Africa (Djouaka *et al.* 2011, Okoye *et al.* 2008, Morgan *et al.* 2010). This contrast of susceptibility could be indicative of restricted gene flow between populations from these regions, although it could also be caused only by the existence of very different local selection forces and enough times may not yet have elapsed to allow the spread of such resistance Africa wide. However the reduced susceptibility observed to DDT in Chikwawa compared to the full susceptibility in Mozambique re-enforces the need of continuous monitoring of susceptibility levels against insecticides in each country before rolling out control interventions instead of relying on data from neighbouring countries or different regions within the same country.

6.2 Metabolic resistance through over-expression of cytochrome P450 genes are controlling carbamate resistance and a possible cross-resistance with pyrethroids

Synergist assays with PBO and genome-wide microarray-based transcription analyses have revealed that metabolic resistance mechanisms through elevated expression of cytochrome P450s are driving the carbamate resistance in Malawi. The nearly full recovery of susceptibility to bendiocarb after pre-exposure to PBO and the consistent over-expression of several P450 genes in the bendiocarb resistant Malawi population support the suggestion previously made by Brooke *et al.* (2001) that carbamate resistance is conferred by P450s just as it is the case also for pyrethroids. Furthermore, the present study has also detected the potential P450s genes involved in the carbamate resistance with the two duplicated P450 genes *CYP6P9a* and *CYP6P9b* being among the best bendiocarb resistance candidate as they were consistently over-expressed in the resistant

mosquitoes from microarray analyses. Other P450 genes also associated include the *CYP6Z1* for which the ortholog in *An. gambiae* is known to have a broad substrate spectrum (Chiu *et al.* 2008), *CYP9J11, CYP6P5* and *CYP9K1*. Interestingly, most of these genes are also associated with pyrethroid resistance suggesting that they could be associated with a cross-resistance between pyrethroids and carbamates. However, these genes will need to be functionally characterised to validate whether or not they are able to confer resistance to carbamates. The consistent over-expression of cuticle protein genes was also a noticeable result suggesting that a reduced penetration mechanism could be operating in bendiocarb resistance, a mechanism less observed for pyrethroid resistance. Future work will need to validate the role of such mechanisms which was previously reported in the *An. funestus* FUMOZ-R resistance strain (Wood *et al*. 2010).

An attempt to further associate the duplicated P450s *CYP6P9a* and *CYP6P9b* to bendiocarb resistance through a comparative analysis of their polymorphism between bendiocarb resistant and susceptible mosquitoes did not provide a clear support for their involvement contrary to what was previously showed for their association with pyrethroid resistance (Riveron *et al*. 2013). The lack of strong association between *CYP6P9a* and *CYP6P9b* polymorphism and bendiocarb could suggest that these genes are not involved in bendiocarb resistance to the same extent than to pyrethroid resistance. However, further analysis is needed to confirm this result notably by increasing the sample size or by using more advanced Next-generation sequencing approaches. For example, a target enrichment (e.g. Agilent Sure select approach) of region spanning these genes such as the rp1 BAC could be performed followed by a deep sequencing with Next-generation sequencing for both susceptible and resistant samples.

It should be noted here that the microarray analysis was carried out for just one population in Malawi and it could be good to analyse other populations across Malawi notably in the northern part to assess whether the resistance profiles and mechanisms remain the same or vary. Such study will further inform the control program for the implementation of any nationwide vector control intervention.

6.3 Patterns of genetic structure across *An. funestus* populations support the presence of barriers of gene flow

The significant contrast observed between patterns of insecticide resistance between populations of *An. funestus* across Africa could be indicative of the presence of barriers

of gene flow between these populations. To test this hypothesis, this study used microsatellite loci markers to assess the genetic structure of *An. funestus* population from various regions of Africa including southern (Mozambique and Malawi), East (Uganda), Central (Cameroon) and West Africa (Ghana and Benin). This study revealed that southern African populations of *An. funestus* are more genetically differentiated to other populations as they always form a different cluster from the Bayesian analyses with STRUCTURE and they exhibited higher F_{st} estimates when compared to populations from other regions. The Population from Uganda appears to be intermediate between southern and West/Central Africa. This result was similar to patterns of genetic structure previously reported for this species by Michel *et al.* (2005).

Patterns of genetic structure observed in this study support the contrast in resistance patterns between populations of *An. funestus* and suggest the presence of barriers of gene flow between populations of this species. The underlying reason for this gene flow barrier between southern populations and others remains unclear although it could be associated with the geographical proximity of both populations, absence of *An. funestus* around the Equatorial belt or the presence of the Rift Valley which has been shown in *An. gambiae* to impact pattern of genetic structure (Lehman *et al.* 1998). However such hypothesis will need to be validated by assessing populations from both sides of the Rift Valley in countries such as Tanzania or Kenya which are split by the Rift Valley.

Furthermore, using new approaches based on more fine-scale genome-wide analysis based on SNPs generated through RAD (Restriction site Associated DNA) sequencing could provide a better view of the genetic structure in this species across Africa. Patterns of gene flow in *An. funestus* described here, give an indication on the risk and speed of spread of genes of interest notably insecticide resistance genes or alleles between these populations. This is further supported by the observation that the resistance alleles L119F in GSTe2 associated with DDT resistance and the A296S RDL allele for dieldrin resistance are completely absent in southern Africa while predominant in West and Central Africa (Wondji *et al.* 2011, Riveron *et al.* 2013). Therefore, such studies could help control programs to predict the spread of resistance between populations of their targeted control areas.

6.4 Pyrethroid and carbamate resistance in southern Africa is associated with a signature of selection

Reduced genetic diversity in southern African populations at two markers (AFUB6 and FUNR) on the 2R chromosome within the pyrethroid resistance QTL rp1, suggested the presence of a selective sweep in this region. This hypothesis was validated in this study by analysing the genetic diversity of 8 microsatellite markers on the 2R chromosome and by performing a fine-scale analysis of the nucleotide diversity around the 120kb rp1 region but also for the full length of the *CYP6P9a* gene. There was evidence of signature of selective sweep in the southern populations characterised by reduced diversity and reduced heterozygosity for microsatellite markers around rp1 followed by a recovery when moving away from both sides of rp1. Similarly, the signature of selection was also evidenced when analysing nucleotide polymorphism of the rp1 genomic region with a significantly reduced diversity around the *CYP6P9a* in both southern African populations of Malawi and Mozambique but not in Cameroon (Central Africa). This signature of selective sweep is similar to that described for the *CYP6G1* P450 gene in *Drosophila melanogaster* (McCart and Ffrench-Constant 2008).

The factors driving this selection remains to be established as it is not known whether it is caused by the recent scale up of control interventions such as LLINS or IRS or rather driven by the selection in the breeding sites from the agricultural usage of insecticides. Future studies should investigate these sources of resistance selection in order to help improve the success of future control programs by reducing the development of resistance. This could involve a partnership with the agricultural sector notably by encouraging the farmers not to use the same insecticides as those used in public health sectors in order to avoid the cross selection of resistance in malaria vectors.

Conclusion

This study has been able to illustrate that the P450 genes and more so the *CYP6P9a* and *CYP6P9b* are driving bendiocarb resistance and also a bendiocarb/permethrin cross resistance in *An. funestus* field population of Malawi since both genes are over expressed in the resistant samples and were able to metabolise bendiocarb (carbamate) and both type I (permethrin) and type II (deltamethrin) pyrethroid. It also demonstrated the possibility that these genes might be shaping the genetic structure of *An. funestus* across southern (Malawi and Mozambique), Central (Cameroon), West (Benin and Ghana) and

East Africa (Uganda) through their resistance profiles. Without barriers to gene flow in natural *An. funestus* population, there is a possibility that the resistance seen in this study will spread across other *An. funestus* populations in Africa which will distort control programs. Suitable resistance management strategies should be considered following the resistance profile and genetic structure of *An. funestus*. This could include the involvement of inhibitors such as PBO into carbmate and pyrethriod formulations to improve the efficacy.

Table S1: NanoDrop™ spectrophotometer readings from 9 RNA pools (3 each for permethrin, bendiocarb and control), using the 4x44 *An. funestus* whole genome array. Permethrin (perm) and Bendiocarb (bend), where mosquitoes that stayed alive after they were exposed to 0.75% permethrin and 0.1% bendiocarb. The Control (C) was mosquitoes that were not exposed to any insecticide.

Total RNANanoDrop™

GROUP	POOL	Ng/µl	260-280
Mal-Perm	1	288.1	2.17
Mal-Perm	2	238.3	2.22
Mal-Perm	3	310.7	2.15
Mal-Bend	1	281.2	2.18
Mal-Bend	2	373.9	2.11
Mal-Bend	3	410.5	2.15
Mal-Cont	1	105.1	2.18
Mal-Cont	2	153.0	2.21
Mal-Cont	3	183.5	2.16

Table S2: NanoDrop™ readings from Cy3 and Cy5labelled for RNA pools for microarray analysis

Cy3/Cy5 LABELLED

Group	Pool	pmol	ng//µl	Specific activity (pmol cy3//cy5µl cRNA)
Mal-Perm CY5	1	3.8	350.7	10.8
Mal-Perm CY5	2	5	432.2	11.6
Mal-Perm Cy5	3	2.8	271	10.3
Mal-Bend CY5	1	7.2	479.1	15
Mal-Bend CY5	2	6.7	445.6	15.1
Mal-Bend Cy5	3	4.7	330.5	14.2
Mal-Cont CY3	1	1.5	165.7	9
Mal-Cont Cy3	2	2	211.9	9.4
Mal-Cont CY3	3	2.3	266.4	8.6

Table S3: Probes from detoxification genes and genes associated with bendiocarb resistance up-regulated in R-C comparison with P<*0.01* and Fold-change (FC)>2 ranked per gene families. Significant expression in other comparisons for bendiocarb and for permethrin are also indicated.

Probes	GeneName	FC R-C	P Value	R-S	C-S	R-C Perm	R-S perm	Description
				FC	FC	FC	FC	
Detoxification								
CUST_9786_PI406199769	combined_c4956 (CYP4G16)	9.1	0.0022		-9.5	9.6		cytochrome p450
CUST_9787_PI406199769	combined_c4956 (CYP4G16)	8.7	0.0016		-9.1	10.1		cytochrome p450
CUST_45_PI406199775	CYP6z1	7.3	0.0014	2.9		7.8	2.9	cytochrome p450
CUST_13272_PI406199769	combined_c6791	6.8	0.0013	10.8			9.3	cytochrome p450
CUST_44_PI406199775	CYP6z1	6.7	0.0021	2.7			2.7	cytochrome p450
CUST_13273_PI406199769	combined_c6791	5.3	0.0010	10.8			8.6	cytochrome p450
CUST_26_PI406199775	CYP6P9a	5.2	0.0026	12.6	5.6	7.9	23.2	cytochrome p450
CUST_43_PI406199775	CYP6z1	5.0	0.0023	2.5	-2.6	6.1	3.1	cytochrome p450
CUST_29_PI406199775	CYP6P9b	3.1	0.0054	4.2			5.4	cytochrome p450
CUST_33_PI406199775	Pseudo_P450_between_6AA2_and_6P9a	3.0	0.0030	2.4				cytochrome p450
CUST_4104_PI406201128	AGAP002418-RA	2.8	0.0034					cytochrome p450
CUST_5107_PI406201128	AGAP002204-RA	2.7	0.0030					cytochrome p450 CYP325D1
CUST_157_PI406199798	AGAP000088-RA	2.7	0.0016					cytochrome p450

CUST_16709_PI406201128	AGAP008212-RA	2.5	0.0053					cytochrome p450
CUST_25_PI406199775	CYP6P9a	2.4	0.0067	17.3	7.8	3.4	37.1	cytochrome p450
CUST_28_PI406199775	CYP6P9b	2.4	0.0058	3.9			4.8	cytochrome p450
CUST_17_PI406199775	CYP6P4a	2.1	0.0032	3.1				cytochrome p450
CUST_23_PI406199775	CYP6P5	2.1	0.0065	3.5				cytochrome p450
CUST_7029_PI406199769	combined_c3556	2.1	0.0093					cytochrome p450
CUST_1946_PI406199798	AGAP002197-RA	2.1	0.0029					cytochrome p450
CUST_369_PI406199788	gb-CYP4H18	2.0	0.0021					cytochrome p450 4d1
CUST_5009_PI406199772	BU038937	3.0	0.0065	-3.1	-5.9			cytochrome p450 4g15
CUST_5907_PI406199769	combined_c2988	2.9	0.0044		-2.9			glucosylglucuronosyltransferases
CUST_633_PI406199788	gb-GSTO1	3.7	0.0015					glutathione s-transferase
CUST_9502_PI406199769	combined_c4812	2.8	0.0062	3.9				short-chain dehydrogenase
CUST_5640_PI406199769	combined_c2855	2.7	0.0044					short-chain dehydrogenase
CUST_8727_PI406199769	combined_c4419	2.9	0.0022		-2.7			abc transporter
CUST_22183_PI406201128	AGAP006226-RA	2.4	0.0022					aldehyde oxidase
CUST_7696_PI406199798	AGAP008141-RA	4.7	0.0042					argininosuccinatelyase

Peroxidases

CUST_1998_PI406199772	CD664210.1	3.6	0.0064					glutathione peroxidase
CUST_3883_PI406199798	AGAP004247-RA	3.1	0.0056	-2.4	-4.5		2.8	glutathione peroxidase

CUST_15573_PI406201128	AGAP009033-RA	2.4	0.0056				chorion peroxidase
CUST_2536_PI406199772	CD578133.1	3.8	0.0034		-5.4		oxidase peroxidase
CUST_722_PI406199788	gb-PX5A	2.0	0.0098				oxidase peroxidase

Cuticular protein genes

CUST_3736_PI406199772	CD577515.1	3.1	0.0066	4.2	3.9		cuticle protein
CUST_3727_PI406199772	CD577519.1	3.0	0.0039	3.3			cuticle protein
CUST_3731_PI406199772	CD577517.1	2.9	0.0039	3.0			cuticle protein
CUST_4047_PI406199772	CD577343.1	2.9	0.0075				cuticle protein
CUST_4100_PI406199772	CD577317.1	2.8	0.0023	2.8			cuticle protein
CUST_3729_PI406199772	CD577518.1	2.6	0.0027	2.6			cuticle protein
CUST_3395_PI406199772	CD577694.1	2.5	0.0030	2.9			cuticle protein
CUST_1977_PI406199772	CD664220.1	2.5	0.0014	3.5			cuticle protein
CUST_3734_PI406199772	CD577516.1	2.3	0.0060	3.3			cuticle protein
CUST_4102_PI406199772	CD577316.1	2.2	0.0044	2.8			cuticle protein

Mitochondrial genes

CUST_4508_PI406199772	CD577111.1	7.9	0.0023	-2.7		9.2	atp synthase f0 subunit 6
CUST_4607_PI406199772	CD577060.1	6.4	0.0072	-3.2	-9.9		atp synthase f0 subunit 6
CUST_4583_PI406199772	CD577072.1	6.4	0.0069	-3.3	-9.8		atp synthase f0 subunit 6

Odorant receptor genes

CUST_15404_PI406201128	AGAP006917-RA	2.5	0.0034					gustatory receptor (agap006917-pa)
CUST_8869_PI406199798	AGAP009393-RA	2.4	0.0063					odorant receptor 13a
Salivary gland proteins								
CUST_593_PI406199772	EE589616.1	20.0	0.0021	4.7		38.2	9.6	d7-related 1 protein
CUST_359_PI406199772	EE589855.1	16.7	0.0010	4.3	-12.1	26.1		d7-related 1 protein
CUST_717_PI406199772	EE589504.1	16.0	0.0008	4.6	-9.9	28.4	6.9	d7-related 1 protein
CUST_1413_PI406199772	EE589737.1	15.5	0.0019	4.9		31.8		d7-related 1 protein
CUST_810_PI406199772	EE589416.1	4.1	0.0018		-9.7			d7-related 2 protein
CUST_356_PI406199772	EE589850.1	3.9	0.0087		-7.1	5.5		d7-related 2 protein
CUST_1736_PI406199772	EE589383.1	4.6	0.0054		-13.9			d7-related 3 protein
CUST_1719_PI406199772	EE589407.1	12.6	0.0009		-30.2			ge rich salivary gland protein
CUST_290_PI406199772	EE589911.1	10.3	0.0014			11.4		ge rich salivary gland protein
CUST_1592_PI406199772	EE589544.1	9.2	0.0008					ge rich salivary gland protein
CUST_923_PI406199772	EE589310.1	3.3	0.0078		-8.1			gsg7 salivary protein
CUST_759_PI406199772	EE589462.1	3.1	0.0022					sg2a salivary protein
Heat shock proteins								
CUST_5373_PI406199798	AGAP004582-RA	3.6	0.0023		-2.6			heat shock protein 70 b2
CUST_5662_PI406199798	AGAP004581-RA	3.1	0.0078					heat shock protein 70 b2
CUST_3090_PI406199772	CD577854.1	3.1	0.0058					hsp70 binding protein

CUST_17251_PI406201128	AGAP008392-RA	2.4	0.0091				hsp70 binding protein
CUST_3256_PI406199772	CD577765.1	7.3	0.0015		-5.1		lethal essential for life l2efl
CUST_5095_PI406199772	BU038894	6.1	0.0017		-6.8		lethal essential for life l2efl
CUST_5096_PI406199772	BU038894	5.9	0.0018		-6.3		lethal essential for life l2efl
CUST_3255_PI406199772	CD577765.1	4.4	0.0066		-5.7		lethal essential for life l2efl
CUST_3440_PI406199772	CD577671.1	6.6	0.0011		-3.6	7.6	peritrophin a
CUST_5118_PI406199772	BU038883	6.1	0.0017		-3.9	7.3	peritrophin a
CUST_3438_PI406199772	CD577672.1	2.8	0.0035		-4.9		peritrophin a

Immunity genes

CUST_5673_PI406199769	combined_c2871	3.9	0.0064		-4.9		tep1 protein
CUST_13266_PI406199769	combined_c6787	2.4	0.0093	-2.9	-7.4		clipa6 protein
CUST_14660_PI406199769	combined_c7718	2.3	0.0093				clipb17 protein
CUST_6151_PI406199769	combined_c3111	2.3	0.0053		-4.3		clipb5 protein
CUST_4033_PI406199769	combined_c2039	2.7	0.0066				ctlma3 protein
CUST_924_PI406199798	AGAP000871-RA	2.2	0.0038				ctlma3 protein

Proteases

CUST_799_PI406199769	combined_c404	3.6	0.0010				serine protease
CUST_7044_PI406199769	combined_c3564	3.0	0.0063		-4.3		serine protease
CUST_15844_PI406199769	combined_c8405	2.5	0.0090				serine protease

Probe name	Gene Name	FC	P-Value					Description
CUST_16055_PI406199769	combined_c8512	5.7	0.0035			4.1		serine protease 14
CUST_1331_PI406199769	combined_c674	5.1	0.0062		-3.1			serine protease 14
CUST_692_PI406199798	AGAP000572-RA	4.8	0.0089					serine protease desc4
CUST_9088_PI406199798	AGAP006416-RA	4.3	0.0088					serine protease sp24d
CUST_2463_PI406199772	CD578169.1	8.7	0.0022			12.7		trypsin
CUST_9482_PI406199798	AGAP008292-RA	6.5	0.0024	2.7	-2.9		4.3	trypsin
CUST_5111_PI406199772	BU038886	5.9	0.0039	2.5	-3.3	10.3	3.8	trypsin
CUST_2288_PI406199772	CD578260.1	5.6	0.0014	2.4	-3.7		3.5	trypsin
CUST_6793_PI406199769	combined_c3435	2.8	0.0024		-2.7			trypsin

Table S4: Probes from detoxification genes and genes associated with bendiocarb resistance differentially up-regulated in R-S comparison with P<*0.01* and Fold-change (FC)>2 ranked per gene families.

Probe name	Gene Name	FC	P-Value	Description
CUST_27_PI406199775	CYP6P9a	24.35	0.0472	cytochrome p450
CUST_30_PI406199775	CYP6P9b	17.76	0.0373	cytochrome p450
CUST_15331_PI426302897	Afun015331	14.06	0.0412	cytochrome p450 307a1
CUST_7_PI426302915	CYP6M3.seq	10.18	0.0224	cytochrome p450
CUST_26_PI406199775	CYP6P9a	7.98	0.0444	cytochrome p450
CUST_22_PI426302915	CYP6S2.seq	7.54	0.0349	cytochrome p450
CUST_26_PI426302915	CYP6Y2_rvcpl.seq	6.03	0.0329	cytochrome p450
CUST_7369_PI426302897	Afun007369	4.79	0.0165	cytochrome p450
CUST_3394_PI426302897	Afun003394	4.76	0.0420	cytochrome p450

CUST_9584_PI426302897	Afun009584	4.71	0.0497	cytochrome p450 6a8
CUST_8_PI426302915	CYP6M3.seq	4.50	0.0052	cytochrome p450
CUST_12197_PI426302897	Afun012197	4.36	0.0170	cytochrome p450
CUST_7469_PI426302897	Afun007469	3.56	0.0399	cytochrome p450
CUST_7769_PI426302897	Afun007769	3.50	0.0442	cytochrome p450 cyp9k1
CUST_28_PI426302915	CYP6Z1_rvcpl_fixed.seq	3.15	0.0208	cytochrome p450
CUST_7861_PI426302897	Afun007861	3.00	0.0424	cytochrome p450
CUST_27_PI426302915	CYP6Z1_rvcpl_fixed.seq	2.97	0.0289	cytochrome p450
CUST_48_PI406199775	CYP6z3	2.48	0.0430	cytochrome p450
CUST_29_PI406199775	CYP6P9b	2.37	0.0433	cytochrome p450
CUST_12342_PI426302897	Afun012342	2.35	0.0381	cytochrome p450 4d1
CUST_10936_PI426302897	Afun010936	2.23	0.0349	cytochrome p450
CUST_1096_PI406199769	combined_c557	8.87	0.0352	trypsin
CUST_7674_PI426302897	Afun007674	3.46	0.0350	late trypsin
CUST_5111_PI406199772	BU038886	3.24	0.0358	trypsin
CUST_2288_PI406199772	CD578260.1	3.16	0.0454	trypsin
CUST_13921_PI426302897	Afun013921	14.05	0.0402	chymotrypsin 1
CUST_8354_PI426302897	Afun008354	5.40	0.0180	glutathione transferase (agap004382-pa)
CUST_45_PI426302897	Afun000045	2.75	0.0308	glutathione-s-transferasegst
CUST_7499_PI426302897	Afun007499	2.35	0.0092	glutathione transferase
CUST_1458_PI406199769	combined_c738	3.75	0.0245	short-chain dehydrogenase
CUST_1378_PI406199798	AGAP001405-RA___2R	2.35	0.0186	short-chain dehydrogenase
CUST_9593_PI426302897	Afun009593	2.01	0.0101	short-chain dehydrogenase
CUST_7008_PI426302897	Afun007008	3.72	0.0487	atp-binding cassette sub-family a member
CUST_13662_PI426302897	Afun013662	3.36	0.0450	atp-binding cassette sub-family a member
CUST_7302_PI426302897	Afun007302	2.27	0.0325	heat shock 70 kda protein cognate 4
CUST_2081_PI406199798	AGAP002076-RA___2R	2.08	0.0042	heat shock protein cognate isoform a
CUST_3698_PI406199769	combined_c1867	3.19	0.0225	odorant-binding protein

Probe Name	Gene Name	FC	p-Value	Description
CUST_493_PI426302897	Afun000493	3.74	0.0232	aldehyde oxidase
CUST_13333_PI406199769	combined_c6826	2.06	0.0419	esterase b1
CUST_7008_PI426302897	Afun007008	3.72	0.0487	atp-binding cassette sub-family a member
CUST_13662_PI426302897	Afun013662	3.36	0.0450	atp-binding cassette sub-family a member
CUST_899_PI406199772	EE589329.1	3.10	0.0458	d7-related 1 protein
CUST_1087_PI406199772	EE590086.1	2.56	0.0473	d7-related 1 protein
CUST_1564_PI406199772	EE589574.1	2.11	0.0314	d7-related 1 protein
CUST_360_PI406199772	EE589855.1	2.00	0.0368	d7-related 1 protein
CUST_6348_PI426302897	Afun006348	3.08	0.0384	sugar transporter
CUST_4241_PI406199769	combined_c2146	17.11	0.0314	AGAP005468-PA [Anopheles gambiae str. PEST]
CUST_11049_PI426302897	Afun011049	14.33	0.0162	AGAP000604-PA [Anopheles gambiae str. PEST]
CUST_9232_PI426302897	Afun009232	6.14	0.0390	AGAP004236-PA [Anopheles gambiae str. PEST]

Table S5: Probes from detoxification genes and genes associated with bendiocarb resistance differentially up-regulated in C-S comparison with P<*0.01* and Fold-change (FC)>2 ranked per gene families.

Probe Name	Gene Name	FC	p-Value	Description
CUST_27_PI406199775	CYP6P9a	69.19	0.0009	cytochrome p450
CUST_30_PI406199775	CYP6P9b	26.56	0.0011	cytochrome p450
CUST_26_PI406199775	CYP6P9a	5.63	0.0021	cytochrome p450
CUST_21_PI406199775	CYP6P4b	2.71	0.0066	cytochrome p450
CUST_309_PI406199788	gb-CYP49A1	2.22	0.0076	cytochrome p450
CUST_5817_PI406199769	combined_c2943	16.41	0.0074	protein
CUST_8240_PI406199769	combined_c4173	10.38	0.0034	glycoprotein 93
CUST_12787_PI406199769	combined_c6526	7.51	0.0083	protein smoothened-like
CUST_5289_PI406199798	AGAP003873-RA___2R	7.47	0.0028	protein crumbs
CUST_1459_PI406199769	combined_c738	6.91	0.0056	short-chain dehydrogenase

CUST_1459_PI406199769	combined_c738	6.91	0.0056	short-chain dehydrogenase
CUST_2860_PI406199769	combined_c1442	3.18	0.0029	short-chain dehydrogenase
CUST_2860_PI406199769	combined_c1442	3.18	0.0029	short-chain dehydrogenase
CUST_1378_PI406199798	AGAP001405-RA	3.93	0.0085	short-chain dehydrogenase
CUST_3511_PI406199769	combined_c1773	4.45	0.0035	odorant binding protein 4
CUST_3510_PI406199769	combined_c1773	4.17	0.0023	odorant binding protein 4
CUST_7429_PI406199769	combined_c3760	3.67	0.0030	chymotrypsin 1
CUST_3736_PI406199772	CD577515.1	3.99	0.0079	cuticle protein
CUST_3835_PI406199772	CD577459.1	3.50	0.0072	cuticle protein
CUST_3828_PI406199772	CD577463.1	3.18	0.0094	cuticle protein
CUST_9165_PI406201128	AGAP003385-RA	3.26	0.0082	cuticle protein
CUST_3827_PI406199772	CD577463.1	2.91	0.0089	cuticle protein
CUST_3828_PI406199772	CD577463.1	3.18	0.0094	cuticle protein
CUST_3837_PI406199772	CD577458.1	2.84	0.0062	cuticle protein
CUST_3408_PI406199769	combined_c1722	2.78	0.0026	pupal cuticle protein
CUST_918_PI406199772	EE589313.1	3.77	0.0069	gsg7 salivary protein
CUST_15340_PI406199769	combined_c8151	3.46	0.0077	gsg6 salivary protein
CUST_809_PI406199769	combined_c410	3.31	0.0085	gsg6 salivary protein
CUST_810_PI406199769	combined_c410	2.99	0.0085	gsg6 salivary protein
CUST_70_PI406199772	EE590122.1	2.18	0.0090	gsg7 salivary protein
CUST_4410_PI406201128	AGAP001777-RA___2R	2.77	0.0069	atp-binding cassette transporter
CUST_2081_PI406199798	AGAP002076-RA___2R	2.58	0.0026	heat shock protein cognate isoform a
CUST_2056_PI406199769	combined_c1037	2.05	0.0015	heat shock cognate 70
CUST_3489_PI406199769	combined_c1762	2.54	0.0090	abc transporter
CUST_528_PI406199772	EE589690.1	3.03	0.0039	salivary protein sg1b
CUST_3249_PI406199769	combined_c1641	2.51	0.0042	kda salivary protein
CUST_14376_PI406199769	combined_c7513	2.36	0.0033	glutathione transferase

Table S6: Probes from detoxification genes or resistance associated genes up-regulated in R-S comparison for bendiocarb resistance with the new 8x60k *An. funestus* chip

ProbeName	Systematic name	FC R-S Bendiocarb	P value	Description	
CUST_493_PI426302897	Afun000493	3.7386973	0.02321149	aldehyde oxidase	AGAP006225-PA
CUST_7646_PI426302897	Afun007646	2.6603148	0.013568931	aldehyde oxidase	AGAP006225-PA
CUST_9227_PI426302897	Afun009227	31.928698	0.015973605	argininosuccinatelyase	AGAP008141-PA
CUST_7008_PI426302897	Afun007008	3.7248065	0.04871036	atp-binding cassette sub-family a member	AGAP012156-PA
CUST_13662_PI426302897	Afun013662	3.3553543	0.045043226	atp-binding cassette sub-family a member	AGAP006379-PA
CUST_5378_PI426302897	Afun005378	2.4533205	0.04983525	atp-binding cassette transporter	AGAP003221-PA
CUST_13921_PI426302897	Afun013921	14.0480995	0.04021558	chymotrypsin 1	AGAP006709-PA
CUST_27_PI406199775	CYP6P9a	24.35316	0.047168925	cytochrome p450	
CUST_30_PI406199775	CYP6P9b	17.761494	0.0373442	cytochrome p450	
CUST_7_PI426302915	CYP6M3.seq	10.183469	0.022436336	cytochrome p450	
CUST_26_PI406199775	CYP6P9a	7.9845333	0.044412024	cytochrome p450	
CUST_22_PI426302915	CYP6S2.seq	7.536029	0.03493809	cytochrome p450	
CUST_25_PI426302915	CYP6Y2_rvcpl.seq	6.1177783	0.028771743	cytochrome p450	
CUST_26_PI426302915	CYP6Y2_rvcpl.seq	6.0280824	0.03294737	cytochrome p450	

CUST_7369_PI426302897	Afun007369	4.7884607	0.016491504	cytochrome p450	AGAP002865-PA
CUST_3394_PI426302897	Afun003394	4.761112	0.042007115	cytochrome p450	AGAP000284-PA
CUST_8_PI426302915	CYP6M3.seq	4.5030403	0.005205322	cytochrome p450	
CUST_12197_PI426302897	Afun012197	4.3602757	0.017035333	cytochrome p450	AGAP003066-PA
CUST_7469_PI426302897	Afun007469	3.5591946	0.039896663	cytochrome p450	AGAP012296-PA
CUST_28_PI426302915	CYP6Z1_rvcpl_fixed.seq	3.1530364	0.02081737	cytochrome p450	
CUST_7861_PI426302897	Afun007861	3.0003326	0.04236047	cytochrome p450	AGAP008219-PA
CUST_27_PI426302915	CYP6Z1_rvcpl_fixed.seq	2.971755	0.028902346	cytochrome p450	
CUST_48_PI406199775	CYP6z3	2.4813251	0.042995602	cytochrome p450	
CUST_29_PI406199775	CYP6P9b	2.3670835	0.043271676	cytochrome p450	
CUST_10936_PI426302897	Afun010936	2.2333164	0.034911714	cytochrome p450	AGAP009241-PA
CUST_15331_PI426302897	Afun015331	14.055244	0.04118577	cytochrome p450 307a1	AGAP001039-PB
CUST_12342_PI426302897	Afun012342	2.3531232	0.038100097	cytochrome p450 4d1	AGAP008358-PA
CUST_9584_PI426302897	Afun009584	4.7071857	0.049704045	cytochrome p450 6a8	AGAP008214-PA
CUST_7769_PI426302897	Afun007769	3.5021703	0.044243805	cytochrome p450 cyp9k1	AGAP000818-PA
CUST_899_PI406199772	EE589329.1	3.1043382	0.04579553	d7-related 1 protein	
CUST_1087_PI406199772	EE590086.1	2.5627787	0.04732069	d7-related 1 protein	
CUST_1564_PI406199772	EE589574.1	2.1090512	0.031356342	d7-related 1 protein	

CUST_360_PI406199772	EE589855.1	2.0003612	0.036782302	d7-related 1 protein	
CUST_13332_PI406199769	combined_c6826	2.4047394	0.04543381	esterase b1	
CUST_13333_PI406199769	combined_c6826	2.064859	0.041889735	esterase b1	
CUST_7499_PI426302897	Afun007499	2.3483977	0.00924583	glutathionetransferase	AGAP004164-PA
CUST_8354_PI426302897	Afun008354	5.400496	0.017975274	glutathionetransferase (agap004382-pa)	AGAP004382-PA
CUST_45_PI426302897	Afun000045	2.7475302	0.030849837	glutathione-s-transferasegst	AGAP009194-PA
CUST_8241_PI406199769	combined_c4173	3.8323588	0.033405844	glycoprotein 93	
CUST_8240_PI406199769	combined_c4173	3.7012727	0.04501862	glycoprotein 93	
CUST_7302_PI426302897	Afun007302	2.2659736	0.0325283	heat shock 70 kda protein cognate 4	AGAP002076-PA
CUST_2081_PI406199798	AGAP002076-RA___2R	2.0839603	0.004172586	heat shock protein cognate isoform a	AGAP002076-RA___2R
CUST_1834_PI406199769	combined_c926	2.6075165	0.0324468	larval cuticle protein lcp-30	
CUST_7674_PI426302897	Afun007674	3.463132	0.034963116	late trypsin	AGAP006385-PA
CUST_3253_PI406199772	CD577766.1	3.854749	0.03916988	lethal essential for life l2efl	
CUST_3699_PI406199769	combined_c1867	3.3009863	0.018138004	odorant-binding protein	
CUST_3698_PI406199769	combined_c1867	3.1888664	0.022498274	odorant-binding protein	
CUST_1458_PI406199769	combined_c738	3.7480102	0.024549251	short-chain dehydrogenase	
CUST_1378_PI406199798	AGAP001405-RA___2R	2.3475215	0.018598432	short-chain dehydrogenase	AGAP001405-RA___2R
CUST_9593_PI426302897	Afun009593	2.0119352	0.010092365	short-chain dehydrogenase	AGAP012513-PA

CUST_1096_PI406199769	combined_c557	8.874661	0.035192452	trypsin	
CUST_1097_PI406199769	combined_c557	7.1126614	0.043315627	trypsin	
CUST_9482_PI406199798	AGAP008292-RA_Trypsin-4	4.0359397	0.013484117	trypsin	AGAP008292-RA_Trypsin-4
CUST_5111_PI406199772	BU038886	3.242788	0.03577211	trypsin	
CUST_2288_PI406199772	CD578260.1	3.1604838	0.04543381	trypsin	
CUST_19_PI426302897	Afun000019	3.1319296	0.04327803	trypsin	AGAP008294-PA
CUST_5112_PI406199772	BU038886	3.0948198	0.0421227	trypsin	
CUST_2287_PI406199772	CD578260.1	3.0053563	0.04335648	trypsin	

References

Adugna, N., and Petros, B. (1996). Determination of the human blood index of some anopheline mosquitos by using ELISA. Ethiopian medical journal, 34(1), 1-10.

Akey, J. M., Zhang, G., Zhang, K., Jin, L., & Shriver, M. D. (2002). Interrogating a high-density SNP map for signatures of natural selection. Genome research, 12(12), 1805-1814.

Akogbéto MC, Padonou GG, Gbénou D, Irish S, Yadouleton A: Bendiocarb, (2010) a potential alternative against pyrethroid resistant *Anopheles gambiae* inBenin, West Africa. Malar J, 9:204.

Allendorf, F. W., and L. W. Seeb (2000). "Concordance of genetic divergence among sockeye salmon populations at allozyme, nuclear DNA, and mitochondrial DNA markers." Evolution**54**(2): 640-651.

Amenya, D. A., Naguran, R., Lo, T., Ranson, H., Spillings, B. L., Wood, O. R., ... & Koekemoer, L. L. (2008). Over expression of a cytochrome P450 (*CYP6P9*) in a major African malaria vector, *Anopheles funestus*, resistant to pyrethroids. Insect molecular biology, 17(1), 19-25.

Anthony, N., T. Unruh, and D. Ganser (1998). "Duplication of the Rdl GABA receptor subunit gene in an insecticide-resistant aphid, Myzus persicae." Molecular and General Genetics MGG **260**(2-3): 165-175.

Arensburger, P., Hice, R. H., Wright, J. A., Craig, N. L., & Atkinson, P. W. (2011). The mosquito *Aedes aegypti* has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs. BMC genomics, 12(1), 606.

Ashley, M. V., and B. D. Dow (1994). "The use of microsatellite analysis in population biology: background, methods and potential applications." Molecular ecology and evolution: approaches and applications Birkhäuser Basel: 185-201.

Awolola, T. S., Oyewole, I. O., Koekemoer, L. L., & Coetzee, M. (2005). Identification of three members of the *Anopheles funestus* (Diptera: Culicidae) group and their role in malaria transmission in two ecological zones in Nigeria. Transactions of the Royal Society of tropical medicine and Hygiene, 99(7), 525-531.

Ayala,D.,Fontaine,M.C.,Cohuet,A.,Fontenille,D.,Vitalis,R.,andSimard, F.(2011a).Chromosomalinversions,naturalselectionandadaptationin the malariavector Anopheles funestus. Mol. Biol. Evol. 28, 745–758.doi: 10.1093/molbev/msq248

Ayala, D., Costantini, C., Ose, K., Kamdem, G. C., Antonio-Nkondjio, C., Agbor, J. P., ... & Simard, F. (2009). Habitat suitability and ecological niche profile of major malaria vectors in Cameroon. Malar J, 8(1), 307.

Baer, C. F. (1999). "Among-locus variation in *F-st*: Fish, allozymes and the Lewontin-Krakauer test revisited." Genetics**152**(2): 653-659.

Bariami, V., Jones, C. M., Poupardin, R., Vontas, J., & Ranson, H. (2012). Gene amplification, ABC transporters and cytochrome P450s: unraveling the molecular basis of pyrethroid resistance in the dengue vector, *Aedes aegypti*. PLoS neglected tropical diseases, 6(6), e1692.

Berzosa, P. J., Cano, J., Roche, J., Rubio, J. M., Garcia, L., Moyano, E., ... & Benito, A. (2002). Malaria vectors in Bioko Island (Equatorial Guinea): PCR determination of the members of *Anopheles gambiae Giles* complex (Diptera: Culicidae) and pyrethroid knockdown resistance (kdr) in *An. gambiae sensu stricto*. Journal of vector ecology: journal of the Society for Vector Ecology, 27(1), 102-106.

Bishop, C. M. (2006). Pattern recognition and machine learning: springer New York.

Biswas, S. and J. M. Akey (2006). "Genomic insights into positive selection." Trends In Genetics**22**(8): 437-446.

Boccolini D, S. N., and Toure YT (1998). "Chromosomal polymorphism in *Anopheles funestus* and description of new inversions in Burkina Faso and Mali." Parassitologia**40:14**.

Boccolini, D., Sabatini, A., Sanogo, E., Sagnon, N., Coluzzi, M.,Costantini, C., (1994). Chromosomal and vectorial heterogeneities in *Anopheles funestus* from Burkina Faso, West Africa. Parassitologia36, 20.

Boccolini, D., Carrara, G.C., Cani, P.J., Costantini, C., (2002). Preliminary data on *Anopheles funestus* chromosomal polymorphism in a peri-urban site of Western Angola. Parassitologia 44, 22.

Bonning, B. and J. Hemingway (1991). "The efficacy of Acetylcholinesterase in organophosphorus and carbamate resistance in *Culex pipiens* L. from Italy." Pest.Biochem.Physiol.**40**: 143-148.

Braginets, O. P., Minakawa, N., Mbogo, C. M., & Yan, G. (2003). Population genetic structure of the African malaria mosquito *Anopheles funestus* in Kenya. The American journal of tropical medicine and hygiene, 69(3), 303-308.

Brogdon, W. G. and J. C. McAllister (1998). "Insecticide resistance and vector control." Emerg Infect Dis**4**(4): 605-613.

Brogdon, W. G., McAllister, J. C., Corwin, A. M., & Cordon-Rosales, C. (1999). Oxidase-Based DDT–Pyrethroid Cross-Resistance in Guatemalan< i>*Anopheles albimanus*</i>. Pesticide Biochemistry and Physiology, 64(2), 101-111.

Brooke, B. D., Hunt, R. H., Chandre, F., Carnevale, P., & Coetzee, M. (2002). Stable chromosomal inversion polymorphisms and insecticide resistance in the malaria vector mosquito *Anopheles gambiae* (Diptera: Culicidae). Journal of medical entomology, 39(4), 568-573.

Brooke, B. D., Kloke, G., Hunt, R. H., Koekemoer, L. L., Tem, E. A., Taylor, M. E., ... & Coetzee, M. (2001). Bioassay and biochemical analyses of insecticide resistance in

southern African *Anopheles funestus* (Diptera: Culicidae). Bulletin of entomological research, 91(04), 265-272.

Brooker, S., Hay, S. I., Issae, W., Hall, A., Kihamia, C. M., Lwambo, N. J., ... & Bundy, D. A. (2001). Predicting the distribution of urinary schistosomiasis in Tanzania using satellite sensor data. Tropical Medicine & International Health, 6(12), 998-1007.

Bruce-Chwatt, L. J. (1963). "A longitudinal survey of natural malaria infection in a group of West African adults." West Afr. Med. J**12**: 199-217.

Bruce-Chwatt, L. J. and C. W. Gockel (1960). "A study of the blood-feeding patterns of *Anopheles* mosquitoes through precipitin tests." Bull.World Health Organ**22**: 685.

Bryan, J. H., Di Deco, M. A., Petrarca, V., & Coluzzi, M. (1982). Inversion polymorphism and incipient speciation in Anopheles gambiae s. str. in the Gambia, West Africa. Genetica,59(3), 167-176.

Callen, D. F., Doggett, N. A., Stallings, R. L., Chen, L. Z., Whitmore, S. A., Lane, S. A., ... & Sutherland, G. R. (1992). High-resolution cytogenetic-based physical map of human chromosome 16. Genomics, 13(4), 1178-1185.

C. Chen, E. D., F. Forbes and O. François (2007). "Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study." Molecular Ecology Notes**7(5)**: 747-756.

Carino, F. A., Koener, J. F., Plapp Jr, F. W., & Feyereisen, R. (1994). Constitutive overexpression of the cytochrome P450 gene< i>*CYP6A1*</i> in a house fly strain with metabolic resistance to insecticides. Insect biochemistry and molecular biology, 24(4), 411-418.

Casimiro, S., Coleman, M., Hemingway, J., & Sharp, B. (2006). Insecticide resistance in *Anopheles arabiensis* and *Anopheles gambiae* from Mozambique. Journal of medical entomology, 43(2), 276-282.

Casimiro, S. L., Hemingway, J., Sharp, B. L., & Coleman, M. (2007). Monitoring the operational impact of insecticide usage for malaria control on *Anopheles funestus* from Mozambique. Malaria journal, 6(1), 142.

Chanda, E., Hemingway, J., Kleinschmidt, I., Rehman, A. M., Ramdeen, V., Phiri, F. N., and Coleman, M. (2011). Insecticide resistance and the future of malaria control in Zambia. PLoS One, 6(9), e24336.

Chandre, F., Darrier, F., Manga, L., Akogbeto, M., Faye, O., Mouchet, J., & Guillet, P. (1999). Status of pyrethroid resistance in *Anopheles gambiae* sensu lato. Bulletin of the World Health Organization, 77(3), 230-234.

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genetical research, 70(02), 155-174.

Mendis, C., Thompson, R., Begtrupl, K., Cuamba, N., & Dgedge, M. (1998). Cordon sanitaire or laissez faire: differential dispersal of young and old females of the malaria vector *Anopheles funestus Giles* (Diptera: Culicidae) in southern Mozambique. African Entomology, 6(1).

Chan, Y. S., Huen, D. S., Glauert, R., Whiteway, E., & Russell, S. (2013). Optimising homing endonuclease gene drive performance in a semi-refractory species: the Drosophila melanogaster experience. PloS one, 8(1), e54130.

Charlwood, J. D. and M. D. R. Jones (1979). "Mating-Behavior in the Mosquito, *Anopheles gambiae* Sl .1. Close Range and Contact-Behavior." Physiological Entomology**4**(2): 111-120.

Chen, X. (1993). "Comparison of inbreeding and outbreeding in her hermaphroditic *Arianta arbustorum* (L) (land snail)." Heredity71: 456–461.

Chiu, T. L., Wen, Z., Rupasinghe, S. G., & Schuler, M. A. (2008). Comparative molecular modeling of *Anopheles gambiaeCYP6Z1*, a mosquito P450 capable of metabolizing DDT. Proceedings of the National Academy of Sciences, 105(26), 8855-8860.

Christian, R. N., Strode, C., Ranson, H., Coetzer, N., Coetzee, M., & Koekemoer, L. L. (2001). "Microarray analysis of a pyrethroid resistant African malaria vector, *Anopheles funestus* from southern Africa." Pesticide Biochemistry and Physiology**99**(2): 140-147.

Clement, M., Posada, D. C. K. A., & Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. Molecular ecology, 9(10), 1657-1659.

Clustal, W. (1994). improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice Thompson, Julie D.; Higgins, Desmond G.; Gibson, Toby J. Nucleic Acids Research, 22(22), 4673-80.

Cockerham, C. C. (1973). Analyses of gene frequencies. Genetics, 74(4), 679-700.

Coetzee, M., & Koekemoer, L. L. (2013). Molecular systematics and insecticide resistance in the major African malaria vector *Anopheles funestus*. Annual review of entomology, 58, 393-412.

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

Coetzee, M., Craig, M., & Le Sueur, D. (2000). Distribution of African Malaria Mosquitoes Belonging to the< i>*Anopheles gambiae*</i> Complex. Parasitology today, 16(2), 74-77.

Coetzee, M. and D. Fontenille (2004). "Advances in the study of *Anopheles funestus*, a major vector of malaria in Africa." Insect Biochem Mol Biol**34**(7): 599-605.

Cohuet, A., Dia, I., Simard, F., Raymond, M., & Fontenille, D. (2004). Population structure of the malaria vector *Anopheles funestus* in Senegal based on microsatellite and cytogenetic data. Insect molecular biology, 13(3), 251-258.

Cohuet, A., Dia, I., Simard, F., Raymond, M., & Fontenille, D. (2004). Population structure of the malaria vector *Anopheles funestus* in Senegal based on microsatellite and cytogenetic data. Insect molecular biology, 13(3), 251-258.

Cohuet, A., Dia, I., Simard, F., Raymond, M., Rousset, F., Antonio-Nkondjio, C., ... & Fontenille, D. (2005). Gene flow between chromosomal forms of the malaria vector *Anopheles funestus* in Cameroon, Central Africa, and its relevance in malaria fighting. Genetics, 169(1), 301-311.

Cohuet, A., Simard, F., Berthomieu, A., Raymond, M., Fontenille, D., & Weill, M. (2002). Isolation and characterization of microsatellite DNA markers in the malaria vector *Anopheles funestus*. Molecular Ecology Notes, 2(4), 498-500.

Cohuet, A., Simard, F., TOTO, J. C., Kengne, P., COETZEE, M., & Fontenille, D. (2003). Species identification within the *Anopheles funestus* group of malaria vectors in Cameroon and evidence for a new species. The American journal of tropical medicine and hygiene, 69(2), 200-205.

Coluzzi, M. (1982). Spatial distribution of chromosomal inversions and speciation in anopheline mosquitoes. Mechanisms of speciation. C. Bargozzi. New York., Alan R. Liss**:** 143-153.

Coluzzi, M. (1994). "Malaria and the Afrotropical ecosystems: impact of man-made environmental changes." Parassitologia **36**(1-2): 223-227.

Coluzzi, M., Petrarca, V., & di Deco, M. A. (1985). Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. Italian Journal of Zoology, 52(1-2), 45-63.

Coluzzi, M., Sabatini, A., della Torre, A., Di Deco, M. A., & Petrarca, V. (2002). A polytene chromosome analysis of the *Anopheles gambiae* species complex. Science, 298(5597), 1415-1418.

Coluzzi, M., Sabatini, A., Petrarca, V., & Di Deco, M. A. (1979). Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. Transactions of the Royal Society of Tropical Medicine and Hygiene, 73(5), 483-497.

Cook, P. E., and S. P. Sinkins (2010). "Transcriptional profiling of *Anopheles gambiae* mosquitoes for adult age estimation." Insect molecular biology **19**(6): 745-751.

Corander, J. and P. Marttinen (2006). "Bayesian identification of admixture events using multilocus molecular markers." Molecular Ecology**15**(10): 2833-2843.

Corander, J., Waldmann, P., Marttinen, P., & Sillanpää, M. J. (2004). BAPS 2: enhanced possibilities for the analysis of genetic population structure. Bioinformatics, 20(15), 2363-2369.

Corander, J., Waldmann, P., & Sillanpää, M. J. (2003). Bayesian analysis of genetic differentiation between populations. Genetics, 163(1), 367-374.

Costantini, C., Sagnon, N., Ilboudo-Sanogo, E., Coluzzi, M., & Boccolini, D. (1999). Chromosomal and bionomic heterogeneities suggest incipient speciation in *Anopheles funestus* from Burkina Faso. Parassitologia, 41(4), 595-611.

Corbel, V., Akogbeto, M., Damien, G. B., Djenontin, A., Chandre, F., Rogier, C., ... & Henry, M. C. (2012). Combination of malaria vector control interventions in pyrethroid resistance area in Benin: a cluster randomised controlled trial. The Lancet infectious diseases, 12(8), 617-626.

Crawford, J. E. and B. P. Lazzaro (2010). "The demographic histories of the M and S molecular forms of *Anopheles gambiae s.s.*" Mol Biol Evol**27**(8): 1739-1744.

Creasey, S. S. and A. D. Rogers (1999). "Population genetics of bathyal and abyssal organisms." Advances in Marine Biology, Vol 35**35**: 1-151.

Cuamba, N., Morgan, J. C., Irving, H., Steven, A., & Wondji, C. S. (2010). High level of pyrethroid resistance in an *Anopheles funestus* population of the Chokwe District in Mozambique. PLoS One, 5(6), e11010.

Curtis, C. F., Jana-Kara, B., & Maxwell, C. A. (2002). Insecticide treated nets: impact on vector populations and relevance of initial intensity of transmission and pyrethroid resistance. Journal of vector borne diseases, 40(1-2), 1-8.

Curtis, C. F., Maxwell, C. A., Finch, R. J., & Njunwa, K. J. (1998). A comparison of use of a pyrethroid either for house spraying or for bednet treatment against malaria vectors. Tropical Medicine & International Health, 3(8), 619-631.

Cygler, M., Schrag, J. D., Sussman, J. L., Harel, M., Silman, I., Gentry, M. K., and Doctor, B. P. (1993). Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. Protein Sci. 2, 366–382

Daborn, P. J., Lumb, C., Harrop, T. W., Blasetti, A., Pasricha, S., Morin, S., ... & Batterham, P. (2012). Using< i>*Drosophila melanogaster*</i> to validate metabolism-based insecticide resistance from insect pests. Insect biochemistry and molecular biology, 42(12), 918-924.

Daborn, P. J., Yen, J. L., Bogwitz, M. R., Le Goff, G., Feil, E., Jeffers, S., ... & Wilson, T. G. (2002). A single P450 allele associated with insecticide resistance in *Drosophila.* Science, 297(5590), 2253-2256.

Dakin, E. E. and J. C. Avise (2004). "Microsatellite null alleles in parentage analysis."

Heredity**93**(5): 504-509.

Daoust, S. P., Mader, B. J., Maure, F., McLaughlin, J. D., Thomas, F., & Rau, M. E. (2010). Experimental evidence of size/age-biased infection of< i> Biomphalaria glabrata</i>(P*ulmonata: Planorbidae*) by an incompatible parasite species: Consequences for biological control. Infection, Genetics and Evolution, 10(7), 1008-1012.

Darsie, R. E. (2005). "Key to the pupae of the mosquitoes (Diptera : Culicidae) of Florida." Proceedings of the Entomological Society of Washington **107**(4): 892-902.

David, J. P., Strode, C., Vontas, J., Nikou, D., Vaughan, A., Pignatelli, P. M., ... & Ranson, H. (2005). The *Anopheles gambiae* detoxification chip: a highly specific microarray to study metabolic-based insecticide resistance in malaria vectors. Proceedings of the National Academy of Sciences of the United States of America, 102(11), 4080-4084.

Davies, T. G. E., Field, L. M., Usherwood, P. N. R., & Williamson, M. S. (2007). DDT, pyrethrins, pyrethroids and insect sodium channels. IUBMB life, 59(3), 151-162.

Davidson, G. and G. B. White (1972). "The crossing characteristics of a new, sixth species in the *Anopheles gambiae* complex." Trans.R.Soc.Trop.Med.Hyg.**66**: 531-532.

de Gruijter JM1, L. O., Vermeulen M, Xue Y, Woodwark C, Gillson CJ, Coffey AJ, Ayub Q, Mehdi SQ, Kayser M, Tyler-Smith C (2011). "Contrasting signals of positive selection in genes involved in human skin-color variation from tests based on SNP scans and resequencing." Investig Geneti**2**(24).

De Meillon B, Van Eeden G, Coetzee L, Coetzee M, MeiswinkelR, Du Toit C, Hansford C, 1977. Observations on a species of the *Anopheles funestus* subgroup, a suspected exophilic vector of malaria parasites in north-eastern Transvaal, South Africa. Mosq News 37: 657–661.

De Jong, R., Liang, C.C., and Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of effects of spatial stimulus-response correspondence. Journal of Experimental Psychology: Human Perception and Performance, 20, 731-750.

Pitts, P. M., & Deininger, R. L. (1954). S-R compatibility:

Devine, G. J. and G. F. Killeen (2010). "The potential of a new larviciding method for the control of malaria vectors." Malar.J.**9:142.**: 142.

Dia, I., Boccolini, D., Antonio-Nkondjio, C., Costantini, C., & Fontenille, D. (2000). Chromosomal inversion polymorphism of *Anopheles funestus* from forest villages of South Cameroon. Parassitologia, 42(3/4), 227-230.

Dia, I., Guelbeogo, M. W., and Ayala, D. (2013). Advances and Perspectives in the Study of the Malaria Mosquito *Anopheles funestus*.

Dimopoulos, G., Christophides, G. K., Meister, S., Schultz, J., White, K. P., Barillas-Mury, C., & Kafatos, F. C. (2002). Genome expression analysis of *Anopheles gambiae*: responses to injury, bacterial challenge, and malaria infection. Proceedings of the National Academy of Sciences, 99(13), 8814-8819.

Ding, Y., Hawkes, N., Meredith, J., Eggleston, P., Hemingway, J., & Ranson, H. (2005). Characterization of the promoters of Epsilon glutathione transferases in the mosquito *Anopheles gambiae* and their response to oxidative stress. Biochem. J, 387, 879-888.

Ding, Z., Wang, C., Chen, S., & Yu, S. (2011). Diversity and selective sweep in the OsAMT1; 1 genomic region of rice. BMC evolutionary biology, 11(1), 61.

Djouaka, R., Irving, H., Tukur, Z., & Wondji, C. S. (2011). Exploring mechanisms of multiple insecticide resistance in a population of the malaria vector *Anopheles funestus* in Benin. PLos one, 6(11), e27760.

Djouaka, R. F., Bakare, A. A., Coulibaly, O. N., Akogbeto, M. C., Ranson, H., Hemingway, J., & Strode, C. (2008). Expression of the cytochrome P450s, *CYP6P3* and *CYP6M2* are significantly elevated in multiple pyrethroid resistant populations of *Anopheles gambiae* ss from Southern Benin and Nigeria. BMC genomics, 9(1), 538.

Doannio, J. M., Dossou-Yovo, J., Diarrassouba, S., Chauvancy, G., Darriet, F., Chandre, F., ... & Carnevale, P. (1998). [Efficacy of permethrin-impregnated Olyset Net mosquito nets in a zone with pyrethroid resistant vectors. I--Entomologic evaluation]. Medecine tropicale: revue du Corps de sante colonial, 59(4), 349-354.

Dong, Y., Aguilar, R., Xi, Z., Warr, E., Mongin, E., & Dimopoulos, G. (2006). *Anopheles gambiae* immune responses to human and rodent Plasmodium parasite species. PLoS pathogens, 2(6), e52.

Du, W., Awolola, T. S., Howell, P., Koekemoer, L. L., Brooke, B. D., Benedict, M. Q., ... & Zheng, L. (2005). Independent mutations in the Rdl locus confer dieldrin resistance to *Anopheles gambiae* and *An. arabiensis*. Insect molecular biology, 14(2), 179-183.

Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P., and Trent, J. M. (1999). Expression profiling using cDNA microarrays. Nature genetics, 21, 10-14.

Durand, O. F. a. E. (2010). "Spatially explicit Bayesian clustering models in population genetics." Molecular Ecology Resources, (Special Issue on Spatial Genetics), in press.

Dupanloup, I., Schneider, S., & Excoffier, L. (2002). A simulated annealing approach to define the genetic structure of populations. Molecular Ecology, 11(12), 2571-2581.

Edi, Constant V., Djogbénou, Luc, Jenkins, Adam M., Regna, Kimberly, Muskavitch, Marc A. T., Poupardin, Rodolphe, Jones, Christopher M., Essandoh, John, Kétoh,

Guillaume K., Paine, Mark, Koudou, Benjamin, Donnelly, Martin, Ranson, Hilary and Weetman, David (2014) 'CYP6 P450 Enzymes and ACE-1 Duplication Produce Extreme and Multiple Insecticide Resistance in the Malaria Mosquito *Anopheles ga*mbiae'. PLoS Genetics, Vol 10, Issue 3, e1004236.

Eisen, J. A., 1998 A phylogenomic study of the MutS family of proteins. Nucleic Acids Res. 26: 4291–4300.

Enayati, A. and J. Hemingway (2010). "Malaria Management: Past, Present, and Future." Annual Review of Entomology**55**: 569-591.

Endler JA, Geographic variation, speciation and clines. May RM, editor. Princetown, New Jeysey: Princetown University Press; 1977.

Epplen, C., Frank, G., Gomolka, M., Albert, E., Nurnberg, P. &Epplen, J.T. (1994a) Dinucleotide repeat polymorphism in the human IL1A gene. Human Molecular Genetics, 3, 1710.

Estevinho, L., Pereira, A. P., Moreira, L., Dias, L. G., & Pereira, E. (2008). Antioxidant and antimicrobial effects of phenolic compounds extracts of Northeast Portugal honey. Food and Chemical Toxicology, 46(12), 3774-3779.

Estoup, A. and B. Angers (1998). "Microsatellites and minisatellites for molecular ecology: Theoretical and empirical considerations." Advances in Molecular Ecology**306**: 55-86.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology, 14(8), 2611-2620.

Favia, G., A. della Torre, & Coluzzi. M. (1997). "Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation." Insect Molecular Biology**6**(4): 377-383.

Ferguson, M. M. and R. G. Danzmann (1998). "Role of genetic markers in fisheries and aquaculture: useful tools or stamp collecting?" Canadian Journal of Fisheries and Aquatic Sciences**55**(7): 1553-1563.

Ferguson, H. M., Dornhaus, A., Beeche, A., Borgemeister, C., Gottlieb, M., Mulla, M. S., and Killeen, G. F. (2010). Ecology: a prerequisite for malaria elimination and eradication. PLoS medicine, 7(8), e1000303.

Feyereisen, R. (2005). Insect cytochrome P450. *Comprehensive Molecular Insect Science*. Oxford, Elsevier. **4:** 1-77.

ffrench-Constant, R. H. (1993). "Cloning of a putative GABAA receptor from cyclodiene-resistant *Drosophila*: a case study in the use of insecticide-resistant mutants to isolate neuroreceptors." EXS**63**: 210-223.

Fontenille, D., Cohuet, A., Awono-Ambene, P., Kengne, P., Antonio-Nkondjio, C., Wondji, C., & Simard, F. (2005). [Malaria vectors: from the field to genetics. Research in Africa]. Revue d'epidemiologie et de sante publique, 53(3), 283-290.

Fontenille, D. and L. Lochouarn (1999). "The complexity of the malaria vectorial system in Africa." Parassitologia**41**: 267-271.

Foster WA, Walker ED (2002). Mosquitoes (Culicidae). In:Mullen G, Durden L (eds) Medical and Veterinary Entomology. Academic Press: New York. Chapter 12, pp 204–256.

Francois O, A. S., Guillot G (2006). "Bayesian clustering using hidden Markov random fields in spatial population genetics." Genetics**174**: 805-816.

Fransen, V. (2012). "Permanet 3.0." from http://www.vestergaard-frandsen.com/permanet/permanet-3.

Fu, Y. X. and W. H. I. A. L. Li (1993). "Maximum-Likelihood-Estimation Of Population Parameters." Genetics**134**(4): 1261-1270.

Garros, C., Koekemoer, L. L., Coetzee, M., Coosemans, M., & Manguin, S. (2004). A single multiplex assay to identify major malaria vectors within the African *Anopheles funestus* and the Oriental An. minimus groups. The American journal of tropical medicine and hygiene, 70(6), 583-590.

Gatton, M. L., Chitnis, N., Churcher, T., Donnelly, M. J., Ghani, A. C., Godfray, H. C. J., and Lindsay, S. W. (2013). The importance of mosquito behavioural adaptations to malaria control in Africa. Evolution, 67(4), 1218-1230.

Gillies, M. T. and M. Coetzee (1987). A supplement to the Anophelinae of Africa south of the Sahara (Afrotropical region). Johannesburg, South African Institute for medical research.

Gillies, M. T. and B. DeMeillon (1968). The Anophelinae of Africa South of the Sahara, The South African Institute for Medical Research.

Gillies, M. T. C., M. (1987). A supplement to: The Anophelinae of Africa South of the Sahara, The South African Institute for Medical Research.

Goodman, S. J. (1997). RST Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. Molecular Ecology, 6(9), 881-885.

Gould, F.. 2010. Applying evolutionary biology: from retrospective analysis to direct manipulation. *In* M. A. Bell, D. J. Futuyma, W. F. Eanes, and J. S. Levinton, eds. *Evolution* since Darwin: the first 150 years. Sinauer Associates, Inc., Sunderland.

Goudet, J., Raymond, M., de Meeüs, T., & Rousset, F. (1996). Testing differentiation in diploid populations. Genetics, 144(4), 1933-1940.

Graham K., Kayedi M.H., Maxwell C., Kaur H., Rehman H., Malima R., Curtis C.F., Lines J.D., Rowland M.W. Multi-country field trials comparing wash-resistance of PermaNet and conventional insecticide-treated nets against anopheline and culicine mosquitoes. Med. Vet. Entomol. 2005;19:72–83.

Green, C. A., and R. H. Hunt (1980). "Interpretation of variation in ovarian polytene chromosomes of *Anopheles Funestus Giles*, *A. Parensis Gillies, and A. Aruni?*." " Genetica**51**(3): 187-195.

Green, C. A. (1982). "Cladistic analysis of mosquito chromosome data (*Anopheles* (Cellia) *Myzomyia*." J.Hered.**73**(1): 2-11.

Gregory, R., Darby, A. C., Irving, H., Coulibaly, M. B., Hughes, M., Koekemoer, L. L., ... & Wondji, C. S. (2011). A de novo expression profiling of *Anopheles funestus*, malaria vector in Africa, using 454 pyrosequencing. PLoS One, 6(2), e17418.

Guelbeogo, W. M., Grushko, O., Boccolini, D., Ouedraogo, P. A., Besansky, N. J., Sagnon, N. F., & Costantini, C. (2005). Chromosomal evidence of incipient speciation in the Afrotropical malaria mosquito *Anopheles funestus*. Medical and veterinary entomology, 19(4), 458-469.

Guillemaud, T., Lenormand, T., Bourguet, D., Chevillon, C., Pasteur, N., & Raymond, M. (1998). Evolution of resistance in *Culex pipiens*: allele replacement and changing environment. Evolution, 443-453.

Harbach RE. (2012) Mosquito Taxanomic Inventory, http://mosquito-taxonomic-inventory.info/, accessed on November 2012. Available from: http://mosquito-taxonomic-inventory.info/.

Haddow, A. J. (1943). "Measurements of temperature and light in artificial pools with reference to the larval habitat of *Anopheles* (M*yzomyia*) *gambiae Giles* and A. (M.) *funestus Giles*." Bull.Entomol.Res.**34**: 89.

Hartl, D. L. (1988). A primer of population genetics. Sinauer Associates, Inc..

Hamon, J. (1963). "[Study of the physiological age of Anopheles females in DDT-treated and non-treated areas of the Bobo-Dioulasso district, Haute-Volta.]." Bull World Health Organ**28**(1): 83-109.

Hardstone, M. C., Leichter, C. A., & Scott, J. G. (2009). Multiplicative interaction between the two major mechanisms of permethrin resistance, kdr and cytochrome P450-monooxygenase detoxification, in mosquitoes. Journal of evolutionary biology, 22(2), 416-423.

Hardstone, M. C., Leichter, C., Harrington, L. C., Kasai, S., Tomita, T., & Scott, J. G. (2007). Cytochrome P450 monooxygenase-mediated permethrin resistance confers limited and larval specific cross-resistance in the southern house mosquito,< i>*Culex pipiensquinquefasciatus*</i>. Pesticide biochemistry and physiology, 89(3), 175-184.

Hargreaves, K., Koekemoer, L. L., Brooke, B. D., Hunt, R. H., Mthembu, J., & Coetzee, M. (2000). *Anopheles funestus* resistant to pyrethroid insecticides in South Africa. Medical and veterinary entomology, 14(2), 181-189.

Harr, B., Kauer, M., & Schlötterer, C. (2002). Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences, 99(20), 12949-12954.

Hassall, K. A. (1990). The Biochemistry and Uses of Pesticides. Macmillan press Limited, London, 536 pp.

Hayes, J. D. and C. R. Wolf (1998). "Role of glutathione transferase in drug resistance". In: Sies H, Ketterer B (eds) Glutathione conjugation: Mechanisms and Biological Significance. Academic Press, London, pp. 315-355.

Hayes, J. D. and C. R. Wolf (1990). "Molecular mechanisms of drug resistance." Biochem.J.**272**(2): 281-295.

Hemingway, J. (1985). "Malathion carboxylesterase enzymes in *Anopheles arabiensis* from Sudan." Pesticide Biochemistry and Physiology**23**(3): 309-313.

Hemingway, J. and I. I. Bates (2003). "Malaria: past problems and future prospects - After more than a decade of neglect, malaria is finally back on the agenda for both biomedical research and public health politics." Embo Reports**4**: S29-S31.

Hemingway, J., Field, L., & Vontas, J. (2002). An overview of insecticide resistance. Science, 298(5591), 96-97.

Hemingway, J., Hawkes, N. J., McCarroll, L., & Ranson, H. (2004). The molecular basis of insecticide resistance in mosquitoes. Insect biochemistry and molecular biology, 34(7), 653-665.

Hemingway, J. and S. H. Karunaratne (1998). "Mosquito carboxylesterases: a review of the molecular biology and biochemistry of a major insecticide resistance mechanism." Med Vet Entomol**12**(1): 1-12.

Hemingway, J. and H. Ranson (2000). "Insecticide resistance in insect vectors of human disease." Annu. Rev. Entomol.**45**: 371-391.

Hillis, D. M. (1998). "Taxonomic sampling, phylogenetic accuracy, and investigator bias." Systematic Biology 3-8.

Hoelzel, A. R., Hancock, J. M., & Dover, G. A. (1991). Evolution of the cetacean mitochondrial D-loop region. Molecular Biology and Evolution, 8(4), 475-493.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics, 65-70.

Hosie, A., Sattelle, D., & Aronstein, K. (1997). Molecular biology of insect neuronal GABA receptors. Trends in neurosciences, 20(12), 578-583.

Huang, H. S., Hu, N. T., Yao, Y. E., Wu, C. Y., Chiang, S. W., & Sun, C. N. (1998). Molecular cloning and heterologous expression of a glutathione *S*-transferase involved in insecticide resistance from the diamondback moth, *Plutella xylostella*. Insect Biochemistry and Molecular Biology, 28(9), 651-658.

Hudson, R. R., Kreitman, M., & Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. Genetics, 116(1), 153-159.

Hunt, R. H., Edwardes, M., & Coetzee, M. (2010). Pyrethroid resistance in southern African *Anopheles funestus* extends to Likoma Island in Lake Malawi. Parasit Vectors, 3(122), 122.

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), 231-235.

Ihle, S., Ravaoarimanana, I., Thomas, M., & Tautz, D. (2006). An analysis of signatures of selective sweeps in natural populations of the house mouse. Molecular Biology and Evolution, 23(4), 790-797.

Ihle, Sonja (2004) Detecting genes involved in selective sweeps within populations of the house mouse species ... PhD thesis, Universität zu Köln.

Irving H, Riveron JM, Ibrahim SS, Lobo NF, Wondji CS (2012) Positional cloning of rp2 QTL associates the P450 genes *CYP6Z1, CYP6Z3* and *CYP6M7* with pyrethroid resistance in the malaria vector *Anopheles funestus*. *Heredity (Edinb)*109, 383-392.

Itokawa, K., Komagata, O., Kasai, S., Masada, M., & Tomita, T. (2011). Cis-acting mutation and duplication: History of molecular evolution in a P450 haplotype responsible for insecticide resistance in Culex quinquefasciatus. Insect biochemistry and molecular biology, 41(7), 503-512.

Jakobsson, M. and N. A. Rosenberg (2007). "CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure." Bioinformatics**23**(14): 1801-1806.

Jarne, P. and P. J. L. Lagoda (1996). "Microsatellites, from molecules to populations and back." Trends In Ecology & Evolution**11**(10): 424-429.

Jimenez, J. A., K. A. Hughes, G. Alaks, L. Graham, and R. C. Lacy (1994). "An experimental study of inbreeding depression in a natural habitat." Science**266**: 271–273.

Jones, C. M., Sanou, A., Guelbeogo, W. M., Sagnon, N., Johnson, P. C., & Ranson, H. (2012). Aging partially restores the efficacy of malaria vector control in insecticide-resistant populations of *Anopheles gambiae sl* from Burkina Faso. Malar J, 11, 24.

Kane, N. C., & Rieseberg, L. H. (2007). Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*.

Genetics, 175(4), 1823-1834.

Kamau, L., Hunt, R., & Coetzee, M. (2002). Analysis of the population structure of *Anopheles funestus* (Diptera: Culicidae) from western and coastal Kenya using paracentric chromosomal inversion frequencies. Journal of medical entomology, 39(1), 78-83.

Kamau, L., Munyekenye, G. O., Koekemoer, L. L., Hunt, R. H., & Coetzee, M. (2003). A survey of the *Anopheles funestus* (Diptera: Culicidae) group of mosquitoes from 10 sites in Kenya with special emphasis on population genetic structure based on chromosomal inversion karyotypes. Journal of medical entomology, 40(5), 664-671.

Kauer, M. O., Dieringer, D., & Schlötterer, C. (2003). A microsatellite variability screen for positive selection associated with the "out of Africa" habitat expansion of *Drosophila melanogaster*. Genetics, 165(3), 1137-1148.

Kayser, M., Silke Brauer, and Mark Stoneking (2003). "A genome scan to detect candidate regions influenced by local natural selection in human populations." Molecular Biology and Evolution**20**(6): 893-900.

Kent RJ, Coetzee M, Mharakurwa S, Norris DE (2006). Feeding and indoor resting behaviour of the mosquito *Anopheles longipalpis* in an area of hyperendemic malaria transmission in southern Zambia. Medical and veterinary Entomology. (Research Support, N.I.H, Extramural Research Support, Non-U.S. Gov't) Dec;20(4):459-63

Killeen GF, Smith TA, Ferguson HM, Mshinda H, Abdulla S. (2007) Preventing childhood malaria in Africa by protecting adults from mosquitoes with insecticide-treated nets. PLoS Med 4: e229. doi:10.1371/journal.pmed.0040229.

Kim, Y. and W. Stephan (2002). "Detecting a local signature of genetic hitchhiking along a recombining chromosome." Genetics**160**(2): 765-777.

Kimura, M. and N. Takahata (1983). "Selective constraint in protein polymorphism: study of the effectively neutral mutation model by using an improved pseudosampling method." Proc Natl Acad Sci U S A**80**(4): 1048-1052.

Kloke, R. G., E. Nhamahanga, E., Hunt, R.H., & Coetzee, M. (2011). "Vectorial status and insecticide resistance of *Anopheles funestus* from a sugar estate in southern Mozambique." Parasites Vectors**4**: 16.

Koekemoer, L. L., Kamau, L., Hunt, R. H., & Coetzee, M. (2002). A cocktail polymerase chain reaction assay to identify members of the *Anopheles funestus* (Diptera: Culicidae) group. The American journal of tropical medicine and hygiene, 66(6), 804-811.

Kohn, M. H., Pelz, H. J., & Wayne, R. K. (2000). Natural selection mapping of the warfarin-resistance gene. Proceedings of the National Academy of Sciences, 97(14), 7911-7915.

Kryukov, G. V., Len A. Pennacchio, and Shamil R. Sunyaev (2007). "Most rare missense alleles are deleterious in humans: implications for complex disease and association studies." The American Journal of Human Genetics **80**(4): 727-739.

Krzywinski, J. and N. J. Besansky (2003). "Molecular systematics of Anopheles: from subgenera to subpopulations." Annu Rev Entomol**48**: 111-139.

Krzywinski, J. and N. J. I. Besansky (2003). "Molecular systematics of *Anopheles*: From subgenera to subpopulations." Annual Review Of Entomology**48**: 111-139.

Kawada, H., Dida, G. O., Ohashi, K., Komagata, O., Kasai, S., Tomita, T., and Takagi, M. (2011). Multimodal pyrethroid resistance in malaria vectors, *Anopheles gambiae ss*, *Anopheles arabiensis*, and *Anopheles funestus* ss in western Kenya. PloS one, 6(8), e22574.

Lande, R. (1998). "Anthropogenic, ecological and genetic factors in extinction and conservation." Researches on Population Ecology **40**(3): 259-269.

Lanzaro, G. C., Touré, Y. T., Carnahan, J., Zheng, L., Dolo, G., Traoré, S., ... & Taylor, C. E. (1998). Complexities in the genetic structure of *Anopheles gambiae* populations in west Africa as revealed by microsatellite DNA analysis. Proceedings of the National Academy of Sciences, 95(24), 14260-14265.

Lee KS, Divis PC, Zakaria SK, Matusop A, Julin RA, Conway DJ,Cox-Singh J, Singh B. (2011).*Plasmodium knowlesi*: reservoir hosts and tracking the emergence in humans and macaques. PLoS Pathog.7:e1002015. doi:10.1371/journal.ppat.1002015.

Le Goff, G., Boundy, S., Daborn, P. J., Yen, J. L., Sofer, L., Lind, R., ... & Madi-Ravazzi, L. (2003). Microarray analysis of cytochrome P450 mediated insecticide resistance in< i>Drosophila</i>. Insect biochemistry and molecular biology, 33(7), 701-708.

Lehmann, T., Blackston, C. R., Besansky, N. J., Escalante, A. A., Collins, F. H., & Hawley, W. A. (2000). Brief communication. The Rift Valley complex as a barrier to gene flow for *Anopheles gambiae* in Kenya: the mtDNA perspective. Journal of Heredity, 91(2), 165-168.

Lehmann, T., Hawley, W. A., Grebert, H., & Collins, F. H. (1998). The effective population size of *Anopheles gambiae* in Kenya: implications for population structure. Molecular Biology and Evolution, 15(3), 264-276.

Lehmann, T., Hawley, W. A., Grebert, H., Danga, M., Atieli, F., & Collins, F. H. (1999). The Rift Valley complex as a barrier to gene flow for *Anopheles gambiae* in Kenya. Journal of Heredity, 90(6), 613-621.

Lehmann, T., Licht, M., Elissa, N., Maega, B. T. A., Chimumbwa, J. M., Watsenga, F. T., ... & Hawley, W. A. (2003). Population structure of *Anopheles gambiae* in Africa. Journal of Heredity, 94(2), 133-147.

Lengeler C (2004): Insecticide-treated bed nets and curtains for preventingmalaria. Cochrane Database Syst Rev, 2(2):CD000363.

Lepoivre, T., Treilhaud, M., Auffray-Calvier, E., Rigal, J. C., & Blanloeil, Y. (2003, May). [Posterior reversible encephalopathy syndrome: about 2 cases related to the cyclosporine]. In Annales francaises d'anesthesie et de reanimation (Vol. 22, No. 5, pp. 466-469).

Lindsay, S. W., Snow, R. W., Armstrong, J. R. M., & Greenwood, B. M. (1989). Permethrin-impregnated bednets reduce nuisance arthropods in Gambian houses. Medical and veterinary entomology, 3(4), 377-383.

Lines, J. D., J. Myamba, and C. F. Curtis. (1987). Experimental hut trials of permethrin-impregnated mosquito nets and eave curtains against malaria vectors in Tanzania. *Med Vet Entomol* 1:37–51.

Livak, K. J. (1984). "Organization and mapping of a sequence on the *Drosophila melanogaster* X and Y chromosomes that is transcribed during spermatogenesis." Genetics**107**(4): 611-634.

Lochouarn, L., Dia, I., Boccolini, D., Coluzzi, M., & Fontenille, D. (1998). Bionomical and cytogenetic heterogeneities of *Anopheles funestus* in Senegal. Transactions of the Royal Society of Tropical Medicine and Hygiene, 92(6), 607-612.

Lumjuan, N., McCarroll, L., Prapanthadara, L. A., Hemingway, J., & Ranson, H. (2005). Elevated activity of an Epsilon class glutathione transferase confers DDT resistance in the dengue vector,< i> Aedes aegypti</i>. Insect biochemistry and molecular biology, 35(8), 861-871.

Luro, F. L., Costantino, G., Terol, J., Argout, X., Allario, T., Wincker, P., ... & Morillon, R. (2008). Transferability of the EST-SSRs developed on Nules clementine (Citrus clementina Hort ex Tan) to other Citrus species and their effectiveness for genetic mapping. BMC genomics, 9(1), 287.

Lynd, A., Weetman, D., Barbosa, S., Yawson, A. E., Mitchell, S., Pinto, J., ... & Donnelly, M. J. (2010). Field, genetic, and modeling approaches show strong positive selection acting upon an insecticide resistance mutation in Anopheles gambiae ss. Molecular biology and evolution, 27(5), 1117-1125.

Mantel, N. A. (1967). "The detection of disease clustering and a generalized regression approach." <u>Cancer Res</u>**27**: 209–220.

Martinez-Torres, D., Chandre, F., Williamson, M. S., Darriet, F., Berge, J. B., Devonshire, A. L., ... & Pauron, D. (1998). Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector *Anopheles gambiae* ss. Insect molecular biology, 7(2), 179-184.

Matsumura, F. (1975). "Toxicology of Insecticides." <u>Plenum Press, New York</u>: 325-354.

Maynard-Smith, J. and J. Haigh (1974). "Hitch-Hiking Effect of a Favorable Gene." Genetics Research**23**(1): 23-35.

Mbogo, C. N. M., Baya, N. M., Ofulla, A. V. O., Githure, J. I., and Snow, R. W. (1996). The impact of permethrin-impregnated bednets on malaria vectors of the Kenyan coast. Medical and veterinary entomology, 10(3), 251-259.

McCaffery, A. R. (1998). "Resistance to insecticides in heliothine Lepidoptera: a global view." Philosophical Transactions Of The Royal Society Of London Series B-Biological Sciences **353**(1376): 1735-1750.

McCart, C. and R. H. Ffrench-Constant (2008). "Dissecting the insecticide-resistance-associated cytochrome P450 gene *Cyp6g1*." Pest Manag Sci**64**(6): 639-645.

McDonald, J. H. and M. I. A. F. Kreitman (1991). "Adaptive protein evolution at the *Adh* locus in *Drosophila*." Nature **351**(6328): 652-654.

Mendis, C., Jacobsen, J. L., Gamage, Mendis, A., Bule, E., Dgedge, M., Thompson, R., ... & Høgh, B. (2000). *Anopheles arabiensis* and *An. funestus* are equally important vectors of malaria in Matola coastal suburb of Maputo, southern Mozambique. Medical and veterinary entomology, 14(2), 171-180.

Michel, A. P., Guelbeogo, W. M., Grushko, O., Schemerhorn, B. J., Kern, M., Willard, M. B., ... & Besansky, N. J. (2005). Molecular differentiation between chromosomally defined incipient species of *Anopheles funestus*. Insect molecular biology, 14(4), 375-387.

Michel, K., Suwanchaichinda, C., Morlais, I., Lambrechts, L., Cohuet, A., Awono-Ambene, P. H., ... & Kafatos, F. C. (2006). Increased melanizing activity in *Anopheles gambiae* does not affect development of Plasmodium falciparum. Proceedings of the National Academy of Sciences, 103(45), 16858-16863.

Milner, E., Sousa, J., Pybus, B., Auschwitz, J., Caridha, D., Gardner, S., ... & Dow, G. S. (2012). Ketotifen is an antimalarial prodrug of norketotifen with blood schizonticidal and liver-stage efficacy. European journal of drug metabolism and pharmacokinetics, 37(1), 17-22.

Mitchell, S. N., Stevenson, B. J., Müller, P., Wilding, C. S., Egyir-Yawson, A., Field, S. G., ... & Donnelly, M. J. (2012). Identification and validation of a gene causing cross-resistance between insecticide classes in *Anopheles gambiae* from Ghana. Proceedings of the National Academy of Sciences, 109(16), 6147-6152.

Miyazaki, M., Ohyama, K., Dunlap, D. Y., & Matsumura, F. (1996). Cloning and sequencing of the para-type sodium channel gene from susceptible and kdr-resistant German cockroaches (*Blattella germanica*) and house fly (*Musca domestica*). Molecular and General Genetics MGG, 252(1-2), 61-68.

Moiroux, N., Gomez, M. B., Pennetier, C., Elanga, E., Djènontin, A., Chandre, F., and Corbel, V. (2012). Changes in *Anopheles funestus* biting behavior following universal coverage of long-lasting insecticidal nets in Benin. Journal of Infectious Diseases, 206(10), 1622-1629.

Morey, J. S., Ryan, J. C., & Van Dolah, F. M. (2006). Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. Biological procedures online, 8(1), 175-193.

Morgan, J. C., Irving, H., Okedi, L. M., Steven, A., & Wondji, C. S. (2010). Pyrethroid resistance in an *Anopheles funestus* population from Uganda. PLoS One, 5(7), e11872.

Morjan, C. L., & Rieseberg, L. H. (2004). How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Molecular Ecology, 13(6), 1341-1356.

Mulamba, C., Riveron, J. M., Ibrahim, S. S., Irving, H., Barnes, K. G., Mukwaya, L. G., and Wondji, C. S. (2014). Widespread Pyrethroid and DDT Resistance in the Major Malaria Vector *Anopheles funestus* in East Africa Is Driven by Metabolic Resistance Mechanisms. PloS one, 9(10), e110058.

Müller, P., Warr, E., Stevenson, B. J., Pignatelli, P. M., Morgan, J. C., Steven, A., ... & Donnelly, M. J. (2008). Field-caught permethrin-resistant *Anopheles gambiae* overexpress *CYP6P3*, a P450 that metabolises pyrethroids. PLoS Genetics, 4(11), e1000286.

Murray CJ, R. L., Lim SS, Andrews KG, Foreman KJ, Haring D, Fullman N, Naghavi M, Lozano R, Lopez AD (2012). "Global malaria mortality between 1980 and 2010: a systematic analysis." Lancet**379**(413-31).

Nair, S., Williams, J. T., Brockman, A., Paiphun, L., Mayxay, M., Newton, P. N., ... & Anderson, T. J. (2003). A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. Molecular biology and evolution, 20(9), 1526-1536.

Najera, J. A. and M. Zaim (2002). Malaria Vector Control. Decision Making Criteria and Procedures for Judicious use of Insecticides. Geneva, WHO. WHO/CDS/WHOPES/2002.5: 1-116.

Nei, M. (1977). F-statistics and analysis of gene diversity in subdivided populations. Annals of human genetics, 41(2), 225-233.

Nei, M. (1987). Molecular Evolutionary Genetics. New York, Columbia University Press.

Nei, M. and T. Gojobori (1986). "Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions." Mol Biol Evol**3**(5): 418-426.

Nelson, D. R. (2013). "A world of cytochrome P450s." Philosophical Transactions of

the Royal Society B." Biological Sciences **368**(1612): 20120430.

N'Guessan, R., Corbel, V., Akogbéto, M., & Rowland, M. (2007). Reduced efficacy of insecticide-treated nets and indoor residual spraying for malaria control in pyrethroid resistance area, Benin. Emerging infectious diseases, 13(2), 199.

Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., & Clark, A. G. (2007). Recent and ongoing selection in the human genome. Nature Reviews Genetics, 8(11), 857-868.

Nielsen R. (2001). Statistical tests of selective neutrality in the ageof genomics. Heredity. 86:641–647.

Nikou, D., Ranson, H., & Hemingway, J. (2003). An adult-specific *CYP6* P450 gene is overexpressed in a pyrethroid-resistant strain of the malaria vector,< i>*Anopheles gambiae*</i>. Gene, 318, 91-102.

Okello, P. E., Van Bortel, W., Byaruhanga, A. M., Correwyn, A., Roelants, P., Talisuna, A., ... & Coosemans, M. (2006). Variation in malaria transmission intensity in seven sites throughout Uganda. The American journal of tropical medicine and hygiene, 75(2), 219-225.

Okoye, P. N., Brooke, B. D., Koekemoer, L. L., Hunt, R. H., & Coetzee, M. (2008). Characterisation of DDT, pyrethroid and carbamate resistance in *Anopheles funestus* from Obuasi, Ghana. Transactions of the Royal Society of Tropical Medicine and Hygiene, 102(6), 591-598.

Oosterhout, G. R., Huntington, C. W., Nickelson, T. E., & Lawson, P. W. (2005). Potential benefits of a conservation hatchery program for supplementing Oregon coast coho salmon (*Oncorhynchus kisutch*) populations: a stochastic model investigation. Canadian Journal of Fisheries and Aquatic Sciences, 62(8), 1920-1935.

P, C. (1990). "Markov random fields in statistics. In: Disorder in Physical Systems. ." A Volume in Honour of John M. Hammersley (eds Grimmett GR, Welsh DJA)(Oxford University
Press, Oxford, UK.): 19–32.

Peters, W. (1965). Drug resistance in< i>*Plasmodium berghei*</i>. I. Chloroquine resistance. Experimental parasitology, 17(1), 80-89.

Payseur, B. A., Cutter, A. D., & Nachman, M. W. (2002). Searching for evidence of positive selection in the human genome using patterns of microsatellite variability. Molecular Biology and Evolution, 19(7), 1143-1153.

Pearson, J. V., Huentelman, M. J., Halperin, R. F., Tembe, W. D., Melquist, S., Homer, N., and Craig, D. W. (2007). Identification of the genetic basis for complex disorders by use of pooling-based genome wide single-nucleotide–polymorphism association studies. The American Journal of Human Genetics, 80(1), 126-139.

Pedra, J. H. F., McIntyre, L. M., Scharf, M. E., & Pittendrigh, B. R. (2004). Genome-

wide transcription profile of field-and laboratory-selected dichlorodiphenyltrichloroethane (DDT)-resistant *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America, 101(18), 7034-7039.

Perera, M. D. B., Hemingway, J., & Karunaratne, S. P. (2008). Multiple insecticide resistance mechanisms involving metabolic changes and insensitive target sites selected in anopheline vectors of malaria in Sri Lanka. Malar J, 7(1), 168.

Pluess, B., Tanser, F. C., Lengeler, C., & Sharp, B. L. (2010). Indoor residual spraying for preventing malaria. Cochrane Database Syst Rev, 4(4).

Pombi, M., Caputo, B., Simard, F., Di Deco, M. A., Coluzzi, M., della Torre, A., ... & Petrarca, V. (2008). Chromosomal plasticity and evolutionary potential in the malaria vector *Anopheles gambiae sensu stricto*: insights from three decades of rare paracentric inversions. BMC evolutionary biology, 8(1), 309.

Pool, J. E., Corbett-Detig, R. B., Sugino, R. P., Stevens, K. A., Cardeno, C. M., Crepeau, M. W., ... & Langley, C. H. (2012). Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. PLoS genetics, 8(12), e1003080.

Powell, J. R., Petrarca, V., Della Torre, A., Caccone, A., & Coluzzi, M. (1999). Population structure, speciation, and introgression in the *Anopheles gambiae* complex. Parassitologia, 41(1-3), 101-113.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.

Provost, G. P. (1972). "Hospital-Based Pharmacy Service for Ambulatory Patients." American Journal of Hospital Pharmacy**29**(5): 393-&.

Provost, G. P. (1973). "Antibiotics and Rational Drug-Therapy." American Journal of Hospital Pharmacy**30**(2): 115-115.

Puinean, A. M., Foster, S. P., Oliphant, L., Denholm, I., Field, L. M., Millar, N. S., ... & Bass, C. (2010). Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. PLOS genetics, 6(6), e1000999.

R, F. (2011). "Arthropod CYPones illustrate the tempo and mode in P450 evolution." Biochem. J., Biophys. Acta**1814**.

Ranson, H., & Hemingway, J. (2005). Mosquito glutathione transferases. Methods in enzymology, 401, 226-241.

Ranson, H., Jensen, B., Vulule, J. M., Wang, X., Hemingway, J., & Collins, F. H. (2000). Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan *Anopheles gambiae* associated with resistance to DDT and pyrethroids. Insect molecular biology, 9(5), 491-497.

Ranson, H., N'Guessan, R., Lines, J., Moiroux, N., Nkuni, Z., & Corbel, V. (2011). Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control?. Trends in parasitology, 27(2), 91-98.

Ranson, H., Nikou, D., Hutchinson, M., Wang, X., Roth, C. W., Hemingway, J., & Collins, F. H. (2002). Molecular analysis of multiple cytochrome P450 genes from the malaria vector, *Anopheles gambiae*. Insect molecular biology, 11(5), 409-418.

Ray , D. E. (1991). "Pesticides derived from plants and other organisms in Hayes, W.J, Wayland."

Raymond, M. I. C. H. E. L., Fournier, D. I. D. I. E. R., Berge, J. E. A. N., Cuany, A., Bride, J. M., & Pasteur, N. (1985). Single-mosquito test to determine genotypes with an acetylcholinesterase insensitive to inhibition to propoxur insecticide. J. Am. Mosq. Control Assoc, 1(4).

Raymond, M., Fournier, D., Bride, J. M., Cuany, A., Berge, J., Magnin, M., & Pasteur, N. (1986). Identification of resistance mechanisms in *Culex pipiens* (Diptera: Culicidae) from southern France: insensitive acetylcholinesterase and detoxifying oxidases. Journal of economic entomology, 79(6), 1452-1458.

Raymond, M. and F. Rousset (1995). "GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism." J Hered**86**(3): 248-249.

Reyes-Villanueva, F., Garza-Hernandez, J. A., Garcia-Munguia, A. M., Tamez-Guerra, P., Howard, A. F., & Rodriguez-Perez, M. A. (2011). Dissemination of Metarhizium anisopliae of low and high virulence by mating behavior in *Aedes aegypti*. Parasit Vectors, 4, 171.

Richard H. ffrench-Constant, Phillip J. Daborn, Gaelle Le Goff. (2004). "The genetics and genomics of insecticide resistance." Trends In Genetics 20(3): 163-170.

Rieseberg, L. H. (2001). "Chromosomal rearrangements and speciation." Trends In Ecology & Evolution**16**(7): 351-358.

Rieseberg, L. H., & Burke, J. M. (2001). A genic view of species integration. Journal of Evolutionary Biology, 14(6), 883-886.

Riveron, J. M., Irving, H., Ndula, M., Barnes, K. G., Ibrahim, S. S., Paine, M. J., & Wondji, C. S. (2013). Directionally selected cytochrome P450 alleles are driving the spread of pyrethroid resistance in the major malaria vector *Anopheles funestus*. Proceedings of the National Academy of Sciences, 110(1), 252-257.

Riveron, J. M., Irving, H., Ndula, M., Barnes, K. G., Ibrahim, S. S., Paine, M. J., & Wondji, C. S. (2013). Directionally selected cytochrome P450 alleles are driving the spread of pyrethroid resistance in the major malaria vector *Anopheles funestus*. Proceedings of the National Academy of Sciences, 110(1), 252-257.

Riveron, J. M., Yunta, C., Ibrahim, S. S., Djouaka, R., Irving, H., Menze, B. D., and Wondji, C. S. (2014). A single mutation in the GSTe2 gene allows tracking of

Ranson, H., N'Guessan, R., Lines, J., Moiroux, N., Nkuni, Z., & Corbel, V. (2011). Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control?. Trends in parasitology, 27(2), 91-98.

Ranson, H., Nikou, D., Hutchinson, M., Wang, X., Roth, C. W., Hemingway, J., & Collins, F. H. (2002). Molecular analysis of multiple cytochrome P450 genes from the malaria vector, *Anopheles gambiae*. Insect molecular biology, 11(5), 409-418.

Ray , D. E. (1991). "Pesticides derived from plants and other organisms in Hayes, W.J, Wayland."

Raymond, M. I. C. H. E. L., Fournier, D. I. D. I. E. R., Berge, J. E. A. N., Cuany, A., Bride, J. M., & Pasteur, N. (1985). Single-mosquito test to determine genotypes with an acetylcholinesterase insensitive to inhibition to propoxur insecticide. J. Am. Mosq. Control Assoc, 1(4).

Raymond, M., Fournier, D., Bride, J. M., Cuany, A., Berge, J., Magnin, M., & Pasteur, N. (1986). Identification of resistance mechanisms in *Culex pipiens* (Diptera: Culicidae) from southern France: insensitive acetylcholinesterase and detoxifying oxidases. Journal of economic entomology, 79(6), 1452-1458.

Raymond, M. and F. Rousset (1995). "GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism." J Hered**86**(3): 248-249.

Reyes-Villanueva, F., Garza-Hernandez, J. A., Garcia-Munguia, A. M., Tamez-Guerra, P., Howard, A. F., & Rodriguez-Perez, M. A. (2011). Dissemination of Metarhizium anisopliae of low and high virulence by mating behavior in *Aedes aegypti*. Parasit Vectors, 4, 171.

Richard H. ffrench-Constant, Phillip J. Daborn, Gaelle Le Goff. (2004). "The genetics and genomics of insecticide resistance." Trends In Genetics 20(3): 163-170.

Rieseberg, L. H. (2001). "Chromosomal rearrangements and speciation." Trends In Ecology & Evolution**16**(7): 351-358.

Rieseberg, L. H., & Burke, J. M. (2001). A genic view of species integration. Journal of Evolutionary Biology, 14(6), 883-886.

Riveron, J. M., Irving, H., Ndula, M., Barnes, K. G., Ibrahim, S. S., Paine, M. J., & Wondji, C. S. (2013). Directionally selected cytochrome P450 alleles are driving the spread of pyrethroid resistance in the major malaria vector *Anopheles funestus*. Proceedings of the National Academy of Sciences, 110(1), 252-257.

Riveron, J. M., Irving, H., Ndula, M., Barnes, K. G., Ibrahim, S. S., Paine, M. J., & Wondji, C. S. (2013). Directionally selected cytochrome P450 alleles are driving the spread of pyrethroid resistance in the major malaria vector *Anopheles funestus*. Proceedings of the National Academy of Sciences, 110(1), 252-257.

Riveron, J. M., Yunta, C., Ibrahim, S. S., Djouaka, R., Irving, H., Menze, B. D., and Wondji, C. S. (2014). A single mutation in the GSTe2 gene allows tracking of

metabolically based insecticide resistance in a major malaria vector. Genome biology, 15(2), R27.

Roh, S. W., Kim, K. H., Nam, Y. D., Chang, H. W., Park, E. J., & Bae, J. W. (2009). Investigation of archaeal and bacterial diversity in fermented seafood using barcoded pyrosequencing. The ISME journal, 4(1), 1-16.

Rozas, J. P. L., J.C. Sánchez-Del Barrio, X. Messeguer and R. Rozas (2010). "DnaSP Version 5 Help Contents " [Help File]. Available with the program at http://www.ub.edu/dnasp/.

Rozendaal JA. (1989) Impregnated mosquito nets and curtains for self protection and vector control. Tropical disease bulletin, 7:R1-R41.

Russell, T. L., Govella, N. J., Azizi, S., Drakeley, C. J., Kachur, S. P., and Killeen, G. F. (2011). Increased proportions of outdoor feeding among residual malaria vector populations following increased use of insecticide-treated nets in rural Tanzania. Malar J, 10(80), 80.

Saavedra Rodriguez, K., Urdaneta-Marquez, L., Rajatileka, S., Moulton, M., Flores, A. E., Fernandez Salas, I., and Black, W. (2007). A mutation in the voltage gated sodium channel gene associated with pyrethroid resistance in Latin American *Aedes aegypti*. Insect molecular biology, 16(6), 785-798.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., ... & Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. Nature, 419(6909), 832-837.

Salvan, M. and J. Mouchet (1994). "*Aedes-Albopictus* and *Aedes-Aegypti* on the Island of La Reunion." Annales de la Societe Belge de Medecine Tropicale **74**(4): 323-326.

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 270(5235), 467-470.

Schlenke, T. A. and D. J. Begun (2004). "Strong selective sweep associated with a transposon insertion in *Drosophila simulans*." Proceedings of the National Academy of Sciences **101**(6): 1626-1631.

Schlötterer, C. (2001). "Genealogical inference of closely related species based on microsatellites." Genetical Research **78**(3): 209-212.

Schlötterer, C. (2002). "A microsatellite-based multilocus screen for the identification of local selective sweeps." Genetics **160**(2): 753-763.

Schlötterer, C. and M. Agis (2002). "Microsatellite analysis of Drosophila melanogaster populations along a microclimatic contrast at lower Nahel Oren canyon, Mount Carmel, Israel." Mol Biol Evol **19**(4): 563-568.

Schlötterer c, P. J. (1994). "The use of microsatellites for genetic analysis of natural

populations." <u>Pub Med</u>**178**(3): 239-248.

Scholte, E. J., Njiru, B. N., Smallegange, R. C., Takken, W., & Knols, B. G. (2003). Infection of malaria (*Anopheles gambiae ss*) and filariasis (*Culex quinquefasciatus*) vectors with the entomopathogenic fungus *Metarhizium anisopliae*. Malaria Journal, 2(1), 29.

Schug, Z. T., da Fonseca, P. C., Bhanumathy, C. D., Wagner, L., Zhang, X., Bailey, B., ... & Joseph, S. K. (2008). Molecular characterization of the inositol 1, 4, 5-trisphosphate receptor pore-forming segment. Journal of Biological Chemistry, 283(5), 2939-2948.

Scott, J. A., Brogdon, W. G., & Collins, F. H. (1993). Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. The American journal of tropical medicine and hygiene, 49(4), 520-529.

Scott, J. G. W. (1999). "Cytochromes P450 and insecticide resistance." <u>Insect Biochemistry And Molecular Biology</u>**29**(9): 757-777.

Sérandour, J., Willison, J., Thuiller, W., Ravanel, P., Lempérière, G., & Raveton, M. (2010). Environmental drivers for Coquillettidia mosquito habitat selection: a method to highlight key field factors. Hydrobiologia, 652(1), 377-388.

Sharakhov, I. V., Serazin, A. C., Grushko, O. G., Dana, A., Lobo, N., Hillenmeyer, M. E., ... & Besansky, N. J. (2002). Inversions and gene order shuffling in *Anopheles gambiae* and A. funestus. Science, 298(5591), 182-185.

Sharakhov, I. V., Sharakhova, M. V., Mbogo, C. M., Koekemoer, L. L., & Yan, G. (2001). Linear and spatial organization of polytene chromosomes of the African malaria mosquito *Anopheles funestus*. Genetics, 159(1), 211-218.

Sharakhov, I., Braginets, O., Grushko, O., Cohuet, A., Guelbeogo, W. M., Boccolini, D., ... & Besansky, N. J. (2004). A microsatellite map of the African human malaria vector *Anopheles funestus*. Journal of Heredity, 95(1), 29-34.

Sharp, B. L., Kleinschmidt, I., Streat, E., Maharaj, R., Barnes, K. I., Durrheim, D. N., ... & Barreto, A. (2007). Seven years of regional malaria control collaboration—Mozambique, South Africa, and Swaziland. The American journal of tropical medicine and hygiene, 76(1), 42-47.

Silva, A. X., Bacigalupe, L. D., Luna-Rudloff, M., & Figueroa, C. C. (2012). Insecticide resistance mechanisms in the green peach aphid Myzus persicae (Hemiptera: Aphididae) II: Costs and benefits. PloS one, 7(6), e36810.

Silva, R. O., Heloisa HG Silva, and Christian Luz (2008). "Effect of *Metarhizium anisopliae* isolated from soil samples of the central Brazilian cerrado against Aedes aegypti larvae under laboratory conditions." <u>Revista de Patologia Tropical</u> **33**(2).

Simard, F., Lehmann, T., Lemasson, J. J., Diatta, M., & Fontenille, D. (2000). Persistence of *Anopheles arabiensis* during the severe dry season conditions in Senegal: an indirect approach using microsatellite loci. Insect molecular biology, 9(5), 467-479.

Sinkins, S. P., B. J. Hackett, et al. (2000). "Isolation of polymorphic microsatellite loci from the malaria vector *Anopheles funestus*." Mol Ecol**9**(4): 490-492.

Slatkin, M. (1985). "Gene flow in natural populations." Annu.Rev.Ecol.Syst.**16**: 393-430.

Slatkin, M. (1994). "Linkage disequilibrium in growing and stable populations." Genetics**137**(1): 331-336.

Slatkin, M. (1995). "A measure of population subdivision based on microsatellite allele frequencies." Genetics**139**(1): 457-462.

Slatkin, M. and T. Wiehe (1998). "Genetic hitch-hiking in a subdivided population." Genetics Research**71**(2): 155-160.

Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y., & Hay, S. I. (2005). The global distribution of clinical episodes of *Plasmodium falciparum* malaria. Nature, 434(7030), 214-217.

Sparks, T. C., Lockwood, J. A., Byford, R. L., Graves, J. B., & Leonard, B. R. (1989). The role of behavior in insecticide resistance. Pesticide science, 26(4), 383-399.

Spillings, B. L., Brooke, B. D., Koekemoer, L. L., Chiphwanya, J., Coetzee, M., & Hunt, R. H. (2009). A new species concealed by *Anopheles funestus Giles*, a major malaria vector in Africa. The American journal of tropical medicine and hygiene, 81(3), 510-515.

 Stepien, C. A., and Joseph E. Faber (1998). ""Population genetic structure, phylogeography and spawning philopatry in walleye (*Stizostedion vitreum*) from mitochondrial DNA control region sequences" Molecular Ecology Notes**7**(12): 1757-1769.

Strode, C., Wondji, C. S., David, J. P., Hawkes, N. J., Lumjuan, N., Nelson, D. R., ... & Ranson, H. (2008). Genomic analysis of detoxification genes in the mosquito< i>*Aedes aegypti*</i>. Insect biochemistry and molecular biology, 38(1), 113-123.

Surtees, G. (1970). "Effects of Irrigation on Mosquito Populations and Mosquito-Borne Diseases in Man, with Particular Reference to Ricefield Extension." International Journal of Environmental Studies**1**(1): 35-42.

Sutherst, R. W. and R. S. Tozer (1995). "Control of Buffalo Fly (*Haematobia-Irritans Exigua De Meijere*) on Dairy and Beef-Cattle Using Traps." Australian Journal of Agricultural Research **46**(2): 269-284.

Tabachnick, W. J. t., and W. C. Black Iv (1995). "Making a case for molecular population genetic studies of arthropod vectors." Parasitology Today **11**(1): 27-30.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., ... & Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Research, 24(16), 3189-3194.

Tajima, F. (1983). "Evolutionary relationship of DNA sequences in finite populations." Genetics **105**: 437-460.

Tajima, F. (1996). "The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites." Genetics **143**: 1457-1465.

Tajima, F. I. A. A. (1989). "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." Genetics **123**(3): 585-595.

Takken, W. and B. G. J. Knols (1999). "Odor-mediated behavior of Afrotropical malaria mosquitoes." Annual Review of Entomology **44**: 131-157.

Takken, W. (2002). Do insecticide treated bednets have an effect on malaria vectors?. Tropical Medicine & International Health, 7(12), 1022

Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular biology and evolution, 24(8), 1596-1599.

Tamura K, P. D., Peterson N, Stecher G, Nei M, Kumar S (2011). "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods." Mol Biol Evol **28**(10): 2731–2739.

Temu, E. A., Hunt, R. H., & Coetzee, M. (2004). Microsatellite DNA polymorphism and heterozygosity in the malaria vector mosquito< i>*Anopheles funestus*</i>(Diptera: Culicidae) in east and southern Africa. Acta tropica, 90(1), 39-49.

Teschke, M., Mukabayire, O., Wiehe, T., & Tautz, D. (2008). Identification of selective sweeps in closely related populations of the house mouse based on microsatellite scans. Genetics, 180(3), 1537-1545.

Thompson, M., Steichen, J. C., & Ffrench-Constant, R. H. (1993). Conservation of cyclodiene insecticide resistance-associated mutations in insects. Insect molecular biology, 2(3), 149-154.

Tijet, N., Helvig, C., & Feyereisen, R. (2001). The cytochrome P450 gene superfamily in< i>*Drosophila melanogaster*</i>: Annotation, intron-exon organization and phylogeny. Gene, 262(1), 189-198.

Itokawa, K., Komagata, O., Kasai, S., Masada, M., & Tomita, T. (2011). < i> Cis</i>-acting mutation and duplication: History of molecular evolution in a P450 haplotype

responsible for insecticide resistance in< i>*Culex quinquefasciatus*</i>. Insect biochemistry and molecular biology, 41(7), 503-512.

Tomita, T. and J. G. Scott (1995). "cDNA and deduced protein sequence of *CYP6D1*: the putative gene for a cytochrome P450 responsible for pyrethroid resistance in house fly." Insect Biochem Mol Biol**25**(2): 275-283.

Torre, A. D., Fanello, C., Akogbeto, M., Dossou-yovo, J., Favia, G., Petrarca, V., & Coluzzi, M. (2001). Molecular evidence of incipient speciation within *Anopheles gambiae.ss* in West Africa. Insect molecular biology, 10(1), 9-18.

Toure, Y. (1982). "Study of Anopheles funestus and *Anopheles gambiae s.l.* susceptibility to insecticides in rural areas of Sudan savanna in Mali." Cashiers ORSTOM, Series Entomologie Medicale et Parasitologie**20**: 125-131.

Touré, Y.T., Dolo, G., Petrarca, V., et al., (1998). Mark-release-recapture experiments
with *Anopheles gambiae s.l.* in Banambani Village, Mali, to determine population size and structure. *Medical and Veterinary Entomology,* 12 (1), 74-83.

Townson, H., Nathan, M. B., Zaim, M., Guillet, P., Manga, L., Bos, R., & Kindhauser, M. (2005). Exploiting the potential of vector control for disease prevention. Bulletin of the world health organization, 83(12), 942-947.

Tseng, G. C., Oh, M. K., Rohlin, L., Liao, J. C., & Wong, W. H. (2001). Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. Nucleic acids research, 29(12), 2549-2557.

Utzinger, J., Tozan, Y., & Singer, B. H. (2001). Efficacy and cost-effectiveness of environmental management for malaria control. Tropical Medicine & International Health, 6(9), 677-687.

Vais, H., Williamson, M. S., Devonshire, A. L., & Usherwood, P. N. R. (2001). The molecular interactions of pyrethroid insecticides with insect and mammalian sodium channels. Pest management science, 57(10), 877-888.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P., & Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes, 4(3), 535-538.

Vaughan, A. and J. W. A. R. Hemingway (1995). "Mosquito Carboxylesterase Est-Alpha-2(1)(A(2)) - Cloning And Sequence Of The Full-Length Cdna For A Major Insecticide Resistance Gene Worldwide In The Mosquito *Culex-Quinquefasciatus*." Journal Of Biological Chemistry**270**(28): 17044-17049.

Vigouroux, Y., McMullen, M., Hittinger, C. T., Houchins, K., Schulz, L., Kresovich, S., ... & Doebley, J. (2002). Identifying genes of agronomic importance in maize by

screening microsatellites for evidence of selection during domestication. Proceedings of the National Academy of Sciences, 99(15), 9650-9655.

Volkan-Salanci, and G. E. Utine, . (2007). "Chromosomal abnormalities presented with seizures." Chromosome Research**15**: 84-84.

Vontas, J., Blass, C., Koutsos, A. C., David, J. P., Kafatos, F. C., Louis, C., ... & Ranson, H. (2005). Gene expression in insecticide resistant and susceptible *Anopheles gambiae* strains constitutively or after insecticide exposure. Insect molecular biology, 14(5), 509-521.

Vontas, J., Small, G., & Hemingway, J. (2001). Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in *Nilaparvata lugens*. Biochem. J, 357, 65-72.
Vulule, J. M., Beach, R. F., Atieli, F. K., McAllister, J. C., Brogdon, W. G., Roberts, J. M., ... &

Hawley, W. A. (1999). Elevated oxidase and esterase levels associated with permethrin tolerance in *Anopheles gambiae* from Kenyan villages using permethrin-impregnated nets. Medical and veterinary entomology, 13(3), 239-244.

Wahlund, S. (1928). Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas, 11(1), 65-106.

Walker , K . (2002) A Review of Control Methods for African Malaria Vectors . Activity Report 108 , Environmental Health Project, Office of Health, Infectious Diseases and Nutrition, Bureau for Global Health, US Agency for International Development , Washington, DC .

Ware, K. E., Marshall, M. E., Heasley, L. R., Marek, L., Hinz, T. K., Hercule, P., ... & Heasley, L. E. (2010). Rapidly acquired resistance to EGFR tyrosine kinase inhibitors in NSCLC cell lines through de-repression of FGFR2 and FGFR3 expression. PLoS One, 5(11), e14117.

Wattier, R., Engel, C. R., SAUMITOU-LAPRADE, P., & Valero, M. (1998). Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). Molecular Ecology, 7(11), 1569-1573.

Wei SH, C. A., Syvanen M (2001). "Identification and cloning of a key insecticide-metabolizing glutathione S-transferase (MdGST-6A) from a hyper insecticide-resistant strain of the house-fly *Musca domestica*." Insect Biochem Mol Biol**31**: 1145-1153.

Weill, M., and O. Duron. (2003). "[Insecticide resistance in the mosquito *Culex pipiens*]." Med Sci (Paris)**19**(12): 1190-1192.

Weill, M., Lutfalla, G., Mogensen, K., Chandre, F., Berthomieu, A., Berticat, C., ... & Raymond, M. (2003). Comparative genomics: Insecticide resistance in mosquito vectors. Nature, 423(6936), 136-137.

Weir, B. S. (1996). <u>Genetic Data Analysis II</u> Sinauer Associates, Sunderland, Mass.

Weir, B. S. and C. C. Cockerham (1984). "Estimating *F*-statistics for the analysis of population structure."

Weir, B. S. and C. C. Cockerham (1984). "Estimating F-Statistics for the Analysis of Population-Structure." <u>Evolution</u>**38**(6): 1358-1370.

White, G. B. (1972). "The *Anopheles gambiae* complex and malaria transmission around Kisumu, Kenya." <u>Trans.R.Soc.Trop.Med.Hyg.</u>**66**(4): 572-581.

White, G. B. and P. Rosen (1973). "Comparative studies on sibling species of the *Anopheles gambiae* complex ecology of species A and B in savanna around Kaduna, Nigeria, during transition from wet to dry season." <u>Bull.Entomol.Res.</u>**62**: 613-625.

WHO (1957). "Malaria Section." <u>Bull.World Health Organ</u>**16** 874-899.

WHO (2007). WHO recommended insecticides for indoor residual spraying against malaria vectors. W. P. E. Scheme.

WHO (2011). World Malaria Report. W. H. Organisation.

WHO (2012). Global Plan for Insecticide Resistance Management in Malaria Vectors. Geneva, World Health Organization.

Wiehe, T. (1998). "The Effect of Selective Sweeps on the Variance of the Allele Distribution of a Linked Multiallele Locus: Hitchhiking of Microsatellites." <u>Theoretical Population Biology</u>**53**(3): 272-283.

Wilkes, T. J., Matola, Y. G., & Charlwood, J. D. (1996). *Anopheles rivulorum*, a vector of human malaria in Africa. Medical and veterinary entomology, 10(1), 108-110.

Williamson, M. S., Martinez-Torres, D., Hick, C. A., & Devonshire, A. L. (1996). Identification of mutations in the houseflypara-type sodium channel gene associated with knockdown resistance (kdr) to pyrethroid insecticides. Molecular and General Genetics MGG, 252(1-2), 51-60.

Wondji, C. S., Coleman, M., Kleinschmidt, I., Mzilahowa, T., Irving, H., Ndula, M., ... & Hemingway, J. (2012). Impact of pyrethroid resistance on operational malaria control in Malawi. Proceedings of the National Academy of Sciences, 109(47), 19063-19070.

Wondji, C. S., Dabire, R. K., Tukur, Z., Irving, H., Djouaka, R., & Morgan, J. C. (2011). Identification and distribution of a GABA receptor mutation conferring dieldrin resistance in the malaria vector< i>*Anopheles funestus*</i> in Africa. Insect biochemistry and molecular biology, 41(7), 484-491.

Wondji, C. S., Hemingway, J., & Ranson, H. (2007). Identification and analysis of single nucleotide polymorphisms (SNPs) in the mosquito *Anopheles funestus*, malaria vector. BMC genomics, 8(1), 5.

Wondji, C. S., Irving, H., Morgan, J., Lobo, N. F., Collins, F. H., Hunt, R. H., ... & Ranson, H. (2009). Two duplicated P450 genes are associated with pyrethroid resistance in *Anopheles funestus,* a major malaria vector. Genome research, 19(3), 452-459.

Wood, O. R., Hanrahan, S., Coetzee, M., Koekemoer, L. L., & Brooke, B. D. (2010). Cuticle thickening associated with pyrethroid resistance in the major malaria vector *Anopheles funestus*. Parasit Vectors, 3(67), 1-7.

Wootton, J. C., Feng, X., Ferdig, M. T., Cooper, R. A., Mu, J., Baruch, D. I., ... & Su, X. Z. (2002). Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. Nature, 418(6895), 320-323.

Wren, J. D., Kulkarni, A., Joslin, J., Butow, R. A., & Harold, R. G. (2002). Cross-hybridization on PCR-spotted microarrays. Engineering in Medicine and Biology Magazine, IEEE, 21(2), 71-75.

Wright, J. M., and Paul Bentzen (1995). "Microsatellites: genetic markers for the future." Molecular genetics in fisheries: 117-121.

Wright, S. (1965). "The interpretation of population structure by F-statistics with special regard to system of mating." Evolution**19**(395-420).

Yan, G., Chadee, D. D., & Severson, D. W. (1998). Evidence for genetic hitchhiking effect associated with insecticide resistance in *Aedes aegypti*. Genetics, 148(2), 793-800.

Yang, Z., Nielsen, R., Goldman, N., & Pedersen, A. M. K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics, 155(1), 431-449.

Young, S. J., Gunning, R. V., & Moores, G. D. (2005). The effect of piperonyl butoxide on pyrethroid-resistance-associated esterases in *Helicoverpa armigera* (Hübner)(Lepidoptera: Noctuidae). Pest management science, 61(4), 397-401.

Yukich, J. O., Lengeler, C., Tediosi, F., Brown, N., Mulligan, J. A., Chavasse, D., ... & Sharp, B. (2008). Costs and consequences of large-scale vector control for malaria. Malaria journal, 7(1), 258.

Zahar, A. R. (1985). Vector biomomics in the epidemiology and control of malaria. Part 1. The W.H.O African region and the Southern W.H.O Eastern Mediterranean region. Section III: vector bionomics, Malaria Epidemiology and control by geographical areas. Geneva, W.H.O. VBC/85.3-MAP/85.3.

Zhang, J., Rosenberg, H. F., & Nei, M. (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proceedings of the National Academy of Sciences, 95(7), 3708-3713.

Zhu, F., Parthasarathy, R., Bai, H., Woithe, K., Kaussmann, M., Nauen, R., ... & Palli, S. R. (2010). A brain-specific cytochrome P450 responsible for the majority of deltamethrin resistance in the QTC279 strain of *Tribolium castaneum*. Proceedings of the National Academy of Sciences, 107(19), 8557-8562.