Investigating the genetic basis of pyrethroid resistance in two members of the *Anopheles gambiae* complex

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

Submitted by

Claudia Anna Elisabeth Witzig

March 2012

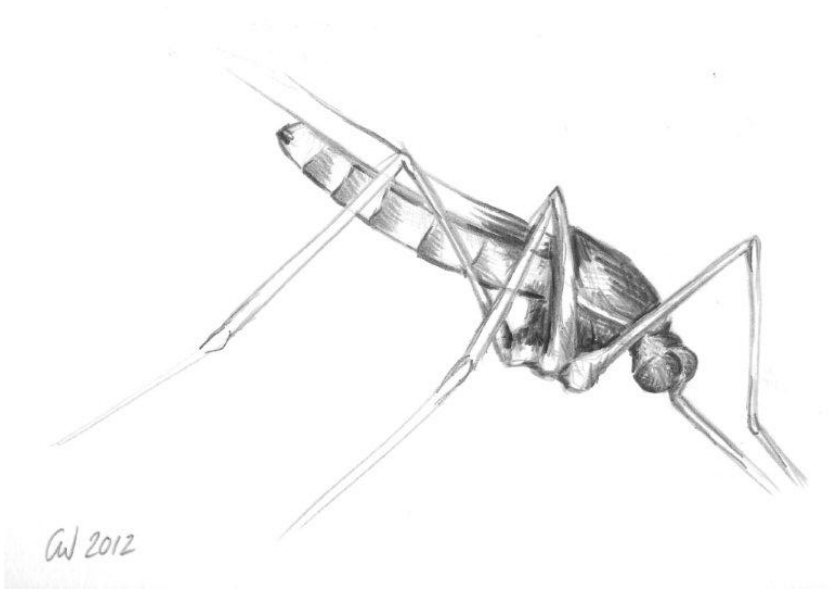
Acknowledgements

I would like to express my sincere gratitude to my supervisors Prof. Hilary Ranson, Dr. Charles Wondji and Dr. Clare Strode for the guidance, support and encouragement they have given me throughout the course of my project. I am especially thankful to Hilary for giving me the opportunity to work in the group and her patience and optimism. Charles' enthusiasm and positive thinking were always very inspiring.

I am grateful to all members of the Vector Group who have helped me over the past few years and have made my time in Liverpool enjoyable. I am particularly grateful to Sara Mitchell and Amy Lynd for helping with lab work and being patient with me whenever I was seeking help and advice. Helen Irving, Keith Steen, Chris Jones and Rodolphe Poupardin have also been of great help. I greatly appreciate the advice and discussions kindly offered by Dave Weetman. Special thanks go to Gwen Finnegan for always finding a way.

I owe a great deal to my parents for their continuous support, encouragement and belief in me.

Finally I acknowledge the moral support and encouragement from all my friends in Liverpool. I owe special thanks to Paul and Michael who have always offered support and retreat in stressful times and kept me sane.



Abstract

Investigating the genetic basis of pyrethroid resistance in two members of the *Anopheles gambiae* complex

Chemical control of mosquito vectors, via indoor residual spraying or insecticide treated bed nets, is an integral component of malaria control strategies. Limited availability of insecticides licensed for public health and the rapid development of resistance in mosquito populations to these insecticides, in particular to some pyrethroids, may compromise vector control efforts.

With the exception of mutations in the insecticide target sites, relatively little is known about the genetics of pyrethroid resistance in malaria vectors. In some populations candidate effector genes, e.g. *cyp6p3* or *cyp6m2* in *An. gambiae s.s.* from Akron, Benin, have been identified as being over expressed in resistant strains but the underlying mechanisms responsible for the increased expression remain unknown. In this study, a combination of quantitative PCR, genetic mapping and microarray tools were used to investigate the mechanisms responsible for pyrethroid resistance in two African major malaria vectors, *Anopheles gambiae* and *An. arabiensis*.

The current work was unable to confirm an association of these known candidates in either a laboratory colony established from Akron or in recent field caught material. Therefore a genetic mapping approach was adopted using field collected mated females to generate F_2 isofemale lines. A major QTL on chromosome 3R was identified which coincides with a genomic region previously implicated in pyrethroid resistance in East African populations. This is the first genetic mapping of insecticide resistance using natural out-bred populations of *Anopheles* and the advantages and limitations of this approach are discussed.

In a second experiment, genetic loci involved in permethrin resistance in *An. arabiensis* were mapped by establishing genetic crosses between a permethrin resistant strain from Chad and a susceptible strain from Mozambique. A single QTL on chromosome 2R was identified in the F_2 progeny that accounts for ~24% of the phenotypic variance. This QTL coincides with a large cluster of detoxification genes. Pyrethroid resistance is not associated with target-site mutations in this population.

Finally, microarrays were used to identify genes differentially expressed between a backcross population, generated by crossing the F_1 population from the resistant Chad strain and the susceptible Mozambique strain of *An. arabiensis* back to the parental resistant strain, with the susceptible strain. A number of candidate genes were identified, including the P450 genes *cyp4h24* and *cyp9j5*, but neither of these were located within the boundaries of the QTL on 2R.

These findings support the presence of metabolic resistance in this population and fine mapping of the identified QTL as well as further investigation of the microarray hits is warranted.

Table of contents

List of Figures

List of Tables

List of abbreviations

AChE	acetylcholinesterase
Ae.	*Aedes*
An.	*Anopheles*
cDNA	complementary DNA
cM	centi Morgan
Cx.	Culex
D.	*Drosophila*
dd	double distilled
DDT	1,1,1-trichloro-2,2-bis (p-chlorophenyl) ethane
DEPC	diethyl pyrocarbonate
EDTA	ethylenediaminetetraacetic acid
gDNA	genomic DNA
GST	glutathione-S-transferase
IRM	Insecticide resistance management
IRS	indoor residual spraying
ITN	Insecticide-treated (bed) net
kdr	knockdown resistance
LLIN	long-lasting insecticidal net
LSTM	Liverpool School of Tropical Medicine
M	molar
M.	*Musca*
$MgCl_2$	magnesium chloride
MIM	Multiple Interval Mapping
MR4	Malaria Research and Reference Resource Center
NaCl	sodium chloride
Na_v	voltage-gated sodium channel
P.	*Plasmodium*
P450	cytochrome P450
PBO	piperonyl butoxide
PCR	polymerase chain reaction
QTL	quantitative trait locus
RBM	Roll Back Malaria
RFLP	restriction fragment length polymorphism
RT	room temperature
SDS	sodium dodecyl sulfate
s.l.	*sensu lato*
s.s.	*sensu stricto*
SNP	single nucleotide polymorphism
TDR	Special Programme for Research and Training in Tropical Diseases
Tris-HCl	Tris(hydroxymethyl)aminomethane hydrochloride
WHO	World Health Organisation
WHOPES	WHO pesticide evaluation scheme

Chapter 1. Introduction

1.1 Malaria

Malaria is caused by the *Plasmodium* parasite, which is transmitted by *Anopheles* mosquitoes. Despite extensive control efforts, over half the world's population remain at risk and the disease has a massive impact on health and economic development, particularly in Africa (WHO 2011c). There were an estimated 216 million malaria episodes in 2010 with over 80% of these occurring in Africa. These cases resulted in an estimated 655,000 malaria deaths worldwide of which 91% were in Africa. Children under five are at greatest risk (WHO 2011c). A more recent analysis however suggests that malaria mortality may be larger than previously estimated, with up to 1,134,000 malaria related deaths in Africa in 2010 (Murray *et al.* 2012).The malaria burden however goes beyond death or acute illness. Malaria infection increases the risk of other illnesses leading to malnutrition and infant mortality, and low birth weight in newborns.

Mosquitoes are important vectors of many diseases including malaria, dengue, and yellow fever. Only mosquitoes of the genus *Anopheles* are capable of transmitting the pathogens that cause human malaria (*Plasmodium falciparum, P. vivax, P. ovale* and *P. malariae*). Furthermore, of the around 430 species of *Anopheles* only about 70 are malaria vectors and of these probably only 40 are of medical importance (Service 2004).

1.2 Malaria vectors

Anopheline malaria vectors are present on nearly every single continent (Kiszewski *et al.* 2004). However, the majority of the worldwide malaria cases occur in Sub-Saharan Africa (WHO 2011c), where the three main African malaria vectors co-exist. Three species of *Anopheles*, *An. gambiae* Giles 1902, *An. arabiensis* Patton 1905 and *An. funestus* are responsible for most of the malaria transmission in this continent. The highly anthropophilic behaviour of these vectors make these species efficient malaria vectors. Other secondary vectors however, play an important role in certain ecological niches (Service 2004).

1.2.1 The *Anopheles gambiae* complex

Many of the anopheline vectors of malaria parasites are members of cryptic species complexes. Species “complexes of *Anopheles* typically include both vector and non-vector species, with two or more member species often being found sympatrically.” (Collins & Paskewitz 1996). The existence of an *An. gambiae* species complex was first noted by Davidson in 1964 (Davidson 1964) and the names *An. gambiae* Giles, 1902 and *An. arabiensis* Patton, 1905 were assigned to the two most abundant of its members (White 1974). *Anopheles gambiae sensu lato* is now known to contain at least seven species, namely *An. gambiae sensu stricto*, *An. arabiensis*, *An. melas*, *An. merus*, *An. bwambae*, *An. quadriannulatus* and *An. quadriannulatus* species B (Coluzzi *et al.* 1979). By definition cryptic species are reproductively isolated from each other, but they are morphologically indistinguishable from each other.

Five of the members of the *An. gambiae* complex (*An. gambiae*, *An. arabiensis*, *An. quadriannulatus*, *An. merus*, and *An. melas*) can be distinguished from each other by a species-specific RFLP (restriction fragment length polymorphism) (Collins *et al.* 1987) or by varying product size following multiplex PCR (Scott *et al.* 1993).

Members of the *An. gambiae* complex are native to tropical Africa. *Anopheles gambiae s.s.* and *An. arabiensis* have the most widespread distribution and live in sympatry over much of Africa (Figure 1.1) with *An. gambiae s.s.* favouring areas with high nocturnal humidity and *An. arabiensis* thriving in relatively arid savannas and steppes (White 1975).

Together with *An. funestus*, they are the major African malaria vectors. They live closely associated with humans (Service 2004) and their larvae are found in a wide range of water sources such as swamps, rice fields, edges of streams or rivers as well as in puddles and water filled hollows such as hoof prints. *Anopheles gambiae* is generally very anthropophagic (prefers to feed on human blood) and endophilic (prefers to rest indoors after taking a blood meal) whereas *An. arabiensis* is partially zoophagic (will feed on animals) and more exophilic (Service 2004). These mosquitoes, however, seem to be rather opportunistic and are able to

adapt their behaviour if environmental factors change, e.g. host availability for blood feeding behaviour, or insecticide sprayed houses for resting behaviour (Collins & Besansky 1994; Killeen *et al.* 2001). *An. quadriannulatus* and *An. quadriannulatus* B (non-malaria vectors) are freshwater breeding and feed predominantly on cattle. *An. bwambae* females are endophagic and partially endophilic, but the species is exclusively known to mineral water springs in Uganda (White 1974). The East and West African salt water breeders *An. merus* and *An. melas* are more minor vectors due to their limited distribution (Coetzee *et al.* 2000; Service 2004). They are predominantly zoophilic but will feed on humans in the absence of animals (White 1974).

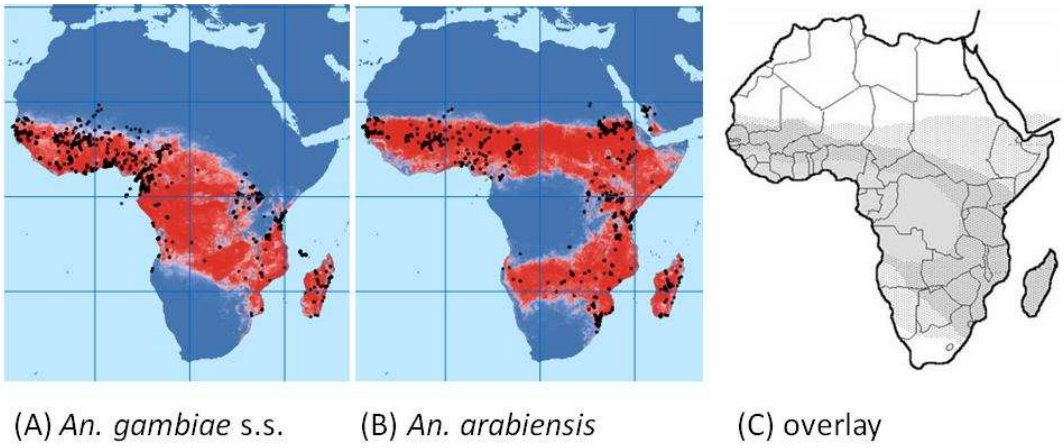


Figure 1.1: Predicted species range of *An. gambiae s.s.* and *An. arabiensis*

(A) *An. gambiae s.s.* and (B) *An. arabiensis* with black dots representing record of occurrence (Sinka *et al.* 2010), and (C) an overlay map of both *An. gambiae s.s.* (grey) and *An. arabiensis* (stippled) showing the extended north and south boundaries of *An. arabiensis* range; adapted from Donnelly *et al.*, 2001.

Originally *An. quadriannulatus* was thought to be the best candidate for the ancestral An. *gambiae* (Coluzzi *et al.* 2002; Coluzzi *et al.* 1979) but opinion was later changed to favour *An. arabiensis* (Ayala & Coluzzi 2005). *An. gambiae* s.s. seems to be an unlikely candidate for the ancestral An *gambiae* as the evolution of this highly anthropophilic species must have been driven by human impact (Coluzzi *et al.* 2002).

The current hypothesised topologies of the *An. gambiae* complex give conflicting relationships between *An. gambiae* s.s. and *An. arabiensis* (White *et al.* 2011a). Figure 1.2 shows the conflict arising from the question if the sister species to *An. gambiae* s.s. is *An. merus* (a) or *An. arabiensis* (b). In addition, contrary to expectations, the two salt water species *An. merus* and *An. melas* and the two antropophilic species *An. gambiae* and *An. arabiensis* are not sister taxa but seem to be the result of convergent evolution (Coluzzi & Sabatini 1969).

The frequency and extent of genetic introgression differs depending on the genomic region under investigation. This genetic introgression between the member species of the *Anopheles gambiae* complex and in particular between *An. gambiae* s.s. and *An. arabiensis* make it difficult to reconstruct the evolutionary relationships (White *et al.* 2011a).

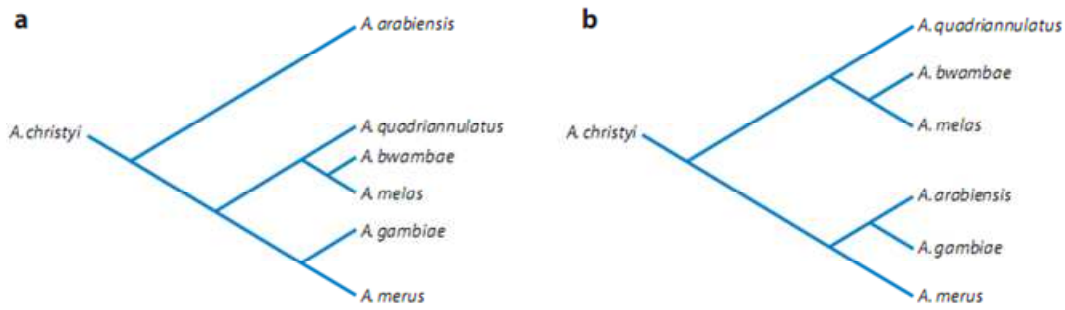


Figure 1.2: Evolutionary relationships within the *Anopheles gambiae* complex showing sister species to *An. gambiae* to be *An. merus* (a) or *An. arabiensis* (b).

Trees are based on (a) X-linked sequences and fixed inversion differences (Coluzzi & Sabatini 1969) or (b) autosomal sequences and mtDNA. *An. christyi* belongs to the same subgenus as the other shown species and is used as an outgroup to tentatively place the root of the trees. (White *et al.* 2011a).

1.2.2 Chromosomal and molecular forms of *An. gambiae s.s.*

Members of the *An. gambiae* complex show ecological and behavioural diversification and this behaviour is associated, in part, with fixed patterns of fixed chromosomal inversions which can be used to facilitate the identification of field specimens (Coluzzi *et al.* 2002; Coluzzi *et al.* 1979). In *An. gambiae s.s.* the greatest density of inversions is found on chromosome arm 2R (Pombi *et al.* 2008) and the inversions most frequently observed are inversions j, b, c, d and u on 2R and inversion 2La (Figure 1.3).

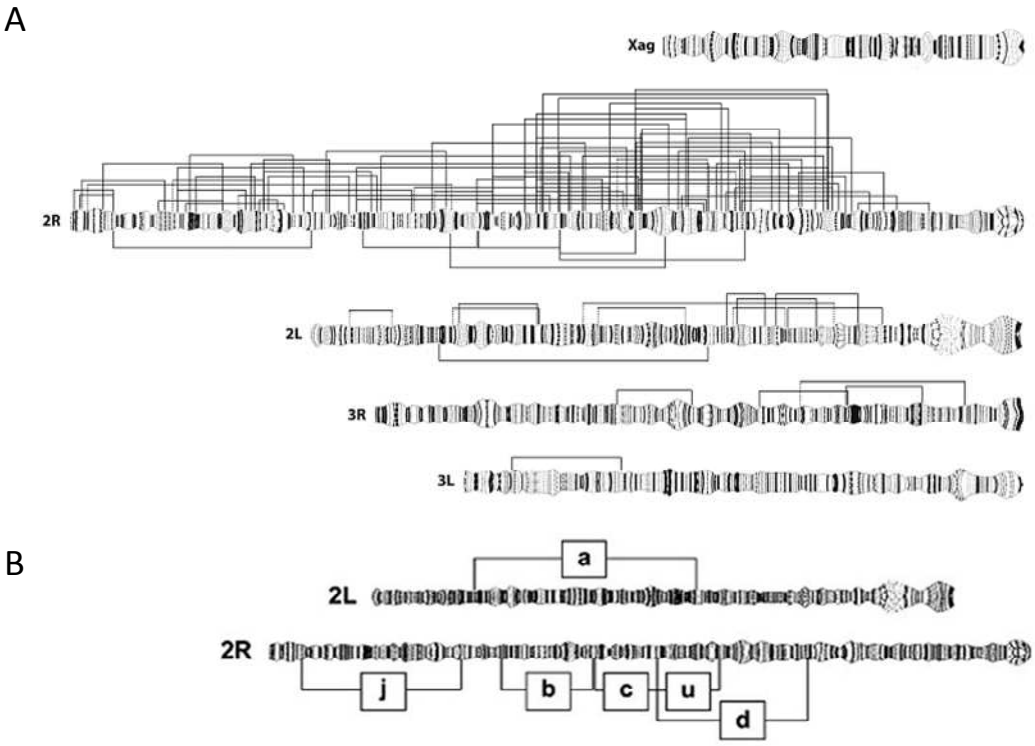


Figure 1.3: Chromosomal Inversions in *An. gambiae s.s.*

Paracentric chromosomal inversions of *An. gambiae s.s.*. (A) Location of 82 rare chromosomal inversions (above) and 7 common chromosomal inversions (below) on the *An. gambiae* polytene chromosome complement (Pombi *et al.* 2008). (B) Close up of chromosome arms 2R and 2L showing the six most common chromosomal inversions in *An. gambiae s.s.* (Torre della *et al.* 2005).

Originally *An. gambiae s.s.* was classified into different chromosomal forms (Bryan J. H. *et al.* 1982; Favia *et al.* 1997; Torre della A. *et al.* 2001) based on specific inversion karyotypes arising from combinations of intra-specific chromosomal inversions. Five partly isolated sub-populations

were defined and named as chromosomal forms Forest, Bissau, Bamako, Savannah and Mopti (Bryan J. H. *et al.* 1982; Coluzzi *et al.* 1985; Torre della A. *et al.* 2001). Various combinations of these inversions are associated with ecotypic differences, and each form is adapted to a different ecology (Coluzzi *et al.* 1979; Touré *et al.* 1994), indicating that selection played a role in those differences (Lehmann & Diabate 2008). Populations from forest areas, for example, are characterized by high frequencies of chromosome 2Rb and 2La standard arrangements whereas in savannah populations, inversion polymorphisms 2Rb/+ and 2La/+ are more common and correlate with the degree of aridity (Coluzzi *et al.* 1985). The association of 2La/+ to increased tolerance to thermal stress in larvae is supported by a recent microarray study (Cassone *et al.* 2011). Polymorphic chromosomal inversions restrict recombination locally by suppressing the pairing of homologous chromatids. Hence inversions are thought to be hotspots for differentiation as recombination suppression in these regions facilitates functional divergence of the inverted and the wild type arrangements (Neafsey *et al.* 2010).

In addition, *An. gambiae s.s.* exists as two sympatric sub-taxa known as molecular forms, namely M and S (Gentile *et al.* 2001; Torre della A. *et al.* 2001). Originally M and S-form were thought to covary with certain chromosomal forms (Torre della *et al.* 2005; Torre della A. *et al.* 2001). Most recent data however suggests that the different 2R karyotypes are not directly responsible for the evolution of M and S-forms and probably pre-date the division of M and S molecular form (Costantini *et al.* 2009; Simard *et al.* 2009; White *et al.* 2009).

M and S-forms have diverged in larval ecology and reproductive behaviour (Lehmann & Diabate 2008) and results of various studies suggest an incipient speciation of *An. gambiae* M and S molecular forms (Lawniczak *et al.* 2010; Torre della *et al.* 2005). The two forms exist largely in sympatry throughout West Africa (geographically and microspacially) and gene flow between forms does continue as the build-up of barriers of gene flow is not yet completed (Torre della *et al.* 2005). M and S-form F_1 hybrids are fully viable and fertile, not showing reduced fitness, at least under

laboratory conditions (Diabaté *et al.* 2007). Considerable levels of hybridization in the wild are observed but to varying degrees (Caputo *et al.* 2008; Tripet *et al.* 2001), and it is suggested that reproductive isolation between M and S-form varies across Africa, being higher in the west of the continent (Marsden *et al.* 2011; Weetman *et al.* 2012).

One important example of gene flow between those molecular forms is the introgression of the knock down resistance (*kdr*) allele (Milani 1954). This is a single-nucleotide mutation in the gene for the Sodium channel which causes an amino acid change in the protein. This mutation termed *kdr* confers resistance to DDT and pyrethroid insecticides and is therefore extensively studied in *An. gambiae* populations across Africa. The mutation probably originated in the S-form first (Fanello *et al.* 2003) and was later acquired by the M-molecular form through introgression (Weill M *et al.* 2000) rather than through an independent *de novo* mutation.

It was long thought that reproductive isolation was due to changes at a small number of loci called speciation islands (near centromeres) (Turner *et al.* 2005) and that hybridization lead to M-S genome homogenization in all except those few small regions. More recently however, results from a SNP genotyping (Neafsey *et al.* 2010) and a whole genome sequencing study (Lawniczak *et al.* 2010) support a different model. This suggests that gene flow between forms is much lower and the process of speciation more advanced than thought although the use of samples from lab colonies in the study of Lawniczak *et al.* may have introduced bias (Weetman *et al.* 2012). The process by which species or genomes evolve and the key genes driving differentiation in the case of *An. gambiae s.s.* remain largely unknown.

Morphologically M and S-forms are indistinguishable at all life stages (Torre della A. *et al.* 2001) but can genetically be distinguished by specific single nucleotide polymorphisms (SNPs) in the spacer regions of ribosomal DNA (rDNA) (Fanello *et al.* 2002) or by the amplification of a SINE (short interspersed element) retrotransposon on the X chromosome (*SINE200*) (Santolamazza *et al.* 2008b).

1.3 Malaria and vector control

The Global Malaria Action Plan promoted by the international Roll Back Malaria (RBM) initiative outlines a global strategy to control malaria and eventually eliminate malaria country by country. Recommended control strategies are rapid diagnosis and prompt treatment as well as effective malaria prevention.

The introduction of rapid diagnostic tests have improved detection of malaria, and improved treatment regimes, involving artemisinin-based combination therapy, are being scaled up across Africa (WHO 2008). However, drug resistance is an increasing threat (Kappe *et al.* 2010). Extensive efforts have been directed at developing a vaccine against malaria and the most advanced vaccine candidate is RTS,S/AS01, produced by GlaxoSmithKline, which is currently undergoing phase 3 studies in seven African countries (The RTS 2011). Initial results are promising with >59% vaccine efficacy against all malaria episodes in infants, but trials are ongoing (Asante *et al.* 2011).

Malaria prevention relies extensively on vector control and this can take different forms. Local success has been achieved by biological control, for example using larvivorous fish, and environmental control via draining and reducing breeding sources. Genetic methods are also being explored. However, currently chemical control using insecticides is the most widely applied method.

1.3.1 Use of insecticides in vector control

In most malaria control programs, vector control is directed at adult mosquitoes. The main methods applied are indoor residual spraying (IRS) and the use of insecticide-treated bed nets (ITN). These reduce mosquito density but more importantly, they directly reduce disease transmission by shortening the mosquito's life span thereby reducing its vectorial capacity. Larval control measures, in contrast reduce disease transmission indirectly by reducing overall vector density (Walker & Lynch 2007). In an environment where mosquito breeding sites are permanent and few as well as easy to treat, larvicides to control vector density directly may be

used to complement IRS and ITNs. Another method is space spraying which is mainly used to control urban epidemics of mosquito transmitted disease and is of minor importance for malaria control (WHO 2011a).

1.3.1.1 Indoor residual spraying

Indoor residual spraying (IRS) is “the application of long-acting chemical insecticides indoors, in order to kill the adult vector mosquitoes that land and rest on these surfaces”. (WHO 2006). After having taken a blood-meal the main malaria vectors in Africa frequently rest indoors. If walls are treated with insecticide, the mosquito may be exposed to a toxic dose and may die. IRS aims to reduce disease transmission and to reduce the density of vector populations. Some insecticides may also have an additional effect in repelling mosquitoes from entering a treated room (WHO 2006). Currently twelve insecticides belonging to four chemical groups are recommended for IRS by WHO (WHO 2011c).

Effective implementation and good planning of spray operations are crucial for the success of IRS. This involves monitoring of vector susceptibility, choice of insecticide and timing of spraying prior to implementation and effective monitoring and evaluation post spraying.

IRS is also one of the main foci of the US President’s Malaria Initiative launched in 2005, which provides international funding and coordination for the control of malaria (USAID-CDC Interagency Working 2005). According to WHO, in Africa IRS now protects 11% of the people at risk (WHO 2011c) (Figure 1.4).

1.3.1.2 Insecticide-treated bed nets

Insecticide-treated bed nets (ITN) or now more commonly long-lasting insecticidal nets (LLINs) are the mainstay of malaria vector control in Africa. If used and maintained correctly, bed nets protect by acting as a physical barrier between humans and mosquitoes to reduce the man-vector contact.

Incorporating an insecticide into the net extends its protective effect to others. Modest coverage of all adults and children with ITNs can have a good community-wide protective effect (Killeen *et al.* 2007). Various

studies have shown the beneficial effects of using ITNs, by reducing transmission rates, the incidence of malaria and decreasing mortality (Lengeler 2004). The percentage of African households owning at least one ITN has risen from 3% in 2000 to 50% in 2011and ITN coverage in sub-Saharan Africa is higher than IRS coverage, especially in West Africa (WHO 2011c) (Figure 1.4).

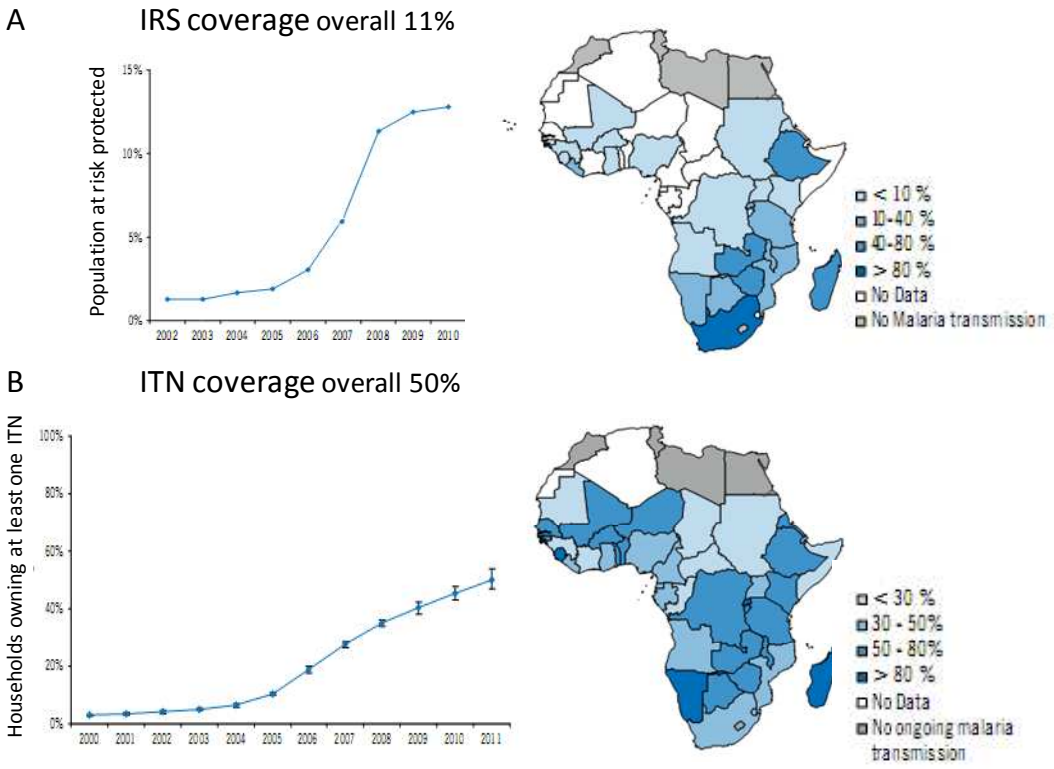


Figure 1.4: Scale of IRS and ITN coverage in Africa

(A) Trend in estimated population at risk of malaria protected by IRS across Sub-Saharan Africa, 2002-2010 and estimated proportion of population at risk of malaria protected by IRS in the different countries of Sub-Saharan Africa. (B) Trend in estimated proportion of households with at least one ITN in Sub-Saharan Africa, 2000-2011; and estimated proportion of households with at least one ITN in the different countries of Sub-Saharan Africa, June 2011. Graphs (WHO 2011c), source (Flaxman *et al.* 2010). Note: colour coding between maps of ITN coverage and IRS coverage does vary slightly.

1.3.2 Chemical classes of insecticides and modes of action

Most insecticides act on the nervous system, where they interfere with the transmission of nervous impulses. Very few insecticide classes are available for use in public health.

1.3.2.1 Organochlorides

The first insecticide used on a large scale in public health was the organochloride DDT (1,1,1-trichloro-2,2-bis (p-chlorophenyl) ethane, Figure 1.5) (Zeidler 1874). Easy and cheap to produce, with a long lasting toxicity, DDT revolutionized malaria control, when in 1957 the World Health Organisation (WHO) embarked on a global initiative to eradicate malaria vectors, relying on indoor residual spraying with DDT. DDT is the only organochloride insecticide used in vector control today (WHO 2011a).

DDT is a neurotoxin, targeting the peripheral nervous system of arthropods. It acts by paralysing the sodium channels impeding channel closing, causing the neurons to fire spontaneously leading to spasm and eventually death.

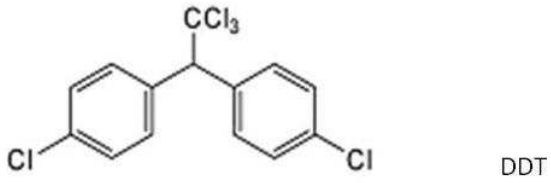


Figure 1.5: Chemical structure of the organochloride DDT
1,1,1-trichloro-2,2-bis (p-chlorophenyl) ethane (Zeidler 1874) adapted from (Davies *et al.* 2007b).

1.3.2.2 Organophosphates and carbamates

Organophosphates are esters of phosphoric acid, and are structurally characterised by a reactive phosphate bonded to a terminal oxygen, two lipophilic groups (generally methyl or ethyl) and a leaving group (Fukuto 1990). Carbamates are esters of carbamid acid, NH_2COOH, an unstable compound. Most carbamate insecticides are methylcarbamates, derivates

of methylcarbamic acid, $NHCH_3COOH$, containing a phenol ring (Fukuto 1990).

Organophosphates and carbamates inhibit acetylcholinesterase (AChE) by binding to it either irreversibly (organophosphates) or reversibly (carbamates). Inhibition of acetylcholinesterase results in the build up of acetylcholine at neuromuscular junctions which causes repetitive firing of neurones, leading to paralysis and eventual death (Fukuto 1990).

Organophosphorus and carbamate insecticides exist as a variety of chemical structures with different physical and chemical properties. Examples of members of these classes recommended for use in public health include the organophosphates malathion, fenitrothion, pirimiphos-methyl, and the larvicide temephos and the carbamates bendiocarb and propoxur (http://www.who.int/whopes/Insecticides_IRS_Malaria_ok.pdf) (Figure 1.6).

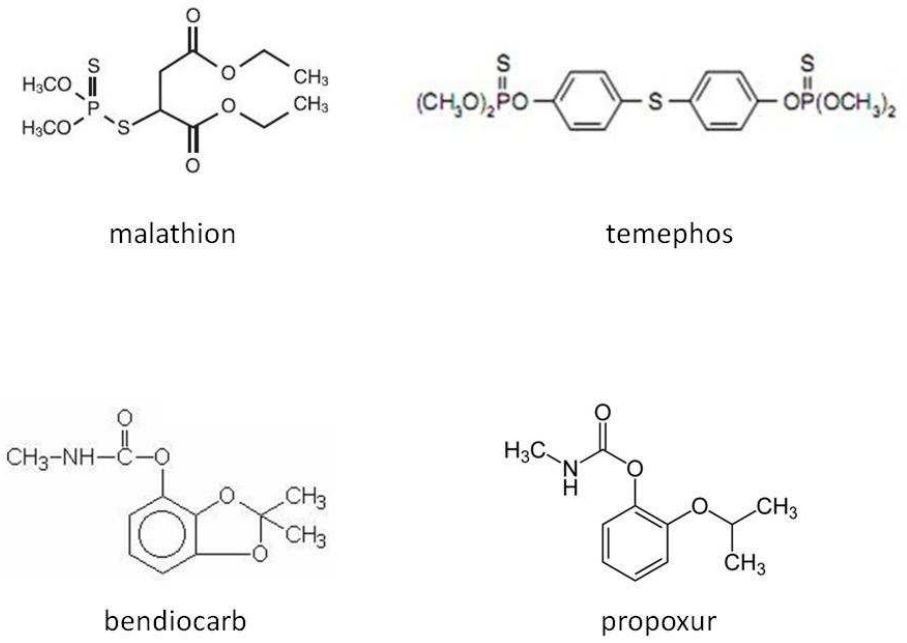


Figure 1.6: Chemical structures of examples of organophosphates and carmabates. The organophosphate insecticides, malathion and temephos, showing one or two reactive phosphorus-ester groups, and the carbamates, bendiocarb and propoxur, showing the methylcarbamic group, NH_2COOH (Fukuto 1990).

1.3.2.3 Pyrethroids

Synthetic pyrethroids are the major class of insecticides used in public health. Pyrethroids are the only insecticide class recommended for impregnation of mosquito nets, but are also used for IRS, space spraying, and larviciding.

Pyrethroids are analogues of the naturally occurring insecticidal esters of chysanthemic acid (pyrethrins I) and pyrethric acid (pyrethrins II) (Davies *et al.* 2007b). These natural compounds were not cost effective in production and lacked photostability. Structural modifications soon led to photostable chemicals with high insecticidal, yet low mammalian toxicity.

Pyrethroids cause effects similar to those of DDT: overstimulation of nerve cells by arresting the sodium channels in an open state that will eventually cause paralysis and death. However, with pyrethroids, the nerve cells will continue to function in a state of abnormal hyperexcitability, and the insect suffers an effect known as “knock-down”. This is a sub lethal effect, from which the insect can recover. Type II pyrethroids which contain a cyano group generally kill more efficiently (Bloomquist 1996). This is because “the duration of modified sodium currents by Type I compounds lasts only tens or hundreds of milliseconds, whilst those of Type II compounds last for several seconds or longer” (Davies *et al.* 2007b).

Permethrin, type I, and deltamethrin, type II, (Figure 1.7) are some of the most commonly used pyrethroids. Other pyrethroids used for public health purposes are alpha-cypermethrin, bifenthrin, cyfluthrin, etofenprox, and lambda-cyhalothrin.

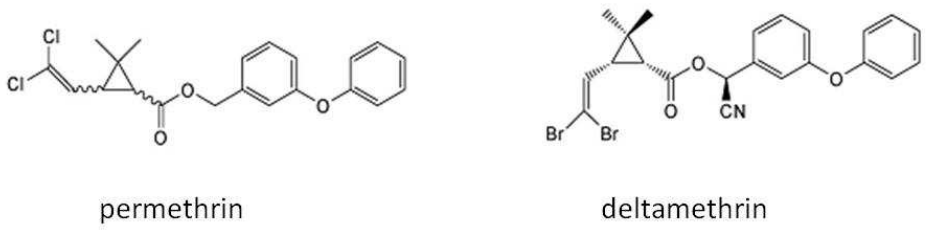


Figure 1.7: Chemical structures of pyrethroids

(left), permethrin a Type I pyrethroid and (right), deltamethrin a type II pyrethroid bearing the cyano group, CN (Davies *et al.* 2007b).

1.3.3 The use of insecticides for malaria control in Africa

Insecticides from the above four chemical classes are used in three main applications for malaria control: indoor residual spraying, treatment of nets and larviciding.

A recent review of insecticide use in public health provided information on the use of each insecticide class for malaria control in Africa, based on a 10-year average (2000-2009) (WHO 2011a):

(1) Indoor residual spraying

- DDT accounts for over 90% (by tons) of the insecticide used even though all four chemical classes can be applied
- When measured in spray coverage, 35% of IRS is with DDT and 62% with pyrethroids (pyrethroids are generally effective at an application rate 60 times lower than DDT).

(2) Bed net impregnation

- accounts for 45% of overall insecticide used for vector control (in spray coverage, i.e. the surface treated with a specific insecticide)
- pyrethroids are the only insecticide class used for net treatment and make up 100% of insecticides used in this area.

(3) Larviciding

- Insecticides used for space spraying or larviciding account for a very small fraction of less than 1%

The pie chart (A) shows the percentage in spray coverage of insecticides used for the different control methods in Africa. The pie chart (B) is based on spray coverage by insecticide class and shows the importance of the pyrethroid classes (WHO 2011a) (Figure 1.8).

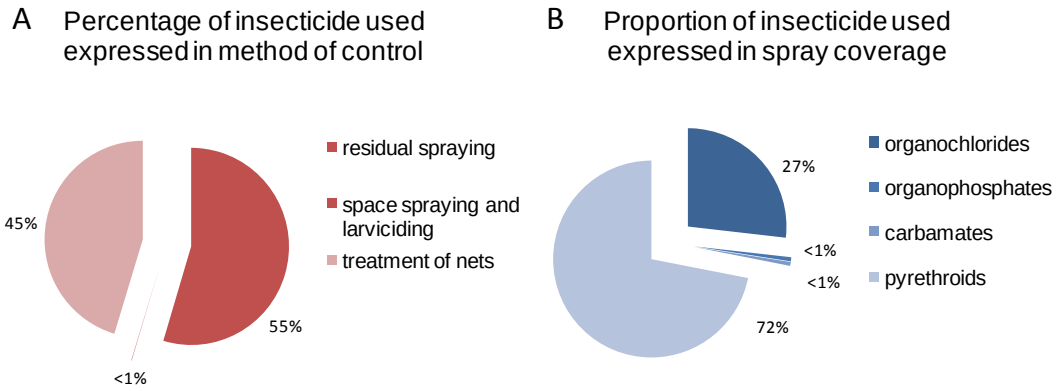


Figure 1.8: The use of insecticides in Africa based on spray coverage

The charts show (A) percentages of insecticides expressed in method of control and (B) percentage of insecticide used overall; data 10-year average 2000-2009 (WHO 2011a).

Figure 1.9 shows a trend of DDT and pyrethroid usage in vector control in Africa over the last 10 years (2000-2009). Use of pyrethroids has escalated in recent years. Organophosphate use has decreased over this time period and carbamate use has remained relatively stable (WHO 2011a).

Given that IRS and ITNs are the mainstay of malaria control programmes, insecticide resistance and pyrethroid resistance particularly to the pyrethroid class, clearly poses a major threat to malaria control.

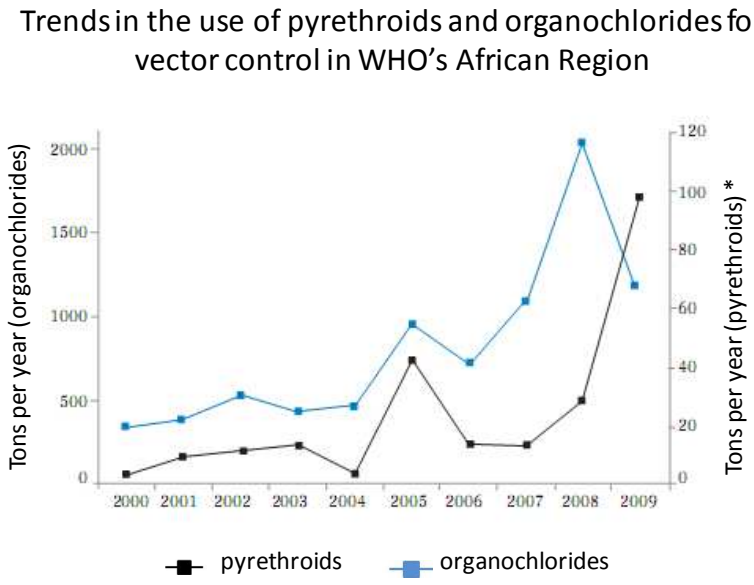


Figure 1.9: Trends for the use of organochlorines and pyrethroids

for vector control in the WHO African region shown in tons per year. Note that pyrethroids achieve approximately 60-times higher spray coverage * (WHO 2011a).

1.4 Insecticide resistance

Insecticide resistance is defined by WHO as an inherited trait that allows the insect to survive a dose of a pesticide, which would normally be lethal. With maintained selection pressure, the eventual result is reduced effect of the respective chemical to control the target-organism or even complete failure to do so.

Resistance has occurred against the four main classes of insecticides sanctioned for public health (Hemingway & Ranson 2000) and is of growing concern. Currently there is a lack of alternative, cost-effective and safe insecticides to use against resistant populations.

1.4.1 Emergence of resistance

Mosquitoes reproduce fast and produce a large number of offspring, and they can adapt fast to environmental changes. As a consequence of this and the widespread use of insecticides in agriculture and public health, resistance has arisen relatively rapidly in malaria vectors.

Insecticide resistant phenotypes are favoured where mosquitoes are exposed to sub-lethal doses of the pesticide. Under these conditions resistant individuals have a better chance to survive until reproduction. If insecticides target young mosquitoes, susceptible individuals are more likely to be killed before they reproduce; this means selection pressure towards resistant populations. Such conditions can result from vector control through insecticide decay (on treated walls or nets) or bad spraying technique.

The use of insecticides for control of agricultural pests has been linked with the development of insecticide resistance in mosquito vectors (Brooke *et al.* 2001; Chandre *et al.* 1999; Diabate *et al.* 2002; Yadouleton *et al.* 2009). Other chemicals and factors aside from insecticides may create a selective environment, which gives rise to resistant populations (Poupardin *et al.* 2008). For example urban pollution with xenobiotics from human waste or oil contamination has been suggested to be the major selective force for pyrethroid resistance in Southern Benin (Djouaka *et al.* 2008).

Resistance to pyrethroids was first reported in *An. gambiae s.s.* in 1993 in Cote d'Ivoire (Elissa *et al.* 1993). Initially resistance appeared largely restricted to West Africa but in recent years, pyrethroid resistant *Anopheles* populations have been reported throughout Africa (Balkew *et al.* 2010; Chandre *et al.* 1999; Diabate *et al.* 2002; Fanello *et al.* 2003; Kerah-Hinzoumbe *et al.* 2008) (Figure 1.10). As DDT and pyrethroids have the same target site, cross-resistance between these classes frequently occurs (Dabiré *et al.* 2009; Ranson *et al.* 2011).

Resistance to carbamate (Djogbénou *et al.* 2009; N'Guessan *et al.* 2003; Ranson *et al.* 2009) and organophosphate (Ranson *et al.* 2009) insecticides has been reported in *An. gambiae s.l.* and, although less prevalent than resistance to DDT and pyrethroids, resistance is increasing, particularly in West Africa (Chouaïbou *et al.* 2008; Corbel *et al.* 2007).

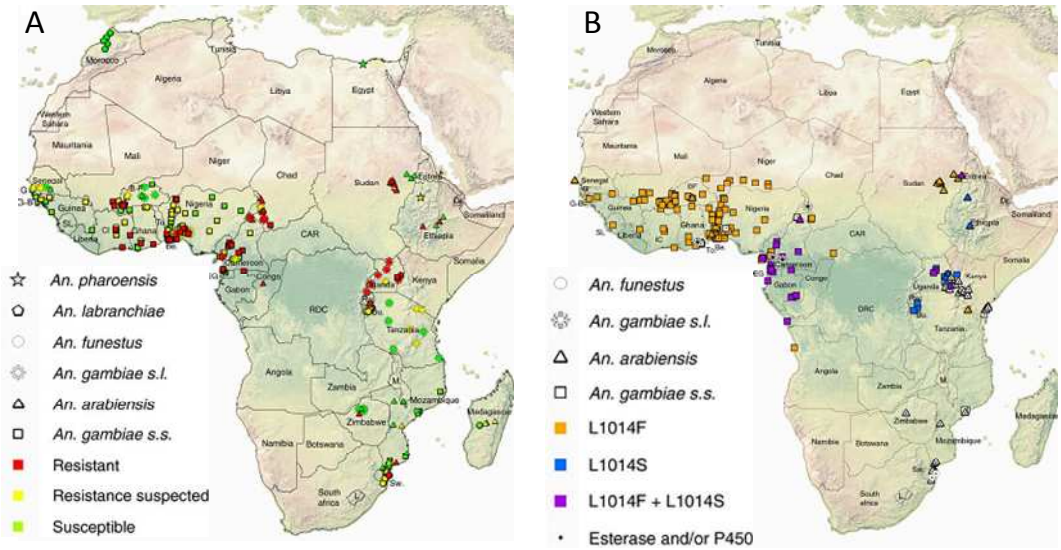


Figure 1.10: Pyrethroid resistance in malaria vectors across Africa

Map of Africa showing (A) the distribution of pyrethroid resistance in the main malaria vectors and (B) their underlying mechanisms (Ranson *et al.* 2011).

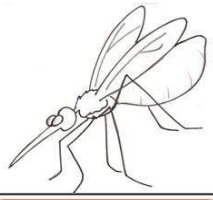
Specifically targeting mosquitoes responsible for disease transmission, older mosquitoes, could considerably reduce this selection pressure for insecticide resistance. In theory this may be an easier task as the older, potentially infected mosquitoes, are generally more susceptible to insecticides (Jones *et al.* 2012a; Rajatileka *et al.* 2010). This led to the

hypothesis (Glunt *et al.* 2011) that in areas, where a high spray coverage can be realized and maintained, application of low insecticide concentrations (lower than currently suggested by WHO) would achieve reduction of malaria transmission in the absence of the strong selection pressure of current practice as both, resistant and susceptible mosquitoes, would have the same chance to reproduce.

1.4.2 Resistance mechanisms

Resistance mechanisms are generally classed as target-site resistance, metabolic resistance, cuticular, or behavioural. Studies usually focus on target-site and metabolic resistance (Figure 1.11) and to a lesser extent on cuticular and behavioural resistance.

The two major resistance mechanisms and their relative role in conferring resistance to each insecticide class are shown in Figure 1.11. Cross resistance may occur due to *kdr* conferring resistance to both pyrethroids and DDT, and *ace-1* based resistance to carbamates and organophosphates. Less is understood about metabolic cross resistance patterns (Mitchell *et al.* 2012).



	Mechanisms of resistance				
	Metabolic resistance			Target-site resistance	
	Esterases	Monoxygenases (P450s)	Glutathion S-transferases	*Kdr*	*ace-1*
Pyrethroids	+	++	++	++	
DDT		+	+	++	
Carbamates	+	+			++
Organophosphates	++	+	+		++

Figure 1.11: Major mechanisms conferring resistance to important classes of insecticides in mosquitoes.

++ indicates an important resistance mechanism; + means this mechanism has been described but is considered to be of lesser importance (adapted from (Nauen 2007)).

1.4.2.1 Target-site resistance

Target-site resistance results from amino acid substitutions in the insecticide target-site, which reduce the binding of the insecticide but maintain the function of the protein. The four classes of insecticide primarily target three proteins essential for the functioning of the mosquito's nervous system (Figure 1.12): acetylcholinesterase, the voltage-gated sodium channel and to a lesser extend (not shown) γ-aminobutyric acid receptors, ligand-gated chloride channels.

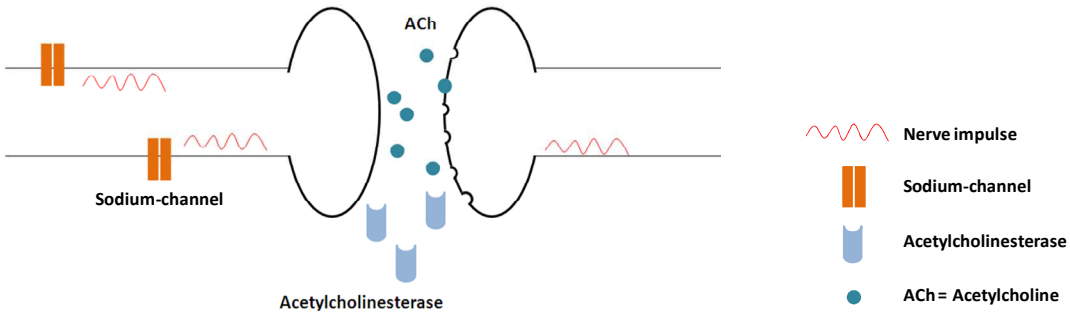


Figure 1.12: Illustration of the nerve synapse including the main target sites for DDT and pyrethroid insecticides (voltage-gated sodium channel) and carbamate insecticides (Acetylcholinesterase) are depicted.

Acetylcholinesterase

Acetylcholinesterase (AChE) is the target for organophosphate and carbamate insecticides. Amino acid substitutions in AChE can result in an enzyme with decreased sensitivity or complete insensitivity to these insecticides (Russell *et al.* 2004).

The *ace* locus (*ace-2*) coding for AChE has been mapped in *Drosophila melanogaster* (Hall & Spierer 1986) and several point mutations have been identified each conferring low levels of resistance (Mutero *et al.* 1994; Pralavorio & Fournier 1992).

However, the *ace-2* locus does not code for the insensitive AChE identified in resistant mosquitoes, and a second locus, *ace-1*, encoding the insecticide target AChE was identified in *An. gambiae* and *Culex pipiens* (Weill 2002). A single glycine to serine substitution at position 119 (G119S), resulting from a point mutation GGC to AGC in the gene *ace-1*

(Weill *et al.* 2003) is responsible for highly insensitive acetylcholinesterase in resistant strains of these mosquito species (Djogbénou *et al.* 2011). The resistance allele for *ace-1* is thought to have a high fitness cost in the absence of insecticide.

Voltage-gated sodium channel

The gene for the voltage-gated sodium channel (Na_v) was initially cloned and sequenced in *Drosophila* (Loughney *et al.* 1989). The pore-forming α-subunit consists of four internally homologous hydrophobic domains (I-IV) (Figure 1.13). Each domain consists of six transmembrane helices, five hydrophobic (S1, S2, S3, S5, and S6) and one positively charged segments (S4). Altogether these form a central ion-conducting pore controlled by a voltage sensing part.

Organophosphates, such as DDT, and pyrethroid insecticides act on the Na_v and prevent it from functioning normally. Some insect pest species have evolved modifications of the sodium channel protein resulting in a resistant phenotype. This resistance phenotype was termed knock-down resistance (Milani 1954).

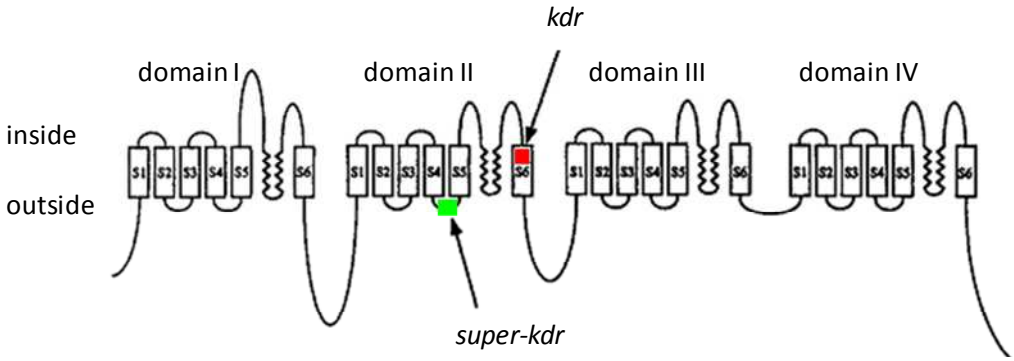


Figure 1.13: The transmembrane topology of the voltage-gated sodium channel.
The pore-forming α-subunit consists of a single polypeptide chain with four internally homologous domains (I – IV), each having six transmembrane helices (S1 – S6) adapted from (Davies *et al.* 2007a). The red and green boxes indicate the position of the amino-acid changes termed *kdr* and *super-kdr*.

Kdr is not unique to mosquitoes and was first recognized in the housefly *Musca domestica* in the 1950s (Milani 1954). The genetic linkage of *kdr* to

the insect sodium channel genes was shown in the housefly in the early 1990s (Knipple *et al.* 1994; Williamson *et al.* 1993). The *kdr* mutation in houseflies results from an single point mutation (TTA to TTT) causing an amino acid change from leucine to phenylalanine at amino acid residue 1014 (L1014F) in the S6 segment of the sodium channel protein (Miyazaki *et al.* 1996; Williamson *et al.* 1996). The phenotype known as *super-kdr* confers much higher resistance (Sawicki 1978) to pyrethroids. It is a combination of the original L1014F mutation combined with a methionine to threonine alteration at position 918 (M918T) in the S4-S5 loop which has been found in the housefly (Williamson *et al.* 1996).

Many other mutations in the sodium channel have been reported including an alternative substitution at 1014 resulting in an amino acid change from leucine to serine (L1014S) (Martinez-Torres *et al.* 1999a; Martinez-Torres *et al.* 1999b; Ranson *et al.* 2000a). O'Reilly and colleagues (O'Reilly *et al.* 2006) give a summary of the different mutations and locations identified in the insect sodium channel that have been associated with insecticide resistance. Modelling of the sodium channel (O'Reilly *et al.* 2006) enables predictions of residues in the insecticide binding pocket and facilitates prediction of mutations likely to be associated with resistance.

In *An. gambiae s.l.*, the two mutations 1014F and 1014S have been found. The 1014S mutation was first found in East Africa (Ranson *et al.* 2000a), whereas the 1014F mutation was first found in West Africa (Martinez-Torres D *et al.* 1998). Both alleles now have a wide distribution across Africa. Both the L1014F and L1014S *kdr* mutations are present in M and S-forms of *An. gambiae s.s.* and in *An. arabiensis* throughout their range (Donnelly *et al.* 2009; Ranson *et al.* 2011) and the frequency of 1014F is approaching fixation in the S-form of *An. gambiae* in many parts of West Africa.

Recently a new mutation N1575Y has been identified in *An. gambiae s.s.* which occurs on only a single long-range haplotype also bearing the *kdr* mutation 1014F. The 1014F-1575Y haplotype was found in both M and S-molecular forms of *An. gambiae* in West/Central Africa (Jones *et al.* 2012b).

Kdr-alleles can be detected in both homozygous and heterozygous states using a range of molecular techniques including allele specific PCR (Martinez-Torres D *et al.* 1998; Ranson *et al.* 2000a), Hot Ligation Oligonucleotide Assay (HOLA) (Lynd *et al.* 2005), TaqMan real-time assay (Bass *et al.* 2007) or SNP detection using pyrosequencing.

In *An. gambiae*, where *kdr* resistance is well investigated, the association between the *kdr* mutations and the pyrethroid and/ or DDT-resistance phenotype has been shown in several studies using QTL mapping (Ranson *et al.* 2004) and genotype/phenotype association approaches (Chandre *et al.* 2000; N'Guessan *et al.* 2007). There is a clear link between *kdr* genotype and resistance phenotype, although this association in *An. gambiae* seems to be strongest for *kdr* and DDT, less strong for *kdr* and permethrin and even less strong for *kdr* and deltamethrin (Brooke & Koekemoer 2010). The *kdr* genotype may only explain part of the variance in a resistance phenotype (Brooke & Koekemoer 2010; Donnelly *et al.* 2009; Saavedra-Rodriguez *et al.* 2008) and the full contribution of *kdr* to the efficacy of pyrethroids against *An. gambiae* is not clear. It is possibly for this reason that studies that try to link *kdr* frequency to effectiveness of control intervention yield contradictory results as they omit the factor metabolic resistance (next section).

γ-aminobutyric acid receptors (GABA)

The receptors for the neurotransmitter γ-aminobutyric acid (GABA) are implicated as site of action for chlorinated hydrocarbons other than DDT, such as dieldrin (Nauen 2007). In *An. gambiae s.s.* an alanine to glycine mutation in this receptor confers resistance to dieldrin, *rdl*, whereas an alanine to serine mutation at the same position has been associated with resistance to dieldrin in *An. arabiensis* (Du *et al.* 2005). Dieldrin is no longer used in public health although other insecticides, such as fipronil, which have the same target site, are important in agriculture (Caboni *et al.* 2003).

1.4.2.2 Metabolic resistance

Metabolic detoxification of insecticides is a common resistance mechanism (Hemingway & Ranson 2000). Metabolic resistance can be inferred from studies with synergists or using biochemical assays which measure the levels of enzyme families associated with insecticide metabolism. This resistance mechanism is prevalent in malaria vectors (Awolola *et al.* 2009; Djouaka *et al.* 2008; Müller *et al.* 2008b; Verhaeghen *et al.* 2008; Wondji *et al.* 2007).

Metabolic resistance is based on an increased rate of insecticide metabolism due to enhanced activity of detoxifying enzymes which can degrade insecticides before they reach their targets. This can be either due to enhanced expression levels of the respective enzyme (enhanced expression or gene duplication), or mutations in the enzymes. Three large multigene enzyme families are responsible for insecticide metabolism: monooxygenases (cytochrome P450s), carboxylesterases (COEs) and glutathione S-transferases (GSTs) (Ranson *et al.* 2002a).

One noticeable feature that these three gene families have in common in insects is the high number of recent duplication events, which tends to result in closely related genes being clustered in the genome (Figure 1.14) (Ranson *et al.* 2002a).

Species	Gene class	Chromosome arm					
		X	2L	2R	3L	3R	Total
A. gambiae	GSTs	5	1	14	1	10	31
	COEs	1	23	11	15	1	51
	P450s	10	11	47	11	32	111

Figure 1.14: Cytological locations of *An. gambiae* GSTEs, COEs and P450s.
(from (Ranson *et al.* 2002a)).

Monooxygenase based resistance

The P450 enzymes (also known as mixed function oxidases and cytochrome P450 monooxygenases) are a diverse family, one of the largest and oldest, of haem containing enzymes found from bacteria to mammals. They fulfil a range of important tasks, from synthesis to degradation. In insects they are associated with detoxification of plant defence chemicals against herbivores (Gould 1984) and certain subfamilies are also involved in insecticide resistance (Ronis & Hodgson 1989).

P450s are both diverse in function, due to a wide range in structure, and large in number. This diversity arose from extreme gene duplications that formed large gene clusters of up to 15 P450 genes per cluster (Ranson *et al.* 2002a; Werck-Reichhart & Feyereisen 2000) with *CYP4* and *CYP6* being the largest gene families in *Drosophila* (Tijet *et al.* 2001) and *An. gambiae* (Ranson *et al.* 2002a). Many insect species carry about a hundred different P450 genes, e.g. there are 111 in *An. gambiae* (Ranson *et al.* 2002a), and all evolved from one ancestral gene (Feyereisen 1999).

With such a number and diversity, some P450s may adopt a new function linked to insecticide detoxification whilst the others are able to maintain the normal metabolic functions. Possibly small conformational changes of cytochrome P450s may increase the affinity or the metabolic turnover of insecticides, leading to resistance, although resistance via elevated transcript levels is more common (Li *et al.* 2007; Scott 1999). In *An. gambiae*, the two P450 genes most strongly implicated in pyrethroid resistance, *cyp6p3* and *cyp6m2*, were identical or near identical in sequence to those in susceptible strains (Müller *et al.* 2008b; Stevenson *et al.* 2011) but expression is elevated in resistant strains. Gene duplication of P450s has been implicated in resistance in *An. funestus* (Wondji *et al.* 2009), *Ae. aegypti* (Bariami *et al.* 2012) and the crop pest *Myzus persicae* (Puinean *et al.* 2010).

The *CYP6* family is exclusive to insects (Tijet *et al.* 2001) and has been linked to insecticide resistance in *Drosophila* (Cohen & Feyereisen 1995)

and various other insects including *An. gambiae*. In a Kenyan lab strain of *An. gambiae s.s.* resistance to pyrethroids has been mapped to a genetic region that contains clusters of the P450 families, *CYP4*, *CYP6* and *CYP12* (Ranson *et al.* 2004). In microarray studies other members of the *CYP6* gene family have been shown to be involved in DDT or pyrethroid resistance. Increased transcript levels of *cyp6z1* and *cyp6z2* (David *et al.* 2005; Müller *et al.* 2007; Nikou *et al.* 2003; Vontas J 2005), and *cyp6p3* and *cyp6m2* (Djouaka *et al.* 2008; Müller *et al.* 2008a) have been found in resistant *An. gambiae* populations. This alone does not establish a causal relationship with resistance and a biochemical link with enhanced insecticide metabolism has to be ascertained. The recombinant *An. gambiae* CYP6P3 (Müller *et al.* 2008b) and CYP6M2 proteins (Stevenson *et al.* 2011) are able to metabolise pyrethroids *in vitro*.

Carboxylesterase based resistance

Carboxylesterases (COE) is a collective term for enzymes that hydrolyse carboxylic esters into their acid and alcohol moiety. They are involved in metabolism of endogenous as well as exogenous compounds and their target substrates are diverse. No unified classification system exists for carboxylesterases.

Esterases can target insecticides that contain ester bonds, such as organophosphates (most contain a phosphotriester bond), carbamates and pyrethroids (carboxylester bonds). There are generally two mechanisms by which carboxylesterases can contribute to insecticide resistance: (i) increase in catalytic efficiency through overexpression, (ii) high levels of carboxylesterase that act as suicide inhibitors and delay or prevent interaction between toxin and target site (Wheelock *et al.* 2005). Overexpression of non-specific carboxylesterases that hydrolyse the insecticide before it reaches the target molecule AChE is a common mechanism of organophosphate resistance in insects (Devonshire *et al.* 1998). Resistance to organophosphates in *Culex quinquefasciatus* is associated with increased production of two non-specific

carboxylesterases, ESTα2 and ESTβ2 (Vaughan *et al.* 1997) and a DDT resistant strain of *An. stephensi* had elevated levels of one COE (Vontas *et al.* 2007).

COEs have also been associated with pyrethroid resistance in *An. gambiae* (Vontas J 2005); however, they do not seem to be the major mechanism causing resistance against this class of insecticides.

Glutathione S-transferase based resistance

Glutathione transferases (GSTs) are a diverse family of dimeric, mainly cytosolic proteins with strong ligand binding properties found in almost all living organisms. They play a central role in the detoxification of endogenous and xenobiotic compounds, as well as in transport, biosynthesis of hormones and protection against oxidative stress (Enayati *et al.* 2005). Mammalian GSTs are classified into eight cytosolic and one microsomal class, whereas in insects GSTs are classed into six groups (delta, epsilon, zeta, sigma, omega and theta, with the exception of a few GSTs that are unclassified) of which the delta and epsilon class are insect specific (Ranson *et al.* 2002a). The delta and epsilon classes have expanded independently in *D. melanogaster* and *An. gambiae*, suggesting a role in specific adaptation for these enzymes. The majority of GSTs implicated in xenobiotic metabolism belong to the delta or epsilon classes (Ranson *et al.* 2002a).

GSTs catalyze the transfer of a thiol group to xenobiotic compounds, neutralizing electrophilic sites and rendering the products more water-soluble (Habig *et al.* 1974). In the context of insecticide metabolism insect this helps excretion of the toxic compounds. GSTs of *An. gambiae* and other disease vectors such as *Aedes aegypti* are of special interest because of their role in DDT resistance (Lumjuan *et al.* 2005; Ortelli *et al.* 2003; Wang *et al.* 2008). Studies on DDT resistant *An. gambiae* (Ding *et al.* 2005) and *Ae. aegypti* (Lumjuan 2005) strains have revealed overexpression of *gste2*, an epsilon class GST. The structure of the *An. gambiae gste2* has recently been determined (Wang *et al.* 2008) and

modelling of DDT into the putative binding pocket of the enzyme supports a role for this enzyme in the conversion of DDT to the non-toxic metabolite DDE (1,1-dichloro-2,2-bis-(p-chloro-phenyl)ethylene). *Gste2* and two other GST epsilon class genes, *gste5* and *gste7,* have more recently been shown to confer resistance to DDT and the pyrethroid deltamethrin in *Ae. aegypti* (Lumjuan *et al.* 2011). At present no direct involvement of GSTs in pyrethroid resistance has been shown (Lumjuan *et al.* 2011). However, it is postulated that activity of GSTs contributes indirectly to pyrethroid resistance by offering protection against oxidative stress (Vontas *et al.* 2001).

1.4.2.3 Reduced penetration

Insects are covered by a hard outer layer, called an exoskeleton. This provides a very tough, non-mineral covering for the insect, maintains the body's structure, protecting it from mechanical injury and dehydration. The exoskeleton is made up of different layers, with the top layer being the cuticle. The insect cuticle consists of protein and the polysaccharide chitin (Rebers & Riddiford 1988). Arthropod cuticular proteins (Magkrioti *et al.* 2004) are a very diverse group that may play an important role in insecticide resistance. The majority of cuticular proteins belong to a family with the R&R consensus, first identified by Rebers and Riddiford (Rebers & Riddiford 1988; Rebers & Willis 2001) named CPRs. The R&R consensus is a domain of about 35-36 amino acids, which is extensively conserved suggesting that it binds to chitin (Rebers & Willis 2001). In *An. gambiae* 156 of such CPRs have been identified (Cornman *et al.* 2008). However, there are other families of cuticular proteins such as the CPLCs (cuticular proteins with low-sequence complexity), of which 18 members have been annotated in *An. gambiae* (He *et al.* 2007).

Insecticides need to penetrate the cuticle to reach their target site. Hence the rate of insecticide absorption has an important impact on insecticide toxicity. Decreased absorption of insecticides could protect the insect, or at least encourage resistance by allowing detoxification mechanisms more time to act (Plapp 1976).

Cuticular resistance was first reported in *M. domestica* in 1968 (Hoyer & Plapp JR. 1968). Delayed cuticular penetration involved in enhanced pyrethroid resistance has been shown in a number of insect pests (Ahmad *et al.* 2006; Puinean *et al.* 2010; Valles *et al.* 2000). Recent studies on insecticide resistant *An. gambiae* and *An. stephensi* using microarray experiments have found two *cplcs*, namely *cplcg3* and *cplcg4*, significantly overexpresssed in the resistant mosquito strains (Awolola *et al.* 2009; Djouaka *et al.* 2008; Vontas *et al.* 2007).

1.4.2.4 Behavioural resistance

It is known that avoidance of insecticide treated surfaces is a behaviour displayed by insects (Lockwood *et al.* 1985). Knowledge about behavioural resistance of mosquito vectors, however, is very limited. This behaviour can take the form of hypersensitivity to a particular insecticide or a lowered sensitivity threshold or both.

Insecticides, too, can have a deterrent effect on mosquitoes. Insecticide irritancy can be demonstrated by a strong stimulation to take off and fly, mosquitoes exiting from a treated house, or both (Pates & Curtis 2005). Behavioural resistance in some countries has developed in response to spraying programs. This may include shifts in biting times, or a shift towards outdoor biting. For example *An. farauti* on the Solomon Islands showed a change in mosquito behaviour towards outdoor biting as a result of indoor use of insecticides (Taylor 1975). A more recent paper on malaria vectors in Kenya showed a small shift in biting times for *An. arabiensis* but not for *An. funestus* (Mathenge *et al.* 2001).

There is a lack of comprehensive studies to assess whether behavioural resistance affects vector control and whether behavioural changes may be an immediate response to the irritant insecticides, or a genetic trait evolved under the selection of the presence of insecticides in houses (Pates & Curtis 2005).

1.4.3 Impact of resistance on vector control

Insecticide resistance was first reported in malaria vectors in the 1950s (Elliott & Ramakrishna 1956) and resistance to DDT and pyrethroids is now widespread (Ranson *et al*. 2011). Resistance may be expected to impair malaria control efforts but evidence from field studies is limited and potentially conflicting.

An experimental hut trial conducted in Southern Benin showed, that a high frequency of *kdr* correlates to reduced efficacy of pyrethroid-based vector control efforts using insecticide-treated bed nets and indoor residual spraying (N'Guessan *et al.* 2007). However, in North-Benin pyrethroid-treated nets still remained effective against *An. gambiae* in a pyrethroid-resistance area (Corbel *et al.* 2004). Pyrethroid resistance has been implicated in the failure of IRS on Bioko Island and specifically linked to *kdr* (Sharp *et al.* 2007).

Perhaps the best example of resistance negatively impacting malaria control is in Kwazulu Natal in South Africa, where over 30 years of malaria case data are available (Craig *et al.* 2004). Here, a change in insecticide used for IRS from DDT to pyrethroids resulted in the reappearance of *An. funestus* from neighbouring Mozambique and a dramatic increase in malaria cases. The *An. funestus* population was highly resistant to pyrethroids and found resting inside sprayed houses. A later change in policy to reintroduce DDT was accompanied by a decrease in malaria incidence (Hargreaves *et al.* 2000).

A systematic review is in preparation trying to quantify the impact of resistance on malaria control (Dr. Clare Strode, LSTM, personal communication).

Emergence of resistance however may be delayed or even prevented through good practice, e.g. choice of insecticide, rotation of insecticides, combination with synergists, or mosaic spraying. The current spread of insecticide resistance in the major malaria vectors stresses the need for frequent and careful monitoring of resistance in disease vectors.

1.4.4 Monitoring of insecticide resistance

Surveillance to monitor the emergence and spread of resistance is an essential step in insecticide resistance management (IRM) providing baseline data for programme planning and pesticide selection (Black IV *et al.* 2008). Effective resistance monitoring can improve the efficacy of vector control and may also delay or prevent the onset and spread of resistance.

Insecticide resistance is commonly assessed through bioassays by exposing mosquitoes to a diagnostic dose of insecticide using standard protocols published by the World Health Organization (WHO 2011b). Knockdown time is an indicator for resistance. However, if resistance alleles are partially or fully recessive, like *kdr* (Chandre *et al.* 2000) bioassays will only detect resistance, when alleles have already reached a frequency high enough for resistant homozygotes to occur.

Detection of resistance at the molecular level is more sensitive and can provide early warning of resistance. Target-site mutations can, once characterised, be detected via PCR based methods (Bass *et al.* 2007; Bass *et al.* 2010; Du *et al.* 2005).

Metabolic resistance is more complex and there is no quick and easy field compatible approach to identify the mechanisms or enzymes involved. Biochemical assays, whilst relatively low cost and simple to perform (Hemingway 1989; Penilla *et al.* 1998) lack specificity. Microarrays, e.g. ‘detox chips’ can identify genes putatively conferring resistance but require specialist equipment and training and are relatively expensive (David *et al.* 2005; Strode *et al.* 2007).

1.5 Techniques to investigate resistance mechanisms

1.5.1 Gene expression analysis

Microarray and quantitative RT-PCR are RNA based assays used to assess transcription levels of genes of a number of samples, tissues or conditions.

A microarray experiment is a targeted expression analysis to create a snapshot of transcriptional activity. Expression profiles of large numbers of genes can be measured simultaneously in two samples through probe-target hybridization, to assess the relative abundance of sequences in the target (Schulze & Downward 2001). This process seeks to identify candidate genes which are differentially expressed in a sample of interest, i.e. in resistant mosquito populations (David *et al.* 2005; Djouaka *et al.* 2008; Müller *et al.* 2008a).

Small scale microarray platforms, known as 'detox-chips' were developed to specifically study expression profiles of detoxification genes (primarily cytochrome P450s, GSTs and COEs) in both *Anopheles* and *Aedes* mosquitoes (David *et al.* 2005; Strode *et al.* 2007). These have been extensively used to study insecticide resistance in mosquitoes (Awolola *et al.* 2009; Christian *et al.* 2011; David *et al.* 2005; Djouaka *et al.* 2008; Müller *et al.* 2008a; Müller *et al.* 2007; Poupardin *et al.* 2008; Strode *et al.* 2007; Strode 2006).

More recently, use of these specific microarray platforms has been largely replaced by whole genome microarrays (Mitchell *et al.* 2012; Poupardin *et al.* 2012), which have the advantage of being able to identify additional mechanisms that may play an important role in insecticide resistance.

The microarray process is an initial screen which will generate a list of candidate genes which need to be validated by further steps. A microarray screen of all known P450s in *Drosophila* led to the identification of a single gene (*cyp6g1*) conferring insecticide resistance to DDT (Daborn *et al.* 2002). Since then a cis-acting regulatory element of *cyp6g1* has been shown to cause tissue specific over-transcription of this gene (Chung *et al.* 2007). Similarly *cyp6p3* and *cyp6m2* were identified by microarray as

overtranscribed in permethrin resistant field populations of *An. gambiae* (Djouaka *et al.* 2008; Müller *et al.* 2008b).

Candidate genes can be investigated further by quantitative RT-PCR to confirm elevated expression levels of a number of candidate genes (Müller *et al.* 2008b; Munhenga & Koekemoer 2011) or measure tissue related expression (Stevenson *et al.* 2011). As a typical follow-up for microarray experiments, quantitative RT-PCR is cheaper and quicker to carry out and a larger number of samples can be assayed in parallel for comparison. The method is based on the principle that DNA amplification can be measured in real-time by measuring the increase in fluorescence as a chemical compound is incorporated into a newly synthesised double stranded DNA fragment (Morrison *et al.* 1998).

Expression analysis however lacks the potential to establish a functional link between gene candidates and the phenotype.

1.5.2 Functional validation of candidate genes

Functional assessment is an important part of candidate validation. An *in vitro* approach to functional validation is heterologous protein expression, subsequent purification and biochemical assays to investigate the metabolic properties of candidate enzymes. Two *An. gambiae* cytochrome P450 genes, *cyp6p3* and *cyp6m2*, were initially highlighted as candidates for pyrethroid resistance by microarray experiments (Djouaka *et al.* 2008; Müller *et al.* 2008b). Subsequently their ability to metabolise pyrethroids was validated *in vitro* (Mitchell *et al.* 2012; Stevenson *et al.* 2011).

RNA interference (RNAi) is an *in vivo* approach (it may also be carried out *in vitro* in cell lines), to the functional validation of expression candidates. This technique is a genetic interference tool (Fire *et al.* 1998) which creates a loss-of-function phenotype by depleting a chosen transcript. The first description of *An. gambiae* immune genes that have an antagonistic effect on *Plasmodium* development was carried out with RNAi (Osta *et al.* 2004). RNAi induced gene silencing in a pyrethroid resistant strain of *Ae. aegypti* led to increased susceptibility to deltamethrin (Lumjuan *et al.*

2011) and *in vivo* knock down of cytochrome P450 reductase increased permethrin susceptibility in *An. gambiae* (Lycett *et al.* 2006).

1.5.3 QTL Mapping

Expression analysis will fail to identify regulatory factors that change a phenotype due to structural changes in proteins rather than overexpression, such as target-site mutations or enzymes with enhanced substrate affinity. Quantitative trait locus mapping has the potential to identify these factors.

Quantitative Trait Loci (QTL) mapping is a whole genome based approach that involves the study of associations between phenotype and genotype in order to understand and dissect the regions of the genome that affect complex traits (Doerge 2002). The main requirements for QTL mapping are a population of a measured phenotype and availability of genetic markers. The population is then genotyped for a chosen set of those markers followed by construction of a genetic map which is used to identify loci linked to the trait of interest.

Microsatellite markers are a type of genetic markers commonly applied for QTL mapping. A genetic map consisting of 131 microsatellite markers with an average resolution of 1.6 cM was constructed for *An. gambiae* many years before the genome sequence of this species was determined (Wang *et al.* 1999; Zheng *et al.* 1996; Zheng *et al.* 1993). With a draft genome now complete (Holt *et al.* 2002) there is a vast array of putative markers that can be used for genetic mapping studies.

QTL mapping has successfully been applied in vectors of disease. Examples include mapping of insecticide resistance in various mosquito species such as *An. gambiae* (Ranson *et al.* 2000b; Ranson *et al.* 2004), *An. funestus* (Wondji *et al.* 2007) and *Ae. aegypti* (Saavedra-Rodriguez *et al.* 2008) as well as susceptibility to *Plasmodium* infection in *An. gambiae* (Niare *et al.* 2002; Riehle *et al.* 2006).

Previous work on genetic mapping of the genes responsible for insecticide resistance in mosquitoes all used experimental crosses between two inbred lab strains of mosquitoes, a resistant and a susceptible population,

to generate mapping populations (Ranson *et al.* 2004; Saavedra-Rodriguez *et al.* 2008; Wondji *et al.* 2007). QTL studies can also be used for continuous traits in unmanipulated natural populations (Slate 2005). Two studies on *Plasmodium* infection in *An. gambiae s.s.* (Niare *et al.* 2002; Riehle *et al.* 2006) used phenotypic variation in natural populations of *An. gambiae* to map allelic variants effecting parasite development.

1.6 Aims of the study

This study investigates the genetic basis of permethrin resistance in two of the most important African malaria vectors.

(A) Permethrin resistance and QTL mapping of permethrin resistance in isofemale families of *Anopheles gambiae s.s.* from Benin, West Africa.

Pyrethroid resistance is prevalent in South Benin (Corbel *et al.* 2007; Djegbe *et al.* 2011; Djogbénou *et al.* 2010). Previous studies had demonstrated the presence of the 1014F *kdr* alleles but also implicated both metabolic and cuticular genes in the resistance phenotype (Djouaka *et al.* 2008).

In this study genetic mapping was used to identify the genetic basis of pyrethroid resistance in field populations of *An. gambiae s.s.* in Benin. The specific objectives were to:

1. Identify the main QTLs associated with permethrin resistance using F_2 isofemale lines from Benin
2. Assess how much target-site and metabolic resistance contribute to the overall variance in the resistant phenotype
3. Validate the association of the candidate genes with resistance using qPCR

(B) Investigating the genetics of permethrin resistance in *Anopheles arabiensis* using a combination of microarray study and QTL mapping

Anopheles arabiensis is the most important malaria vector in Chad, Central Africa (Kerah-Hinzoumbe *et al.* 2009). Resistance to pyrethroids is prevalent in the south of the country (Kerah-Hinzoumbe *et al.* 2008; Ranson *et al.* 2009) but this resistance is not associated with target-site mutations.

This study aimed to identify the genes conferring resistance to permethrin in this vector. Specifically the objectives were to:

1. Identify the major loci associated with pyrethroid resistance in the Ndja laboratory strain
2. Identify genes differentially expressed in the resistant strain relative to a laboratory susceptible strain

Chapter 2. General Materials and Methods

2.1 Mosquito samples

A range of mosquito samples from three countries across West, Central and East Africa was used in this study. These include lab colonies as well as larval and adult field collections belonging to two species of the *Anopheles gambiae* complex, *An. gambiae s.s.* and *An. arabiensis*, both important African malaria vectors.

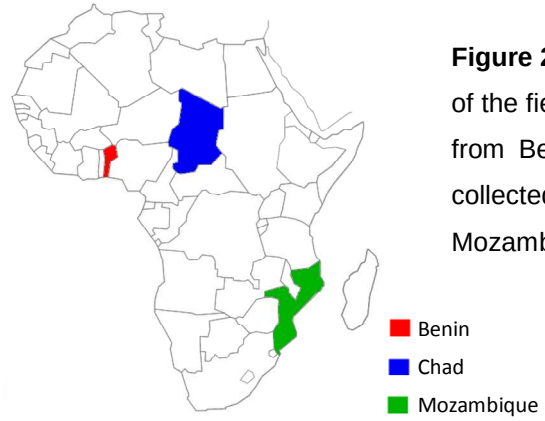


Figure 2.1: **Outline of Africa with countries of origin** of the field samples used in this work. *An. gambiae s.s.* from Benin (red) in West Africa, and *An. arabiensis* collected from Chad (blue) in Central Africa and Mozambique (green) in the south-east of the continent.

2.1.1 *Anopheles gambiae s.s.*

2.1.1.1 Lab colonies

During maintenance in the insectaries adult mosquitoes were kept at T=26±1℃ and relative humidity=75±5% (photoperiod o f 12 hours) in cubic cages (25 cm), supplied with 1% sugar solution applied to cotton balls on top of the cage. Mosquitoes were allowed to mass mate for around one week, when females were blood-fed on human blood using artificial membranes and electric feeder arms (blood heated to 37℃). Gr avid females were provided with cups half filled with water and lined with filter paper for egg laying. The eggs were then transferred to clean filter paper, rinsed with 1% bleach for surface-cleansing, followed by rinsing multiple times with distilled water to remove remaining bleach. Eggs were stored on moist filter paper in a plastic bag for one day until they were set up for hatching in distilled water. This procedure seemed to result in more synchronised hatching. Larvae were fed finely ground tetra flake fish food (Tetra GmbH, Germany) until the

pupal stage, and 1st and 2nd instar larvae were supplemented with yeast extract (Holland & Barrett, Burton Upon Trent, UK). Larvae were kept to a density of about 1 larva/ ml in plastic basins (20x25x5 cm) containing distilled water to a depth of 4 cm. Pupae were removed from the water trays with plastic pipettes before emerging and transferred into cages.

Akron lab colony

The Akron lab colony is an insecticide resistant strain of *An. gambiae s.s.*, originating from Porto-Novo, Benin, which was established in MR4 (Malaria Research and Reference Resource Center, USA) in June 2008. This colony is described by MR4 as being *An. gambiae* M-form and containing the *ace-1* and 1014F *kdr* alleles and being resistant to carbamates. Initial characterisation of the strain once received at LSTM confirmed the carbamate resistance phenotype (0% mortality after 1 hour exposure to bendiocarb 0.1%) and also a low level of resistance to permethrin (32.5 % mortality after 20 minutes exposure to permethrin 0.75%).

The colony was split into two. One part was subjected to selection for permethrin resistance, henceforth referred to as Akron-permethrin-selected or Pakron strain. Selections with permethrin (0.75%) were performed on one to two-day old adults, survivors of each assay were pooled and allowed to mass mate for the next generation, non-survivors were discarded. The initial exposure time of 20 minutes (for generation F_5) was gradually increased to 60 minutes.

The second colony was subjected to selection for bendiocarb resistance, henceforth referred to as Akron-carbamate strain. Selections were performed sporadically by exposing 3-5 day old adults to bendiocarb (0.1%) for one hour. Survivors were allowed to mass mate for the next generation, non-survivors were discarded.

Ngousso lab colony

The Ngousso lab colony is an insecticide susceptible strain of *An. gambiae s.s.*, M molecular form. This strain was colonised in 2006 in Ngousso, a district of Yaoundé, Cameroon, West Africa, and was kindly provided by Antonio Christophe, OCEAC, Cameroon.

2.1.1.2 *An. gambiae s.s.* field samples from Benin, West Africa

Larval collections

The material for the study consisted of *An. gambiae* collected in 2009 between September 17th and 23rd.

Larvae and pupae were collected on three different days from various breeding pools between the vegetable fields, filtered, put into fresh water and transferred to the laboratory (Figure 2.2). Larvae were fed with finely ground dog food and raised to adults. The emerging adults were retained until they were 2-5 days old and then used in bioassays. All insecticide bioassays described in this report were carried out according to standard WHO specifications using WHO insecticide susceptibility kits and insecticide treated papers (permethrin 0.75%, DDT 4% and bendiocarb 0.1%) (WHO 1998). After an exposure of 1 hour, mosquitoes were maintained on honey water in insecticide free tubes for 24 hours. Dead and alive individuals were stored in silica gel for further analysis. Survivors from a subset of the permethrin bioassays were killed 24 hours after exposure and stored in RNA later ® (Sigma-Aldrich, UK).

Figure 2.2: Larval collection, cleaning and rearing for bioassay experiments.

Another set of field samples collected by Dr. Rousseau Djouaka, IITA, Benin, from Akron in 2008 was already available. The set consisted of 69 *An. gambiae s.s.* specimens (34 alive, 35 dead; exposed to 0.75% permethrin for one hour). Some of these samples were used for the sequencing of the candidate genes and SNP genotyping.

Adult mosquito collections and forced egg laying

The material for this part of the study consisted of *An. gambiae* collected between September 17th and 23rd, 2009. Adult mosquitoes were collected inside houses in Akron either in the early morning hours (between 6 and 8 am) or during the night (between 12 midnight and 6 am). Mosquitoes were collected in a cage, transferred to the laboratory and only female *Anopheles* mosquitoes were retained. The *Anopheles* mosquitoes were blood fed on two to three subsequent nights using guinea pigs that were physically immobilised and left in the cage overnight.

Four to six days after blood feeding gravid females were individually transferred into 1.5 ml Eppendorf tubes (Morgan *et al.* 2010) (Figure 2.3) containing a square of moistened filter paper lining the inside of the tube. The tubes also had two small holes in the lid and further down the tube to provide the insect with oxygen. Tubes with female mosquitoes were put upright into racks and left to stand in the insectaries.

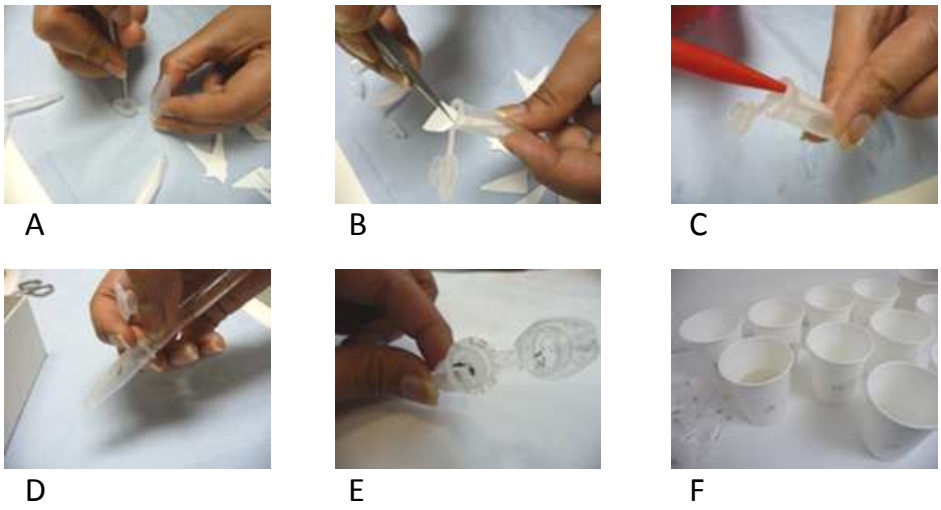


Figure 2.3: Preparation of 1.5 ml Eppendorf tubes for forced egg laying
(A-C), individualisation of gravid females (D), female with eggs (E) and raising of larvae (F) (pictures John Morgan).

Tubes were checked for eggs daily and those females which had oviposited were cooled down in the fridge for 1-2 hours to reduce the motility of the female. Each female was then separated from her eggs and stored in a tube with silica gel for further molecular analysis.

Eggs were stored in the original Eppendorf tubes until transported to LSTM, where they were set up in distilled water and raised to adults as described above (Lab colonies). Their F_1 generations were left to randomly intermate to produce the second generation. The F_2 offspring were then phenotyped and used for QTL mapping experiments.

2.1.1.3 Resistance phenotyping of F_2 offspring

Three isofemale families (families 15, 17, and 25) yielded sufficient progeny to be phenotyped. WHO insecticide treated papers (permethrin 0.75%) were used to phenotype F_2 offspring of family 17. Due to very high survival rates on these papers F_2 offspring of families 15 and 25 were phenotyped on self made 2% permethrin papers. These were made by dissolving 100 mg of permethrin (Chem Service, West Chester; purity 46% cis – 52% trans) in 5 ml of acetone and an equal volume of Dow Corning 556 Silicone Fluid (BDH/ Merck). Aliquots of 1.4 ml were evenly spread on 12x15 cm pieces of Whatman No. 1 (1 mm) paper until all paper was covered. Papers were dried overnight, wrapped in tin foil and stored at 4℃.

After exposure mosquitoes were maintained on sugar water in insecticide free tubes for 24 hours. Dead and alive individuals were counted, sexed and put on silica gel for further analysis.

2.1.2 *Anopheles arabiensis* from Chad and Mozambique

2.1.2.1 Field collections, establishment and characterization of mosquito colonies

Gravid or blood fed resting *An. gambiae s.l.* adult females were collected by the National Malaria Control Programme in houses in N'Djamena (12° 6' 47" North, 15° 2' 57" East), capital city of Chad, in September 2009 between 6 and 12 am. They were left in cages until they had laid eggs which were then transported to the Liverpool School of Tropical Medicine (LSTM). This work was done by Clément Kerah-Hinzoumbé (Programme National de Lutte contre le Paludisme, NDjaména, Tchad) who kindly provided the Ndja eggs to LSTM. The eggs were reared and fed with Tetramin™ baby fish food and a colony, named Ndja, was established. Additionally, a similar sampling of females *An. gambiae s.l.* was carried out from Chokwe (24° 33' 37" S, 33° 1' 20" E) in Southern Mozambique in July 2009 by Charles Wondji (LSTM) and Nelson Cuamba (National Institute of Health, Maputo, Mozambique). Eggs were also transported to LSTM for rearing and establishment of a colony named Moz. All field collected females used to establish the laboratory colony were morphologically identified as belonging to the *An. gambiae* species complex according to the key of (Gillies & Coetzee 1987) and identified as *An. arabiensis* by PCR (Scott *et al.* 1993).

To characterize the newly established colonies, insecticide susceptibility assays were carried out using 3-5 day old F_1 adults from both colonies following the WHO protocol (WHO 1998). The following insecticides were tested: the pyrethroids permethrin (0.75%), and deltamethrin (0.05%); the carbamate bendiocarb (0.1%); the organophosphate malathion (5%); and the organochlorines DDT (4%) and dieldrin (4%). The work in LSTM was carried out by Dr. Charles Wondji and Matthew Parry, LSTM.

As P450 monooxygenases have previously been implicated in pyrethroid resistance in *An. arabiensis* (Müller *et al.* 2008b), their potential involvement in the resistance was assessed in the Ndja strain using the P450 inhibitor PBO (piperonyl butoxide) (Scott 1999). The synergist PBO effect was analyzed in combination with 0.75% permethrin and also 4% DDT to

compare between the two insecticide classes. 100 female mosquitoes were pre-exposed to 4% PBO paper for 1h and immediately exposed to 0.75% permethrin or 4% DDT for 1h. Final mortality was assessed after 24h and compared to the results obtained without PBO. Resistant mosquitoes were stored at -80℃ for subsequent DNA and RNA extraction, and dead mosquitoes were stored with silica gel.

Results of insecticide susceptibility assays with a range of insecticides for the established strains Ndja and Moz are shown in Table 2.1. The Ndja colony shows resistance to the type I pyrethroid, permethrin (64.9% mortality after a 1 hour exposure) and reduced susceptibility to the type II pyrethroid deltamethrin (90.6% mortality for females). A mortality of 100% was restored by exposure to 4% piperonyl butoxide, PBO, prior to 0.75% permethrin. PBO increased DDT mortality from 69.1 % to 84%.

Insecticide (Concentration %)	Females		Males		Total	
	n	% Mortality	n	% Mortality	n	% Mortality
Ndja						
Permethrin (0.75)	81	54.3	131	75.6	212	64.9
Deltamethrin (0.05)	85	90.6	120	98.3	204	95.5
DDT (4.0)	97	69.1	114	91.2	211	81.0
Bendiocarb (0.1)	83	100	72	100	155	100
Malathion (5.0)	80	95.0	73	100	153	97.4
Dieldrin (4.0)	81	90.1	67	100	148	94.6
Moz						
Permethrin (0.75)	109	100	70	100	179	100
Deltamethrin (0.05)	93	100	77	100	170	100
DDT (4.0)	114	89.5	81	95.1	195	91.7
Bendiocarb (0.1)	87	100	55	100	142	100
Malathion (5.0)	81	100	62	100	143	100
Dieldrin (4.0)	65	100	58	100	123	100

exposure time 1 hour

Table 2.1: WHO susceptibility test results for 2-5 day old *An. arabiensis* from the colonies Ndja, and Moz; the median number of mosquitoes genotyped per insecticide is 84 females and 73 males (data from Dr. Charles Wondji LSTM).

The *An. arabiensis* Moz strain is susceptible to all insecticides apart from DDT where a reduced susceptibility is observed with 89.5 and 95.1% mortality respectively for females and males (Table 2.1).

A subset of 92 mosquitoes was genotyped for both L1014F and L1014S *kdr* mutations, comprising of 82 from the Ndja strain (25 permethrin resistant females, 15 permethrin resistant males, 9 deltamethrin resistant females and 13 DDT resistant females, as well as 15 females and 5 males that were permethrin susceptible) and 10 permethrin susceptible females from Moz strain. Additionally, all the mosquitoes alive after bioassays with the organophosphate malathion were screened for the presence of the acetylcholinesterase target-site mutation G119S (*ace-1*) associated with carbamate and organophosphate resistance using the pyrosequencing method. No *kdr* mutation was found in the Ndja strain or in the Moz strain. No *ace-1* mutation was observed for the 50 mosquitoes genotyped (25 Ndja and 25 Moz) with all samples showing the wild type susceptible allele in both strains. This work was carried out by Dr. Charles Wondji and Matthew Parry, LSTM.

2.1.2.2 Establishment of genetic crosses and experimental populations

Reciprocal crosses between the pyrethroid resistant (Ndja) and susceptible (Moz) colonies were established by Dr. Charles Wondji and Matthew Parry. Pupae from each strain were individually placed in a tube for emergence and newly emerged male and female adults of one strain were placed in mating cages with the opposite sex of the other strain (Ndja (RR) males with Moz (SS) females and *vice versa*). Mosquitoes were allowed to mate for 4 days after which they were blood fed for oviposition. Eggs from each crossing were collected and reared separately to adults. One part of the resulting F_1 hybrid generation was then intercrossed to generate the F_2 generation; the other part was backcrossed to the resistant parent, the NDJA strain, to generate the backcross generation B_1 (Figure 2.4). This set up provided us with two sets of mosquito samples: (a) a backcross B_1 generation, providing RNA for microarray analysis as well as qPCR assays and (b) F_2 progeny from reciprocal crosses to yield DNA for QTL mapping.

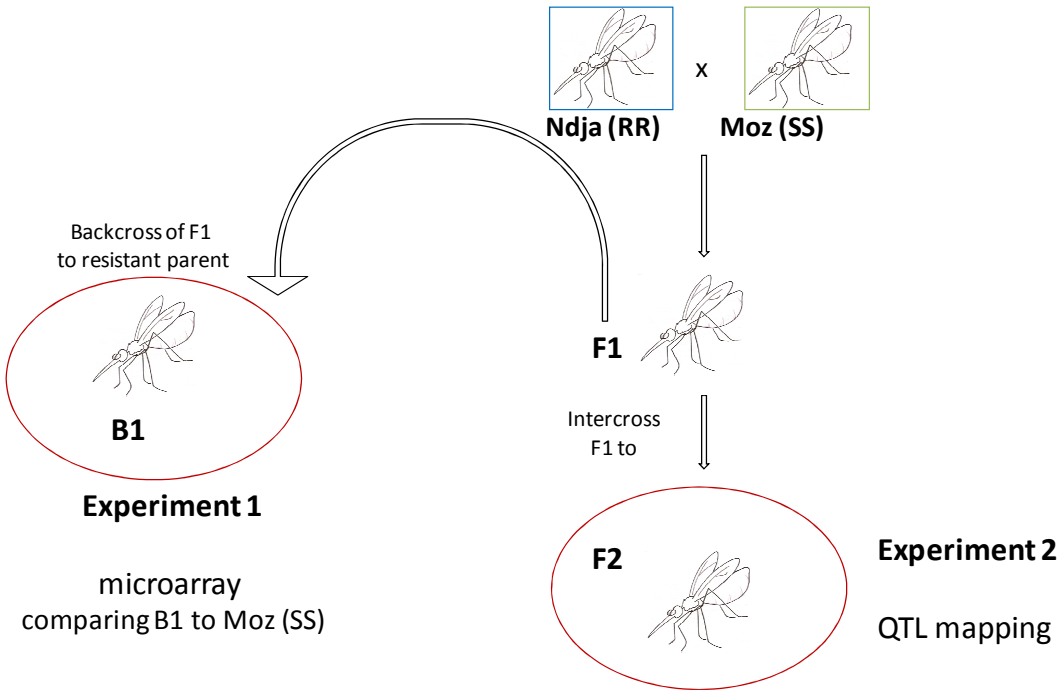


Figure 2.4: Experimental crosses of *An. arabiensis*.

Shown is a schematic overview of the experimental set-up of the genetic crosses of the two *An. arabiensis* strains Ndja and Moz yielding progeny B_1 and F_2 for the two different experiments.

2.1.2.3 Resistance profile of genetic crosses

WHO Bioassay were carried out using 0.75% permethrin for 1h, as described above for susceptibility assays on the F_1 progeny, the F_2 progeny from F_1 intercross as well as the B_1 backcross of F_1 with the parental Ndja strain.

Permethrin induced mortalities for the reciprocal crosses family 2 and 3 in the F_1 and F_2 generations were similar (72-78 %) (Table 2.2). Backcrossing to the parental resistant Ndja strain reduced the mortality to 27%.

Survivors and non-survivors of the F_2 samples were dried and stored on silica gel for later DNA extraction and are further referred to as alive and dead; survivors of the B_1 cross were snap frozen and stored at -80℃ for later RNA extraction. DNA of the F_2 samples was used for QTL mapping (Chapter 5) and RNA of the B_1 samples was used for microarray and qPCR experiments (Chapter 6).

Cross	n tested	mortality [%]
Parental strains		
Ndja (RR)	213	61.0± 10.2
Moz (SS)	270	100±0.0
F_1 from reciprocal parental crosses		
F_1-Family 2 (SS ♂ x RR ♀)	129	78.3± 5.5
F_1-Family 3 (SS ♀ x RR ♂)	168	75.0± 6.3
Backcross to RR		
B1 (F_1(♀) family 3 X RR(♂))	148	27.0± 6.0
F_2 from F_1 intercross		
F_2-family 2 (F_1(♀) x F_1(♂))	168	77.4± 8.4
F_2-family 3 (F_1(♀) x F_1(♂))	534	72.1± 11.0

Table 2.2: Survival of experimental crosses of *An. arabiensis*

Shown is mortality [%] 24hr post exposure following exposure to 0.75% permethrin for 1 h (data from Dr. Charles Wondji, LSTM).

2.2 Genetic characterisation of mosquitoes

Genomic DNA was extracted from single mosquitoes using the protocol by Livak (Livak 1984). Each mosquito was ground in 100 µl buffer (0.5% SDS/0.08 M NaCI/0.16 M sucrose/0.06 M EDTA/0.I2 M Tris-HCI, pH 9) in a 1.5 ml Eppendorf tube, and incubated at 65°C for 30 min. A volume of 14 µl of potassium acetate (8M) was added and the homogenate incubated on ice for 30 min. The mixture was then centrifuged for 15 min (13,000 rcf) to collect debris and precipitated protein. The supernatant was transferred to a clean tube and nucleic acid precipitated by adding 200 µl of ice cold ethanol (100%). The mixture was vortexed for a couple of seconds and then left on ice for 20-60 min. The pellet was collected by centrifuging for 15 min (13,000 rcf) and washed twice in ice cold ethanol (70%). It was then dried, re-dissolved in 100 µl sterile water and kept at -20℃.

2.2.1 Species identification and molecular cytotyping

Molecular identification to species level was performed for all mosquito samples using the PCR protocol developed by (Scott *et al.* 1993). For specimens identified as *An. gambiae s.s.* the molecular form of each sample was confirmed using the PCR protocol based on the amplification of a SINE transposon (Santolamazza *et al.* 2008b) as well as the karyotype for the 2La inversion (White *et al.* 2007) (Table 2.3). The 2Rb karyotype was determined in all parental females and a small number of F_1 offspring (Lobo *et al.* 2010) (Table 2.3).

2.2.2 Genotyping of the kdr and ace-1 mutation

Single mosquitoes were genotyped for the 1014F *kdr* allele using either the TaqMan assay (Bass *et al.* 2007) or a pyrosequencing assay (developed by Charles Wondji, LSTM). The *ace-1* allele was genotyped using a pyrosequencing assay (Charles Wondji, LSTM). For the three field collected *An. gambiae s.s.* parental females (families 15, 17 and 25) from Akron the *kdr* genotype was additionally confirmed through sequence analysis (sequencing service, Macrogen, Korea).

Primer name	Primer sequence 5'-3'	Prod. size [bp]	$MgCl_2$ [mM]	T_A [°C]	Primer source and comments
Species ID					
UN_F	GTGTGCCCCTTCCTCGATGT		1.75	50	(Scott *et al.* 1993)
GA_R	CTGGTTTGGTCGGCACGTTT	390			
ME_R	TGACCAACCCACTCCCTTGA	464/466			
AR_R	AAGTGTCCTTCTCCATCCTA	315			
QD_R	CAGACCAAGATGGTTAGTAT	153			
Inversion Karyotyping 2Rb					
2Rb_For	CGGGAGCAAAGATAAGTAGCA	429 or	1.75	58	(Lobo *et al.* 2010)
2Rb_Rev	AACCCTACCATATACCAGTACCAACG	630			
2Rb+_Rev	CCGGATAATCGACGCTCTAC				
Inversion Karyotyping 2La					
Universal	CTCGAAGGGACAGCGAATTA	207 or	2.5	60	(White *et al.* 2007)
2La	ACACATGCTCCTTGTGAACG	492			
2L+[a]	GGTATTTCTGGTCACTCTGTTGG				
Molecular Form Genotyping					
SINE200_F	TCGCCTTAGACCTTGCGTTA	M: 479;	1.5	54	(Santolamazza *et al.* 2008b)
SINE200_R	CGCTTCAAGAATTCGAGATA	S: 249			
1014F *kdr* Taqman Assay					
kdr-Forward	CATTTTTCTTGGCCACTGTAGTGAT				(Bass *et al.* 2007)
kdr-Reverse	CGATCTTGGTCCATGTTAATTTGCA				
probe WT	CTTACGACTAAATTTC				VIC labelled
probe 1014F	ACGACAAAATTTC				FAM labelled
probe 1014S	ACGACTGAATTTC				FAM labelled
1014F *kdr* Pyrosequencing Assay					
kdr-PyrAg_F	TTGTGTTCCGTGTGCTATGC	154	1.5	50	Charles Wondji
kdr-PyrAg_R-bio	AAAAACGATCTTGGTCCATGT				(5'Biotin) labelled
kdr-Ag-seq	TGTAGTGATAGGAAAT			T C/T A/T GTCGTAAG ¥	
Ace-1 Pyrosequencing Assay					
AChE-PyrAg_F	CCTGTCCGAGGACTGTCTGT	164	1.5	55	Charles Wondji
AChE-PyrAg_R-bio	ACCACGATCACGTTCTCCTC				(5'Biotin) labelled
AChE-seq	TGTGGATCTTCGGCGG			C A/G GCTTCTACTCC ¥	
Sequencing of Sodium channel					
Ex20_F	AAATGTCTCGCCCAAATCAG	568	3.5	61	(Lynd *et al.* 2010)
Ex20_R	GCACCTGCAAAACAATGTCA				
Exon 29_F	AAATGCTCAGGTCGGTAAACA	330	1.8	61	Hilary Ranson
Exon 29_R	GCCACTGGAAAGAATGGAAA				
EX27-31_F	GAATGCGTTGGTTCAAGCTA	1355	3.0	57	C. Wondji
EX27-31_R	TTTGACGTGCATGAAAAATGA				

¥ sequence to be analysed

Table 2.3: Details of the primers used for species identification and genotyping of *An. gambiae* gDNA samples.

For a subset of the samples additional regions of the sodium channel were amplified using primer sets Ex20, Ex29 and Ex27-31 (Table 2.3) and sequenced by Macrogen, Korea.

2.3 Quantitative RT PCR

2.3.1 RNA extraction and cDNA generation

RNA was extracted from pools of three *An. gambiae s.s.* using the Trizol extraction method or from 10 *An. arabiensis* using the PicoPure™ RNA Isolation Kit (Arcturus®, Applied Biosystems). All mosquito samples were three day-old, unmated female mosquitoes. All mosquito samples with the exception of the *An. gambiae s.s.* Benin field samples and the *An. arabiensis* B_1 samples (described in Chapter VI), were unexposed specimens, snap frozen and stored at -80℃. RNA was also extracted from the Benin field samples that survived a 1 hour exposure to 0.75% permethrin and after a recovery period of 24 hours were snap frozen and stored in RNA*later*® (Ambion, Applied Biosystems).

For the Trizol extraction samples were homogenized in 200µl Trizol using a pestle and centrifuged at 13,000 rcf for 10 minutes (4℃). The clear supernatant was transferred to a fresh tube, left to stand for 5 minutes (RT) and then 40µl of chloroform was added. Samples were vortexed for several seconds, left to stand for 2-15 minutes (RT) and centrifuged at 13,000 rcf for 15 minutes (4℃). The upper aqueous phase was transferred to a new tube, RNA precipitated by adding 100 µl of isopropanol and the mixture left to stand for 5-10 minutes (4℃). RNA was pelleted by centrifuging at 13,000 rcf for 10 minutes (4℃). The supernatant was removed, the pellet washed in 200 µl of ethanol (75%), vortexed and centrifuged at 5000 rcf for 5 minutes (4℃). The pellet was briefly air-dried and re-dissolved in 26 µl of DEPC water. For the extractions with the Pico pure kit, instructions were followed as stated in the handbook by the manufacturer.

For DNase treatment, 1 µl 10x DNase buffer and 1 µl DNase (Promega UK, Southampton) was added to 8 µl RNA (~300ng/µl), the mixture was incubated at 37℃ for half an hour, and the reaction stopped with 1 µl of DNase stop solution (65℃ for 10 minutes). RNA was stored at -80℃ for later use.

An aliquot of RNA served as template for making cDNA by reverse transcription. To obtain single stranded cDNA a reaction mix of 20 µl was

made consisting of 2 µl oligo dTs 18 (50 µM), 1 µl dNTPs (10 mM), 4 µl 5x transcriptase buffer, 1 µl Reverse Transcriptase (RevertAid Premium Reverse Transcriptase, Fermentas), 0.5 µl RNase inhibitor (RiboLock, Fermentas), 1.5 µl dd. H_2O and 10 µl RNA containing up to 5 µg RNA. The reaction mixture was incubated at 50℃ for 30 min followed by 85℃ for 5 minutes. The resulting cDNA was RNase treated to remove RNA, by adding 5 µl 10x RNase buffer, 1 µl RNase (New England Biolabs UK) and 24 µl dd. H_2O. The mixture was incubated at 37℃ for 1 hour followed by heat inactivation at 65℃ for 20 minutes. The cDNA was further cleaned-up by column purification (QIAquick PCR Purification Kit, Qiagen) and eluted in 30 µl dd H_2O.

To check for genomic DNA contamination in the cDNA a PCR reaction using *s7* primers (targeting the transcript for the ribosomal subunit *s7* AGAP010592-RA (Salazar *et al.* 1993), used as a control gene) was carried out. The primers span an intron of 149 bp, hence PCR products from gDNA will be longer than the product from cDNA, a result which can easily be visualized by running the PCR products on a 1.5% agarose gel. Details for *s7* primers are listed in Table 2.4. The 25 µl reaction mixture was as follows: 2.5 µl 10x reaction buffer, 0.1 µl Taq Polymerase (Kapa), 0.5 µl dNTPs (10 mM), 1.25 µl of each primer (10 µM), 18.4 µl H2O and 1 µl cDNA. Cycling conditions are 95℃/ 2‘ (95℃/ 30‘‘,60℃/ 30‘‘,72℃ / 30‘‘)x 40, 72℃/ 2‘.

2.3.2 qPCR primers and experimental set up

The quantitative PCR reaction was performed on the Stratagene Mx3005P qPCR system (Agilent Technologies) and analyzed using Agilent’s qPCR software, MxPro. The qPCR primers for *cyp6p3*, *cyp6m2, cplcg4* and *cpr* were designed against the *An. gambiae* consensus sequence from Akron field samples (Chapter 3.3.2). Primers for *cyp4h24* were designed against the *An. arabiensis* consensus sequence from Ndja and Moz samples (Chapter 2.1.2).

The qPCR assays were optimized with respect to $MgCl_2$ concentration and annealing temperature. The optimal conditions were no additional $MgCl_2$ (concentration in reaction butter not defined by manufacturer) and an

annealing temperature of 60℃ for the primer sets d etecting *cyp6p3*, *cplcg4*, *cyp4h24*, *s7* (Salazar *et al.* 1993) and cytochrome P450 reductase, *cpr*, and 62℃ for those targeting *cyp6m2*. The real-time qPCR reaction mixture was prepared to a final volume of 25 µl using the Maxima® SYBR/ROX qPCR Master Mix (Thermo Fisher Scientific): 10 µl nuclease free water, 0.75 µl of each primer (final concentration 0.3 µM), 12.5 µl reaction mix 2x, and 1 µl template cDNA (diluted to 10 µM). The thermal cycling protocol was as follows: initial denaturation for 10 min at 95 °C followed by 40 cycles of 15 s at 95 °C, 30 s at 60/62 °C and 30 s at 72 °C. The fluorescence signal was measured at the end of each extension step at 72 °C. After amplification, a melting curve analysis with a temperature gradient of 0.1 C/s from 60 to 95 °C was performed to confirm that only the specific products were amplified.

For each sample three biological replicates were run alongside three technical replicates of the standard curves. Primer details are listed in Table 2.4.

Target gene	Primer name	Accession number	Primers 5'-3' F forward and R reverse	Prod. size [bp]		TA [°C]
				gDNA	cDNA	
cyp6p3	P3.7	AGAP002865	F: CGGCGTGTTTACCAATGC R: GGTTGGTTGTAGCTTTGCTCC	192	192	60
cyp6m2	M2.5	AGAP008212	F: TACGATGACAACAAGGGCAAG R: GCGATCGTGGAAGTACTGG	130	130	62
cplcg4	G4	AGAP008447	F: CTTGCCGTCGTCTCGGA R: CAGATAAGCGCCATGGTGA	141	141	60
cyp4h24	H24	AGAP013490	F: GATGTGCTCGGCGTCGATTACC R: CGATAAACGAAACGGGTGGCAGC	121	121	60
control gene						
S7	S7_rtq	AGAP010592	F: AGAACCAGCAGACCACCATC R: GCTGCAAACTTCGGCTATTC	298¥	149	60
cytochrome P450 reductase	CPR_rtq	AGAP000500	F: CCCAGACAGAAACGGAAGTG R: ACGAGTTCTCCACCATCGTC	301¥	207	60

¥ different sizes in gDNA and cDNA due to primers spanning an intron

Table 2.4: Primer details for qPCR assays

for the target genes *cyp6p3*, *cyp6m2*, *cplcg4, cyp4h24* and the control genes s*7* and *cpr*.

2.3.3 Construction of standard curves for plasmid copy number determination

To enable absolute quantification of expression for each of the target and control genes a plasmid containing the targeted fragment was established using pGEM®-T Easy (Promega) or for *s7* pJET 1.2 (Thermo Fisher Scientific). In the case of *cyp6p3* a plasmid pcWompA2 containing an insert of the entire gene was used. All plasmids with one exception contained fragments from *An. gambiae s.s.*; for *cyp4h24* a fragment from *An. arabiensis* was cloned. As *s7* and *cpr* control plasmids were used in *An. arabiensis* experiments also (normalization of *cyp4h24*), *s7* and *cpr* sequences of *An. gambiae* and *An. arabiensis* were compared. Sequences were near identical; no SNPs were detected in the primer binding sites.

The concentration of the plasmid was measured in triplicate using the NanoDrop and then diluted accordingly to a concentration of 1 ng/µl. A ten-fold serial dilution series of each plasmid ranging from 100 pg/µl to 1 fg/µl was used to construct standard curves for each target. Each standard curve was generated by linear regression of the plotted points. From the slope of each curve, the value PCR efficiency E was calculated according to the equation below (Rasmussen 2001) and listed in Table 2.5:

$$E\ (\%)\ =\ \left(10^{-1/slope} - 1\right)\ x\ 100\%$$

The absolute copy number per pg of plasmid DNA was calculated based on the exact size and molecular weight of the plasmid DNA plus insert sequence)Table 2.6:

$$number\ of\ copies\ in\ 1\ ng\ DNA = \frac{1}{mass\ of\ single\ plasmid\ [g] x\ 10^{-9}}$$

$$= \frac{1}{\left(plasmid + insert\ \left[\frac{g}{mol}\right]\right) \Big/ Avogadro's number\ \left[\frac{1}{mol}\right]}$$

experiment	target	efficiency [%]	slope	intercept	r^2
An. gambiae s.s.	***cyp6p3***	100	-3.313	17.390	0.9994
	cyp6m2	102	-3.282	18.663	0.9987
	cplcg4	102	-3.268	19.551	0.9977
	s7	98	-3.357	17.726	0.9942
	cpr	102	-3.265	17.790	0.9987
An. arabiensis	***cyp4h24***	103	-3.249	16.416	0.9977
	s7	115	-3.007	17.067	0.9934
	cpr	102	-3.276	18.884	0.9986

Table 2.5: PCR efficiencies, slopes, intercept and correlation coefficient r^2 of calibration curves for all five qPCR assays.

target	plasmid	insert + plasmid [g/mol]	mass of single plasmid [g]	copies/ng
cyp6p3	pcWompA2	5107528	8.48125E-18	117907175
cyp6m2	pGEM®-T Easy	1935033	3.2132E-18	311216500
cplcg4	pGEM®-T Easy	2320020	3.85248E-18	259572848
s7	pJET 1.2	1929636	3.20424E-18	312086910
cpr	pGEM®-T Easy	1981538	3.29E-18	303912546

Table 2.6: Plasmid sizes and calculated copy number per ng of plasmid DNA.

After calculating all values for the absolute copy numbers for each gene and sample, average values were calculated for the three technical replicates of each sample. The copy numbers of the control genes ((a) ribosomal gene *s7*, accession no. AY380336 (Salazar *et al.* 1993); and (b) *cpr*, *cytochrome P450 reductase*, accession no. AGAP000500) were used to normalize for variation in total cDNA concentration. To normalize the copy number, the value of the target gene is divided by the value of the control gene, as shown below.

$$normalized\ copy\ number = \frac{copy\ number\ of\ target\ gene}{copy\ number\ of\ control\ gene}$$

For each pair of biological replicates the average of the ratios (candidate gene/ control gene) and standard deviations were calculated. To detect statistically significant differences two-tailed t-test were carried out.

2.4 QTL mapping

The genomic DNA of the F_0 parental females, eight F_1 individuals from each family (four females and four males) for *An. gambiae s.s.* and four F_1 individuals per family for *An. arabiensis,* and all F_2 samples was extracted from single mosquitoes using the protocol by Livak (Livak 1984) as described before. Parental females were identified to species level, molecular form (if applicable), genotyped for *kdr* and *ace-1* target-site mutation as well as for chromosomal inversions 2La and 2Rb.

2.4.1 Selection of informative microsatellite and SNP markers

Informative markers were selected on the basis of the genotypes of the parental female F_0 and F_1 individuals. A total of 64 microsatellite loci were screened (51 in *An. gambiae s.s.* and 59 in *An. arabiensis*) (Table 2.7).

To save costs, in initial screens a tailed primer system described by (Oetting *et al.* 1998) was used to label the PCR products with a fluorescent dye. This PCR reaction is performed with three primers: a sequence-specific forward primer with M13-tail at its 5′ end, a sequence-specific reverse primer, and the universal fluorescent-labelled M13 primer. During the reaction cycles the fluorescent labelled forward primer with the M13-sequence is incorporated into the accumulating PCR products. Three different fluorescent dyes were used in total using a specific M13 tail each:

D2-CAC GAC GTT GTA GAA CGA C-3',
D3- GTA GTC GAC AAT CCG TAC G-3'
and D4- ATC GGA CTC GAG CTA AG-3'.

Each PCR reaction was made by mixing 9.7 µl dd H2O, 1.5 µl 10x Taq buffer (1.5 mM $MgCl_2$, 0.2 µl of 10 mM dNTPs, 1 µl of 25mM $MgCl_2$), 0.75 µl of the 10 mM fluorescent labelled primer, 0.325 µl of 10 mM site specific forward and reverse primer, 0.2 µl of Taq Polymerase and 1 µl of gDNA to a final volume of 15 µl. Cycling conditions were as follows: 95℃/5'; (95 ℃ /45''; 57℃ /45''; 72℃ /45'') for 35 cycles; 72℃/ 10'.

5 µl of PCR product was run out on 1.5% agarose gels to confirm successful amplification. The PCR products of up to three loci were then combined and allele sizes determined using capillary gel electrophoresis (Beckman Coulter, Inc., USA). Allele scoring was performed using CEQTM 8000 Genetic Analysis System.

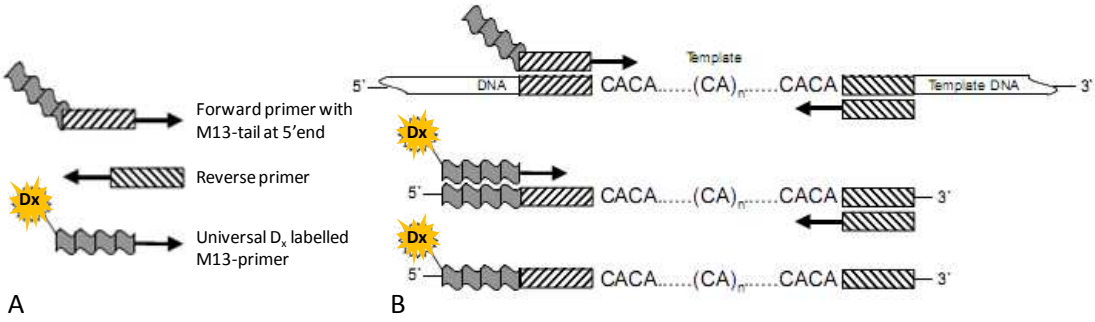


Figure 2.5: Amplification of fluorescent labelled PCR-product via M13 tailing method. (A) represents the three primers used, with the hatched boxes indicating the site specific microsatellite primers, the undulating grey box indicating the M13-tail and the star indicating the Dx-fluorescent label. (B) In the first PCR cycles, the forward primer with the M13-tail is incorporated into the PCR products. These products are then the target for the labelled universal M13-primer, which is incorporated during subsequent cycles. The final product carries the fluorescent label and can be analyzed on a laser detection system (Schuelke 2000).

SNPs were either identified by *ne novo* sequence analysis of amplicons in the various target genes or, existing SNP markers, including the known target-site mutation *kdr*, were tested for informativity based on the genotypes of the parental females F_0 and F_1 individuals. A total of nine different SNP markers were used in the mapping study (5 in *An. gambiae s.s.* and 4 in *An. arabiensis*) (Table 2.8). The *1014 kdr* marker was only informative in one family of *An. gambiae s.s.*, family 17. In order to identify an alternative SNP marker in the sodium channel, sequences for exons 20, exon 29 and an amplicon spanning exons 27-31 were obtained (Macrogen, Korea) and aligned. Primer details and reaction information are listed in Table 2.3.

2.4.2 Genotyping microsatellite and SNP markers

Informative microsatellite markers identified in the initial screen using tailed primers, were re-synthesised with the fluorescent label incorporated directly into the forward primer (unless otherwise stated) (Table 2.7).

Each PCR reaction was made by mixing 11 µl dd H_2O, 1.5 µl 10x Taq buffer (1.5 mM $MgCl_2$), 0.3 µl of 10 mM dNTPs, 0.3 µl of 25mM $MgCl_2$, 0.5 µl each of 10 mM site specific forward and reverse primer, 0.12 µl of Taq Polymerase and 1 µl of gDNA to a final volume of 15 µl. Cycling conditions were as follows: 95℃/5'; (95℃ /45''; 57℃ /45''; 72℃ /45 '') for 35 cycles; 72℃/ 10'. A random subset of PCR products for each primer set was run on 1.5% agarose gels to confirm that the PCR reaction was successful and analyzed as described above.

In *An. gambiae* five SNP markers were genotyped, four by pyrosequencing (kdr, cyp6p3 (P3_1033), cyp6m2 and cplcg4) and one (SODI) through restriction fragment polymorphism. In *An. arabiensis* four SNP markers (cyp4h24, cyp6p3 (P3_520), ex27 and chymotrypsin) were genotyped via pyrosequencing (Table 2.8) and used to complement the microsatellite data (Table 2.7).

						tested in		genotyped in				
						gambiae	*arabiensis*	*gambiae*			*arabiensis*	
	marker	$	distance (vector base)	5'primer (5' --> 3')	3'primer (5' --> 3')			15	17	25	2	3
X Chromosome	cyp4h24*	1	1.5				x				x	x
	503	5	1.8	AGGTTAGAGTGAGCAACCAC	GCACTGCATCTCTCCAATAC	x						
	36	5	1.9	CGTATGTTTGCTAGGGGTGG	GTCAAGAAATGGGCCACAGG		x					
	71	5	3.1	GCGGAGTTATTTCCTGAACC	ACAGGCCAAGCAAATGCAGG	x	x					
	53	5	3.6	GTTTCGGGGCTTGAGAAGTG	CTTCACGTGGCTTTGCTGTG	x	x		x		x	x
	1002	5	5.5	GATCGGTATATGCTTCCCGC	AATAAGCCACGGCGTATCCC		x					
	784	5	6.9	TGGTGAAAGAACAGACCCCG	TGTAACGGGCAAGAAAAGC	x	x					
	80	5	7.0	TGCTCTCTCCTACATCGAGG	GCCAGTGCTCTAGATTAACG	x	x		x		x	x
	99	5	8.4	CGGGAATTTGTTGCTTCCTG	TCGCCCTCTTTCTCCATCTC	x	x					
	49	5	9.3	CAGCGCCTCCATATAGAACG	GATCATTCAGCTGAACCTGC		x					
	711	5	9.9	CCCACAGCAAAACGAGAATG	GACAACTTGCATTTCACTATG		x					
	7	5	11.2	CACGATGGTTTTCGGTGTGG	ATTTGAGCTCTCCCGGGTG	x	x				x	x
	766	5	15.3	CAGGTAGTAGGAGTAGATGC	AATTATGAGCACGGTGGGTG		x					
	678	5	19.2	CCTCTCCCCAGAATCGGTAC	AAGAGCAGAAACAACCGCAG	x	x				x	x
Chromosome 2R	46	5	1.2	CGCCCATAGACAACGAAAGG	TGTACAGCTGCAGAACGAGC	x	x			x		
	803	5	1.9	CTCGATAAATCCCGTCGGTG	GTCGGTTTGAGGTTGTAAAGC	x	x	x	x		x	x
	199	3	3.3	CGATTGCAAGCAGTAAGTCG	CCGGAACCATTTATCATCTCC		x				x	x
	24	3	3.5	TGCCGTATTTCAATGTCAGC	ACAACCACCACCCTAACTGC		x				x	x
	175	5	6.6	AGGAGCTGCATAATTCACGC	AGAAGCATTGCCCGCATTCC	x						

	197	5	10.4	TACCTCTGTGTTCGGTTTCC	GGTGGTATGGCGATGGAAGG	x	x	x	x			
	757	5	14.7	TGATCGCGCCCAATCAATCC	ATCGATCGTACAGATGTGCC	x	x	x	x	x	x	x
	799	5	14.8	TTATGGGCAACTGCGGATGG	CGTGCGTTTGATACATCTACG	x	x			x	x	x
	187	5	17.8	CCGGAGCAGAGATAAACAGC	CACAGACGTACACCTAATGC	x						
	26	5	24.6	GGTTCCTGTTACTTCCTGCC	CCGGCAACACAAACAATCGG		x					x
	60	5	25.0	TGTTTGGGACGGAACCGAG	TCTCGTGACGGATGATACTG		x					
	2R_si_5	3	24.9	TTCTCGAAAGACTGCTGCTG	ATTGGATCGAAAACGGTCTG	x	x	x	x	x		
	P3_520 *	1	28.5				x				x	x
	P3_1033*	1	28.5			x	x	x	x	x		
	590	5	28.9	CGGGAAAGCGAAGTGTACGA	TGCGGCTGGTGAACATTTTC	x	x				x	x
	cyp304c1	3	32.0	CATCATTAACGGGCTCGACT	AGCGTTAGGAAGAGTGCATTG		x				x	x
	720	5	33.2	ATTAGAATCCAAACCAGCGG	ATAAGCTAATGCGCTGCTCC		x				x	
	135	5	34.4	TCATGCACTGTTTGCTCGGC	CTGCCCCATTCAATTGCAGC		x					
	770	5	40.4	CAAGATGGAGGCGCATGATC	GCGTTCCATCGAAATCAGAC	x	x		x	x	x	
	117	5	42.7	CGGAACGCACGGAACAATTG	CGTTGCAGATTTCCCAAACG	x	x		x	x		x
	125	5	45.7	AGGAGCATAACACATCGCCC	CGCTCGTCAAAGAAACTGGC		x					
	786	5	48.3	TGTGAAGCATTTCCTTGGCG	TGCCCTTGAGTCGAGGTAGC	x	x				x	x
	1	5	53.7	CTTTTACACCGAGGGAAAG	CGACCGTACACATAAACAC	x						
Chromosome 2L	325	5	16.0	CCGGTGTCCGTGTTG	GCGCGAAAGCAAATGACACG	x	x					
	kdr *	4	24.2			x	x		x			
	ex27 *	1	24.2				x				x	
	637	5	25.5	TCGAAATGTATGCGAAATGCAG	CCTTCTTTCCTCGATGCATTCC	x	x	x				
	1012	5	25.5	AGTGTTCAGAGCGGGAAAAG	GTACAACCCGCAGGAGAAAC	x						
	787	5	26.8	CGGGTCAAAGAAAACTCAGC	GCATAAGAACGGCACATTGC	x	x		x	x	x	x

	772	5	42.0	TACAGCTGTTTGGGAGTTGG	GGGTCGGCTTTTATTTCCTCG	x	x			x		x
	603	5	42.0	TGCACCGTTGATGCACATGC	GTGGACGATGTGAAAGATAAG	x	x		x		x	x
	675	5	46.9	CGTGACACTTTCAGGACACC	GGCAAAAGGCTGGAAAACCG	x	x	x		x	x	x
Chromosome 3R	776	5	2.5	TGCGGATCATAATCGAGTCC	TCACAAACACGCAACGAGTC	x	x	x	x	x	x	x
	746	5	3.8	TGGGTTCGAAATTCGCCAAC	GACGTGTGCACCCGTTGTG	x	x					
	59	5	4.2	CCCCTATTAAACCCTGGACG	TGTTGTTGCCCTGCGTTACC	x	x	x	x	x	x	x
	812	5	6.2	CTGGCCCATTTTGCATATGC	TGCTCCACCCAAACCACATC	x	x	x	x	x	x	x
	cyp6m2 *	1	6.9			x	x		x	x		
	249	5	8.8	ATGTTCCGCACTTCCGACAC	GCGAGCTACAACAATGGAGC	x	x	x	x	x	x	x
	30C1	2	9.6	GCCAAAAGATTCATTCGCTCG	GCCAAAAGATTCATTCGCTCG	x	x	x	x	x	x	x
	cplcg3 *	1	10.9			x	x	x	x	x		
	6M19	2	13.1	CACAAGGTCACACGCTGAAG	AACACAACACCGAGCTTGC	x	x	x	x	x	x	x
	119	5	14.8	GGTTGATGCTGAAGAGTGGG	ATGCCAGCGGATACGATTCG	x	x	x		x	x	x
	555	5	21.3	GCAGAGACACTTTCCGAAAC	TGTCAACCCACATTTTGCGC	x	x	x				
	158	5	23.0	CTGGCACGATCAATCAATCG	ACGATGGTGTACACGTAACG	x	x			x		
	341	5	23.0	CCCAAAGCAATGAACCTCGC	AGTAGAAGAAGAGGGCAGCG	x	x		x	x		
	6H1	2	25.5	CCTGCTCCACCGGGTTCAAATG	GATAATGCCACCGGTAATGC	x	x	x	x	x		
	33C1	3	25.6	ATGAAACACCACGCTCTCGG	TTGCGCAACAAAAGCCCACG	x	x		x			
	6F5	2	31.6	TCTCTTTTCGCACGCTCTCG	GTTCTTCCTTCCCTTTCCCTTTTA	x	x		x	x		
	88	5	34.7	TGCGGCGGTAAAGCATCAAC	CCGGTAACACTGCGCCGAC	x	x			x		
	30L17	2	41.1	CACCCATTTTCAGCTTGTTCTTC	ATTACGCATGGATATTTTTGTGT	x	x		x	x		
	chymo-trypsin*	1	43.9				x					x

Chromosome 3L	SODI *	3	5.6			x			x			
	811	5	14.8	AACCCACAGTACAGCTCGCG	GTTGCTGCATACTAACCTCG	x	x		x	x	x	x
	127	5	16.9	CCTCTAACTCGATTACCGTG	GTCAGGGAATTGGAAAGAGC	x	x					
	577	3	19.8	TTCAGCTTCAGGTTGGTCTC	GGGTTTTTTGGCTGCGACTG	x	x	x	x		x	
	817	5	31.9	ACTGGTCCGTTGCTGCGCG	ATGAGTGAATGGTGCGCTGG	x	x		x	x		
	242	5	37.1	TTCATTTCCACCGCAGCTGC	GGCGACACTCAATCCTTCC	x	x					
	46C3	2	40.4	AGTCGGCAGGTTCTTTCGGTG	GAAGCTAGAAGCGGGAACAC	x	x				x	x
	46C2	2	41.4	CTGTGGCAGGTGGAATGGAATG	GCCCTTTTTGGAGTGACCTCG	x	x		x	x	x	x
					sum all	**56**	**67**	**18**	**30**	**28**	**31**	**31**
					sum microsatellites	**51**	**59**	**16**	**25**	**25**	**28**	**27**

* SNP marker (details Table 2.8)

$ source 1, C. Witzig (LSTM); 2, H. Ranson (LSTM); 3, D. Weetman (LSTM); 4, Wondji Charles (LSTM); 5, Zheng *et al.* 1996

Table 2.7: Details of all microsatellite markers tested for informativity in the F_1 progeny of the *An. gambiae* and *An. arabiensis* families.

Those markers that proved informative were subsequently genotyped in the F_2 progeny of the mapping populations *An. gambiae* and *An. arabiensis.* The table also includes the physical distances in Mb according to vector base and the names of the genotyped SNP markers (details in Table 2.8) in the order they share with the microsatellite markers.

Markers	primers 5'-3', forward, reverse (biotynilated) and sequence	Product size	alleles ¥	MgCl [mM]	Ta [°C]
X chromosome					
cyp4h24	F_GCGCACCTACTGGTACACGA R_bio_ACACTTCCGAGCTCCTTTCC S_GAAGCTGATCCGACG	128 bp	G/A syn	1.75	56
Chromosome 2					
cyp6p3 (P3_520)	F_CTTACGCCGACGTTCACCT R_bio_GTCGTAAAACGGCCCAGGA S_GCAAAGCTACAACCAA	145 bp	G/T syn	2.3	63
cyp6p3 (P3_1033)	F_CTGGCCGCGTTCATCTTC R_bio_CCTTCGCATGCCCGAACA S_GCATAACTATTGGAAGGA	124 bp	C/A non-syn	2.3	64
kdr	F_TTGTGTTCCGTGTGCTATGC R_bio_AAAAACGATCTTGGTCCATGT S_TGTAGTGATAGGAAAT	154 bp	A/T non-syn	2.75	50
ex27	F_TGAACGATGCTATTGATTCTAGAGA R_bio_TACCGACCTGAGCATTTAATCAT S_AAAACATGTTGTAAACAATT	112 bp	C/T syn	2.0	55
Chromosome 3					
cyp6m2 (M2.2_301)	F_TACGATGACAACAAGGGCAAGC R_bio_CGCGATCGTGGAAGTACTGG S_TCAGTTCGTGAAGCC	130 bp	C/T syn	2.3	50
cplcg4 (G3_1845)	F_ACCAAACACCACCCAACTCAATC R_bio_ACGGAAGCACTCCAGCCT S_GACACAACACTCTCACCT	171 bp	T/A/C Intron	2.3	64
chymotrypsin	F_CCGGGTTGACCGATAGTACATATA R_bio_GCAGCCTTGCGTATTCAAATG S_GGGTTGACCGATAGTACA	111 bp	T/C syn	1.75	56
SODI * (AGAP010517; Variant ID rs3582365	F_GCATCGACGTAGTTGGGTCT R_TCATGGAGGTACGCAAGATG	638 bp	A/G PstI syn	1.95	58

¥ syn= synonymous mutation; non-syn= non synonymous mutation; Intron= SNP in Intron

Table 2.8: Details of SNP markers scored via pyrosequencing or RFLP *

in the F_2 progeny of the mapping populations *An. gambiae* and *An. arabiensis.*

2.4.3 Software used for marker data analysis, linkage mapping, and LOD scoring

The JoinMap 2.0 software package (Stam & van Ooijen 1996) was used to build genetic linkage maps for each individual family and for combined genotyping data. Genotype data for each marker were tested for conformity to Mendelian ratio (Hardy-Weinberg equilibrium) with a $\chi 2$ goodness-of-fit analysis using the JoinMap single locus analysis module (JMSLA) procedure. A $\chi 2$ test for genotype/ phenotype association was performed using Haploview 4.2 (Barrett *et al.* 2005) (full data in Appendix 9.1).

To build the genetic maps, loci were separated into linkage groups using the linkage group assignment and data file splitting modules (JMGRP and JMSPL) procedures with minimum and maximum LOD thresholds of 1.0 and 4.0 respectively and LOD increments of 0.1. The JoinMap recombination estimation module (JMREC) program estimated pairwise centi Morgan (cM) distances between all pairs of informative loci in each linkage group and the map construction module (JMMAP) program estimated the maximum likelihood map using the Kosambi distances. MapChart 2.1 software (Voorrips 2002) was then used to plot the genetic maps.

The JoinMap linkage positions along with the genotype/phenotype data were entered into Windows QTL Cartographer 2.5 (Wang *et al.* 2005). Interval mapping (IM) (Lander & Botstein 1989), composite interval mapping (CIM) (Zeng 1993) and Multiple Interval Mapping (MIM) (Zeng *et al.* 1999) procedures were performed for each family separately and for the combined data. The optimum LOD thresholds were estimated by permutation of trait and marker data 500 times with a walking speed of 1 cM. For Multiple Interval Mapping (MIM) an initial model was created by forward and backward selection on markers with criteria set to probability of partial R^2 of 0.05 for *An. gambiae* and 0.01 for *An. arabiensis*. The model was refined by optimizing QTL positions, searching for new and testing for existing QTLs. The MIM model estimates QTLs and gives map position in centimorgan, nearest marker, additive and dominance effects for single QTLs as well as the entire model. It gives an estimate of the genetic variance expressed as percentage of the phenotypic variance explained by the respective QTL.

SigmaPlot 11.0 (Systat Software Inc.) was used to plot LOD graphs, and MapChart 2.1 software (Voorrips 2002) was used to plot the LOD graphs next to the respective linkage maps for final visualization.

Chapter 3. Preliminary studies on permethrin resistance in *Anopheles gambiae s.s.* from Benin, West Africa

3.1 Introduction

The Republic of Benin is a country in Western Africa. It borders Togo to the west, Nigeria to the east and Burkina Faso and Niger to the north. Its official capital is Porto-Novo in the south near the Atlantic coast; however the seat of government and economical capital is the city of Cotonou. The country's main income sources are subsistence agriculture of fruit and vegetables (prevailing in the south) or cotton farming for export (prevailing in the more arid climate of the north).

Benin covers three climatic zones. From south to north these regions are a Guinean bioclimatic zone, an intermediate zone with Sudano-Guinean climate and a Sudanian semi-arid bioclimatic zone in the north. As the bioclimatic characteristics change within the country, so do the main agricultural practices as well as species distribution of *An. gambiae*. The M-form is generally more prevalent in the Northern Guinean zone with the S-form predominating in the Southern regions. *An. arabiensis* is mainly located in the central regions (Djogbénou *et al.* 2008).

Akron is an agricultural part of Porto-Novo, located on the outskirts of town (Figure 3.1).

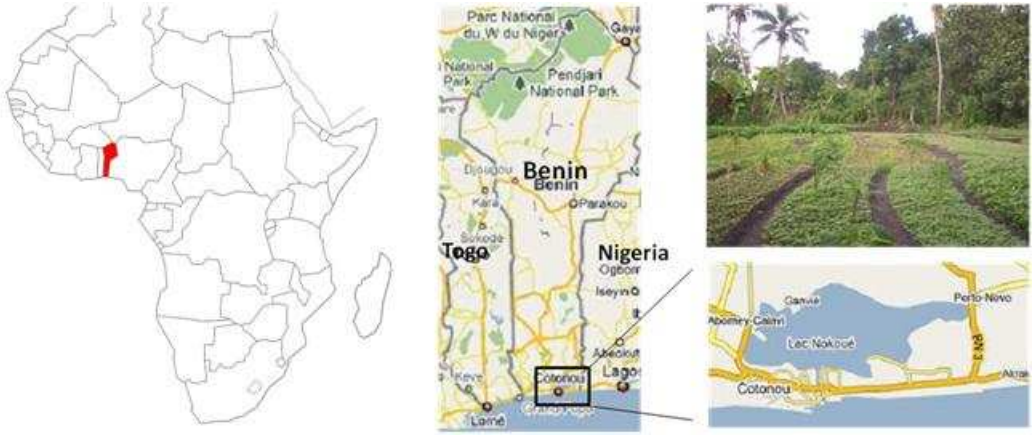


Figure 3.1: Map of Africa and Benin showing location of the collection site Akron. The photo shows a representative view of one of the vegetable fields in Akron near Lake Nokoué. Maps from Google Maps and picture Claudia Witzig.

In Akron, former swamp land is used for growing vegetables. This makes the area a permanent breeding site for mosquitoes. The mosquitoes breed in puddles that build on the paths between the single vegetable fields. The heavy use of agricultural pesticides in this area puts the mosquitoes breeding at this site under strong selection pressure, and has resulted in the emergence of insecticide resistant populations (Corbel *et al.* 2007; Djouaka 2008; Djouaka *et al.* 2008).

Two recent studies clearly confirm the massive and uncontrolled use of pesticides (sometimes unregistered ones) in vegetable farming in Southern Benin (Akogbéto *et al.* 2005; Yadouleton *et al.* 2009) suggests the overuse of pesticides to contribute to the widespread emergence of insecticide resistance in *Anopheles*.

A countrywide survey on insecticide resistance in *Anopheles gambiae s.l.* in Benin provides an overview of the resistance status of this malaria vector (Djogbénou *et al.* 2010). Studies were carried out during 2006-2007, before extensive vector control was undertaken. Bioassays were conducted on field collected mosquitoes from 18 localities using DDT, permethrin, carbosulfan and chlorpyrifos-methyl (belonging to the four chemical classes used in public health, DDT, pyrethroids, carbamates and organophosphates respectively) (Figure 3.2). Resistance to DDT and permethrin was widely distributed in *An. gambiae* in Benin, but stronger in the most southern climatic zone, whereas resistance to carbosulfan and chlorpyrifos-methyl was limited to a few localities in the more northern zones. Resistance to DDT is very prevalent with fully resistant populations in 10 locations, reduced susceptibility in seven locations and only one location that shows full susceptibility (mortality <80% counted as resistant; mortality 80-97% likely resistant and mortality above 98% fully susceptible according to WHO guidelines (WHO 1998)).

Resistance to permethrin is prevalent in the very south of the country with no fully susceptible population tested in the Guinean area. Across the whole of Benin five out of 18 localities showed full susceptibility to permethrin, in 10 locations likely resistance was suspected and three

populations were resistant (including Pahou and Cotonou in the very south).

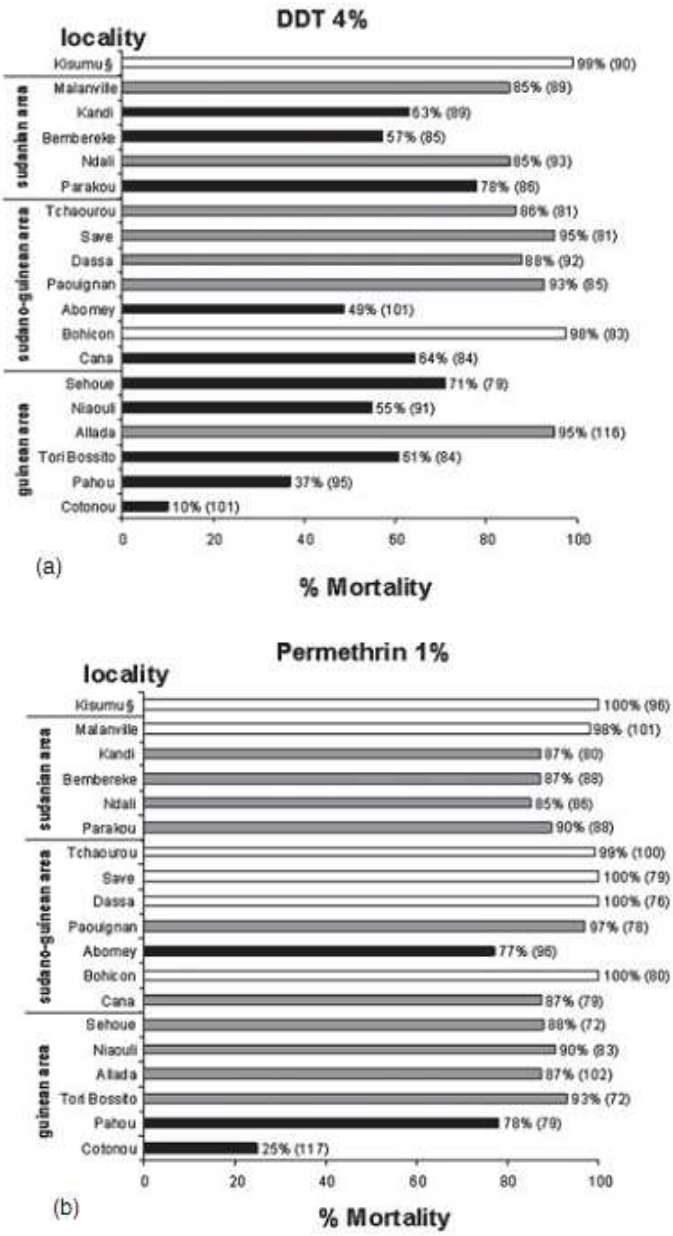


Figure 3.2: Resistance status of *An. gambiae s.l.* in Benin.

Mortality rates and susceptibility/ resistance status of different *Anopheles gambiae s.l.* populations from Benin after 1 h of exposure to (a) DDT 4% and (b) permethrin 1 % (from (Djogbénou *et al.* 2010).

In 2006-07 the 1014F *kdr* allele was present in the M-form in all localities of the Guinean zone, with average frequency of 0.5 and with the highest frequency 0.9 observed in Cotonou. The 1014F *kdr* allele was present in all S samples at frequencies between 0.1 and 0.9. Recent data from a TDR (Special Programme for Research and Training in Tropical Diseases) sponsored study in Benin found an increase in the frequency of *kdr* mutations in the M-form. The 1014F allelic frequency in the M-form in

2010 ranged from 0.47 in Malanville in the North to 0.9 in Cotonou in the South. A small number of *An. arabiensis* with either the 1014F or 1014S allele were also found in Malanville in 2010 (Djegbe *et al.* 2011).

The *ace-1*^{R} allele was not detected in any of the *An. arabiensis* samples, or in the M-form of *An. gambiae s.s.*, and was present at low frequencies in most S-samples in 2006-7 (allele frequency 0.01-0.17).

A number of other studies looked at pyrethroid resistance in *An. gambiae* in Benin to assess the resistance status (Yadouleton *et al.* 2010) study its effects on the efficacy of ITN and IRS (Akogbéto & Yakoubou 1999; N'Guessan *et al.* 2007) or examine the mosquitoes' behaviour when challenged with ITNs (Corbel *et al.* 2004). All of these studies used *kdr* frequency as a marker of resistance. N'Guessan and colleagues (N'Guessan *et al.* 2007) were able to show a loss of efficacy of ITNs and IRS associated with pyrethroid resistance near Cotonou (*kdr* frequency in field population 0.83 and resistant in WHO bioassays).

Another study in 2008 (Djouaka *et al.* 2008) focussed on permethrin resistance in field populations of Southern Benin. Populations of *An. gambiae* from four different locations were collected, bioassays using permethrin were carried out, and *kdr*-frequencies were assessed. High frequencies of *kdr* (0.83-0.86) were found in the resistant populations. This study also included a microarray analysis using the *An. gambiae* detox chip (David *et al.* 2005) to identify genes associated with metabolic resistance to permethrin in M-form *An. gambiae* from Akron (Djouaka *et al.* 2008). Three P450 genes including *cyp6p3* and *cyp6m2* that have been found over expressed in pyrethroid resistant populations previously (Müller *et al.* 2008b) and shown to metabolise pyrethroids *in vitro* (Müller *et al.* 2008b; Stevenson *et al.* 2011) were overexpressed in the Akron population (Figure 3.3).

The study by Djouaka *et al.* also identified two cuticular precursor genes (*cplcg3* and *cplcg4*). Those same two genes were similarly found to be overexpressed in permethrin resistant *An. gambiae* samples from neighbouring Nigeria (Awolola *et al.* 2009).

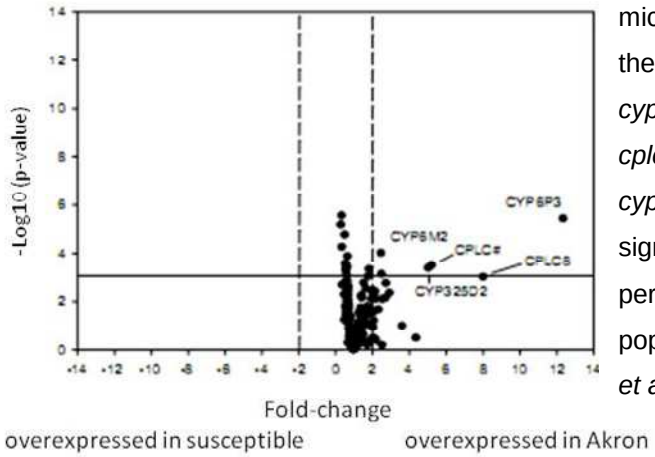


Figure 3.3: Volcano plot of microarray experiment showing the P450 genes *cyp6p3* and *cyp6m2* and the cuticular genes *cplcg3* and *cplcg4*, as well as *cyp325d2* (not further mentioned) significantly overexpressed in the permethrin resistant field population from Akron (Djouaka *et al.* 2008)**.**

Further evidence for metabolic resistance to pyrethroids in Benin comes from biochemical assays on four populations in Benin which show elevated levels of enzyme activity (mixed function oxidase, non-specific esterases and glutathione-S-transferases) (Corbel *et al.* 2007). Synergist assays with the monooxygenase inhibitor piperonyl-butoxide also support a role for elevated P450s in conferring pyrethroid resistance in Benin (Awolola *et al.* 2009; Djegbe *et al.* 2011).

3.2 Aims of the chapter

1. Collect field specimens of *An. gambiae s.s.* from Akron in Southern Benin, an area where this species is known to be resistant to pyrethroid insecticides.
2. Confirm the association of candidate detoxification and cuticular genes with permethrin resistance in field collected populations and laboratory colonised strain of *An. gambiae s.s.* from Akron through quantitative RT-qPCR.
3. Identify sequence polymorphisms in candidate resistance genes and develop assays to genotype these loci.

3.3 Materials and Methods

3.3.1 Lab colonies and Akron field samples

The characteristics of the Akron and Ngousso lab colonies are described in Chapter 2.1.1.1. Details of field collections (adults and larvae), rearing conditions and conditions for bioassays are described in Chapter 2.1.1.2.

3.3.2 Sequencing of candidate genes in Akron field samples

The four candidate genes identified in earlier microarray experiments, *cyp6p3*, *cyp6m2*, *cplcg3* and *cplcg4,* were amplified from six individual *An. gambiae* M-form samples, collected by Dr. Rousseau Djouaka, IITA, Benin in 2008 Table 3.1. These samples include three each of survivors and non-survivors after a 1 hour exposure to 0.75 % permethrin. Details of the primer sequences and amplicon sizes are given in Table 3.1 and Figure 3.4 shows the positions of these primers.

For the P450 genes, amplicons P3-1, P3-2, P3-3, P3-4, M2-1, M2-2, M2-3, M2-A, M2-B, M2-D and M2-E were sequenced (Macrogen, Korea) from the PCR template. Amplicons P3-1.5 and M2-t were cloned into the pGEM®-T Easy vector (Promega) and transformed into *E. coli* DH5α™ Competent Cells (Invitrogen) for amplification. Cells were harvested and the plasmid DNA recovered (QIAprep® Spin Miniprep Kit, Qiagen). The plasmid DNA was used as template for the sequencing analysis. Primers and PCR conditions for amplification are listed in Table 3.5.

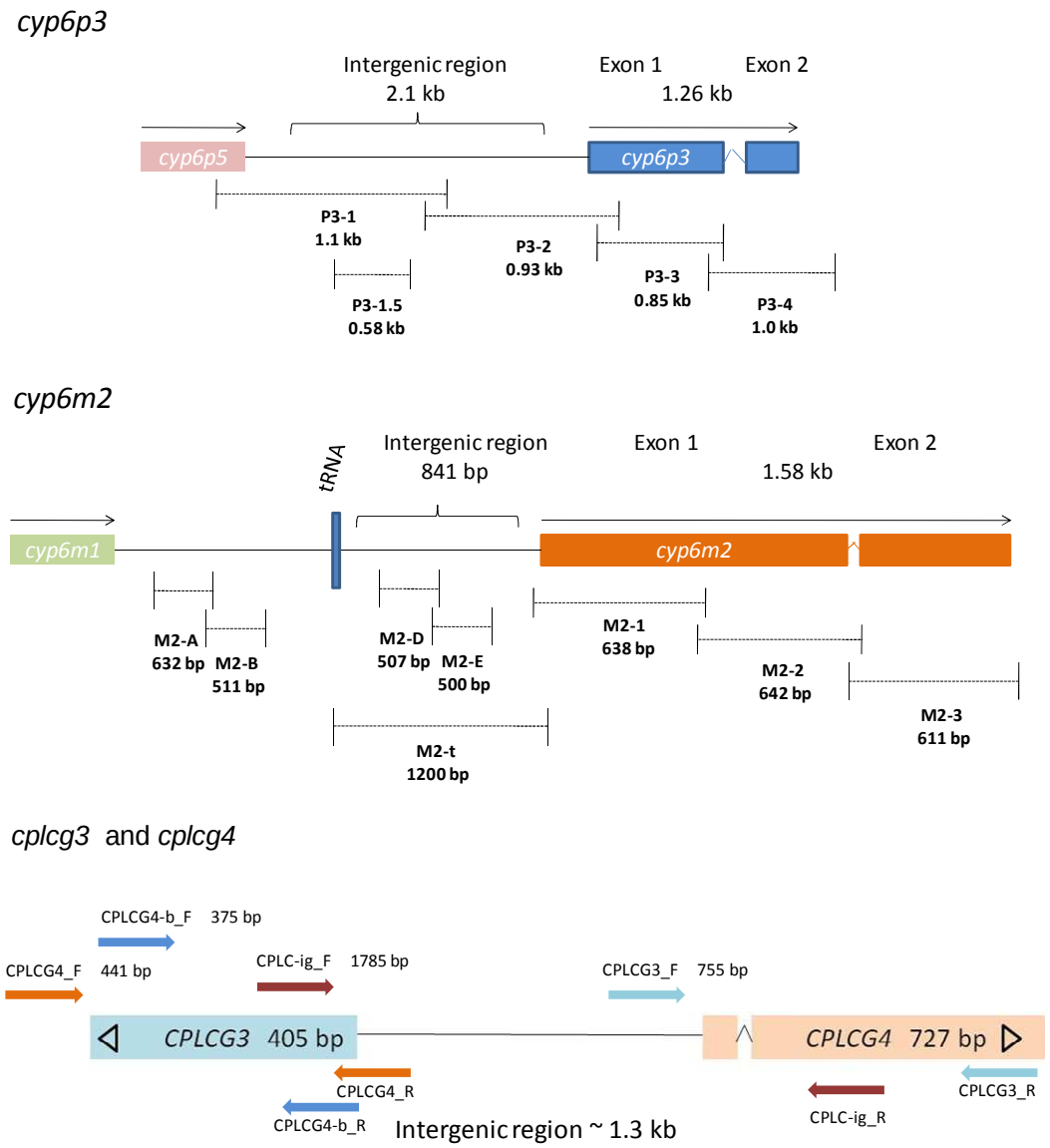


Figure 3.4: Schematic overview of genes *cyp6p3* and *cyp6m2*, as well as *cplcg3* with *cplcg4* indicating introns, UTRs and primer positions. Distances are not to scale.

For the cuticular genes, amplicons G3 and G4 were sequenced (Macrogen, Korea) from the PCR templates. Amplicon ig (for the intergenic region between G3 and G4) was cloned into the pGEM Teasy vector (Promega, Madison, USA) and transformed into *E.coli* DH5α. Cells were harvested and the plasmid DNA recovered (QIAprep® Spin Miniprep Kit, Qiagen). The plasmid DNA was used as template for the sequencing

analysis. Primers and PCR conditions for amplification are listed in Table 3.1.

The DNA sequences for each amplicon were aligned and single nucleotide polymorphisms (SNPs) were identified using SeqMan ^{TM} II sequence analysis software from the Lasergene package (DNASTAR, Madison, USA).

Based on the sequence data pyrosequencing assays were developed for 14 SNPs in the three candidate genes (nine from *cyp6p3*, one from *cyp6m2*, three for *cplcg4* and one for *cplcg3*, Table 3.2) and used to genotype a larger set of 69 permethrin phenotyped field samples collected from Akron in 2008 (Dr. Rousseau Djouaka, IITA). The status of the 1014F allele was determined using another pyrosequencing assay (Table 3.2). PCR reactions to generate the template for the pyrosequencing reaction were performed at the following conditions: 95°C/5'; (95°C/20''; T_a°C/30''; 72°C/20'') for 40 cycles; 72°C/5'.

Test statistics were performed using Haploview 4.2 (Barrett *et al.* 2005). Each polymorphism was tested for by performing a Pearson's chi-squared (χ^2) test, using the observed genotype frequencies obtained from the alive samples to compare them to the observed genotype frequencies in the dead samples.

PCR amplicon	primer name	forward primer 5'-3'	reverse primer 5'-3'	prod. size [bp]	MgCl2 conc. [mM]	annealing temp. [°C]
P3-1	cyp6p3-1	CGATCCGAAGATGTTTACGC	TAGTCATCGGGTGGCGATA	1100	1.75	59
P3-2	cyp6p3-2	TGACCTCTCTGCCAACATCA	CCGAACATTTGCTTCATTCG	930	1.75	59
P3-1.5	cyp6p3_1.5	TGGCTTAAAGAGCTGATTCG	TGAATTCGTTGAGATGTTCCT	582	2.5	56
P3-3	cyp6p3-3	CGGTACGGCGGCATTAG	GTCACCTCTCCACCGTTCTC	852	1.5	59
P3-4	cyp6p3-4	GCTGCAGATCAAGAACAAGG	GGTCGCCAATTGCAGTATCT	1007	2.5	59
M2-1	cyp6m2-1	TCGGTGGACAGTCAAATCAA	GTTACACTCAATGCCGAACG	638	2.5	63
M2-2	cyp6m2-2	GTTCGAGCTGAAGGATCTGC	CAATCCGGGTAAACAAGGTT	642	2.5	65
M2-3	cyp6m2-3	GAAATCCTGCAAAAGCACAA	CTATACAGATCTTAACAGCTAAGTGA	611	3.5	58
M2-A	M2-A	CATACACAAAATGGCCAGAAG	CACGATGTTTCCTCCTGTAGAC	632	1.75	55
M2-B	M2-B	CTGTCGATCCTGGAGAAGC	CGTGTTACTATGCAACATCTCC	511	1.75	60
M2-D	M2-D	TCGTTTCACAGCACGTTAGC	GCTATGAGTCGTTTGCACAA	507	3.5	55
M2-E	M2-E	CACCCTTCTCTAACCCTCCTC	CAGAAATGTGAAATCCAACAAGC	500	3.5	55
M2-t	M2-t	GCTCGAAGCTACGACCCT	CTTAAACGGCACGGTAATGT	1200	2.5	61
G3	CPLCG3	GACGTTGGACGGTGCAG	CCTTTATTCAAGCACTTTGCG	755	1.5	56
G4	CPLCG4_b	CCACAGTGCATGGTAGCTG	CTGCTGCTGGGAGACG	375	1.5	62
ig	CPLC-ig	CGACATGCACAGCACCTC	CAACAACGGCGACCTGT	1785	1.5	66
	CPLC-ig-gap	GGCGACCGATCAGATACC (sequencing primer only)				

Table 3.1: Details of primers and PCR conditions used to amplify the candidate genes from *An. gambiae*.

assay name	gene		sequence 5'-3'	length of template and sequence to analyse	$MgCl_2$ [mM]	T_A [°C]
CYP6P3_787	*cyp6p3*	forward	CGCCACCCGATGACTAGTACAA	107 bp	2.3	61
		reverse	biotin_GGCCAGCAGATAAATCAGCA			
		sequencing 787	GGAACATCTCAACGAAT	TCAC C/T GCA		
		sequencing 812+814	CACACACAAACACACTGA	CA A/G C T/G/A TG		
CYP6P3_926	*cyp6p3*	forward	CTCAGTTTGTCCCGGTGGC	60 bp	3.0	62
		reverse	biotin-CCAGCACCGCGTTAATTAGCTC			
		sequencing	GTTTGTCCCGGTGGC	C/T CGAAT		
CYP6P3_1033	*cyp6p3*	forward	CTGGCCGCGTTCATCTTC	124 bp	2.3	64
		reverse	biotin_CCTTCGCATGCCCGAACA			
		sequencing	GCATAACTATTGGAAGGA	C/A AATGG G/A TTT		
CYP6P3_1067	*cyp6p3*	forward	CAGTGTCGATCGTGTATCTGTTCA	158 bp	3.3	62
		reverse	biotin_GTACAGCTCCTGATGGATGTCG			
		sequencing	CCCGAATCCACACTT	T C/T TGTTCG GGCATGCGAA		
CYP6P3_1211	*cyp6p3*	forward	AGTTTATCGTGCCCTCGGT	208 bp	2.3	66
		reverse	biotin_ACGTCGGCGTAAGCTTCTG			
		sequencing 1211	CGGTCCTGGTGATCG	A C/T CC A/G/C GAG		
		sequencing 1229	CTGAGCTGGCGAAGA	CGAT T/C CT		
		sequencing 1280	CGGCGTGTTTACCAAT	GC A/T AAGGA		

Cyp6M2.2_301	*cyp6m2*	forward	TACGATGACAACAAGGGCAAGC	130 bp	2.3	65
		reverse	biotin_CGCGATCGTGGAAGTACTGG			
		sequencing	CAGTTCGTGAAGCCG	GGTGGCACTGAT C/T		
G3_1602	*cplcg4*	forward	CTTCGTACCATGTCCTTCCTG	124 bp	3.0	61
		reverse	biotin_ CTATTCCGAGCGTTTGGTATG			
		sequencing 1602	GTACCATGTCCTTCCTG	GAGT T/G GGAC C/T ACG G/C		
		sequencing 1651	CTCGGGAGCTGACGT	G/T GACGG T/A CAGG		
G3_1845	*cplcg4*	forward	ACCAAACACCACCCAACTCAATC	171 bp	2.3	64
		reverse	biotin_ACGGAAGCACTCCAGCCT			
		sequencing	GACACAACACTCTCACCT	T T/A/C GA A/C CGCA T/C TAAC G/A CT		
G4_540	*cplcg3*	forward	TCCTGCCGTACGGTGGCT	80 bp	1.5	66
		reverse	biotin_ GGGTGGGCCAGCACATTC			
		sequencing	GGCCCTATGGTCATCT	GGC C/T GCC		

Table 3.2: Details of pyrosequencing primers and PCR conditions; SNPs underlined are the ones genotyped by the respective assays.

3.3.3 qPCR on candidate genes *cyp6p3*, *cyp6m2* and *cplcg4*

3.3.3.1 RNA extraction and cDNA generation

Various mosquito samples were used for the qPCR experiment. These included samples from the Akron colony (permethrin resistant; selected or unselected), Ngousso colony (susceptible) and field samples raised from larval collections. RNA was extracted from pools of three 3-day-old, unmated female mosquitoes using the Trizol extraction method (Chapter 2.3.1). All mosquito samples with the exception of the Benin field samples, were unexposed specimens, snap frozen and stored at -80°C. The Benin field samples are mosquitoes that survived a 1 hour exposure to 0.75% permethrin and after a recovery period of 24 hours were snap frozen and stored in RNA*later*® (Ambion, Applied Biosystems) before RNA extraction.

Materials and Methods for the quantitative PCR were as described in Chapter 2.3. Absolute quantification using standard curves generated with plasmid DNA and data from two control genes, *cpr* and *s7*, was performed.

3.4 Results

3.4.1 Akron field samples

A total number of 74 adult females were collected resting inside houses in Akron, Benin. DNA was extracted and assessed for species ID and molecular form. All specimens were *An. gambiae s.s.* M-form. The 74 specimens were genotyped for *kdr*, and 32 of those for the *ace-1* target-site mutation using a pyrosequencing assay. The 1014F *kdr* frequency was very high (0.92), but the *ace-1* mutation and 1014S *kdr* allele were not detected.

For twelve females exon 20 (including the *kdr* locus at position 1014) and exon 29 of the voltage gated sodium channel (encompassing the F1534C mutation involved in insecticide resistance in *Ae. aegypti* (Harris *et al.* 2010) and *Anopheles* species, N1575Y recently described in *An. gambiae* (Jones *et al.* 2012b)), were sequenced. The 1014 *kdr* genotype was confirmed and no other polymorphisms detected.

Field collected larvae were reared to adults and exposed to permethrin (0.75%,), bendiocarb (0.1%,) or DDT (4%) for one hour. Twenty four hour mortality is shown in Table 3.3.

exposure 1 hour	n	dead	alive	mortality [%]
Permethrin 0.75%	82	43	39	50.6
DDT 4%	60	0	60	0.0
Bendiocarb 0.1%	96	96	0	100.0

Table 3.3: Summary of bioassay results from larval collections in September 2009 in Southern Benin, West Africa.

3.4.2 Akron lab strain

Selecting the original parental Akron strain obtained from MR4 with the insecticide permethrin resulted in an increase in resistance in this strain (Figure 3.5) with a final mortality of approximately 17 % after 1 hour exposure to permethrin 0.75%.

The carbamate resistance phenotype in Pakron was stable at 0% mortality after 1 hour exposure to 0.1% bendiocarb. Selections using bendiocarb were performed at generations F_7 and F_{10} and the phenotype was confirmed through bioassays at generation F_{21}.

Table 3.4 summarizes frequencies of molecular form and the target-site mutations L1014F/S and *ace-1* in various generations of the Akron/ Pakron lab colony. When received from MR4 the Akron colony contained a small proportion of S-form (0.11). This was later confirmed by Paul Howells at MR4 (personal communication).

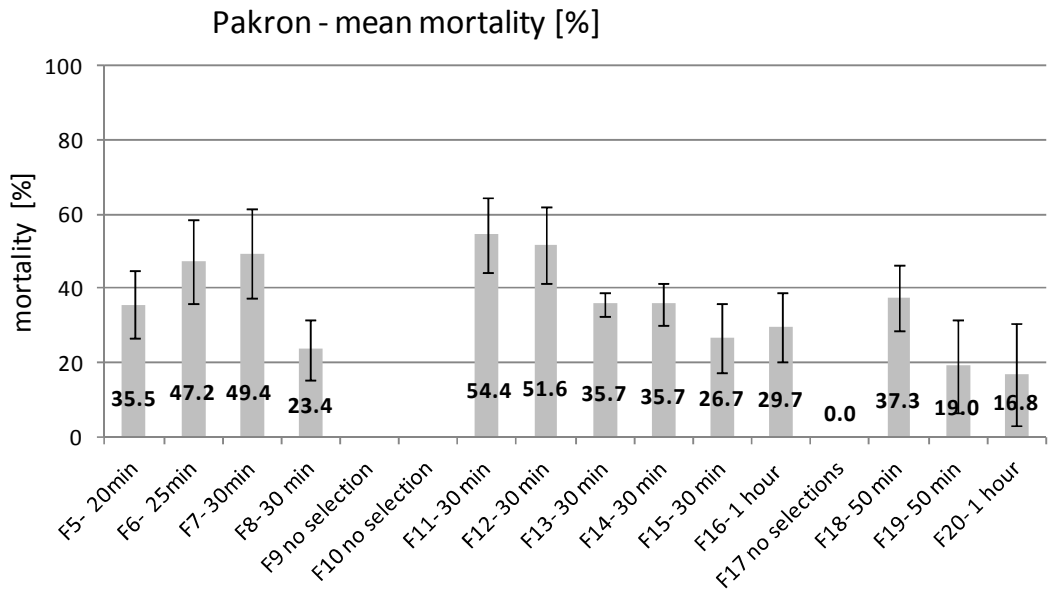


Figure 3.5: Trend in mortality during selection of Akron colony.

Bar diagram showing the change in mortality (±standard deviation) after selections of subsequent generations of the Akron lab strain with 0.75% permethrin. Note the name of the colony was changed from Akron to Pakron due to the selection process which changed the phenotype.

	selections	n	M-form	1014S *kdr*	1014F *kdr*	*ace-1*
Akron F0	none	20	0.89	0	0.60	0.93
Akron F7	permethrin 0.75%	23	0.91	0	0.08¥	0.94
Akron F26	permethrin 0.75%	16	1.0	0	1.0	no data
Akron Carbamate	carbamate 0.1%	16	1.0	0	0.06	1.0

¥ the low *kdr* frequency is due to biased sampling (non-survivors after permethrin exposure).

Table 3.4: Summary for genotyping of target-site mutations in different generations of the Akron colony.

3.4.3 Sequencing of candidate genes in Akron field samples

All 69 field samples (Djouaka 2008) were molecularly confirmed as *An. gambiae s.s.* M-form. Frequencies of 1014F *kdr* were f(R) 0.97 and 0.85 for alive and dead respectively. No resistance alleles for 1014S *kdr* or *ace-1* were observed.

Three samples of each group, dead and alive (six each in the case of *cyp6p3*) were sequenced and sequence alignments produced in order to identify polymorphisms. The numbers of substitutions found in each gene are listed in Table 3.5.

Table 3.6 lists SNP identities, codons, amino acid changes (if applicable) for the identified polymorphisms and genotype data and L1014F *kdr* status in the sequenced samples. Pyrosequencing assays were successfully designed for 13 SNPs. Those polymorphisms were assessed in the 69 field samples and χ^2 and p-values are listed in Table 3.7.

						substitutions/ 1 kb sequence		
gene	**upstream**	**Exon 1¥**	**Intron¥**	**Exon 2¥**	**overall¥**	**overall**	**non coding**	**coding**
cyp6p3	40/ 0.9	28/ 1.1	4/ 0.08	14/ 0.5	86/ 2.66	32	45	26
cyp6m2	57/ 0.9	38/ 1.1	2/ 0.07	9/ 0.4	106/ 2.44	43	63	31
cplcg3	10/ 0.4	7/ 0.4	na	na	17/ 0.85	20	25	18
cplcg4	31/ 1.3	17/ 0.7	na	na	48/ 1.98	24	24	24

¥ number of substitutions found / length of DNA sequence in kb

Table 3.5: Number of substitutions found per kb of candidate gene sequences.

gene	*cyp6p3*								
assay	787	812	814	926	1033	1039	1067	1229	1280
position	-153 UTR	-128	-126	-15	93 Exon 1	99	127	289	340
codon change¥					GAA > GAC Glu > Asp non syn	GGA > GGG Gly syn	TTG > CTG Leu syn	ATC > ATT Ile syn	GCA > GCT Ala syn
vector base	C	G	G	T	C	A	T	C	
samples§									
A1	C/C	G/G	G/G	T/T	C/C	A/A	T/T	C/C	
A2	T/T	G/G	G/G	C/C	A/A	G/G	T/T		
A3	T/T	G/G	G/G	C/C	A/A	G/G	T/T	C/C	A/A
A4	C/C	A/A	G/G	T/T	C/C	A/A	T/T	T/T	T/T
A5									
A6								T/T	T/T
D1	C/C	G/G	G/G	T/T	C/C	G/G	C/C	C/C	A/A
D2	T/T	G/G	T/T	C/C	A/C	G/G	T/T	C/C	A/A
D3	C/C	G/G	A/G	C/T	A/C	A/A	C/T	C/C	A/A
D4	C/C	G/G	A/A	T/T	C/C	G/G	C/C	C/C	A/A
D5	C/T	G/G	G/G					C/C	A/A
D6								C/C	A/A

gene	*cyp6m2*	*cplcg3*	*cplcg4*		sodium channel
assay	m2.2_301	G4_540	G3_1651	1845	1014F *kdr*
position	255 Exon 1	87 Exon 1	-136	62-74 Intron 1	1014 Exon 20
codon change¥	ATC > ATT Ile syn	CTG > CCG Leu > Pro non syn			TTA > TTT Leu > Phe non syn
vector base	C	T	T	A ¥	A
samples§					
A4	C/C	C/C	T/T	**G**CT**TTCG**CTCTA**T**	T/T
A5	C/C	C/C		GCTTTCGCTCTAT	T/T
A6	T/T	C/T		ACTCCTACTCTAC	T/T
D4	C/C	C/T	A/A	ACTCCTACTCTAC	T/T
D5	C/C	C/T		heterozygous	T/T
D6	C/C	C/T		heterozygous	T/T

¥ non syn = non synonymous mutation; syn = synonymous mutation

§ A = alive; D = dead

Table 3.6: List of SNPs identified in the candidate genes.

SNPs identified based on sequencing data, showing SNP identity and genotypes in the sequenced field samples A=alive and D=dead; all sequenced field samples show the listed insertion relative to the vector base sequence.

locus	gene		codons	χ^2	p-value
kdr	sodium channel	Exon20	TTA-Leu to TTT-Phe	4.66	0.03
P3_787	*cyp6p3*	UTR	G → A	1.44	0.23
P3_812	"	UTR	C → T	0.00	0.98
P3_814	"	UTR	G → A or T	¥	¥
p3_926	"	UTR	A → T	2.96	0.09
P3_1033	"	Exon 1	GAA-Glu →GAC-Asp	2.94	0.09
P3_1039	"	Exon 1	GGA and GGG-Gly	0.77	0.38
p3_1067	"	Exon 1	CTG and TTG-Leu	1.66	0.20
P3_1229	"	Exon 1	ATC and ATT-Ile	0.30	0.58
P3_1280	"	Exon 1	GCA and GCT-Ala	0.92	0.34
G4_540	*cplcg3*	Exon 1	CCG-Pro →CTG-Leu	0.00	0.96
G3_1651	*cplcg4*	intergenic	T → A	0.42	0.52
G3_1845	"	Intron	A → 13 bp insertion	1.11	0.29
m2.2_301	*cyp6m2*	Exon 1	ATC and ATT - Ile	0.55	0.46

Table 3.7: Table showing χ^2 and p-values for the SNP polymorphisms that were identified and genotyped in the field samples from Akron (R F Djouaka 2008; n=69).
¥ not possible to calculate correct values due to multiple substitutions at this locus.

Primers for quantitative RT PCR were based on the sequences of *cyp6p3*, *cyp6m2* and *cplcg3/4* obtained in this section. It was not possible to design exon junction spanning primers due to high levels of polymorphism detected and typical for cytochrome P450 genes (Wilding *et al.* 2009).

3.4.4 qPCR on candidate genes *cyp6p3*, *cyp6m2* and *cplcg4*

Two sets of qPCR experiments were carried out. In the first set, gene expression in the field samples from Benin, survivors after permethrin exposure, were compared to the Akron colony (prior to further selection at LSTM) and the susceptible Ngousso strain. In the second experiment three samples from the Akron colony (generation F_0 before permethrin selection, F_7 after a few generations of selection and F_{24} after several generations of selection) were compared (Table 3.8).

For experiment 1 the ratios of transcript levels of *cyp6p3* and *cyp6m2* are increased in the Benin field samples compared to the lab strains Akron and Ngousso, irrespective of which control genes the data is normalized to. This difference is not significant (Figure 3.6 A and B). The transcript levels for *cyp6p3* and *cyp6m2* between the Akron F_0 and Ngousso strain are similar. Comparing Akron F_0 with Akron F_{24} samples in experiment 2, normalized against *cpr*, the ratios of transcript levels of *cyp6p3* and *cyp6m2* show a trend of increasing level, this however is not significant (Figure 3.6 C and D). This trend is not observed when transcript levels are compared after normalizing to the control gene *s7*.

The absolute copy number of the control genes was compared in an attempt to explain the variable results obtained with the two genes (Figure 3.7). Expression of the *cpr* genes is relatively consistent between samples but the *s7* gene expression is highly variable with expression being particularly low in the pyrethroid exposed field samples. This will account for the higher ratio of expression in field samples compared to the colony samples when normalized against *s7* (Table 3.8 and Figure 3.6 A).

		cyp6p3				***cyp6m2***			
		/s7		*/cpr*		*/s7*		*/cpr*	
experiment 1	**Akron F0**	0.050	0.01	0.218	0	0.143	0.01	0.617	0.16
	Benin field	0.134	0.05	0.299	0.02	0.442	0.18	1.132	0.04
	Ngousso	0.039	0.01	0.205	0.01	0.150	0.01	0.786	0.24
		cplcg4							
		/s7		*/cpr*					
	Akron F0	0.144	0.04	0.622	0.09				
	Benin field	0.285	0.17	0.730	0.17				
	Ngousso	0.214	0.02	1.119	0.49				

		cyp6p3				*cyp6m2*			
		/s7		*/cpr*		*/s7*		*/cpr*	
experiment 2	**Akron F0**	0.079	0.02	0.356	0.04	0.163	0.04	0.731	0.02
	Akron F7	0.056	0.02	0.385	0.14	0.120	0.05	0.835	0.42
	Akron F26	0.070	0.03	0.455	0.10	0.178	0.06	1.158	0.25
		cplcg4							
		/s7		*/cpr*					
	Akron F0	0.168	0.03	0.766	0.16				
	Akron F7	0.167	0.03	1.150	0.36				
	Akron F26	0.159	0.02	1.051	0.04				

Table 3.8: Ratio of gene expression (mean values ± sdev).

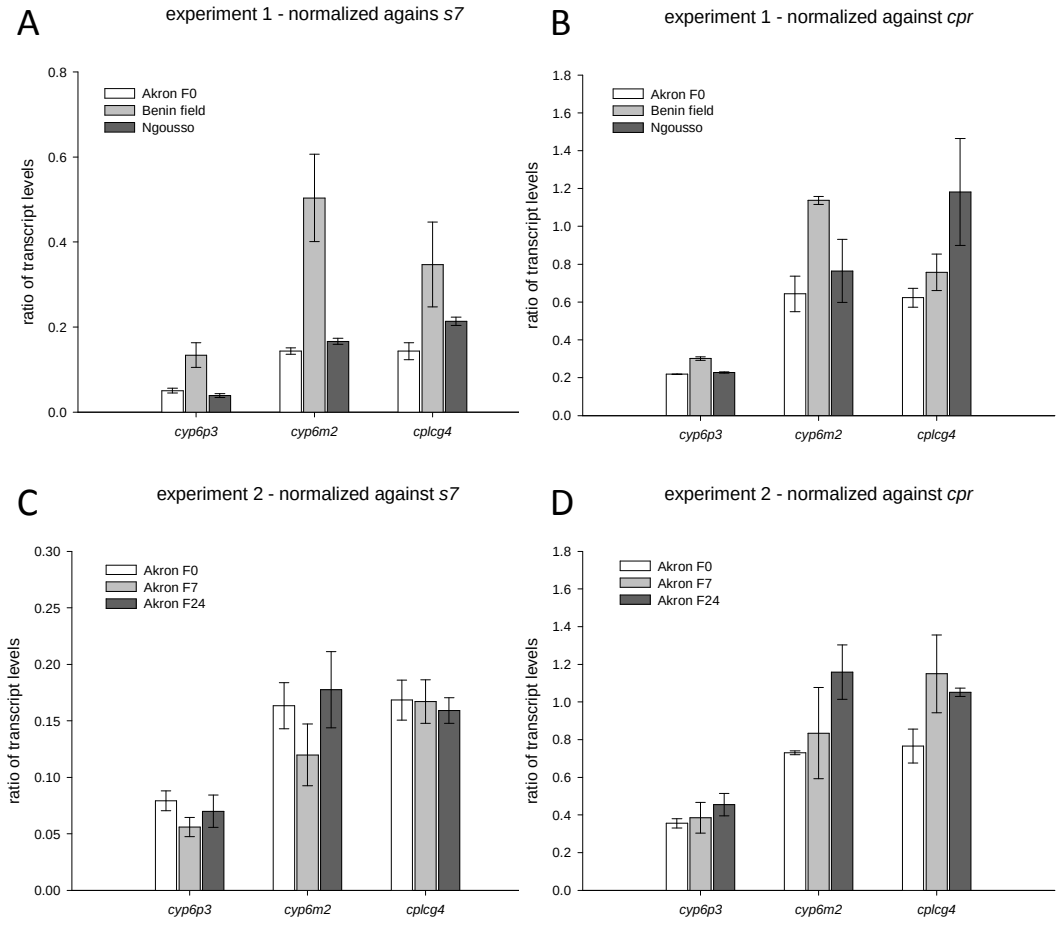


Figure 3.6: Summary of the qPCR (absolute quantification) results on the candidate genes *cyp6p3*, *cyp6m2* and *cplcg4.*

Control genes, *s7* (left column) and *cpr* (right column) (mean±se). (A and B) Compared are the resistant lab colony (Akron F_0), resistant field samples (exposed) and the susceptible lab colony, Ngousso. (C and D) Shown are the results from the second set of data comparing Akron colony samples only.

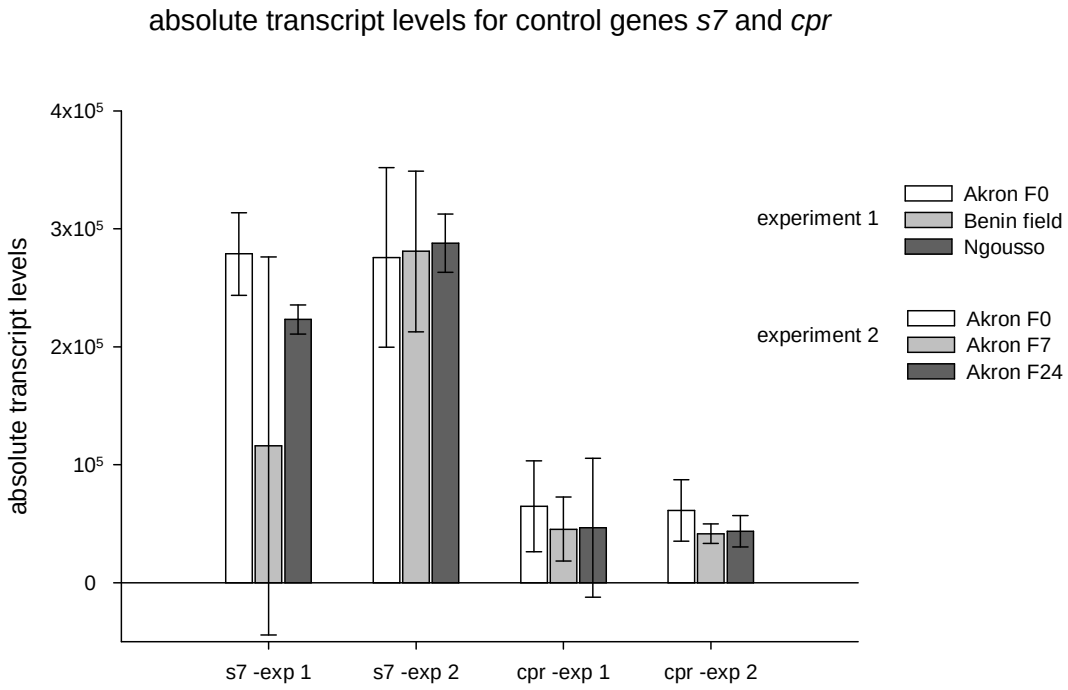


Figure 3.7: Absolute copy numbers of the control genes *s7* and *cpr* for both experiments (mean±CI 95%).

3.5 Discussion

The objective of this chapter was to use association mapping and gene expression techniques to investigate whether any of the candidate genes, identified in an earlier microarray study (Djouaka *et al.* 2008), were associated with permethrin resistance in the field population from Akron or in a lab population colonised from the same site. Furthermore frequencies of the *kdr* and *ace-1* target-site resistance alleles were assessed those populations.

Analysis of the field samples collected from Akron in 2009 confirmed a high frequency of permethrin (and DDT) resistance in the M-form *An. gambiae* from this site. As expected there was a high frequency of the 1014F *kdr* allele. Sequencing of a larger region of the sodium channel, comprising the putative pyrethroid binding site (Bloomquist 1996; Ingles *et al.* 1996) did not identify any further non-synonymous SNPs, i.e. potential additional target-site mutations.

Genotyping the Akron colony for target-site mutations revealed a very different profile to the field populations. The Akron colony contained the 1014F allele but at a lower frequency than the field samples (0.6 vs 0.9). The Akron colony had a high frequency of the 119S *ace-1* allele (> 0.9) but the *ace-1* allele was not detected in the field populations samples although the sample size (n=32) was relatively low.

The 119S *ace-1* allele is thought to exert a relatively high fitness cost (Djogbénou *et al.* 2009; Weill M. *et al.* 2004) and its allele frequency is expected to decline when selection pressure is reduced (Hemingway *et al.* 1997). Carbamate insecticides are one of the main insecticides used to fight insect pests in Akron (Djouaka 2008). These may however be applied in rotation and it is possible that the low frequency of the *ace-1*^{R} allele in the field collections is a result of the timing of the collections. Presumably the Akron strain had been selected extensively with carbamates before deposition in MR4. Selections of the original Akron MR4 strain with permethrin led to an increase in permethrin resistance resulting in a ‘new’ strain named Pakron. The selection resulted in an increase in the 1014F *kdr*

frequency, as expected, but one of the objectives of this chapter was to see if selection also increased expression of any of the candidate genes.

Unfortunately, results of the first qPCR experiment on the candidate genes for populations from both a susceptible and permethrin resistant laboratory strain and the Benin field samples were inconclusive due to the variation obtained with the two control genes in particular *s7*. The presence of SNPs in the primer binding sites was not adequately investigated in all used strains. Another valid approach could have been to use another control gene.

Earlier studies using microarray analysis demonstrate an up-regulation of *cyp6p3* and *cyp6m2* in permethrin resistant field populations of *An. gambiae* from Ghana and the genes showed good concordance between the microarray and quantitative RT-PCR results (Müller *et al.* 2008b).

The individuals from Benin field samples surviving permethrin exposure do show very high transcript levels for the P450 genes; however this may be due to induction of those genes following insecticide exposure (Poupardin *et al.* 2008).

In the second experiment transcript levels of different generations of the Akron/ Pakron lab colony were compared with the expectation that transcript levels of the candidate genes would increase from F_0 to F_7 and F_{24}. This was not observed, even though the observed resistance to permethrin had increased. This may indicate that selection of the lab colony caused the increase of *kdr* frequency but did not select for any of the metabolic resistance mechanisms originally present in the population, but the reasons for this are unclear.

The coding regions of *cyp6p3*, *cyp6m2*, *cplcg3* and *cplcg4* including ~900 bp upstream of each gene were sequenced in a small number of phenotyped field samples from Akron to identify sequence polymorphisms within these samples. The sequences for *cyp6p3* and *cyp6m2* show numerous polymorphisms and the SNP density is high in coding and non coding regions. The *cplcg3* and *cplcg4* sequences are less polymorphic between the sequenced samples (Table 3.5). The cuticular proteins CPLCG3 and CPLCG4 are assembled into fibrils to form part of the cuticule and, as

structural proteins, may be expected to be more conserved than detoxification enzymes which are known to be highly polymorphic (Andersen *et al.* 1995; Rebers & Willis 2001; Wilding *et al.* 2009). Changes in the promoter regions can affect transcriptional activity and have been linked to enhanced expression levels of detoxification genes (Daborn *et al.* 2002; Ding *et al.* 2005). Recently another type of element, called microRNAs, has been shown to play gene regulatory roles (Winter *et al.* 2012). A bioinformatics search using MicroInspector version 1.5 (Rusinov *et al.* 2005) predicted one of these micro RNA-binding sites in the 5'-UTR of the *cyp6p3*, as well as two miRNA-binding sites upstream of *cyp6m2* (overlapping with identified sequence polymorphisms (Table 3.6)) and three in the intergenic region of *cplcg3/g4*. Due to lack of sequence information on the 3'-UTR no search could be performed downstream those genes. The potential role of microRNAs in regulating genes involved in insecticide resistance requires further investigation.

A selection of SNPs identified in these 'candidate' loci were selected for genotyping in the 69 phenotyped field specimens; in an attempt to find statistical linkage to the resistant phenotype. Only the *kdr* locus showed significant linkage. The lack of linkage may be due to the relatively small sample size genotyped and/or due to bias in sample collection. However, the current data do not support an association between the candidate genes identified in earlier studies and resistance to permethrin in Akron.

The lack of association between the candidate genes and the resistance phenotype, using either association mapping or qPCR, may be due to limitations of the techniques such as insufficient sample sizes. Alternatively, the candidates identified in the earlier microarray studies may not be responsible for the resistance observed in the population under study. This result was surprising given the association between several of these candidates and pyrethroid resistance in several independent studies (Amenya *et al.* 2008; Müller *et al.* 2007; Müller *et al.* 2008b; Wondji *et al.* 2009). Nevertheless, as the role of the candidate genes could not be confirmed, a genetic mapping study was undertaken to identify the major loci responsible for resistance in this population.

Chapter 4. QTL mapping of permethrin resistance in isofemale families of *An. gambiae s.s.* from Benin

4.1 Introduction

Quantitative trait locus (QTL) mapping is a forward genetics approach used to identify genome regions associated with traits that govern certain phenotypes. One of the advantages of QTL mapping is that no prior assumptions about mechanisms involved need to be made. QTL mapping can be used to *de novo* discover candidate regions as well as to confirm existing candidate genes. QTL mapping has the power to detect *trans*-acting factors involved in gene regulation, such as regulatory elements and transcription factors, just as well as *cis*-acting factors like mutations in target sites or structural enzymes.

QTL mapping has been successfully applied to vectors of disease. Recent examples include mapping of insecticide resistance in various mosquito species such as *An. gambiae* (Ranson *et al.* 2000b; Ranson *et al.* 2004), *An. funestus* (Wondji *et al.* 2007) and *Ae. aegypti* (Saavedra-Rodriguez *et al.* 2008) as well as susceptibility to *Plasmodium* infection in *An. gambiae* (Niare *et al.* 2002; Riehle *et al.* 2006). A prerequisite for such studies is always the availability of a sufficient number of genetic markers as well as a suitable mapping population.

Microsatellites are the genetic markers widely applied for QTL mapping as they are inherited in mendelian fashion, co-dominant, randomly distributed across the genome, and fast and easy to genotype. They are abundant and highly polymorphic within *anopheline* genomes. A first genetic map consisting of 131 microsatellite markers with an average resolution of 1.6 cM was constructed for *An. gambiae* and published in 1993 (Wang *et al.* 1999; Zheng *et al.* 1996; Zheng *et al.* 1993). This enabled the first QTL studies on *anopheline* mosquitoes.

A very common format of QTL studies in insects is the F_2-design (or an advanced intercross design if it goes beyond generation F_2) based on a cross of two individuals from two different populations representing the

phenotypic extremes one is interested in. The studies of insecticide resistance by Ranson, Wondji and Saavedra-Rodriguez mentioned above have in common, that they used experimental crosses between two inbred lab strains of mosquitoes, a resistant and a susceptible population, to generate mapping populations. This approach is well established and has proved to yield results but has its limitations. It is arguable as to how laboratory colonised insects represent natural breeding populations. These lab strains have often been under selective pressure purely by being kept under laboratory conditions for numerous generations. Through generations of inbreeding they are likely to have adapted to laboratory conditions and this may have altered both genotype and phenotype. As a result, QTL studies using this design may be trying to establish genotype-phenotype associations in the context of a genetic make-up that is very unlikely to occur in nature (Dworkin *et al.* 2005). QTL mapping can be successful in unmanipulated natural populations of mammals (Slate 2005). Similarly studies on *Plasmodium* infection (Niare *et al.* 2002; Riehle *et al.* 2006) have used phenotypic variation in natural populations of *An. gambiae* to map allelic variants affecting parasite development. No QTL studies on unmanipulated natural populations that seek to identify factors involved in insecticide resistance have been published. In this chapter field caught gravid females of *An. gambiae* from Benin were used to generate isofemale lines for a QTL mapping study investigating permethrin resistance.

4.2 Aims of the chapter

Benin was chosen as a study site due to the high level of pyrethroid resistance in the South of the Country (see Chapter 3) and the data, from microarray experiments, suggesting metabolic and possibly cuticular resistance mechanisms were at least partially responsible for this phenotype. The specific objectives of this work are to:

1. Assess how much target-site and metabolic resistance contribute to the overall variance in the resistant phenotype
2. Identify the major loci conferring permethrin resistance in the field population

An. gambiae females generally mate only once (Tripet *et al.* 2001). This and the large number of progeny produced by a single female enable a study design based on isofemale families. The isofemale families were raised from single F_0 females caught in Akron, their F_1 generation intercrossed and the F_2 generation tested for permethrin resistance. F_2 individuals were genotyped with microsatellite and SNP markers.

As mentioned above QTL mapping is a genome wide approach for which no preliminary knowledge about involved factors is required. However, as in this case there is evidence for a role of certain candidate genes, *cyp6p3*, *cyp6m2*, *cplcg3*, and *cplcg4* (Djouaka *et al.* 2008) in this particular resistance phenotype, the initial screen on F_2 samples included markers in the vicinity of these candidate genes.

4.3 Material and Methods

4.3.1 Adult mosquito collections, forced egg laying and generation of mapping families

A detailed description of field work, mosquito samples and establishment of mapping families is found In Chapter 2.1.1.2.

4.3.2 Resistance phenotyping of F_2 offspring

All insecticide bioassays described in this report were carried out according to standard WHO specifications using WHO insecticide susceptibility kits (WHO 1998). Further details are described in Chapter 2.1.1.3.

4.3.3 Genotyping of mapping families

Basic genetic characterization was carried out to standardized procedures described in Chapter 2.2.

Informative markers were selected on the basis of the genotypes of the parental female F_0 and F_1 individuals. A total of 51 microsatellite loci were screened (Table 2.9). Additionally three new SNP markers were identified, one each in the candidate genes *cyp6p3*, *cyp6m2* and *cplcg4*.

Potential informative SNP markers, including known target-site mutations, were tested in the same way on basis of the genotypes of the parental female F_0 and F_1 individuals (four female and four male offspring). The 1014 *kdr* marker was only informative in one family, family 17. In order to identify an alternative SNP marker in the sodium channel that could be genotyped in the remaining families, sequences for exons 20 and exons 27-30 were determined (Macrogen, Korea) and aligned. Primer and PCR details are listed in Table 2.5. Informative microsatellite and SNP markers, identified in the initial screen were then genotyped in all F_2 individuals of the three investigated *An. gambiae* families (38 for family 15; 63 for family 17 and 43 for family 25) (Table 2.7).

4.3.4 Software used for marker data analysis, linkage mapping, and LOD scoring

The JoinMap 2.0 software package (Stam & van Ooijen 1996) was used to build genetic linkage maps, Haploview 4.2 (Barrett *et al.* 2005) for χ^2 testing, Windows QTL Cartographer 2.5 (Wang *et al.* 2005) for LOD analysis and MapChart 2.1 software (Voorrips 2002) for final visualization. More details can be found in Chapter 2.4.3.

4.4 Results

4.4.1 Adult mosquito collections and forced egg laying

A total of 109 females were set up to oviposit in individual Eppendorfs and 56% (64 females) laid eggs. Forty seven of these egg batches were fertile (hatching rate 73%) and 21 of these were kept as separate families.

4.4.2 Resistance phenotyping of F_2 offspring

Of the 21 families, only four produced sufficient F_2 progeny for phenotyping. All four families showed strong permethrin resistance. Families 15, 20 and 25 were initially phenotyped using 0.75% permethrin (the WHO diagnostic dose) but 100 % survived a one hour exposure and mortality rates ranging from 5.3% to 20% mortality were obtained after 90 minutes. Using a higher concentration of permethrin (2%) and an exposure time of 60 minutes resulted in a higher mortality of 46%, 67% and 51% for families 15, 20 and 25 respectively (Table 4.1). The total numbers of phenotyped individuals in those four families ranged from 12 to 201 mosquitoes. Mortality in females was generally lower than mortality in males ranging from 14-50% as compared to 24-100%.

	24 h post permethrin exposure						
	overall		female		male		
family	n	mortality	n	mortality	n	mortality	
15[¥]	68	46%	30	20%	38	66%	*
17[$]	201	18%	111	14%	90	24%	*
20[¥]	12	67%	8	50%	4	100%	
25[¥]	43	51%	24	38%	19	68%	*
positive control[$]	71	99%					
negative control	104	1%					

[¥] 2% permethrin [$] 0.75% permethrin

n = number of mosquitoes phenotyped (overall, male and female)

controls: susceptible *An. gambiae* Ngousso colony

* differences between female and male mortality are significant at the 0.05 level; non parametric test (Mann-Whitney U test)

Table 4.1: Summary of WHO bioassay results from a subset of the F_2 progeny.

Families were raised from single females of *An. gambiae*. Mosquitoes of the susceptible Ngousso colony were used as positive (1 hour exposure to 0.75% permethrin) and negative controls (1 hour exposure to insecticide-free filter paper).

4.4.3 Genotyping of mapping families

4.4.3.1 Species identification and genotyping for target-site mutations

All parental females and F_1 progeny for families 15, 20 and 25 from Akron were confirmed as *An. gambiae s.s.* and M-molecular form. In family 17 the majority of F_1 samples were M-form, however one specimen in the F_1 generation was identified to be an M/S-hybrid (n tested= 43). Microsatellite analysis of this F_1 sample shows foreign alleles at a number of loci, which are not present in any other samples of family 17. To clean up the F_2 samples all M/S-hybrid and S-form specimens were removed. In a second step a microsatellite analysis was performed to find specimens carrying the contaminating M haplotype and all samples positive for the "foreign" alleles were discarded. Altogether 29 samples were removed from the F_2 sample set of family 17, 20 of these from the group of 37 dead samples (54%) and nine of these from the group of 56 alive samples (16%). The remaining F_2 progeny was used for genetic mapping.

The *ace-1* mutation was not present but all tested samples were homozygous for *kdr* (1014F). The *kdr*-genotype was first determined via pyrosequencing and subsequently confirmed through sequence analysis (Macrogen, Korea). Karyotyping for the 2Rb inversion showed all F_0 and F_1 specimens homozygous for the inverted 2Rb+ arrangement. The results for the 2La karyotyping are shown in Table 4.2 below. Both the standard (S) and inverted (I) arrangements are present at frequencies of roughly 0.5 in the F_2 progeny from each family. Family 25 deviates from this with values of f(S)=0.35 and f(I)=0.65.

2La inversion		F_2 progeny family		
		15	17	25
frequency of allele in family	f(S)	0.49	0.47	0.35
	f(I)	0.51	0.53	0.65
association with phenotype	χ^2	0.05	0.58	1.32
	p-value	0.82	0.45	0.25

Table 4.2: Results of molecular karyotyping for the 2La chromosomal inversion.
Included are χ^2 and p-values for comparing the frequencies in the two phenotype groups. Family 20 was excluded due to low numbers of F_2 progeny.

4.4.3.2 Selection of informative markers

Individual F_2 progeny can be either maternal homozygotes, heterozygotes or paternal homozygotes for each marker locus. However, in some cases, it is not possible to assign each allele to a single parent. In this case the marker may be semi informative (if some but not all of the alleles can be tracked back to a specific parent).

Of the 51 tested microsatellites, 34 were informative in at least one of the three families, nine of these were informative in all three families, 14 were informative in two families, and 11 were informative in one family only. Details of informative markers and the range of allele size for each marker are shown in Table 4.4. The 1014 *kdr* allele was informative in family 17 only, in which the paternal genotype was deduced to be heterozygous 1014F/1014L. The other two families were homozygous for the 1014F in all parents and progeny, meaning *kdr* as a marker was non-informative. Sequencing of exon 20 and exons 27-31 of the sodium channel in the F_0 and three F_1 specimens of the three families did not identify any further informative SNPs.

4.4.3.3 Genotyping microsatellite and SNP markers in F_2 samples

Of the four phenotyped families the three families with the largest number of F_2 offspring were selected for genotyping (families 15, 17 and 25). The number of progeny genotyped ranged from 38 to 63 (Table 4.3). Family 20 yielded the lowest number of F_2 individuals with only eight survivors and four susceptible mosquitoes. No genotyping was carried out on F_2 samples of family 20.

	Number of F_2 offspring genotyped						
	dead			**alive**			
family	**male**	**female**	**sum dead**	**male**	**female**	**sum alive**	**sum dead and alive**
15	13	6	19	13	6	19	**38**
17	10	6	16	30	17	47	**63**
25	6	15	21	13	9	22	**43**
sum	29	27	56	56	32	88	144

Table 4.3: Summary of phenotype and sex in each of the three families used for genotyping.

Details of microsatellite and SNP markers successfully genotyped in each family are shown in Table 4.4. The full results for the tests for Hardy-Weinberg equilibrium (dead and alive separate as well as combined) are shown in Appendix 9.1. Single marker analysis was performed for each family to test for association between genotype and resistance phenotype (haploview 4.2). Microsatellite markers for which a significant association was observed are listed in Table 4.5 (full data in Appendices 9.2 and 9.3). The p-values for the SNP markers in candidate genes are also included, even if not significant.

	marker	location ¥ [Mb]	allele size range [bp]	15	17	25
X Chromosome	53	3.6	87-97		x	
	80	7.0	80-108		x	
Chromosome 2	46	1.2	133-142			x
2R	803	1.9	144-146	x	x	
	197	10.4	83-96	x	x	
	757	14.7	79-101	x	x	x
	799	24.9	74-93			x
	2R_si_5	28.5	246-256	x	x	x
	cyp6p3 (P3_1033) $	28.5		x	x	x
	770 *	40.4	174-188		x	x
	117 *	42.7	116-123		x	x
2L	*kdr*	24.2			x	
	637	25.5	102-105	x		
	787 *	26.8	131-135		x	x
	772 *	42.0	131-152			x
	603	42.0	105-109		x	
	675	46.9	101-114	x		x
Chromosome 3	776	2.5	87-102	x	x	x
3R	59	4.2	120-132	x	x	x
	812	6.2	107-131	x	x	x
	cyp6m2 (M2.2_301) $	6.9			x	x
	249	8.8	108-132	x	x	x
	30C1	9.6	155-168	x	x	x
	cplcg4 (G3_1845) $	10.9		x	x	x
	06M19	13.1	160-170	x	x	x
	119	14.8	179-194	x		x
	555	21.3	83-96	x		

	158	22.99	78-90			x
	341	23.02	116-142		x	x
	06H01	25.5	127-156	x	x	x
	33C1	25.6	147-159		x	
	06F05	31.6	174-184		x	x
	88	34.7	184-197			x
	30L17	41.1	71-79		x	x
3L	SODI (RFLP) $	5.6			x	
	811	14.8	124-134		x	x
	577	19.8	80-103	x	x	
	817 *	31.9	143-149		x	x
	46C2	41.4	177-197		x	x
			sum Chromosome x	**0**	**2**	**0**
			sum Chromosome 2	**7**	**10**	**10**
			sum Chromosome 3	**11**	**18**	**18**
			sum	**18**	**30**	**28**

¥ According to vector base (08/2010)

$ SNP markers

***** Markers used as M13-tailed (Oetting *et al.* 1998)

Table 4.4: Details of microsatellite and SNP markers.

Markers found informative in the three tested families based on maternal and F_1 genotypes and subsequently used for genotyping the F_2 samples.

	marker	χ^2	p-value		df
family 15	cyp6p3 (P3_1033) $	0.27	0.602		1
	637	4.07	0.044	**	1
	675	5.32	0.021	**	2
	776	3.27	0.071	*	1
	812	1.94	0.163		2
	249	4.33	0.037	*	2
	30C1	5.18	0.023	**	1
	cplcg4 (G3_1845) $	6.18	0.013	***	1
	6M19	5.67	0.017	**	1
family 17	757	3.59	0.058	*	1
	cyp6p3 (P3_1033) $	0.16	0.692		1
	kdr	0.36	0.547		1
	603	3.42	0.064	*	1
	cyp6m2 (M2.2_301) $	0.48	0.488		1
	817	8.76	0.003	**	2
family 25	cyp6p3 (P3_1033) $	0.38	0.538		1
	787	3.44	0.064	*	2
	cyp6m2 (M2.2_301) $	0.20	0.659		1
	cplcg4 (G3_1845) $	0.01	0.946		1
	811	3.83	0.050	*	2
	817¥	4.35	0.037	**	1

significance levels: *:0.1 **:0.05 ***:0.01

$ SNP markers

¥ those markers were not implemented in the genetic maps

Table 4.5: Chi^2 and p-values for SNP (except SODI) and significant microsatellite markers (haploview 2.0).

4.4.4 Linkage mapping

Very few informative markers were identified on the X chromosome and it was not possible to generate a linkage map for this chromosome.

A summary of the maps for chromosomes 2 and 3, including number of markers and resolution, is shown in Table 4.6. All generated maps for the separate family data are shown in Figure 4.1. Maps using the combined data set based on all three families are shown in Figure 4.2.

		Chromosome 2			Chromosome 3			average res.
		map [cM]	no of markers	average res.	map [cM]	no of markers	average res.	per family
family	**15**	38.8	6	6.5	80.0	11	7.3	7.0
	17	105.7	7	15.1	103.7	16	6.5	9.1
	25	67.3	9	7.5	53.4	14	3.8	5.2
combined	**optimal physical**	67.5	10	6.8	112.5	16	7.0	6.9
	optimal genetic	57.1	10	5.7	110.5	16	6.9	6.4

Table 4.6: Summary of map size and map resolution for all maps with average resolution in cM/ marker.

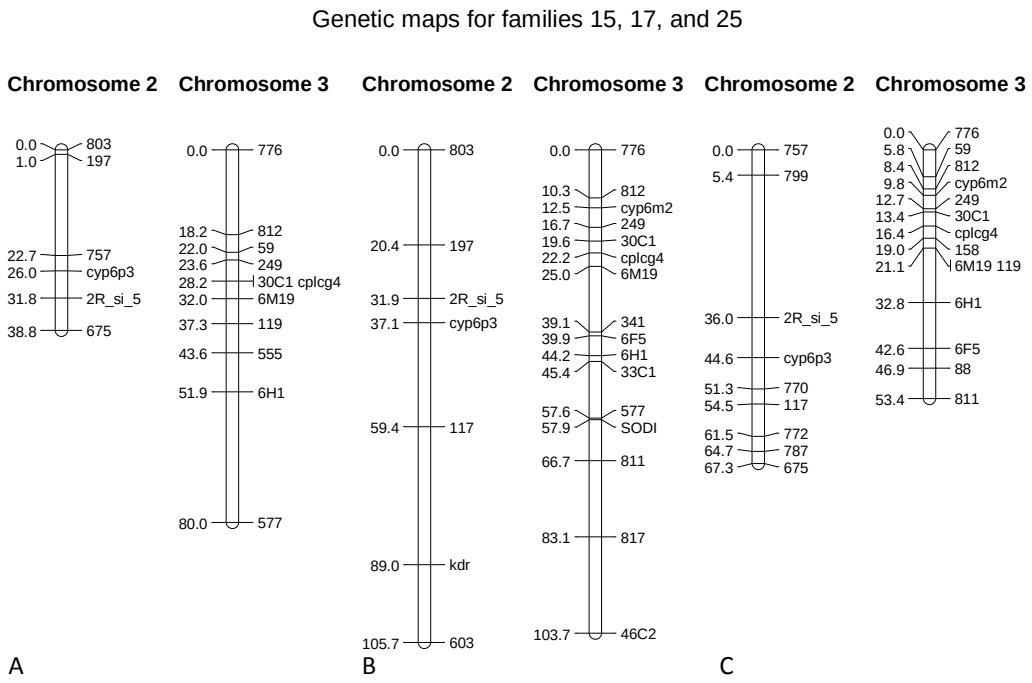


Figure 4.1: Summary of the genetic linkage maps for single families.
Chromosomes 2 and 3 for families (A) 15, (B) 17 and (C) 25 when analysed separately. A scale of genetic distance in centimorgan is shown.

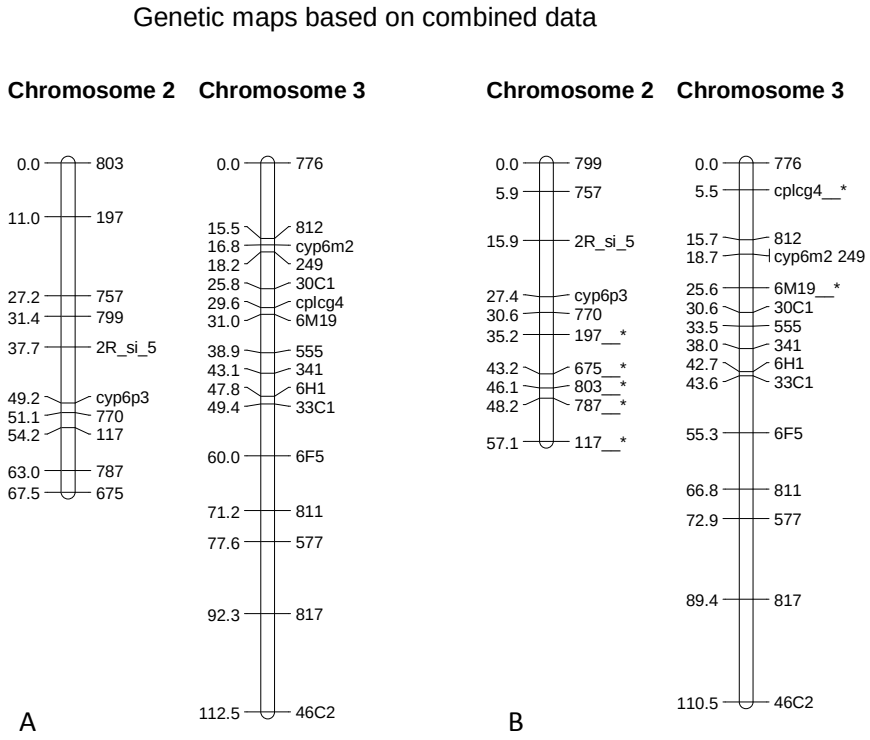


Figure 4.2: Genetic linkage maps using the combined data from the three F_2 intercross families. A scale of genetic distance in centimorgan is shown. (A) Linkage maps of chromosomes 2 and 3 are shown, as based on physical marker order (A) and optimal genetic map (B), with those markers that appear in a different position marked with _*.

Plotting mortality rate as a function of the number of maternal alleles inherited is a simplified means of visualising the genotypic effect on a certain phenotype. Figure 4.3 plots this data for selected markers for chromosomes 2 and 3 for each family.

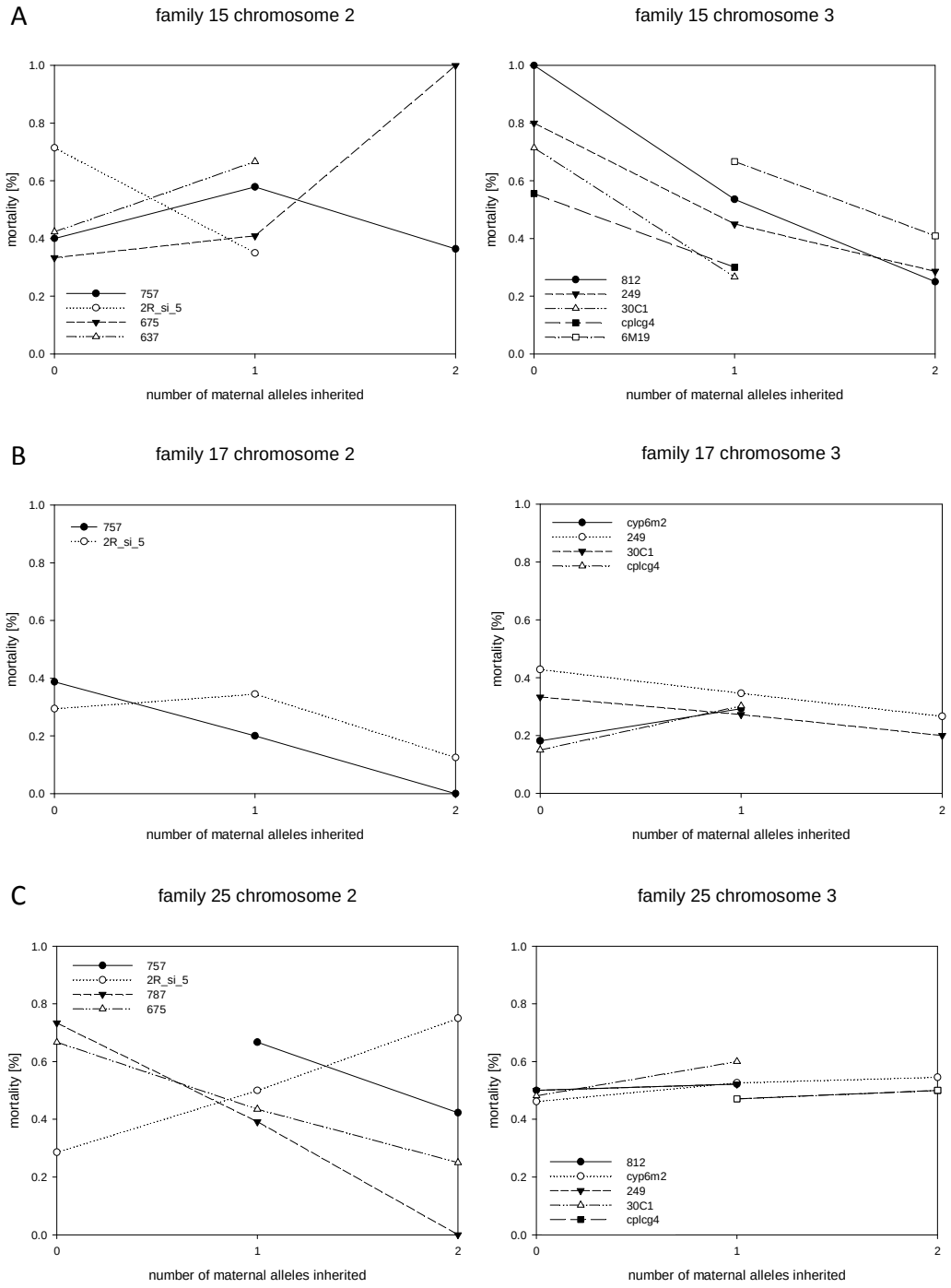


Figure 4.3: Phenotype-genotype association of single markers.

Mortality rate as a function of alleles inherited from the mother for a number of markers on the two chromosomes for families 15 (A), 17 (B) and 25 (C).

Figure 4.3 shows an association between the number of alleles inherited from the maternal parent and the insecticide resistance phenotype for some markers. No unidirectional trend within groups of adjacent markers is obvious for chromosome 2. In contrast, on chromosome 3, a strong association does exist, particularly in family 15. Associated markers are markers 812 and 249 (fully informative) and three adjacent semi-informative markers 30C1, cplcg4 and 6M19. A similar pattern, but less strong, is observed in family 17 for markers 249 and 30C1 (fully informative). The adjacent markers cyp6m2 and cplcg4, however, show a trend in the opposite direction. The same set of markers is shown for family 25, but there is no association between their genotype and the resistance phenotype (mortalities between 47-60%).

4.4.5 QTL analysis

The software Windows QTL cartographer version 2.5 (Wang *et al.* 2005) was applied to test for linkage between resistance phenotype and genotype. Both Interval Mapping (IM) and Composite Interval Mapping (CIM) were performed and the results compared. As each family had a different genetic background, the results from individual families were first analysed individually and then a QTL analysis based on the combined data and maps was performed. Interval Mapping and CIM are largely in agreement for all family data as well as for the combined data.

One QTL was located in family 25 on chromosome 2L reaching a LOD value of 3.1 around 60 cM (only detected with CIM), together with significant χ^2 values for the two microsatellite markers on arm 2L.

On chromosome 3 one QTL was identified in family 15 on arm 3R with a LOD score of 2.5 at around 30 cM (spanning chromosomal divisions 30 A-D). This QTL is detected by both IM and CIM although less pronounced by IM with a lower LOD of 1.4. The χ^2 values for markers in this region are significant.

When combining the data of all three families LOD scores approaching the 2.5cM cut-off are observed on chromosome 2L (Figure 4.6 and 4.7). A peak is also visible on chromosome 3 with a LOD score of approx. 2.0 (CIM). If family 25 is excluded from the analysis, the QTL on chromosome 3 reaches significance (IM LOD_{max} 2.53) (Figure 4.6 B and 4.7 B). The QTL on chromosome 2L remains.

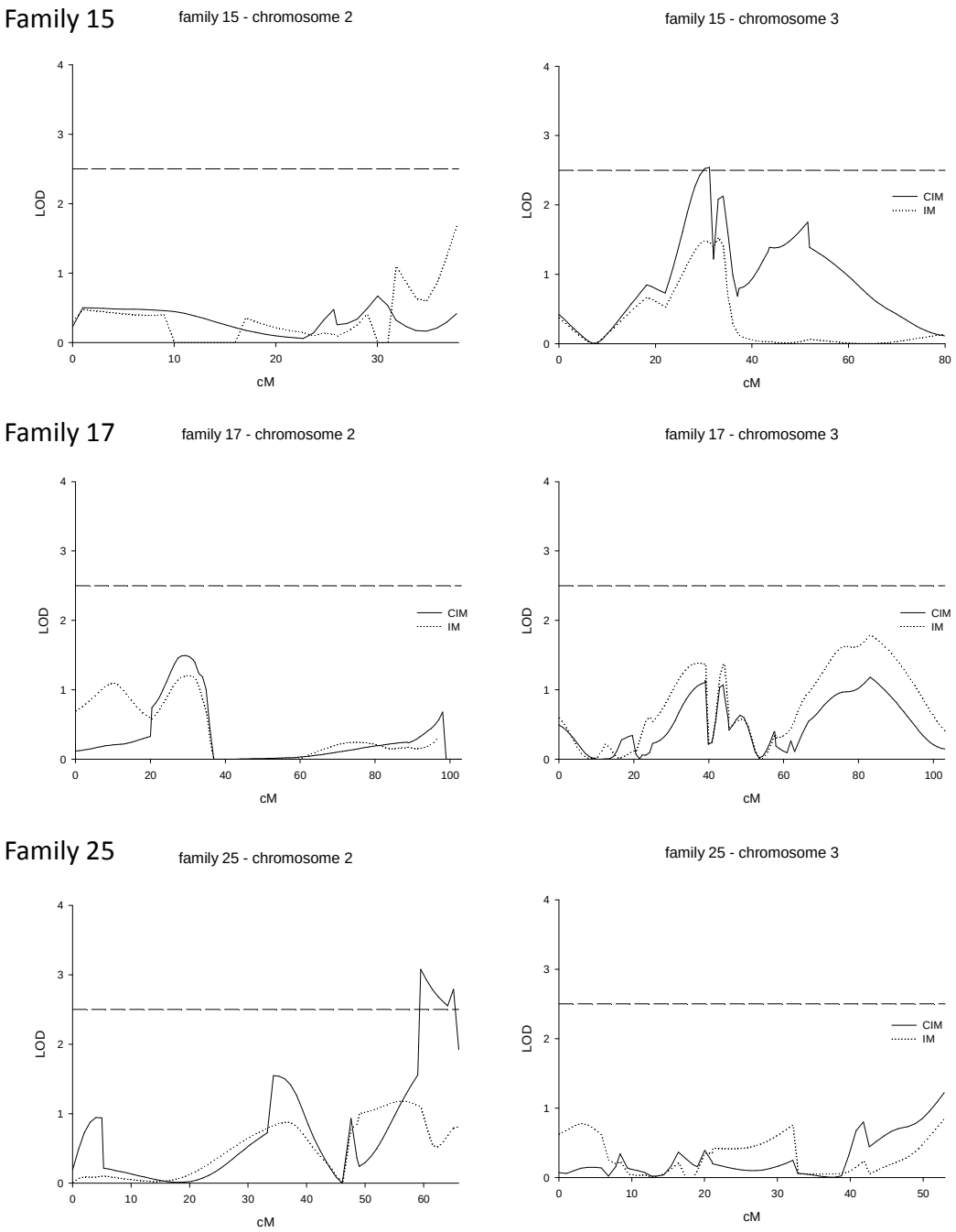


Figure 4.4: Plots of LOD scores associated with permethrin resistance.

Shown are chromosomes 2 and 3 for families 15, 17 and 25. Solid lines represent LOD estimated by composite interval mapping and dotted lines represent LOD estimated by interval mapping. The straight line along the top of each graph represents the threshold value for LOD, set at 2.5 (dashed).

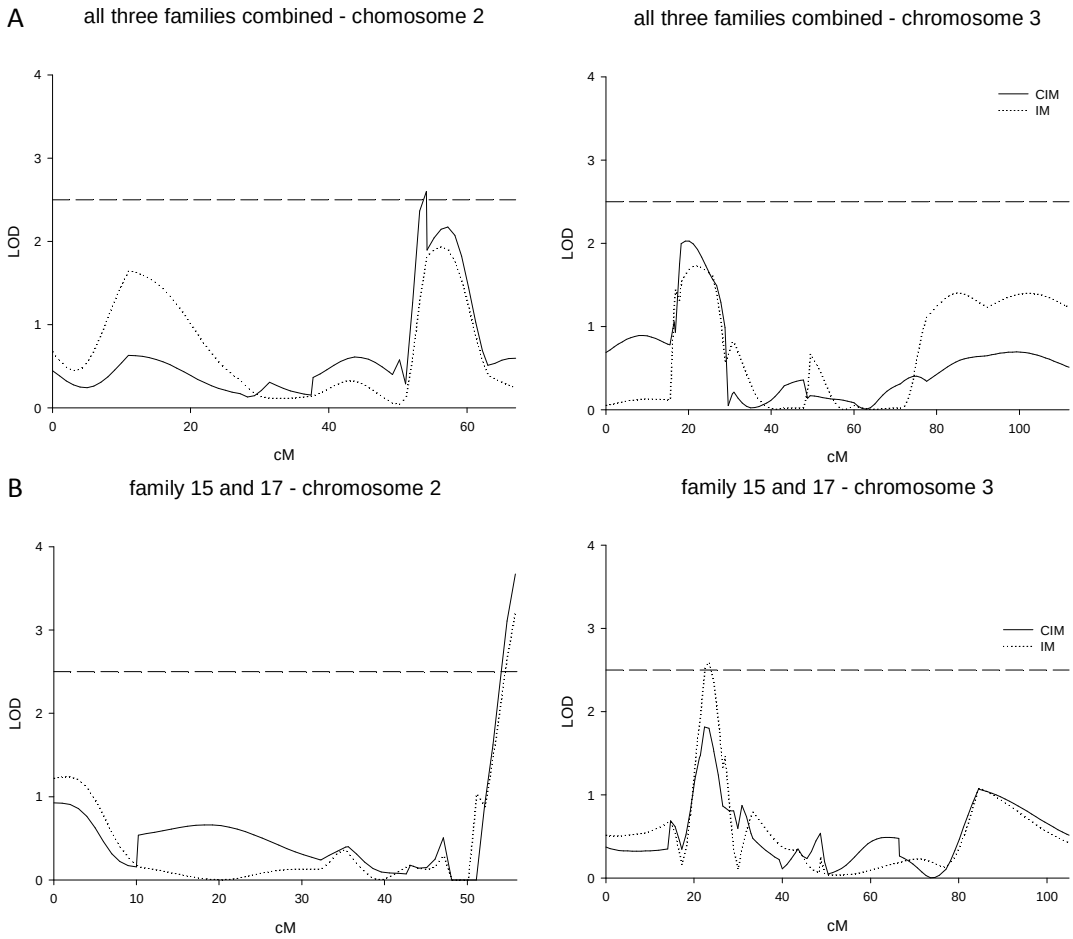


Figure 4.5: LOD plots of combined data.

Plot of LOD scores associated with permethrin resistance on chromosomes 2 and 3 for (A) the combined data of all three families and (B) for data from families 15 and 17 only. Solid lines represent LOD estimated by composite interval mapping and dotted lines represent LOD estimated by interval mapping. The straight line along the top of each graph represents the threshold value for LOD, set at 2.5 (dashed).

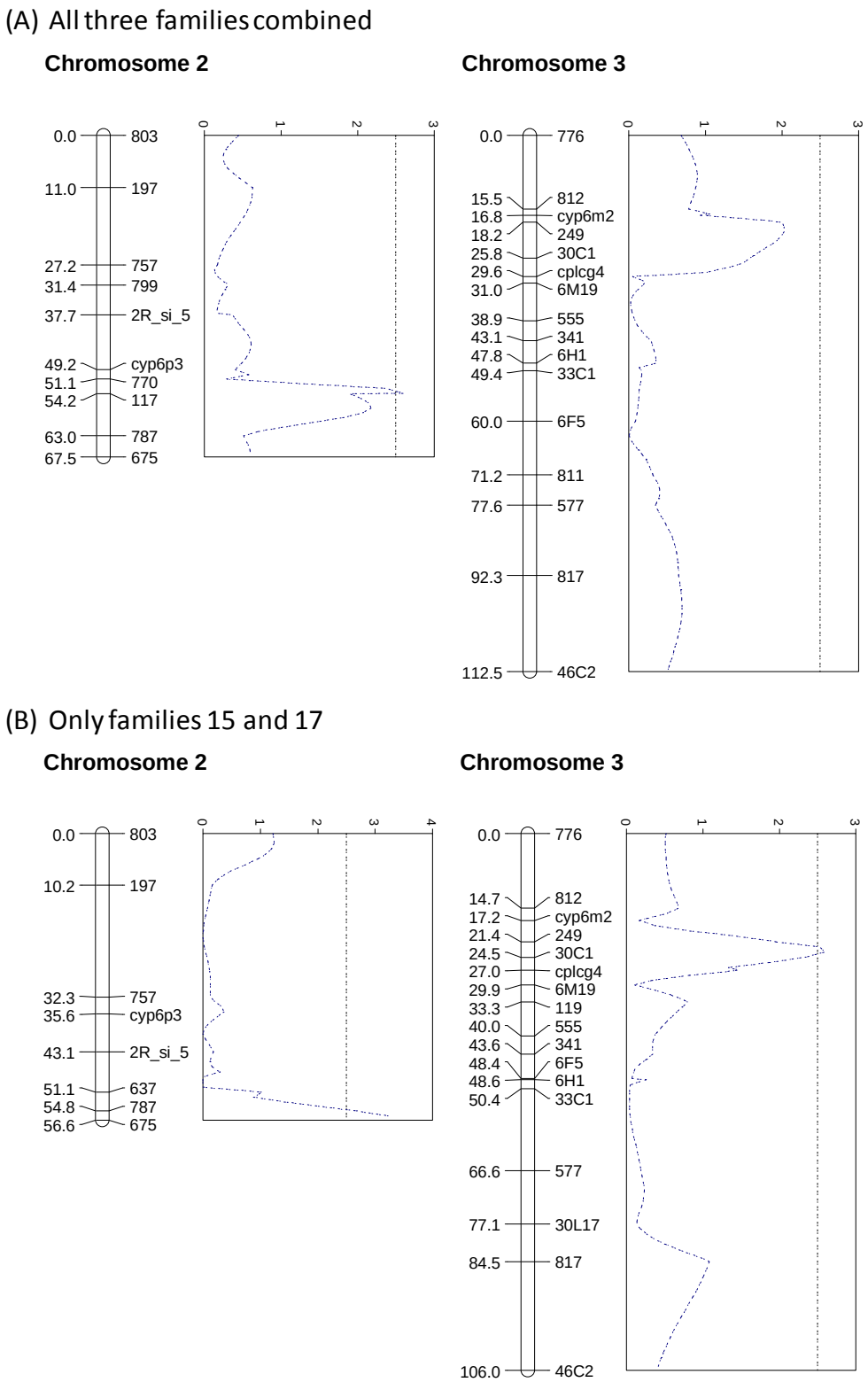


Figure 4.6: LOD plots of combined data alongside the genetic maps.

Plot of LOD scores estimated by (A) composite interval mapping for all three families and (B) interval mapping for families 15 and 17. The straight line along the top of each graph represents the threshold value for LOD, set at 2.5.

Multiple Interval Mapping (MIM) confirms the presence of the QTL on chromosome 3R in family 15 and estimates that its genetic variance (δ_g^2) accounts for 9.1% of the phenotypic variance (δ_p^2) of permethrin resistance (3.6% resulting from additive effect and 5.6% from dominance effect). In families 17 MIM estimates one QTL on chromosome 2R with 8.8% of the genetic variance (δ_g^2) accounting for phenotypic variance (δ_p^2) (8.5% resulting from additive effect and 0.3 from dominant effect). In family 25 MIM estimates that two QTLs on chromosome 2 are responsible for 17.9% of genetic variance. For the combined data MIM estimates two QTLs accounting for 13.2% (all families) and 23.0% (families 15 plus 17) genotypic variance. The QTL close to marker 117 on chromosome 2R is possibly due to *kdr*. Results are summarised in Table 4.7.

sum of δ_g^2 [%]	δ_g^2 [%] per QTL	nearest marker	¥	genetic distance [cM]	effect in %δ_g^2
family 15					
9.1		812	3R	18.3	A: 3.6
					D: 5.6
family 17					
8.8		2R_si_5	2R	32.0	A: 8.5
					D: 0.3
family 25					
17.9	48	2R_si_5	2R	34.4	A: 8.8
					D: -0.1
	52	675	2L	66.7	A: 2.4
					D: 6.9
combined all families					
13.2	56	117*	2R*	54.3	A: 2.2
					D: 5.2
	44	249	3R	18.3	A: 3.8
					D: 2.0
combined families 15 and 17					
23.0	64	787	2R	56.5	A: 9.9
					D: 4.1
	41	30C1	3R	24.6	A: 6.3
					D: 2.8

δ_g^2 [%] genetic variance in % of overall estimated phenotypic variance and per locus;
¥ chromosome and location in genetic distance for nearest marker;
A for additive effect; D for dominant effect;
* associated marker located on chromosome 2R, but effect possibly due to *kdr* on 2L.

Table 4.7: Multiple Interval Mapping estimates of QTL position and associated genetic, variance and additive and dominance effects associated with permethrin survival QTL in *An. gambiae*.

4.5 Discussion

Insecticide resistance is a complex trait and factors involved vary depending on species, insecticide and population. In Southern Benin microarray analysis identified a number of candidates for permethrin resistance aside from the target-site mutations widely described, with genes encoding the two pyrethroid metabolising enzymes CYP6P3 and CYP6M2 being up-regulated (Djouaka *et al.* 2008). This study sought to identify the genomic loci responsible for this resistance.

Raising isofemale families from single egg batches was facilitated by the method developed by John Morgan (Morgan *et al.* 2010). This is an easy and reliable method of retrieving eggs from field caught females. The stress exerted on parental females through forced egg laying may have induced epigenetic changes affecting gene expression in the offspring (Jaenisch 2003). No loss of permethrin resistance, however, was observed in the family offspring, indicating that this is not the case. Yet this kind of stress exerted through experimental conditions on a population needs to be considered when designing similar experiments.

The number of progeny of those families reared to F_2 were rather small and only a very limited F_2 sample size was available for the QTL mapping. This and a restricted number of informative markers are the main limitations of the study.

Insecticide resistance is a quantitative trait and should, as such, be measured quantitatively. However, after exposure to insecticides only two outcomes, alive or dead, were scored. Hence the quantitative trait was scored as a binary trait. A more powerful alternative would be using the extreme phenotyping procedure, whereby the survivors of a longer exposure time and the non-survivors of a short exposure time only are used and insects showing an intermediate phenotype are discarded. However, as in this method only a subset of the phenotyped mosquitoes is used for genotyping, the availability of a large number of samples is required. In this study it was not possible to generate sufficiently large numbers of F_2 generation to do this.

Another alternative to the binary scoring (dead or alive) for phenotyping the mapping population has been used to map pyrethroid resistance in *Ae. aegypti* (Saavedra-Rodriguez *et al.* 2008). The phenotyping for the described study is slightly more complex and classes mosquitoes as knock-down resistant, recovered or dead. This may have helped to overcome the strong *kdr* resistance phenotype to affect the entire QTL mapping however, similarly to the extreme phenotyping this method would also require a larger sample size, which was not available.

Bioassays conducted on adults raised from larvae collected in the same area showed an intermediate level of permethrin resistance (52.4% mortality after 1 hour exposure to 0.75% permethrin) (Chapter 3). Initial phenotyping of F_2 individuals of the four isofemale families however, showed very low mortality with this insecticide concentration (0% mortality in families 15, 20 and 25 after 60 minute exposure and <20% mortality in family 17). Hence for families 15, 20 and 25 a higher insecticide concentration (2%) was used. Observing higher resistance in the F_2 family progeny for some families (15, 20 and 25) than in the Akron field samples was surprising. This effect has been observed before (Ranson *et al.* 2004) and could be due to larger sized mosquitoes caused by optimised larval rearing conditions in the lab. Single family progeny were raised in cleaner conditions (availability of dd H_2O) and at lower densities than the field caught larvae. Furthermore, the observed 1014F *kdr* frequency in the sum of all collected gravid females was very high with f(R)=0.92 and in the four phenotyped families f(R)=0.94. Families 15, 20 and 25 were homozygous for the 1014F, whereas the parental male for family 17 was heterozygous for *kdr*, partly explaining the differences in resistance observed.

In this study female mosquitoes were more likely to survive insecticide exposure than males. This sex bias has been reported previously (Ranson *et al.* 2004) and may be due to the larger size of *Anopheles* females. The possible role of sex associated factors in permethrin resistance in the RSP-ST strain from Kenya was investigated by reciprocal backcross experiments of F_1 males from a R x S or S x R cross with the susceptible strain followed by phenotyping. Phenotyping and QTL analysis of these

crosses showed no major differences between the sexes and an absence of linkage in the four X-linked markers (Ranson *et al.* 2004). The Benin microarray study did not identify candidates on the X chromosome. The data above and the absence of informative markers on the X chromosome was the reason to focus on the two autosomes for genetic analysis.

Prior to genotyping of the F_2 progeny, one single M/S-hybrid sample of F_1 was identified in family 17. If the F_1 generation had been the result of a cross between an M-form parental female (confirmed) and an S-form parental male mosquito the entire F_1 generation would be M/S-hybrids, and similarly, if the parental male had been a hybrid the expected ratio of M:M/S in generation one would have been 3:1. The conclusion therefore is that the M/S specimen was found in the F_1 generation is a contaminant. This hypothesis is supported by microsatellite analysis of those F_1 samples as the contaminating DNA sample carries foreign alleles, which are not present in any other sample, at a number of loci. M/S typing and microsatellite analysis were performed to identify “contaminated” F_2 progeny. Altogether 29 samples were removed from the F_2 sample set of family 17, 54% of the dead samples 16% of the alive samples. Around half of all were detected using M/S-genotyping the other half using the microsatellite approach which is expected, as the same number of contaminating S- (detected by M/S-typing) and M- alleles (only detectable through microsatellite check) must be present in the F_2 generation. The significantly higher percentage of offspring from the contaminating sample in the group of non-survivors suggests that the contaminating mosquito was of susceptible phenotype.

The ratios of dead to alive as well as female to male samples used in the QTL study should ideally be equal to avoid bias. With limitations in sample sizes mentioned above, this was only possible for family 15. In family 25 there was a bias towards more females in the dead progeny. And with an already very limited sample size, conditions for family 17 had to be even less stringent. A low mortality rate in F_2 and a further reduction of the non-survivors due to the “clean-up” shifted the ratio of dead: alive to a final ratio of 16:47.

For chromosome 2 some single markers showed an association between number of alleles inherited from the resistant parent and mortality, but there was no uniform trend between adjacent markers or across families. In families 15 and 25 the two markers closest to the *kdr* locus (namely 637 and 787) both show genotype-phenotype association as well as significant χ^2 values (Table 4.5). For chromosome 3 groups of adjacent markers show a similar trend within families, but not across families. In families 15 and 17 the microsatellite markers close to the detected QTL (812, 249, 30C1 and 6M19) are associated with the phenotype, but no such trend was observed in family 25.

With *kdr* being informative in family 17 a QTL on chromosome 2L attributable to *kdr* was expected. This was not observed. In family 25 a QTL on 2L was observed albeit only by IM. Coherent with this is the significant association of marker 787 near *kdr* with the resistance phenotype (Figure 4.3 and Table 4.5).

For the combined data sets, a QTL on chromosome 2L is observed, between markers 117 (on 2R) and 787 (on 2L). This is most likely attribute to *kdr* but the *kdr* marker itself was not informative in the combined family analysis and the boundaries of the QTL on chromosome 2 are very large. This imprecision in the location of the QTL could be caused by limited genetic resolution in this area due to little recombination around the centromere and the 2La inversion polymorphy. This is additionally influenced by the availability of a low number of informative markers near the *kdr* locus.

The locus attributed to *kdr* on chromosome 2L explains a larger part of the genetic variance (7.4%) than the locus on 3R (5.8%) in the data set of all families. Similarly for the data set of families 15 and 17 MIM estimates two QTLs with the locus on 2L explaining 14 % of genetic variance of and the locus on 3R explaining 9.1%. This is similar to the previous QTL study of permethrin resistance in the RSP-ST strain which found the two main QTLs (equally on 2L and 3R) to explain roughly equal amounts of the genetic variance in one of the families.

Due to the limited genetic resolution it is not impossible that the QTL on chromosome 2 is the result of two separate QTLs that could not be resolved and *kdr* is therefore estimated to contribute a higher percentage of genetic variation in this analysis than it actually does.

Genetic maps for chromosome arm 2L were difficult to establish due to a paracentric chromosomal inversion in *An. gambiae* on the left arm designated 2La. This inversion is highly polymorphic and widespread in *An. gambiae s.s.* (Sharakhov *et al.* 2006). Molecular karyotyping of the 2La inversion through PCR showed that all families are polymorphic for the two arrangements (Table 4.2). Four of the five informative microsatellite markers identified on 2L are located within this inversion (637, 787, 772, and 603) with 603 very close to its distal breakpoint. The *kdr* locus is located close to the inversion.

Only a limited number of microsatellite markers on chromosome 2L were found to be informative. This may be due to selective sweeps adjacent to the *kdr* locus resulting in reduced polymorphism on this chromosome arm. In addition the presence of the inversion would have suppressed recombination in this region and hence may have masked any linkage effect attributable to the *kdr* mutation. Previous genetic mapping studies on permethrin resistance in *An. gambiae* used a laboratory strain (RSP-ST) fixed for the 2L standard chromosomal arrangement to overcome this problem. In this earlier study, a QTL linked to the *kdr* locus was identified (Ranson *et al.* 2004).

The most convincing QTL identified in this study is on chromosome 3R, spanning chromosomal division 30A-D which encompasses a P450 cluster as well as the cuticular candidate genes *cplcg3* and *cplcg4.* This QTL was identified by both CIM and IM methods and found in the single family analysis of family 15 and the combined analysis of family 15 and 17. This QTL is in a similar position to the *rtp2* QTL identified in the RSP-ST strain of *An. gambiae* originally from Kenya (Ranson *et al.* 2004).

The boundaries of the QTL identified on 3R are ~15 cM with the flanking markers 812 and 6M19 spanning a genome region of ~6.9 Mb including

438 annotated genes (vectorbase biomart 03/2012). Enrichment analysis (DAVID) (Huang *et al.* 2009a) shows that serine proteases (1.6% of the genes in the QTL) with a fold enrichment of 27.2 as well as P450s (4.1% of the genes within the QTL boundaries) with a fold enrichment of 4.8 (DAVID enrichment analysis, IPR001128 Cytochrome P450) (Appendix 9.4) are over-represented in this region. Serine proteases can hydrolyse ester bonds and so could conceivably be involved in insecticide metabolism (Hedstrom 2002). Alternatively, elevated protease activity may be a result of metabolic stress induced by insecticide exposure.

Chromosome 3R is densely clustered with detoxification genes and 18 P450 genes are located within the boundaries of the QTL. A cluster of 14 *CYP6* genes (Appendix 9.4) is in division 30A (Ranson *et al.* 2002b) and smaller clusters of *CYP4* and *CYP12* P450s are on neighbouring divisions, but outside the QTL boundaries (Figure 4.7). The *CYP6* cluster includes *cyp6m2,* identified as a candidate resistance gene in microarray experiments (Djouaka *et al.* 2008; Müller *et al.* 2008b) and *cyp6z1* and *cyp6z2*, which have been found associated with pyrethroid resistance in other populations (David *et al.* 2005; Müller *et al.* 2007; Nikou *et al.* 2003).

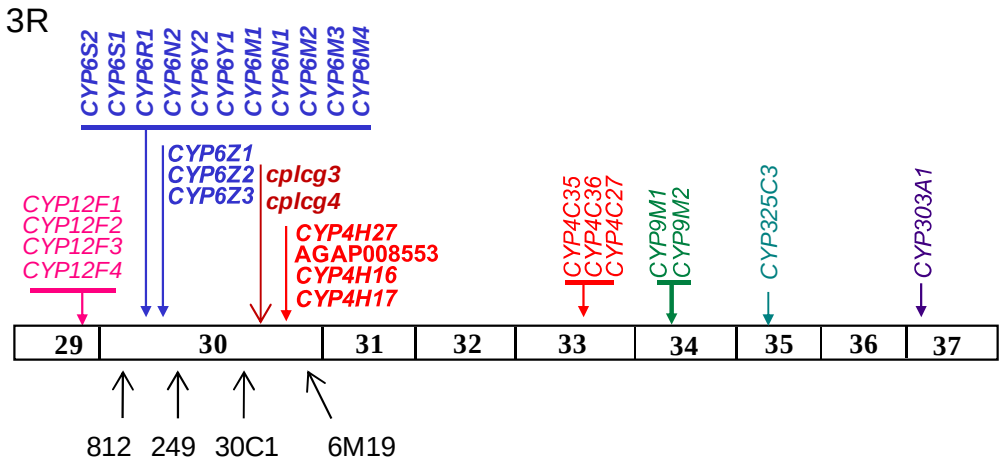


Figure 4.7: P450s and candidate *cplcs* on polytene chromosome arm 3R.

Arrows from top roughly indicate locations of P450 genes and the two *cplcg* candidates in the *An. gambiae s.s.* genome. The black arrows from below indicate locations of the four markers that define the QTL position (adapted from Ranson, unpublished).

In addition 12 of the 27 *cplcg* genes (cuticular proteins with low-sequence complexity) in *An. gambiae* are members of a sequence cluster (Group A) and the entire set is found in a 195kb array on chromosome 3R (Willis 2010) division 30D. Over-expression of cuticular proteins causing reduced cuticular penetration of insecticides has been implicated in insecticide resistance (Puinean *et al.* 2010). Two members of this *cplc* cluster, *cplcg3* and *cplcg4*, were identified as candidates for permethrin resistance in *Anopheles* in earlier microarray experiments (Djouaka *et al.* 2008; Vontas *et al.* 2007).

The current study, using families raised directly from field collected mated females, identified the same QTL as identified in previous laboratory experiments (Ranson *et al.* 2004). This is despite the much stronger resistance phenotype in the Benin population than in the laboratory RSP-ST colony that originated from Kenya and is a significant result, suggesting that there is a conserved 'resistance factor' other than *kdr*, which is found in both East and West African populations of *An. gambiae*. Further fine scale mapping on the laboratory populations using advanced intercross lines failed to identify a candidate locus for this QTL (Ranson, unpublished data). Furthermore, sequencing of candidate genes within this QTL from wild caught individuals (Chapter 3) did not identify a marker linked to the resistance phenotype. Therefore the nature of this QTL remains unresolved.

Mapping experiments on outbred populations have two main difficulties outlined in a review (Slate 2005): firstly, if the parent is not heterozygous at both marker locus and QTL, these families will be un-informative with respect to QTL detection, and secondly, the phase between marker and QTL is not necessarily consistent across families, hence marker effects must be considered independently within each family. And considering, that a population is heterogenous and the genes underlying the phenotype of interest or their contribution to the latter do differ between individuals, i.e. between families, then combining the data of different sibships is not valid. Looking at the example of this study, combining data of families 15 and 25 shows no effect (results not shown) whereas the power increases

when combining the data of chromosome 3 for the other two families only (15 and 17). The latter two families are likely to have a very similar genetic makeup for chromosome 3 and the QTL on 3R becomes more pronounced when the two data sets are combined. The availability of data from more than three families would have possibly further supported these findings.

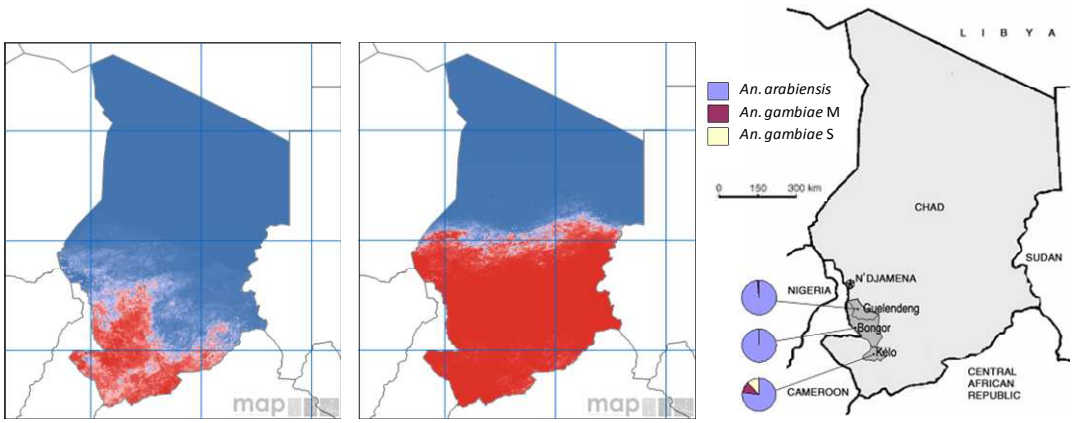
Fine scale mapping to reduce the boundaries of the identified QTL on 3R would be the next step to take but the resolution will likely be restricted by the number of phenotyped progeny available. Further study in this area would require the generation of advanced intercross lines or association mapping approaches on large numbers of phenotyped field samples.

Chapter 5. Mapping a QTL conferring permethrin resistance in the African malaria vector *An. arabiensis*

5.1 Introduction

Anopheles arabiensis is a member of the *Anopheles gambiae* species complex and, although it can share the same breeding habitats with *An. gambiae s.s.* it generally prefers drier locations with a sahelian climate (grassland and savannah). *Anopheles arabiensis* is the second most important malaria vector in the *An. gambiae* complex in Africa and in some parts of Africa it is the sole malaria vector and a big threat to human health.

It is the main malaria vector in Chad, a country in which 95% of the population is at risk of malaria (Kerah-Hinzoumbe *et al.* 2009). Figure 5.1 shows the species distribution of *An. gambiae s.s.* and *An. arabiensis* in Chad illustrating the dominance of *An. arabiensis*.



(A) *An. gambiae* s.s. (B) *An. arabiensis* (C) sympatry

Figure 5.1: Putative species range of *Anopheles gambiae s.s.* and *An. arabiensis* in Chad.

Maps of Chad showing the predicted geographic extent of (A) *An. gambiae s.s.* and (B) *An. arabiensis* (2010 Malaria Atlas Project) as well as (C) relative proportions of the two species (*An. arabiensis* in blue) south of Ndjamena found in a recent study (Kerah-Hinzoumbe *et al.* 2008).

The mechanisms responsible for insecticide resistance in *An. arabiensis* are less understood than for its sibling species *An. gambiae s.s.*. The presence of target site alleles in *An. arabiensis* populations has been detected in various African countries. The 1014F allele producing the *kdr* pyrethroid resistant phenotype was reported for the first time in *An. arabiensis* in 2004, being found at low frequencies in samples from Burkina Faso, West Africa (Diabate *et al.* 2004a). In 2006 Kulkarni and co-authors reported this 1014F *kdr* allele in *An. arabiensis* samples from Tanzania, East Africa (Kulkarni *et al.* 2006). The 1014S *kdr* allele was first reported in a single specimen (one out of 572) of natural populations of *An. arabiensis* from Kenya (Stump *et al.* 2004) and subsequently both *kdr* alleles were detected in the same four individuals in its neighbour country Uganda (Verhaeghen *et al.* 2006). In West Africa the 1014S *kdr* allele was first reported in *An. arabiensis* in Benin (Djegbe *et al.* 2011) and has also reached high frequencies in some urban areas in Burkina Faso (Jones, unpublished). Both alleles now have a wide distribution across Africa (Lynd *et al.* 2010; Ranson *et al.* 2011).

The pattern of emergence of *kdr* mutations in *An. arabiensis* suggests that this mutation is a recent event and its spread is ongoing. A polymorphism upstream of the 1014 *kdr* locus differentiates *An. arabiensis* from *An. gambiae s.s.* and this has been used to determine that the *kdr* mutation in *An. arabiensis* is a *de novo* event rather than the result of introgression from *An. gambiae s.s.* (Diabate *et al.* 2004b).

Most of the above studies are not supported by accompanying bioassay data. A more recent study reports the highest *kdr* allele frequency observed in *An. arabiensis* to date (Yewhalaw *et al.* 2010) with >98%, in a DDT and permethrin resistant population from Ethiopia. In Sudan resistance in *An. arabiensis* to DDT, permethrin and malathion as well as high frequencies of 1014F have been reported (Abdalla *et al.* 2008).

In short, across Africa resistance in *An. arabiensis* to permethrin and DDT has been reported, as well as the presence of both the L1014F and L1014S mutation in the sodium channel.

In 2008 reduced susceptibility to DDT and permethrin in *An. arabiensis* populations from Cameroon were reported in conjunction with total absence of the *kdr* allele (Chouaïbou *et al.* 2008). Furthermore, a study on South African *An. arabiensis* populations found permethrin resistance associated with elevated levels of detoxifying enzymes, supported by synergist bioassays again combined with absence of *kdr* (Mouatcho *et al.* 2009).

Similarly in Chad high resistance to permethrin was found in *An. gambiae* and *An. arabiensis* (Kerah-Hinzoumbe *et al.* 2008). But, whereas in the *An. gambiae* survivors the *kdr* resistance allele was present at a high frequency, all *An. arabiensis* survivors carried the susceptible allele at position 1014. These results suggest that alternative resistance mechanisms such as metabolic mechanisms play a role in insecticide resistance in *An. arabiensis*.

In 2008 Müller used the *An. gambiae* detox chip for a microarray study on pyrethroid resistance in *An. arabiensis* from Northern Cameroon (Müller *et al.* 2008a). A number of genes, including P450s, superoxide dismutases and GSTs, were identified as up-regulated in the resistant populations.

5.2 Aims of the chapter

Chad was chosen as a study site due to the high level of insecticide resistance in absence of any known target-site mutations, suggesting presence of metabolic resistance mechanisms (Chapter 2.1.2).

In this chapter QTL mapping was used to identify the major loci conferring pyrethroid resistance in a population of *An. arabiensis* from Ndjamena, Chad. Crosses were established between susceptible and resistance strains maintained in the laboratory at LSTM.

The specific objective of this work is to identify the major loci conferring permethrin resistance in the Ndja population.

5.3 Methods

5.3.1 Field collections, establishment and characterization of mosquito colonies

General mosquito maintenance techniques and details about field collected *An. arabiensis* mosquitoes, establishment of colonies and genetic crosses as well as phenotypic and genotypic characterization of mosquitoes are described in Chapter 2.1.2.

5.3.2 Selection of informative markers and genotyping of mapping families

Informative markers were selected on the basis of the genotypes of the parental female F_0 and F_1 individuals. Details of microsatellite and SNP markers and PCR protocols are contained in Chapter 2.4.1 and 2.4.2.

5.3.3 Linkage mapping and QTL analysis

Linkage mapping and QTL analysis were performed as described previously in Chapter 2.4.3.

5.4 Results

5.4.1 Insecticide susceptibility profile of field collected strains and the genetic crosses

Mosquito samples for this study were provided by Dr. Charles Wondji, LSTM. Results for insecticide susceptibility assays for the parental strains, and the lines resulting from the genetic crosses, are given in Chapter 2.1.2.

5.4.2 Genotyping of experimental populations

5.4.2.1 Informative markers and genotyping of mapping families

The F_2 progeny of two families, one from each of the resulting reciprocal crosses were selected for genetic mapping. For each family 48 female mosquitoes were genotyped; no male mosquitoes were used. F_0 and four F_1 specimens per family were confirmed to be *An. arabiensis* of the wildtype 2Rb karyotype.

Details for all microsatellite markers and SNP markers chosen for genotyping of F_2 progeny of the *An. arabiensis* families for QTL mapping are shown in Table 2.7 in Chapter 2.4.2.

Each marker was tested for Hardy-Weinberg equilibrium (dead and alive separate as well as combined) and the results are shown in Appendices 9.5 and 9.6. Single marker analysis was performed for each family to test for association between genotype and resistance phenotype (haploview 4.2). Chi^2 and p-values are listed in Table 5.1 below.

family 2		marker	location ¥ [Mb]	χ^2	p-value	sig	df
X chromosome		cyp4h24	1.4	0.70	0.40		2
		53	3.6	0.42	0.52		2
		80	7.0	1.11	0.29		2
		H7	11.2	2.00	0.16		1
		678	19.2	0.18	0.68		1
Chromosome 2	2R	803	1.9	0.46	0.50		1
		199	3.5	3.22	0.07	*	1
		24	3.5	3.22	0.07	*	1
		757	14.7	2.53	0.11		1
		799	24.9	0.82	0.37		1
		2R_si_5	24.6	0.55	0.46		1
		cyp6p3	28.5	0.19	0.66		1
		590	28.9	4.62	0.03	**	2
		cyp304c1	32.0	2.65	0.10	*	1
		720	33.2	0.20	0.65		1
		770 $	40.4	0.18	0.67		1
		786	42.7	0.10	0.75		1
	2L	ex27	24.3	0.45	0.50		1
		787	26.8	0.03	0.87		2
		603 $	42.0	0.41	0.52		2
		675	46.9	0.01	0.91		1
Chromosome 3	3R	776	2.5	0.21	0.65		2
		59	4.2	0.03	0.86		2
		812	6.2	0.27	0.60		2
		249	8.8	0.06	0.81		2
		30C1	9.6	0.14	0.71		2
		6M19	13.1	0.30	0.59		2
		119	14.8	0.10	0.75		2
	3L	811	14.8	0.04	0.85		2
		577	19.8	0.55	0.46		1
		46C3	40.4	0.81	0.37		2
		46C2	41.4	0.03	0.86		2

Table continued

family 3		marker	location ¥ [Mb]	χ^2	p-value	sig	df
X chromosome		cyp4h24	1.4	1.05	0.30		1
		53	3.6	0.09	0.77		1
		80	7.0	0.98	0.32		2
		H7	11.2	2.98	0.08	*	1
		678	19.2	1.08	0.30		1
Chromosome 2	2R	803	1.9	11.92	0.00	***	2
		199	3.5	8.94	0.00	***	2
		24	3.5	8.51	0.00	***	2
		757	14.7	9.08	0.00	***	2
		799 $	24.9	0.88	0.35		1
		26 $	24.6	8.36	0.00	***	1
		cyp6p3 $	28.5	4.04	0.04	**	1
		590	28.9	11.26	0.00	***	2
		cyp304c1	32.0	0.18	0.67		1
		117	40.4	4.94	0.03	**	1
		786	42.7	0.00	1.00		2
	2L	787	26.8	0.17	0.68		2
		772	42.0	4.84	0.03	**	1
		603	42.0	1.86	0.17		1
		675	46.9	0.00	0.99		2
Chromosome 3	3R	776	2.5	0.00	1.00		2
		59	4.2	0.05	0.83		2
		812	6.2	0.68	0.41		2
		249	8.8	0.62	0.43		2
		30C1	9.6	0.02	0.88		1
		6M19	13.1	2.40	0.12		2
		119	14.8	0.17	0.68		2
		chymotrypsin	43.9	0.20	0.65		1
	3L	811	14.8	0.11	0.74		1
		46C3	40.4	0.21	0.64		2
		46C2	41.4	1.62	0.20		2

Significance *:0.1 **:0.05 ***:0.01

$ those markers were not implemented in the genetic maps

¥ According to vector base (08/2010)

df: Degrees of freedom; df(1) semi-informative and df(2) fully informative marker

Table 5.1: A list of χ^2 and p-values for microsatellite and SNP markers.

Markers were tested for significant association with phenotype. The markers are shown in physical order.

5.4.3 Linkage mapping

A summary of the linkage maps for each of the three chromosomes is shown in Table 5.2 and the maps themselves are shown in Figure 5.2. Genetic maps generated from the combined data set from both families are shown in Figure 5.3. Where alternative marker orders were proposed by the linkage software, maps were generated to reflect the optimal physical maps, i.e. to reflect the marker order along the chromosomes according to their physical location (based on the *An. gambiae* genome on vectorbase).

The marker order on the generated maps is very similar for maps generated for the single families and the combined data. The map for chromosome 2 for family 2 correlates with the optimal physical map, and for family 3 it deviates only in the position of markers 772 and 787 on 2L which are expected to be the other way around. The marker order on the combined map for chromosome 2 correlates to the optimal physical order of markers, with the exception of markers 590 and cyp6p3 on 2R which are expected to be the other way around.

There are minor differences when comparing the marker orders of chromosome 3 for the single families and the combined data maps. For family 3 the generated map is identical to the optimal physical map and for family 2 and the combined map, there are some variations in the positions of neighbouring markers (776, 59 and 812 on 3R, and 577, 46C3 and 46C2 on chromosome 3L).

	X Chromosome		**Chromosome 2**		**Chromosome 3**		all
	map [cM]	no of markers	map [cM]	no of markers	map [cM]	no of markers	average resolution¥
family 2	70.3	5	85.8	14	55.7	11	7.1
family 3	54.1	5	31.5	11	57.0	11	5.3
both combined	62.0	5	81.7	14	54.4	12	6.4

¥ Average resolution in cM per marker

Table 5.2: Summary of map sizes and resolution.

The values for single families and the combined data are based on map size in centi Morgan and the number of markers.

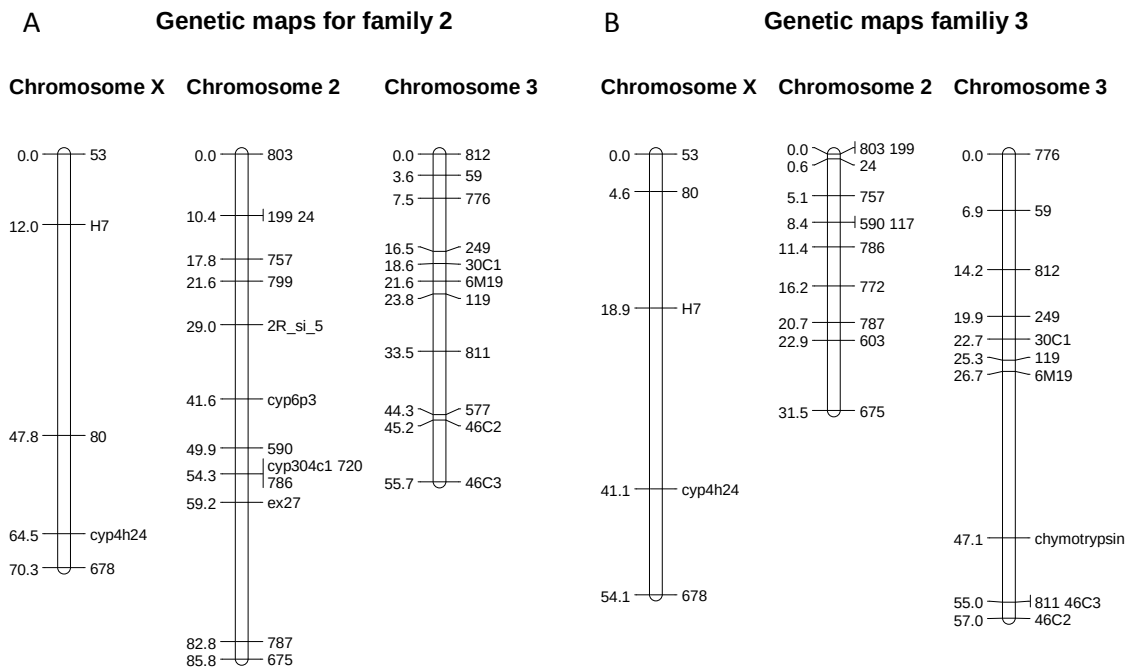


Figure 5.2: Summary of the genetic linkage maps for the separate families.

Chromosomes X, 2 and 3 for families (A) 2 and (B) 3 when analysed separately. A scale of genetic distance in centimorgan is shown.

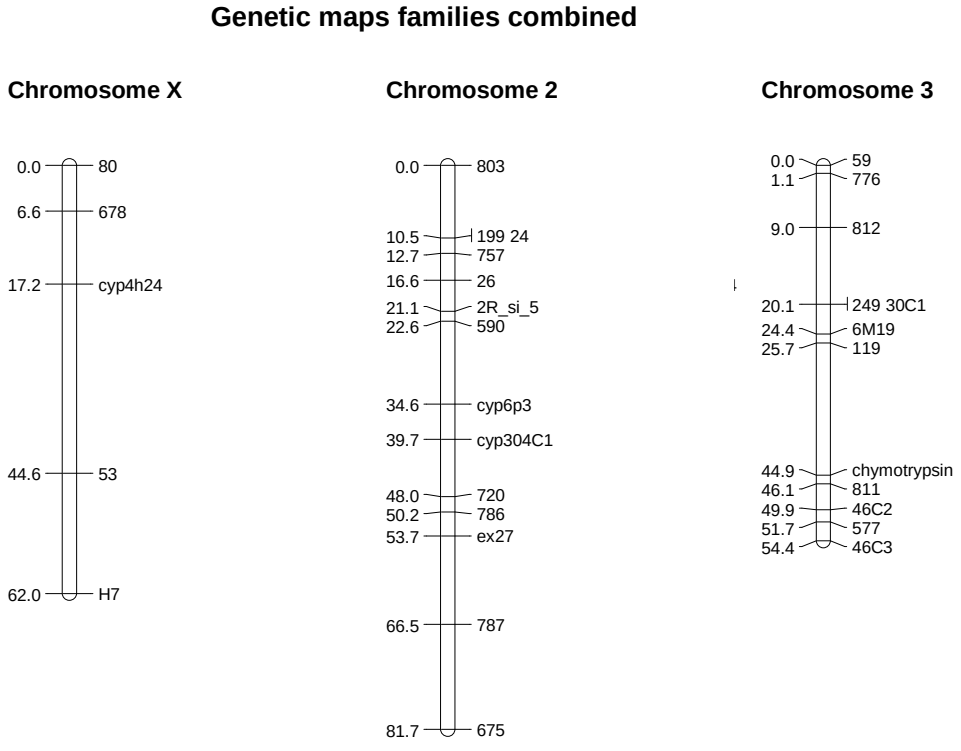


Figure 5.3: Genetic linkage maps using the combined data.

Linkage maps of chromosomes 2 and 3 are shown, as optimal physical map, the map for the X chromosome corresponds to the optimal genetic map. A scale of genetic distance in centimorgans is shown.

Plotting mortality rate as a function of the number of alleles inherited from the resistant parent (the mother in family 2 and the father in family 3) is a simplified means of assessing the association between a given genotype and the resistance phenotype. Figure 5.4 shows plots of the fully informative markers on chromosome arm 2R for both families. Markers 757 and 590 show a very strong association between genotype (i.e. number of a certain allele) and phenotype (here visualized as mortality). Markers 199, 24 and 26 show a similar association to 757 and 590 but are not shown, as they were fully informative in family 3 only. Marker 786 does not show association with mortality rates around 50% irrespective of the number of alleles from resistant parent.

Both markers (757 and 590) associated with resistance on the 2R chromosome have an additive effect in both families, with individuals homozygous for the resistant allele having a greater chance of surviving exposure to permethrin than heterozygotes.

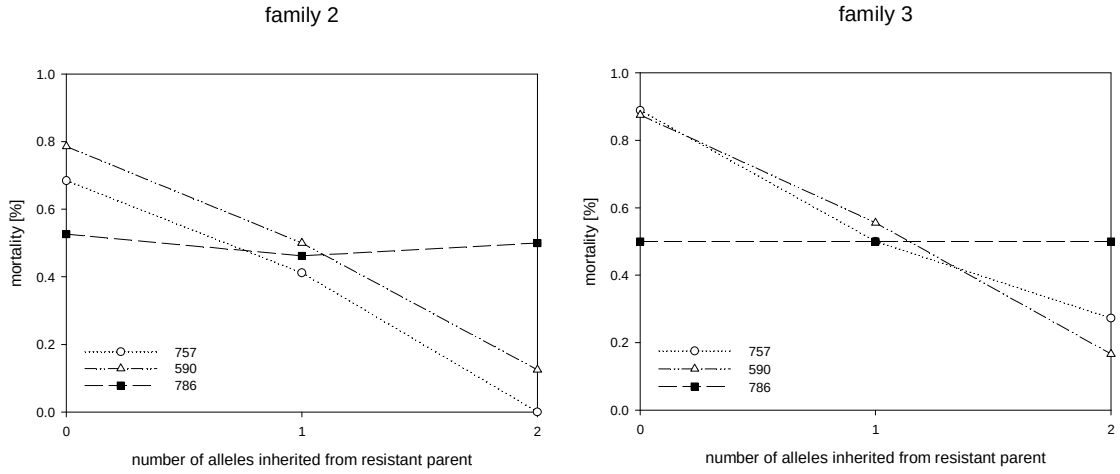


Figure 5.4: Genotype-phenotype association.

Plot of mortality rate as a function of alleles inherited from the resistant parent for the fully informative markers on chromosome 2R for each family.

5.4.4 QTL analysis

The software Windows QTL cartographer version 2.5 (Wang *et al.* 2005) was applied to test for linkage between resistance phenotype and genotype. Both Interval Mapping and Composite Interval Mapping were performed and the results compared. The results from the two families were analysed for QTLs separately first and then a QTL analysis based on the combined data and maps was performed. IM and CIM are largely in agreement for family 3 (with the exception of the X chromosome) as well as for the combined data. Interval mapping and CIM diverge considerably on chromosome X and 2R for family 2. In subsequent discussions, QTL are only considered significant if the LOD score exceeds the value 2.5 and the position of the QTL is supported by both IM and CIM likewise.

No QTL were identified in family 2 (Figure 5.6). However, significant LOD scores were observed on chromosome 2R in family 3 (Figure 5.7) and this QTL was further strengthened in the combined family analysis (Figure 5.8). Figure 5.5 shows the LOD values for CIM for the combined data set including the linkage maps to indicate QTL position.

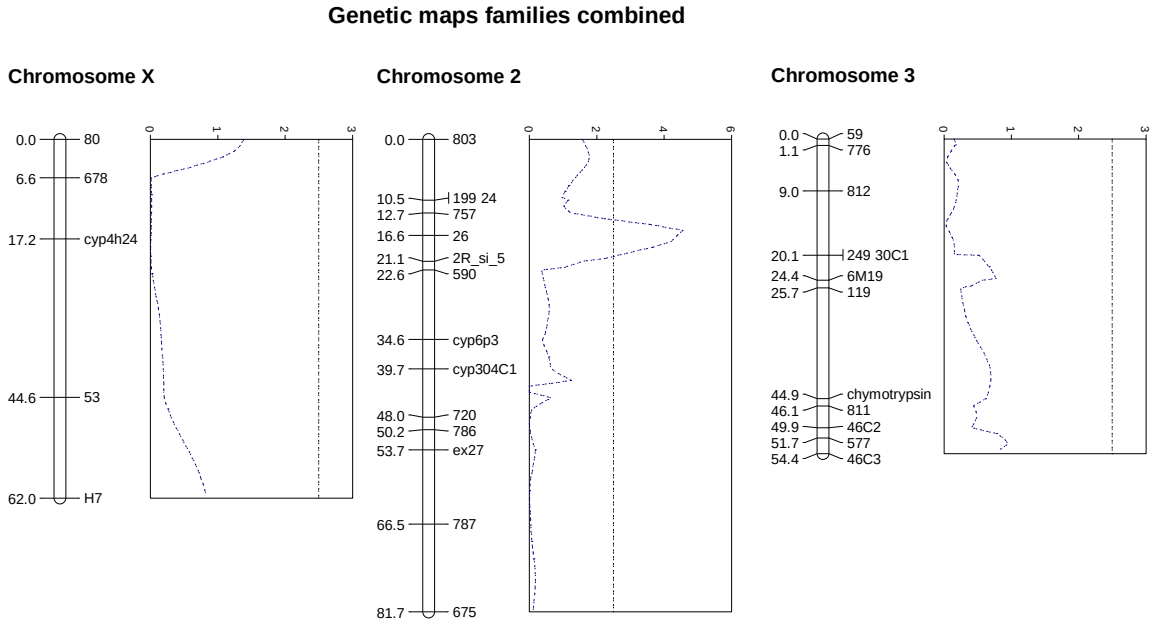


Figure 5.5: Plots of LOD scores estimated by CIM for the combined data.

The linkage maps are included to orientate the QTL positions. The dotted line along the top of each graph represents the threshold value for LOD, set at 2.5.

Family 2

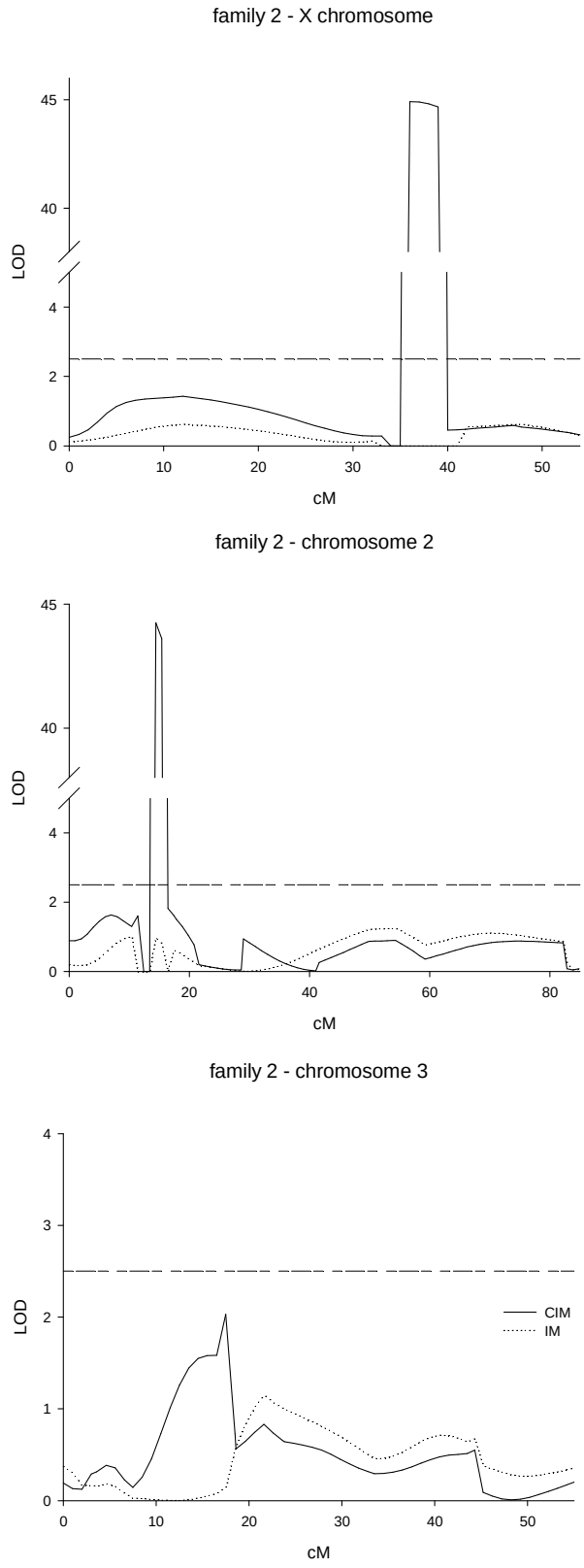


Figure 5.6: Plots of LOD scores associated with permethrin resistance.

Shown are the three chromosomes for *An. arabiensis* for family 2. Solid lines represent LOD estimated by composite interval mapping and dashed lines represent LOD estimated by interval mapping. The straight line along the top of each graph represents the threshold value for LOD, set at 2.5.

Family 3

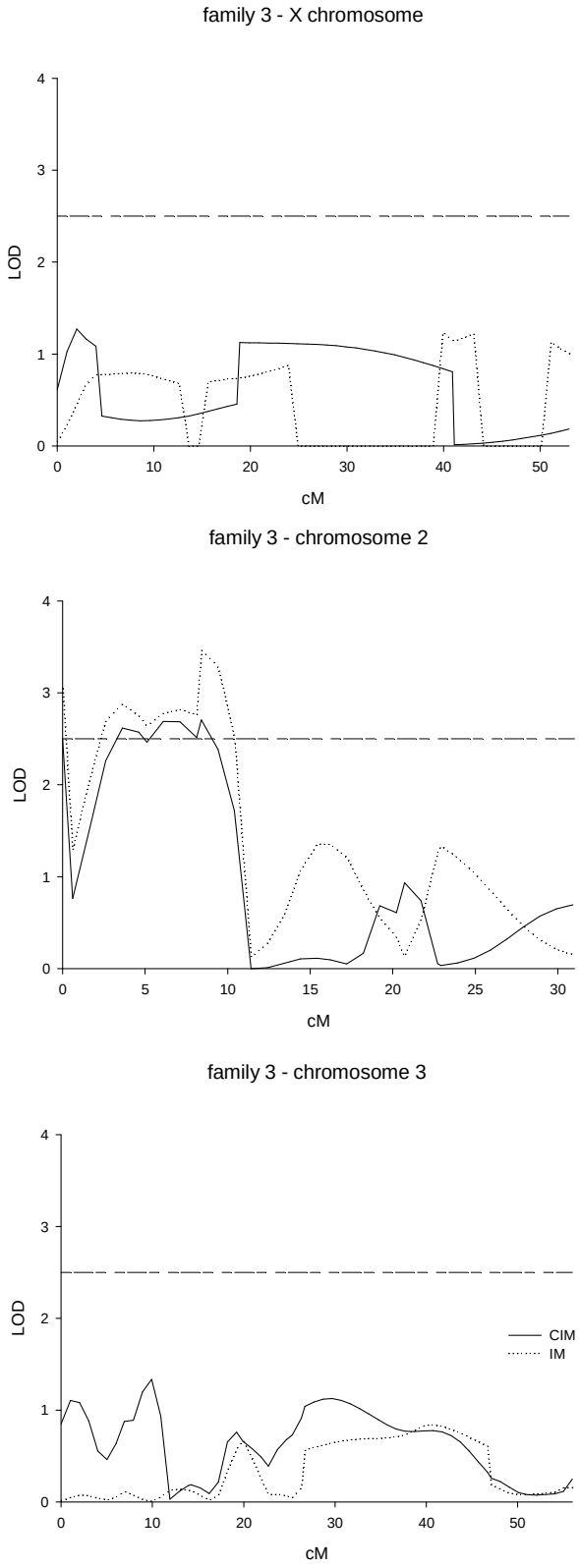


Figure 5.7: Plots of LOD scores associated with permethrin resistance.

Shown are the three chromosomes for *An. arabiensis* for family 3. Solid lines represent LOD estimated by composite interval mapping and dashed lines represent LOD estimated by interval mapping. The straight line along the top of each graph represents the threshold value for LOD, set at 2.5.

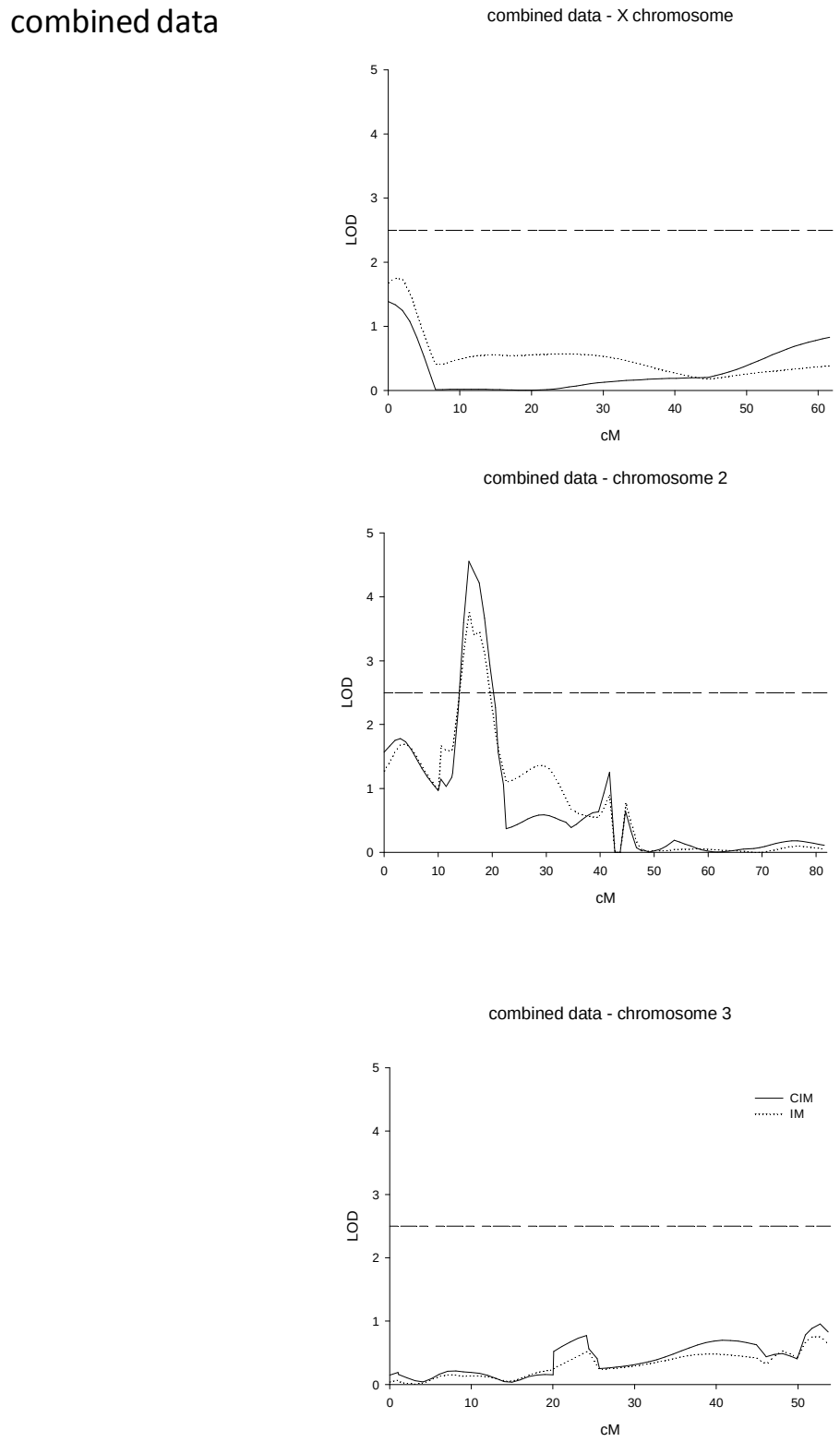


Figure 5.8: Plots of LOD scores associated with permethrin resistance.

Shown are the three chromosomes for *An. arabiensis* for the combined data of both families. Solid lines represent LOD estimated by composite interval mapping and dashed lines represent LOD estimated by interval mapping. The straight line along the top of each graph represents the threshold value for LOD, set at 2.5.

Multiple Interval Mapping (MIM) confirms the presence of a single QTL on chromosome 2R in both families and estimates its genetic variance (δ_g^2) as 14.4 % and 18.9 % of the phenotypic variance (δ_p^2) of permethrin resistance for family 2 and family 3 respectively (Table 5.3). For the combined data this QTL explains 24.4 % of the phenotypic variance (resulting from 0.5% for additive and 23.9% for dominance effect). No additional QTLs were detected using MIM.

	δ_g^2 [%]	nearest marker	¥	genetic distance [cM]	effect in %δ_g^2
family 2					
	14.4	590	2R	50.0	A: -4.8
					D: 19.2
family 3					
	18.9	590/ 117	2R	8.5	A: 22.9
					D:-4.0
combined					
	24.4	26	2R	16.7	A: 0.5
					D: 23.9

δ_g^2 [%] genetic variance in % of overall estimated phenotypic variance;
¥ chromosome and location in genetic distance for nearest marker;
A for additive effect; D for dominant effect

Table 5.3: Multiple Interval Mapping estimates of QTL position and associated genetic and phenotypic variance and additive and dominance effects associated with survival QTL in *An. arabiensis*.

5.5 Discussion

Preliminary analysis, the results of which are presented in Chapter 2.1.2.1, did not identify any known *kdr* alleles in the pyrethroid resistant Ndja population. The objective of this chapter was to identify the loci responsible for the resistance phenotype. At the same time, absence of target-site resistance in the sodium channel could be confirmed.

This is the first QTL mapping experiment to investigate the trait insecticide resistance in *An. arabiensis*. A QTL mapping approach was applied to the F_2 progeny and a single QTL on chromosome arm 2R was identified. The QTL on chromosome 2R spans chromosomal division 10C-13C.

The QTL on chromosome 2R is very pronounced in family 3, but in family 2 this QTL is only detectable by IM, not CIM. After combining the family data the QTL becomes more pronounced, suggesting that this region is involved in resistance in family 2 as well. A possible explanation is that many of the markers in the vicinity of the QTL are only semi informative for family 2, hence reducing the genetic information available. (Table 5.1 degrees of freedom).

Multiple Interval Mapping confirmed the single QTL on 2R in both families and in the combined data set estimating its genetic variance at between 14.4-24.4%. No additional QTLs were identified using MIM. In the combined data set the majority of the genetic variance is due to dominance effect at this locus. A similarly strong dominance effect is estimated for family 2 whereas in family 3 the additive effect is much stronger (Table 5.3). The single marker analysis however suggests that the effect of marker 590 in this locus is additive in both families (Figure 5.4). For a locus of metabolic resistance due to enzymatic up-regulation an additive genetic effect would be expected. The genetic variance of this QTL is estimated to account for around a quarter of the overall phenotypic variance, much less than the >60% genetic variance estimated for a previously identified QTL in *An. funestus* (Wondji *et al.* 2007). It is probable that there are other minor QTL present which could not be detected by the experimental design.

The observed discrepancies in the genetic maps of family 2 and 3 could be due to the presence of different inversions in these families. *An. arabiensis* is fixed for the 2La inversion mentioned in Chapter 4. But inversions are highly concentrated on the right arm of chromosome 2 (2R) in *An. gambiae s.s.* and *An. arabiensis* (Coluzzi *et al.* 2002; Coluzzi *et al.* 1979). In a population where both inversion karyotypes are present, inversions make linkage mapping very difficult as no recombination is possible and markers within the inversion will appear as if in linkage disequilibrium with others when they are not. In case the parental strains, that were crossed to yield the F_2 mapping population, were not fixed for the same chromosomal arrangement, this will undoubtedly have distorted the recombination events on 2R.

Three of the 2R inversions have been molecularly characterized in *An. gambiae s.l.* and the karyotype of these can be determined via a PCR based assay: 2Rj only in *An. gambiae s.s.* (Coulibaly *et al.* 2007), 2Rb (Lobo *et al.* 2010) and 2Rd (Mathiopoulos *et al.* 1998). The boundaries of the identified QTL are very large (~14Mb) and inversion 2Rb may possibly be overlapping with it. Molecular karyotyping however, showed, that all tested individuals (parental females and four F_1 individuals for each family) were fixed for the ancestral 2Rb arrangement.

An interesting result from this mapping experiment is the absence of a QTL corresponding to the sodium channel located on chromosome 2L in *An. gambiae s.l.*. One single QTL was identified on 2R but is unlinked to *kdr* in the sodium channel. Previous studies of genetic mapping of resistance have identified *kdr* as a major QTL in populations possessing target-site and metabolic resistance (Ranson *et al.* 2004; Saavedra-Rodriguez *et al.* 2008). It is possible that minor QTLs are present and have remained undetected in the Njda population, but this is very unlikely for *kdr*. These results suggest that target-site resistance is not involved in conferring the resistance phenotype; a conclusion supported by the absence of any known *kdr* mutations in this *An. arabiensis* strain.

At least 47 cytochrome P450 genes are encoded on chromosome 2R (Ranson *et al.* 2002a) with the majority of these located in three clusters (gene families *cyp4*, *cyp325* and *cyp6*) (see figure 5.9) (Ranson *et al.* 2002b). The boundaries of the identified QTL are ~10cM with the flanking markers 757 and 590 spanning a genome region of ~14 Mb including 910 annotated genes (vectorbase biomart 03/2012). This region includes the three mentioned P450 gene clusters and 25 cytochrome P450s in total (3.2% of the genes within the QTL boundaries) with a fold enrichment of 4.2 (DAVID enrichment analysis, IPR001128 Cytochrome P450) (Appendix 9.7). This affirms the QTL is in a region enriched in P450 genes.

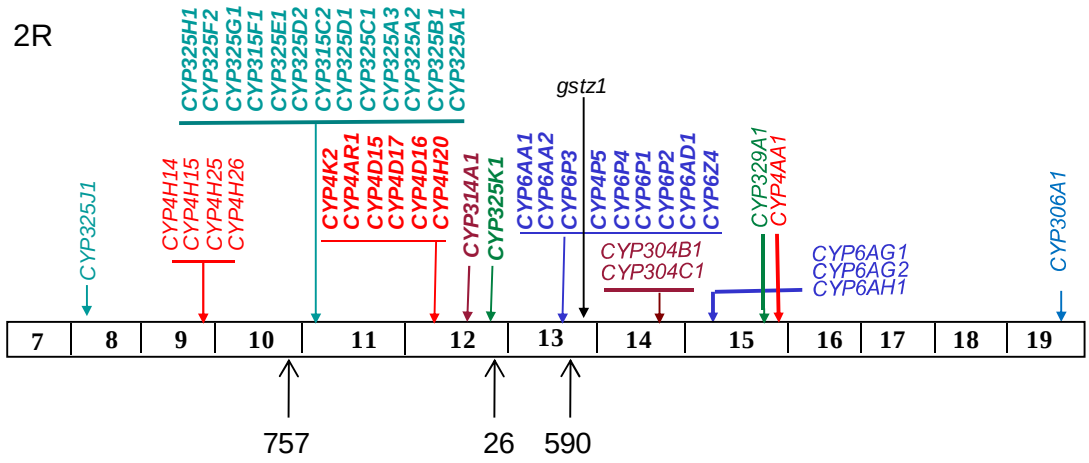


Figure 5.9: P450s and *gstz1* on polytene chromosome arm 2R.

Arrows from top roughly indicate locations of P450 genes in the *An. gambiae s.s.* genome. The black arrows from below indicate locations of the four markers that define the QTL position (adapted from Ranson, unpublished).

The involvement of P450s in the observed pyrethroid resistance in the Ndja strain is supported by the PBO synergist assay in which mortality of 100% was restored by exposing the Ndja strain to 4% PBO prior to 0.75% permethrin (Chapter 2.1.2.1). Biochemical assays also showed increased levels of monooxygenases in the Ndja strain compared to the Moz strain (Charles Wondji and Matthew Parry, unpublished data).

One of the few microarray studies on insecticide resistance in *An. arabiensis* has found four cytochrome P450s overexpressed in the resistant population (Müller *et al.* 2008a). Interestingly three out of the four are located on

chromosome 2R, *cyp6p3*, *cyp325c1* and *cyp6ag1*. Previous studies in mosquito vectors have found a number of P450s located on chromosome 2R to be involved in insecticide resistance. The gene *cyp6p3* has been shown to be overexpressed in a pyrethroid resistant field population of *An. gambiae s.s.* (Müller *et al.* 2008b) and the authors have also shown the enzyme CYP6P3 has the ability to metabolise the insecticide permethrin. A QTL mapping study on crosses of two *An. funestus* lab strains identified a QTL on chromosome 2R (Wondji *et al.* 2007), with the two most closely linked markers in a cluster of genes of the *cyp6* family. This first mapping approach was done on the F_2 progeny and the QTL on 2R was consistently identified in the majority of the families and both reciprocal crosses. The boundaries of the QTL were reduced by fine scale mapping on F_6/F_8 progeny of the original crosses and this confirmed the previous QTL as well as reducing its boundaries resulting to implicate two candidate genes *cyp6p4* and *cyp6p9* (Wondji *et al.* 2009). *Cyp6p9* is the putative ortholog of *cyp6p3* in *An. gambiae*, lending further support for the role of this P450 in pyrethroid resistance.

A number of genes encoding GSTs involved in DDT resistance in *An. gambiae s.s.* (Ranson *et al.* 2001) are found on 2R, but these are located in divisions 18 and 19 outside the identified QTL. One GST lies within the QTL, *gstz1*, located in division 13C (Ding *et al.* 2003). The *gstz1* belongs to the non-insect-specific zeta class of cytosolic GSTs. Zeta-class GSTs have recently been implicated in permethrin resistance in lepidopterans (Yamamoto *et al.* 2009) but there has been no link established between Zeta-class GSTs and insecticide metabolism in other insects.

The QTL detected on chromosome 2R spans a region of approx. 10 cM/ 14 Mb. The presence of three clusters of P450 genes within the boundaries of this QTL is interesting. Using F_2 generations for mapping limits the resolution that can be obtained in the QTL analysis. Unfortunately, as in the previous chapter, the absence of advanced intercross lines in this study limited the resolution that could be obtained by QTL mapping.

Chapter 6. Microarray analysis of permethrin resistant *An. arabiensis* from Chad

6.1 Introduction

In this study a microarray chip based on the *An. gambiae s.s.* sequence will be used for hybridisation with *An. arabiensis* samples. The two species are very closely related sibling species, but sequence polymorphisms between the *An. arabiensis* and the probe sequences on the microarray are expected to reduce hybridisation. Cross species arrays have been successfully performed before using the detox-chip (Müller *et al.* 2008a; Vontas *et al.* 2007) and a number of candidate genes have been identified in *An. arabiensis* and *An. stephensi*. This study will be the first use of the whole genome array, which was designed based on *An. gambiae* sequence data (Mitchell *et al.* 2012) for *An. arabiensis*.

The previous chapter used QTL mapping to identify a major locus associated with pyrethroid resistance in *An. arabiensis* from Chad (Ndja strain). Here we employ microarray in an attempt to identify candidate genes that may be responsible for the resistance trait.

Finding sympatric insecticide susceptible *Anopheles* mosquitoes in Africa is becoming increasingly difficult as pyrethroid resistance is spreading rapidly. As no susceptible strain from the study area was available, a susceptible strain from Mozambique (Moz) was chosen for comparison with Ndja. This strain was expected to have a different genetic background due to the geographical distance. To overcome this, a backcross generation was generated by crossing the F_1 population back to the resistant parental Ndja strain. This backcross generation would be enriched in resistance alleles from Ndja, yet share a common genetic background with the susceptible Moz strain. Microarray experiments compared the expression in the backcross population to the susceptible strain from Mozambique.

6.2 Materials and Methods

6.2.1 Microarray experimental procedure

6.2.1.1 Experimental and microarray design

The genetic crosses between the Ndja and Moz strains of *An. arabiensis* were established by Dr. Charles Wondji, LSTM, and Matthew Parry as described in Chapter 2.1.2.2.

The microarray experiment compared RNA pools of exposed (survivors of the backcross (1 h to 0.75% permethrin according to WHO standards)) to non-exposed (susceptible Moz strain) comprising three independent dye swaps, resulting in six arrays in total. The experiment was performed on the Agilent 8x15K'AGAM_15K'-chip (8 arrays per slide) and scanned on an Agilent two-colour scanner following an experimental protocol developed by Dr. Sara Mitchell. The Agilent chip was custom made to the design of Dr. Sara Mitchell and Dr. Pie Müller (LSTM) based on the *An. gambiae* genome (Mitchell 2011).

The Agilent 8x15K microarray enables high coverage across the whole genome and at the same time reduced costs and increased though-put, with eight arrays on a single slide. The entire coding transcriptome from the latest version of the *An. gambiae s.s.* genome (Ensembl AgamP3.5, 2009) was employed for probe design, with additional probe coverage for the detoxification gene families previously implicated in insecticide resistance (David *et al.* 2005; Ranson *et al.* 2002a). The design process was performed via the Agilent on-line design package eArray (https://earray.chem.agilent.com/earray/). Further details of the array design are contained in (Mitchell 2011).

6.2.1.2 Sample preparation and labelling

RNA extractions, cDNA synthesis and labelling reactions were performed independently for each biological replicate. Total RNA was extracted from pools of 10 mosquitoes (3-day-old unmated females) using a PicoPure™ RNA isolation kit (Arcturus Technologies, USA) according to the

manufacturer's instructions. Total RNA quantity was assessed using the Nanodrop spectrophotometer (NanoDrop™ (Thermo Scientific)) before further use. As potential degradation of RNA is not detectable via NanoDrop readings the quality of the RNA was assessed using a 2100 Bioanalyzer (Agilent Technologies, UK) and RNA 600 Nano Kit (Agilent Technologies, UK) according to the manufacturer's guidelines. RNA pools which had been determined to be of good quality were selected for microarray labelling. The concentration of each pool was re-measured using the NanoDrop™ (Thermo Fisher Scientific, USA) on the day of labelling and 100ng in a maximum volume of 1.5µl were taken from each pool for labelling.

Samples were labelled with both cyanine dyes, Cy3 and Cy5, using the Low Input Quick Amp Labelling Kit, two-color (Agilent Technologies, UK) with the inclusion of Spike-in controls. All sample incubation steps were performed in a thermal cycler (Dyad, Biorad/ MJ Research). Purification of amplified and labelled cRNA was performed using Qiagen RNeasy mini spin columns (Qiagen) and eluted in 30µl of RNase-free water (Invitrogen). Amplification and dye incorporation of samples were assessed using the NanoDrop™ (Thermo Fisher Scientific, USA).
Before hybridization the quality of each labelled cRNA sample was checked using the Agilent 2100 Bioanalyzer.

6.2.1.3 Microarray hybridization

For cRNA fragmentation and hybridisation the Gene Expression Hybridization kit (Agilent Technologies, UK) was used following the manufacturer's protocols. Labelled cRNA samples to be compared on a single array were combined and fragmented for 30 minutes at 60°C. The reaction was stopped by adding 55µl of 2x GEx Hybridization Buffer Hi-RPM (Agilent Technologies, UK) to the samples. Each tube was spun down briefly and kept on ice until hybridisation.

A clean 8x15k gasket slide was placed into a SureHyb chamber base (Agilent Hybridization Chamber Kit) and 100µl of each hybridisation sample was dispensed into the centre of each gasket. A 8x15k array slide

was lowered onto the gaskets and the chamber fully assembled. The array sandwich was then put into the hybridization oven (Agilent Technologies, UK) and rotated at 10 rpm for 17 hours at 65°C.

After hybridisation, the sandwich was disassembled and arrays were washed in GE wash buffers 1 and 2 (Agilent) according to manufacturer's instructions. To protect the arrays against ozone-induced degradation of cyanine dyes and to reduce the background noise, they were additionally treated with acetonitrile (Sigma-Aldrich, UK) followed by a Stabilization and Drying Solution (Agilent Technologies, UK). Arrays were stored in protective boxes (Corning) until scanned.

6.2.1.4 Microarray scanning and feature extraction

Arrays were scanned using G2505C/G2539A Series Microarray Scanner utilising the default profile and settings according to the Agilent Microarray Scanner System User Manual (v.7.0). High (100% photo-multiplier tube(PMT) and low (10% PMT)) extended dynamic range (XDR) scan images were combined and extracted using Feature Extraction (FE) software GE2_10.5_Dec08 (Agilent Technologies, UK) and the custom array grid template (022094_D_F_20081124.XML). QC reports were consulted to give an indication of array quality. A QC score of 12/12 indicates the twelve main array parameters were passed. QC parameters include signals from spike-in controls, spatial distribution of outliers and signals from non-control spots. In addition ".tif" scan images were visualized in the FE software under the 'LOG scale' view to check for artefacts on the array which may have affected spot signal intensities.

6.2.2 Microarray data analysis

The GeneSpring GX version 11.0 software (Agilent Technologies, Cheshire, UK) was used for statistical analysis of the microarray data. Mean expression ratios were submitted to a t-test against Zero with a multiple testing correction (Benjamini-Hochberg false discovery rate). Genes showing both t-test p-values ≤0.01 and a fold change value ≥2-fold were considered significantly differentially expressed between the two samples.

6.2.2.1 Enrichment analysis

Entities considered significantly differentially expressed served as input file for DAVID Gene Functional Classification software v.6.7 to perform an enrichment analysis (Huang *et al.* 2009a, b) with a list of all genes targeted by the probes present on the Agilent chip submitted as background.

The analysis was run using the DAVID default settings and Chromosome as additional general annotation option. Both sets of down-regulated as well as up-regulated genes were compared to the background (genes probed on microchip).

6.2.3 Quantitative RT-PCR on candidate genes

One transcript, *cyp4h24* (AGAP013490), was selected from the genes that were significantly differentially expressed in the microarray analysis for qPCR validation.

Materials and methods for the quantitative PCR were as described in Chapter 2.3. Primers for the *cyp4h24* RT-PCR were designed against the *An. arabiensis* consensus from sequence results from eight samples.

Absolute quantification using standard curves generated with plasmid DNA was performed. Normalization with data from two control genes, *cpr* (AGAP000500) and *s7* (AGAP010592), was carried out. The RNA samples used for the microarray experiment (i.e. backcross survivors and samples belonging to the Moz strain) were used in the qPCR but an additional three biological replicates of pools of 10 females (3-day-old, unmated) from the Ndja strain were also included.

6.3 Results

6.3.1 Microarray

All other arrays passed with an overall QC value of 12 out of 12. After data cleaning, 7961 entities were retained out of an initial 14999. Of these, 281 entities were significantly differentially expressed satisfying a corrected p-value (FDR p) cut-off of ≤0.01 and fold change cut-off ≥2.0. The numbers of down-regulated and up-regulated entities in the backcross samples are shown in Table 6.1. Tables listing the top 50 up- and top 50 down-regulated gene probes can be found in Appendices 9.8 and 9.9.

Two strategies were employed for data analysis: an enrichment analysis and a candidate gene approach.

Data analysis parameters	Experiment:	permethrin
	corrected p-value cut-off:	0.01
	Selected Test :	T Test Against Zero
	p-value computation:	Asymptotic
	Multiple Testing Correction:	Benjamini-Hochberg
Differentially expressed	281 entities out of 7961 satisfied the corrected p-value cut-off of 0.01 and fold change cut-off ≤2	
	up-regulated	144
	down-regulated	137
	Fc[¥] >10	12
	detox-probes[§]	17

[¥] fold change of backcross samples compared to moz control samples

[§] probes targeting genes implicated in metabolic resistance.

Table 6.1: Summary of statistical test parameters and results for the microarray.

6.3.1.1 Functional analysis / Enrichment analysis

Enrichment or cluster analysis is an approach to discover enriched biological themes like a metabolic pathway, or functional-related gene groups. We used the Database for Annotation, Visualization, and Integrated Discovery (DAVID) bioinformatics tool online. The full list of significantly differentially expressed entities (fc≥2 and FDR p-value ≤0.01) was used as an input file for DAVID software to perform an enrichment analysis for functional clustering of potential hits. A full list of genes represented on the Agilent chip served as background.

The top cluster in the functional annotation cluster for up-regulated entities includes 23 entities with the GoTerms ‘peptidase activity’ for molecular function and ‘proteolysis’ for biological process. This annotation cluster has an enrichment score of 3.34 and corrected p-values (Benjamini Hochberg) 1.3E-5. The second cluster with the enrichment score 2.53 and max p-value of 0.006 clusters entities with protein domains (IPR) ‘proteinase inhibitor’ (Figure 6.1 and Appendix 9.10). Cytochrome P450 associated genes and protein domains are listed as the third annotation cluster but are not significantly associated.

For the down-regulated entities the annotation cluster 1 includes terms of the ontology ‘Secondary metabolites biosynthesis, transport, and catabolism’ such as Cytochrome P450 terms and protein domains and the GoTerms for molecular functions associated with P450 metabolism. The enrichment score for this cluster is 2.84 with p-values between 0.01 and 0.03 (Figure 6.1 and Appendix 9.11). There are no other significant enrichment clusters.

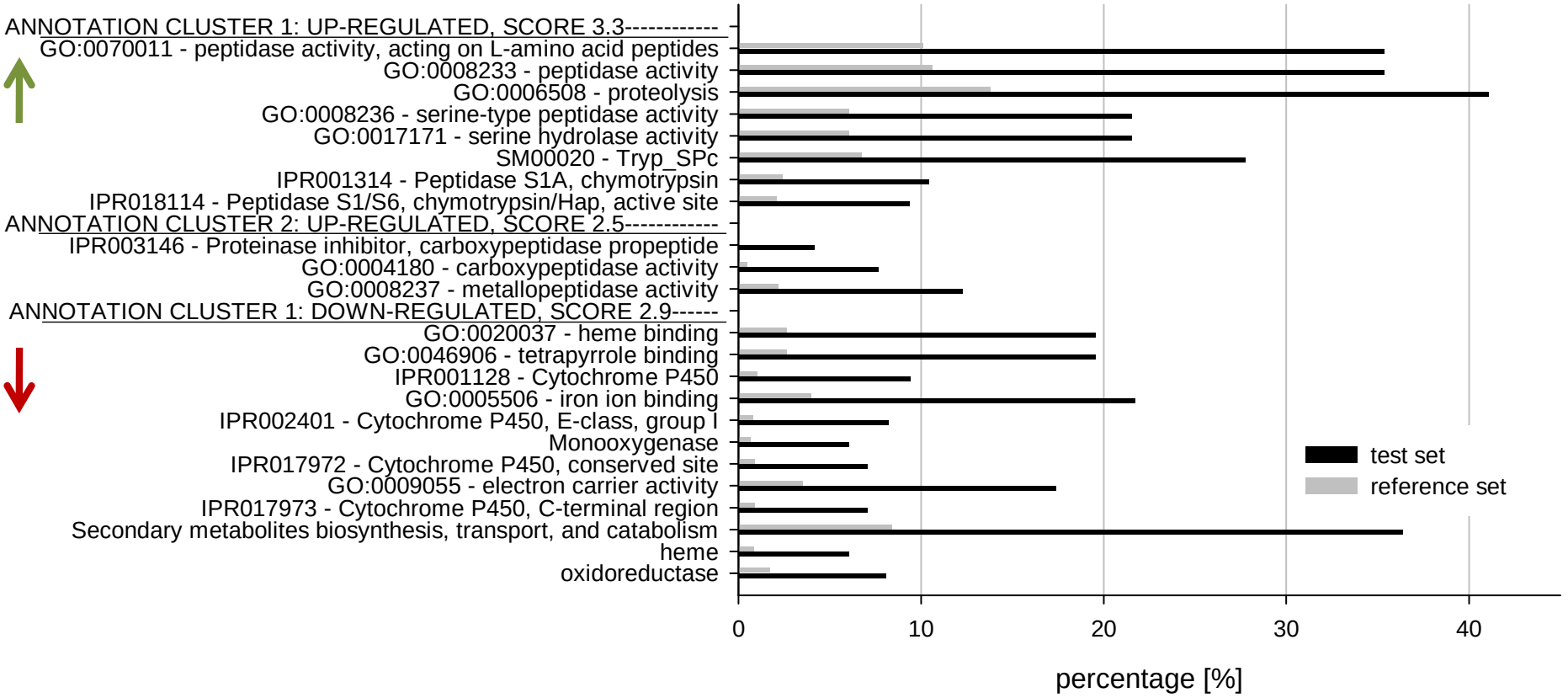


Figure 6.1: Results from the enrichment analysis performed with DAVID software. The figure represents term categories found significantly enriched compared to the reference set (all transcripts present on the microarray) after Benjamini and Hochberg multiple testing correction (pval≤0.05). The test set indicates percentage of differentially regulated genes in a certain term category compared to all differentially regulated genes, while the reference set indicates the percentage of a particular term category compared to all genes on the microarray

6.3.1.2 Candidate gene approach

6.3.1.2.1 Detox genes

A total of 30 probes targeting detox genes were significantly differentially expressed, 11 were up-regulated and 19 down-regulated. These belong to 16 P450s and one GST, listed in Table 6.2. A small set of four P450s is up-regulated and a larger number of 12 different P450s is down-regulated. The probes found in the top 50 hits of each set are indicated.

For each detox gene, there are four probes on the 8 x15K array (some exceptions § in Table 6.2). *Cyp4h24* is the gene with the highest fold change (average 7.5-fold), but only two probes were significantly over-expressed. For genes *cyp9j5*, all four entities were significantly differentially expressed with an average 5.1-fold up-regulation. Similarly three of the probes targeting *gste5* are over-expressed with an average transcript increase of 2.3-fold. Additional up-regulated detox genes *cyp6m1* and *cyp4g16* are listed in Table 6.2. Note that, although in some cases, not all four probes reached significance, each probe for the genes in this list was up-regulated.

The most strongly down-regulated detox-genes are *cyp6m2* and *cyp6p4* with fold changes of 5.3 and 5.0 respectively. Further down-regulated detox genes are *cyp302a1* and *cyp6ag2* (fold-change 2.9 and 2.1 respectively). For all of these down regulated genes multiple probes showed similar patterns of reduced expression adding confidence to the results. For another eight P450 transcripts only one probe is significantly down-regulated, but again, the additional non-significant probes showed the same trend.

description	probe Name	FDR p[$]	FC[¥]	FC[¥] mean	accession number	location
over-expressed in permethrin resistant						
cyp4h24	DETOX_386	4.75E-03	8.2	7.5	AGAP013490	X
	DETOX_387	4.73E-03	6.8			
cyp9j5	DETOX_502	6.85E-03	5.1	5.4	AGAP012296	3L
	DETOX_503	8.70E-03	5.2			
	DETOX_504	6.83E-03	5.1			
	CUST_10851	4.66E-04	5.4			
cyp6m1	DETOX_438	7.49E-03	3.0	3.0	AGAP008209	3R
cyp4g16	DETOX_362	7.49E-03	2.3	2.3	AGAP001076	X
gste5	DETOX_625*	3.68E-03	2.2	2.3	AGAP009192	3R
	DETOX_626	5.66E-03	2.3			
	DETOX_627	9.13E-03	2.4			
under-expressed in permethrin resistant						
cyp6m2	DETOX_439	6.79E-03	5.2	5.3	AGAP008212	3R
	DETOX_440	6.83E-03	5.3			
	DETOX_441	6.07E-03	5.6			
	CUST_11496	3.15E-04	5.3			
cyp6p4	DETOX_463*	5.21E-03	**5**.1	5.0	AGAP002867	2R
	DETOX_464	7.61E-03	4.8			
	CUST_4819	4.07E-04	5.2			
cyp302a1	DETOX_234	6.05E-03	3.0	2.9	AGAP005992	2L
	CUST_1471	2.86E-04	2.7			
cyp6ag2	DETOX_424	5.40E-03	2.1	2.1	AY745224	2R
	DETOX_425	6.37E-03	2.1			
cyp325f1	CUST_3883	6.88E-03	5.0		AGAP002195	2R
P450 §	CUST_3915*	3.71E-03	4.9		AGAP002204	2R
cyp9m1	CUST_12597*	3.68E-03	3.8		AGAP009374	3R
P450 §	CUST_12586	7.67E-03	3.6		AGAP009363	3R
cyp304b1	CUST_5130*	3.97E-03	2.7		AGAP003066	2R
cyp9k1	CUST_8035	5.89E-03	2.3		AGAP000818	x
cyp325a2	CUST_3897*	4.75E-03	2.2		AGAP002209	2R
cyp4aa1	CUST_5838	8.80E-03	2.1		AGAP003608	2R

§ genes targeted by only one probe each

***** probe present in top 50 list, ranked according to p value

$ false discovery corrected p-value (Benjamini Hochberg)

¥ fold change

Table 6.2: Microarray results of significantly (FDP p≤0.01) differentially expressed detoxification genes in the exposed backcross samples compared to the susceptible Moz strain.

Values for significantly differentially expressed probes are listed separately and mean values are given where applicable.

6.3.2 Candidate gene validation

6.3.2.1 *cyp4h24* Real-time quantitative PCR

The partial cDNA sequence of *cyp4h24* in *An. gambiae* (from vectorbase AGAP013490) and *An. arabiensis* sequence (consensus sequence data from Ndja and Moz used in genetic crosses) are aligned in Figure 6.2. For the *An. arabiensis* consensus sequence data from eight specimens was obtained and aligned (the two parental females from the genetic crosses (one Ndja and one Moz female) plus three F_1 progeny from each cross)). Also shown are the binding sites for the microarray probes and the fragment amplified by qPCR. The qPCR primers for *cyp4h24* were designed against this *An. arabiensis* consensus sequence. There are a number of SNPs in the binding sites of the microarray probes. The two sequences show 96.14% sequence similarity (Identities = 846/880 (96%), Gaps = 1/880 (0%)).

The qPCR compared expression in samples from the Moz and Ndja strain (unexposed to insecticides) and samples from the B_1 backcross generation (surviving permethrin exposure, preserved in RNA later® 24 hours after exposure) (Figure 6.3 A). Normalized values for transcript levels of *cyp4h24* are given in Table 6.3. No significant difference was observed between the three populations with either control gene. The absolute transcript levels of the control genes were also compared (Figure 6.3 B), to check for consistent expression of control gene transcripts across strains. No significant difference in expression was observed.

Thus the results from the qPCR do not confirm the microarray data for *cyp4h24*. This may be a limitation of the power of the experiment as, with both control genes, the transcript level was higher in the backcross population relative to the susceptible strain, but the difference did not reach significance.

A

B

```
gambiae         CAGTACGTGCGGGACGTGAAGCGGATGAGCGAGCTGATACTGCTGCGCATTTTCCATGTG 60
arabiensis      ----------------CAGCGGATGAGCGAGCTGATACTGCTGCGGATTTTCCATGTG 42
                              .*************************** ************

gambiae         CTGTCGTCCTTTCCGCGCACCTACTGGTACACGATGCCGAACGCGTGGGAGCAGAGGAAA 120
arabiensis      CTGTCGTCCTTTCCGCGCACCTACTGGTACACGATGCCGAACGCGTGGGAGCAGCGGAAG 102
                ************************************************.****.

gambiae         CTTATCCGACGGCTGCATGCGTTCACGGACACCGTGATCCACAAACGCCGGGAACAGCTG 240
arabiensis      CTGATCCGACGRCTGCATGCGTTCACGGACACCGTGATCCACAAACGYCGGGAACAGCTG 162
                ** ******** ********************************** ************

gambiae         TTGGAAAGGAGCTCGCAAGTGTCTAACGAGCAGGAGTGTCTGGATGAGGAACATCTTTAC 300
arabiensis      TTGGAAAGGAGCTCGGAAGTGTYTCACGAGCAGGAGTGCCTGGATGAGGAACATCTTTAC 222
                *************** ****** *.************ ********************

gambiae         ACAAAGCGTAGGGAAACGTTCTTGGATCTGCTGCTGAACGTTAGGGTGGATGGCAATTCA 360
arabiensis      ACAAAGCGTAGGGAWACGTTCTTGGATCTGCTGCTGAACGTTAGGGTGGATGGCAATTCA 282
                ************** ************************************************

gambiae         CTCAGTGATCTTGATATTCGGGAGGAGGTGGATACCTTCATGTTCGAGGGTCATGATACG 420
arabiensis      CTCAGTGATCTTGATATACGGGAGGAGGTTGATACYTTYATGTTCGAGGGTCATGATACG 342
                *****************:************ ***** ** **********************

gambiae         ACCACTTCGGGCATTGCGTTCACGTTCTATCAACTCGCTAAGCATCCTGAAATACAGGAG 480
arabiensis      ACCACTTCCGGCATTGCGTTCACCTTCTATCAACTTGCTAAGCATCCTGAAATACAGGAG 402
                ******** ************** ************ ************************

gambiae         AAGCTTTATCGGGAGATCCAAGATGTGCTCGGCGGCGAATACCGCCATGTGCCGCTAACC 540
arabiensis      AAGCTTTATCGGGAGATCCAAGATGTGCTCGGCGTCGATTACCGCCATGTGCCGCTCACC 462
                ********************************** ***:*****************.***

gambiae         TACAACACGCTCCAGAACTTCCCGTACCTGGACATGGTAGTGAAGGAGTCGCTGCGACTT 600
arabiensis      TACAACACGCTCCAGAACTTCCCGTACCTCGACATGGTAGTGAAGGAGTCGCTGCGACTG 522
                **************************** ****************************

gambiae         CTACCACCCGTTTCGTTCATCGGACGTCGGCTAGCAGATGACATCGAGATGAACGGTGTA 660
arabiensis      CTGCCACCCGTTTCGTTTATCGGCCGTCGGCTAGCGGATGACATCGAGATGAACGGTGTA 582
                **.************** *****.***********.************************

gambiae         ACTATTCCAGCCGGCACTGACTTTACCATCCCCATCTACGTCATCCACCGCAATCCCGTT 720
arabiensis      ACTATTCCAGCCGGCACTGACTTTACCATCCCCATCTACGTCATCCACCGCAATCCCGTT 642
                ************************************************************

gambiae         GTGTATCCCGACCCGGAACGATTCGATCCGGAGCGCTTCTCGGACGGCAATACGCAACGG 780
arabiensis      GTGTATCCCGACCCRGARCGATTCGATCCGGAGCGCTTCTCGGACGGAAATACGCAACGG 702
                ************** ** ****************************.************

gambiae         CGCGGCCCGTACGACTACATCCCGTTCAGTATCGGCTCGCGGAACTGTATCGGGCAGCGC 840
arabiensis      CGCGGCCCGTACGACTACATCCCGTTCAGTATCGGCACGCGGAACTGTATCGGGCAGCGG 762
                ************************************:***********************

gambiae         TATGCGCTGCTGGAGATGAAGGTCGCTATCGTACGTATGGTTTCCTTTTATAGGATATTG 900
arabiensis      TATGCGCTGCTGGAGATGAAGGTCGCTATCGTACGTATGATTTCCTTTTATAGGATATTG 822
                ************************************.***********************

gambiae         CCCGGTGATACGATGCACGAGATTCGTCTCAAAACGG-ATCTGGTGCTACGTCCGGACAA 960
arabiensis      CCCGGTGATACGATGCACGAGATTCGTCTCAAAACGGAATCTGGTGCTACGTCCGGACA- 881
                ************************************ ********************
```

Figure 6.2: Sequence alignment of *An. gambiae s.s.* and *An. arabiensis cyp4h24*
(A) Schematic of *cyp4h24* alignment with binding sites for microarray probes and qPCR target fragment. Red stars indicate SNPs within these sites. (B) ClustalW alignment of *An. gambiae* (vectorbase AGAP013490) and *An. arabiensis* sequences (consensus for eight individuals from families 2 and 3) (96.14% similarity) for part of *cyp4h24* with binding sites for microarray probes underlined in blue (DETOX_387), red (DETOX_386) or green (DETOX_385 and CUST_7895), qPCR target fragment in brown with primer sites underlined. SNPs within binding sites are marked red. Base ambiguity code: R = A or G; Y = C or T; W = A or T.

	target gene *cyp4h24* not normalized		normalized against control genes			
	transcripts	δ	*/s7*	δ	*/cpr*	δ
backcross	146305	16032	0.3929	0.0639	2.253	0.714
Ndja	111177	14086	0.3655	0.0255	1.637	0.172
Moz	91248	21991	0.2937	0.0440	1.484	0.120

δ standard deviation

Table 6.3: Transcript levels of *cyp4h24* (AGAP013490) and ratios of transcript levels when normalized against the control genes *cpr* and *s7* (mean±sdev).

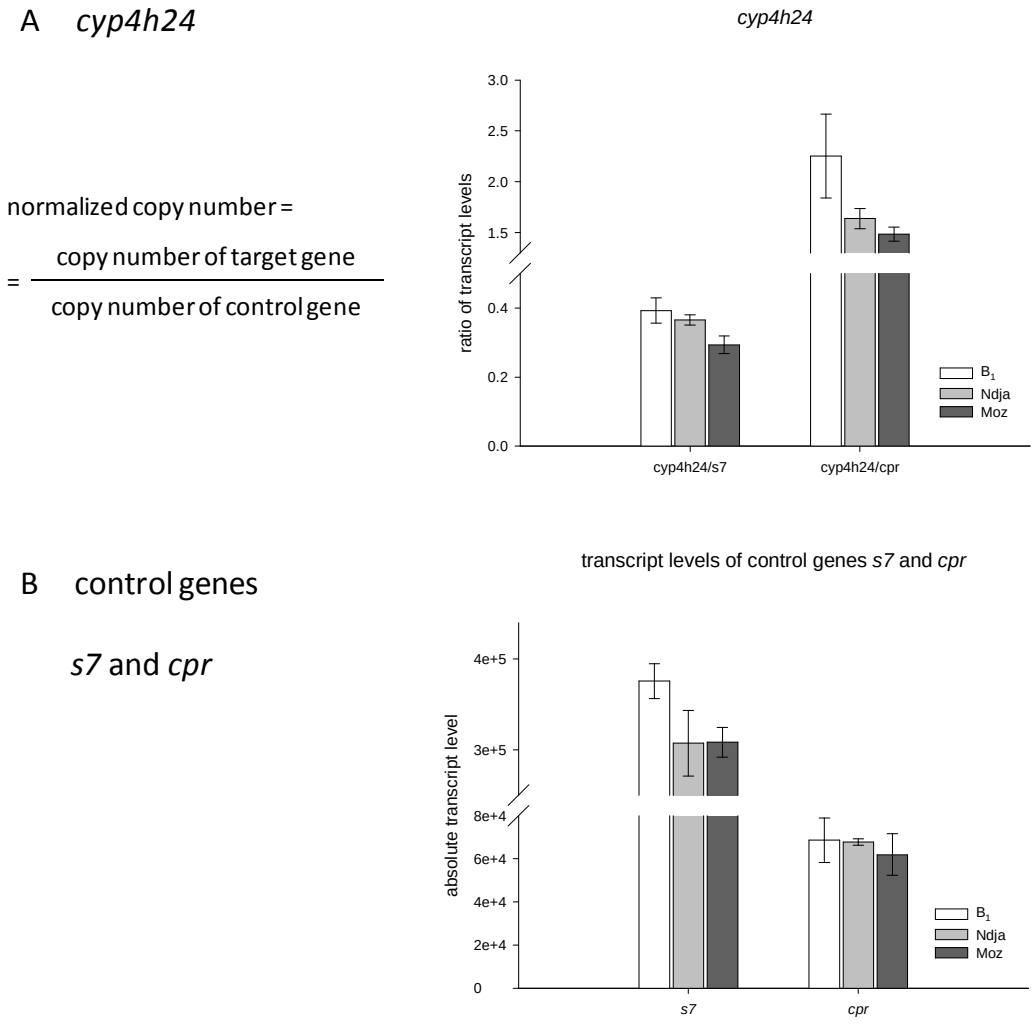


Figure 6.3: (A) Summary of the qPCR (absolute quantification) results on *cyp4h24* using both available control genes, *s7* (left column) and *cpr* (right column) for normalisation (mean±±se). (B) Absolute transcript level of the control genes *s7* (left) and *cpr* (right) expressed as absolute transcript number. Data are presented as mean ± standard error of three biological replicates.

6.4 Discussion

Enrichment analysis is a valuable tool to screen microarray data sets for terms that are associated with either the up-regulated or down-regulated subset of genes. It helps highlight signatures of co-regulation within the data set. By examining networks of genes which are interacting to produce a phenotype it facilitates the transition from data collection to biological meaning. For the current data set, the only terms significantly enriched in the up-regulated set were proteases and peptidases. These enzymes are involved in a multitude of physiological functions and essential in many cellular processes. Although proteases have been found up-regulated in insecticide resistant populations previously (Pedra *et al.* 2004; Vontas J 2005), their role, if any, in resistance is unclear. Serine proteases can hydrolyse ester bonds and so could conceivably be involved in insecticide metabolism (Hedstrom 2002). Alternatively, elevated protease activity may be a result of metabolic stress induced by insecticide exposure.

The terms significantly enriched in the down-regulated set were cytochrome P450s. This was unexpected. A small number of P450s are up-regulated in the backcross population relative to the susceptible strain and to understand the significance of these findings, more needs to be known about the role of individual P450s in insecticide metabolism.

The candidate gene approach complements the enrichment analysis. If the expression of a single gene has a major contribution to the resistance phenotype, the enrichment analysis will not show it. The full lists of 144 significantly up- and 137 significantly down-regulated entities were screened for probes detecting members of the gene families cytochrome P450 monooxygenases, GSTs and carboxylesterases. These are the major enzyme groups responsible for metabolically based resistance to the four classes of insecticides used in public health.

description	location		reference
over-expressed			
cyp4h24	X	↑	Up in permethrin resistant *An. gambiae* (Müller *et al.* 2008a)
		↓	Down in permethrin resistant *An. arabiensis* (Müller *et al.* 2007)
		↑	midgut: female (Baker *et al.* 2011)
cyp9j5	3L	↓	Down in permethrin resistant *An. gambiae* (David *et al.* 2005)
		↑	Up following blood mean in *An. gambiae s.s.* (Marinotti *et al.* 2005)
		↓	ovaries: female; ↑midgut: male (Baker *et al.* 2011)
cyp6m1	3R	↓	Down in permethrin resistant *An. gambiae* (Müller *et al.* 2007)
		↓	ovaries: female; ↑midgut: male (Baker *et al.* 2011)
cyp4g16	X	↑	Up in permethrin resistant *An. arabiensis* (Müller *et al.* 2008a)
		↓	midgut: female; ↑Carcass: male (Baker *et al.* 2011)
gste5	3R	↑	malpighian tubules: male; ↓ head: male (Baker *et al.* 2011)
under-expressed			
cyp6m2	3R	↑	Up in permethrin resistant *An. gambiae s.s.* (Djouaka *et al.* 2008; Müller *et al.* 2007; Stevenson *et al.* 2011) and
		↑	Up in *An. arabiensis* (Munhenga & Koekemoer 2011)
		↓	ovaries: female; ↑malpighian tubules: male; (Baker *et al.* 2011)
cyp6p4	2R	↑	midgut: male; ↓testes: male (Baker *et al.* 2011)
cyp302a1	2L	↑	Male accessory gland: ↓ male; head: male (Baker *et al.* 2011)
cyp6ag2	2R	↑	Up in permethrin resistant *An. gambiae* (Müller *et al.* 2007)
		↓	ovaries: female; ↑midgut: male (Baker *et al.* 2011)
cyp325f1	2R	↑	Up in permethrin resistant *An. gambiae* (Müller *et al.* 2008b)
AGAP002204	2R	↑	Male: head (Baker *et al.* 2011)
cyp9m1	3R	↓	Down as response to thermal stress in 2La (Cassone *et al.* 2011)
AGAP009363	3R		
cyp304b1	2R	↓	Down in pyrethroid resistant *An. funestus* (Christian *et al.* 2011)
cyp9k1	x	↓	ovaries: female; ↑ malpighian tubules: male (Baker *et al.* 2011)
cyp325a2	2R	↓	Down in permethrin resistant *An. gambiae* and *An. arabiensis* (Müller *et al.* 2008a; Müller *et al.* 2007)
cyp4aa1	2R	↓	Down in permethrin resistant *An. gambiae* (Müller *et al.* 2008a)

Table 6.4: Significantly (FDR p ≤0.01; fold change ≥2.0) differentially expressed detoxification genes with references to previous microarray experiments in *An. gambiae s.l.* and *An. funestus*. Significance thresholds in referenced studies vary (fold-change ≥2.0, except (David *et al.* 2005) with fold-change >1.5; FDR p ≤0.001, except (Baker *et al.* 2011) with FDR p < 0.05).

Four cytochrome P450s and one GST were up-regulated and 12 P450s down-regulated. The expression levels of some of these genes have been associated with insecticide resistance in previous studies (*cyp4h24*, *cyp6g16*, *cyp6ag2* and *cyp6m2*) while others (*cyp9j5* and *cyp6m1)* have been negatively correlated with resistance (see Table 6.4) (David *et al.* 2005; Müller *et al.* 2007). Of particular note is *cyp6m2*, which, in the current study, is expressed at lower levels in the resistant population in contrast to several previous studies which had implicated this P450 in permethrin resistance (Djouaka *et al.* 2008; Mitchell *et al.* 2012; Müller *et al.* 2008b; Stevenson *et al.* 2011).

The candidate gene *cyp4h24* was found up-regulated in permethrin resistant field populations of *An. gambiae s.s.* from Ghana (Müller *et al.* 2008a) but expressed at lower levels in an *An. arabiensis* population after insecticide exposure had occurred compared to before insecticide exposure (Müller *et al.* 2008a). It was found up-regulated in permethrin survivors from this microarray study, although this result could not be confirmed by qPCR. The summary in Table 6.4 shows that the observed profile of detox gene expression in this study does not match other previous studies. Follow up is needed to confirm their role in resistance.

In this study relatively few detox genes were found in the top 50 lists of significantly differentially expressed probes (ranked by p-value), only *gste5* in the top 50 up-regulated entities, and three P450s in the top 50 of down-regulated entities (highlighted by * in Table 6.3). Surprisingly a higher density of probes targeting detox genes was found in the list of down-regulated probes than in the list of up-regulated entities. This agrees with the results of the enrichment analysis but is difficult to explain in the context of permethrin resistance.

Four serine proteases are found in the top 50 list of up-regulated genes (AGAP001198, AGAP004900, AGAP008861, AGAP012946) with the transcript for gene AGAP008861 showing the highest fold change with 31.89 (valid for the full set of significantly up-regulated transcripts). Another potentially interesting hit is AGAP002052 on chromosome 2R (fold change 4.4), a transcript coding for a cuticular protein. Thickening of the insect

cuticule may reduce insecticide penetration and some cuticular proteins have been linked to insecticide resistance (Djouaka *et al.* 2008; Puinean *et al.* 2010; Vontas *et al.* 2007). As support for these potential novel candidates from previous studies on insecticide resistance is lacking, they will have to be validated independently. A functional approach via e.g. RNA interference to assess their role in permethrin resistance would be useful.

The microarray platform used in this study was designed on the *An. gambiae s.s.* genome sequence (Mitchell 2011). Although the two species *An. arabiensis* and *An. gambiae* are very closely related, there will inevitably be sequence polymorphisms between the species that may lead to non-specific binding or to samples not hybridising to probes. Comparing sequences of *cyp4h24* from *An. gambiae* and from *An. arabiensis* a high sequence similarity > 96% is observed but between one and three SNPs are found within the binding site of the microarray and these polymorphisms may potentially affect the hybridisation dynamics.

This is an inherent problem with cross-species microarray studies which cannot easily be overcome by processing more biological replicates or by performing a different experimental design.

With some *An. arabiensis* sequence data available (Hittinger *et al.* 2010) and a draft genome sequence expected later this year (http://www.broadinstitute.org/annotation/genome/anopheles) designing a species specific microarray chip for *An. arabiensis* may be a better alternative.

The intention behind the experimental design of the microarray was to gain maximum information whilst performing a minimum of arrays with the available mosquito resources. The backcross design was intended to reduce false positives that may have resulted from comparing two strains from very different geographical regions. This backcross design has not been reported before in microarray experiments of mosquito species, it is however more common in plants (Kirst *et al.* 2004). With issues such as the cross-species hybridisation mentioned above it is hard to judge if this is an ideal comparison. Adding another comparison of un-exposed Ndja samples to the

Moz strain should provide additional information about what genes may be constitutively up-regulated in the resistant strain.

Chapter 7. Conclusions

7.1 Key findings of the current study

Successful application and resistance management of insecticides for vector control relies on efficient use of the available compounds. There are two essential prerequisites for this: potent insecticides and data on presence, level and type of resistance present in the target organism. There are a limited number of insecticides currently available for the use in public health, with no new classes being licensed for adult mosquito control for over 30 years. Furthermore, resistance to the available insecticides is increasing in range and intensity (Djegbe *et al.* 2011; Ranson *et al.* 2009; Santolamazza *et al.* 2008a). This emphasises the importance of monitoring of resistance and understanding underlying mechanisms.

Target-site resistance is one type of resistance mechanism that is reasonably well understood. Once resistance alleles have been characterised, a wide range of molecular detection tools can be applied to enable these alleles to be monitored in population screening programs (Chanda *et al.* 2011; Yewhalaw *et al.* 2011). But the recent discoveries of N1575Y in the sodium channel of *An. gambiae s.s.* (Jones *et al.* 2012b) and F1534C in *Ae. aegypti* (Harris *et al.* 2010) show that researchers cannot rely solely on the present diagnostics and highlight the importance of continuing to screen for new target-site resistance mutations.

Metabolic resistance is more complex at the molecular level. There are no molecular markers for metabolic resistance in mosquitoes. This is partly because there are multiple genes putatively involved in metabolic resistance but also a result of the multitude of different molecular mechanisms by which metabolic resistance could occur. Mutations within the structural genes encoding detoxification enzymes may play a role in enhancing insecticide metabolism if they increase the affinity or specificity of the enzyme for the insecticide and/or increase the rate of turnover. This has been observed in *An. gambiae* glutathione-S-transferases (Mitchell 2011; Wang *et al.* 2008) and in *D. melanogaster cyp6a2* (Amichot *et al.* 2004) but appears less

common in cytochrome P450s. Such allelic variation in resistant genotypes would be easy to monitor by molecular assays once the responsible mutation had been established. Although extensive amino acid substitutions are present in mosquito cytochrome P450 primary sequences, to date these polymorphisms have not been associated with resistance (Feyereisen 1999; Li *et al.* 2007; Müller *et al.* 2008b; Stevenson *et al.* 2011). More commonly metabolic resistance is associated with elevated expression levels of detoxification genes (Amenya *et al.* 2008; David *et al.* 2005; Djouaka *et al.* 2008; Müller *et al.* 2008b; Nikou *et al.* 2003; Wondji *et al.* 2009).

Increased expression may be a result of mutations in the promoter region, other cis-regulatory elements or in trans-acting regulators. For example the presence of an *accord* transposon upstream of *cyp6g1* in *D. melanogaster* is correlated with up-regulation (Chung *et al.* 2007) and a transposable element upstream of *cyp9m10* implicated in larval permethrin resistance in *Cx. quinquefasciatus* acts as a marker for the regulatory motif (Wilding *et al.* unpublished).

At the start of this study, the objective was to confirm the role of the candidate genes that had been implicated in pyrethroid resistance in *An. gambiae* in Benin by earlier microarray studies (Djouaka *et al.* 2008) and identify the causal mutations. Unfortunately, the increased expression of the four candidate genes (*cyp6p3*, *cyp6m2* and *cplcg3/4*) was not detected in a resistant laboratory strain, isolated from the same area as the mosquitoes used in the original microarray. However, field collected mosquitoes that had been collected from the same site as the original Akron colony and survived permethrin exposure, did show elevated activity of candidates *cyp6p3* and *cyp6m2*. Hence, the putative metabolic resistance mechanisms were not maintained in the laboratory population. Selection of this laboratory colony with permethrin did result in an increase in the resistance frequency but this may have been as a result of increased *kdr* frequencies rather than selection of metabolic resistance.

Furthermore, sequencing of the candidate genes, and their upstream regions, failed to identify any sequence polymorphisms associated with

survival to pyrethroids although it must be noted that the sample size used here was very limited.

A genetic mapping approach to identify the major loci associated with permethrin resistance in this Benin population was therefore initiated. The suspected decrease of metabolic resistance levels in the lab strain, and its presence in the field samples is the reason why field collected, non laboratory reared mosquito samples were used for the QTL mapping of this trait in the Akron mosquito population.

Two QTLs for permethrin resistance were identified through genetic mapping of isofemale lines of *An. gambiae s.s.* from Benin. The locus observed on chromosome 2 in *An. gambiae s.s.* from Akron confirms *kdr* as the major QTL accounting for over half of the genetic variance (between 56 and 64%). Another smaller QTL is located on chromosome 3R in the region of a cluster of P450 genes including *cyp6m2* as well as the cuticular candidates from the microarray study although the boundaries of this QTL are too large to enable identification of a single gene. The locus observed on chromosome 2 in *An. gambiae s.s.* from Akron confirms *kdr* as the major QTL possibly masking other minor loci on the same chromosome as well as loci on other chromosomes.

Overall the novel approach of QTL mapping on isofemale families raised directly from field caught females was successful; however, a larger number of families as well as larger progeny number is desirable for future experiments. With the availability of larger sample sets for genetic mapping extreme resistance phenotyping or phenotyping into more classes than dead and alive as shown by Saavedra-Rodriguez and colleagues (Saavedra-Rodriguez *et al.* 2008) is a possibility.

For dissecting the genetics of metabolic resistance, a resistant population lacking any target-site resistance is ideal. Therefore the next step was to examine populations of *An. arabiensis* from Chad which were resistant to pyrethroids but contained no known *kdr* mutations. The more common approach of mapping in F_2 progeny derived from a genetic cross was adopted for this study. The parental populations were recently colonised field

samples, and therefore potentially contain fewer genetic changes resulting from adaptation to prolonged maintenance in colony. The identification of a single QTL on chromosome 2R associated with pyrethroid resistance in *An. arabiensis* is a key result of the current study. The result confirms the absence of *kdr* (the sodium channel gene is located on chromosome 2L) in this population. As to be expected from F_2 mapping populations the resolution of the resulting maps is limited and the nature of the QTL could not be resolved sufficiently to identify candidate genes. However it is interesting to note that this QTL encompasses a large cluster of P450 genes.

Upon completion of the QTL mapping a microarray experiment was initiated in the hope that this would help identify genes differentially expressed in the resistant Ndja strain relative to the susceptible Moz strain, potentially including genes located within the boundaries of the QTL. Several candidate genes were identified but no strong candidates for metabolic resistance were identified within the QTL boundaries. Two transcripts that are present in the top-50 list of significantly up-regulated probes are located within the boundaries of the QTL on chromosome 2R. These belong a novel gene, AGAP001616, and to AGAP001881, a zinc metalloprotease. Another potentially interesting hit is AGAP002052 on chromosome 2R (14.6 Mb) a transcript coding for a cuticular protein.

The possibility that these genes are involved in uncharacterised resistance mechanisms warrants further investigation.

7.2 Alternative approaches and suggestions for further work

Quantitative traits are complex traits and if possible different approaches should be combined in order to obtain robust results and avoid misinterpretation of results or missing out on important aspects.

QTL mapping is an excellent first approach for screening the entire genome but its power is limited. Advanced intercross lines (AIL) enable fine-scale mapping of prior identified QTLs (Darvasi & Soller 1995). AILs are produced from an F_2 population by random intercrossing by which many recombination events are accumulated that provide higher genetic resolution that cannot be reached by increasing marker density only. Once the confidence interval of

the QTL is sufficiently small, underlying genes can be identified by positional cloning (Wondji *et al.* 2009).

The availability of whole genome sequence data for an increasing number of vectors has enabled the design of powerful tools such as whole genome SNP arrays (Neafsey *et al.* 2010; Weetman *et al.* 2012) and microarrays (Mitchell 2011). These allow genome and transcriptome wide comparison between samples. High density SNP analysis has the power to detect and monitor selective sweeps and genetic differentiation in wild populations. A whole genome SNP scan led to the identification of a gene variant potentially increasing *Plasmodium* resistance in *An. gambiae s.s.* (White *et al.* 2011b). Genome wide association mapping of permethrin resistance has confirmed *kdr* as the major effect variant but failed to detect SNPs associated with metabolic variants possibly due to low linkage disequilibrium (Weetman *et al.* 2010).

Next generation sequencing (NGS) enables transcription levels and genetic variation to be analysed simultaneously. Next generation sequencing offers huge potential to uncover the genetics of a certain trait. Questions such as whether the adaption is the result of many loci of small effects or a few loci of large effect, or the nature of the underlying genetic change (e.g. point mutation, gene duplication etc.) can potentially be addressed by this technology. In addition the *de novo* sequencing approach is ideal for any non model species regardless of availability of sequence or marker data, provided a reference sequence from a suitably similar species exists. This technology has only recently become affordable and to date, the full potential of NGS to study traits in mosquitoes has not been explored.

Whole genome transcriptional profiling has the potential to provide information about networks of genes and biological pathways involved in a phenotype. However, microarray approaches can only identify candidates. These need to be validated. Perhaps the best validation of candidates is via replication from independent studies. A meta-analysis of microarray data on insecticide resistance at LSTM is currently being conducted with the objective of identifying genes or pathways repeatedly associated with the resistance trait.

Insecticides remain a key component of malaria control programmes. With resistance threatening their efficacy in many malaria endemic countries, it is necessary to also consider alternative methods of vector control. Research describing novel methods for population reduction, such as sterile insect technique (Catteruccia *et al.* 2005; Harris *et al.* 2011) or means of engineering resistance of vector species to pathogens (White *et al.* 2011b) show promise. However, given that it will be many years before these new methods become operational, and no new insecticides are on the immediate horizon, it is essential that efforts to understand and manage insecticide resistance are maintained.

Chapter 8. Literature

Abdalla H, Matambo TS, Koekemoer LL*, et al.* (2008) Insecticide susceptibility and vector status of natural populations of *Anopheles arabiensis* from Sudan. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **102**, 263-271.

Ahmad M, Denholm I, Bromilow RH (2006) Delayed cuticular penetration and enhanced metabolism of deltamethrin in pyrethroid-resistant strains of *Helicoverpa armigera* from China and Pakistan. *Pest Management Science* **62**, 805-810.

Akogbéto M, Djouaka R, Noukpo H (2005) Use of agricultural insecticides in Benin. *Bulletin de la Societe de pathologie exotique (1990)* **98**, 400-405.

Akogbéto M, Yakoubou S (1999) Resistance of malaria vectors to pyrethrins used for impregnating mosquito nets in Benin, West Africa. *Bulletin de la Société de pathologie exotique* **92**, 123-130.

Amenya DA, Naguran R, Lo TCM*, et al.* (2008) Over expression of a cytochrome P450 (CYP6P9) in a major african malaria vector, *Anopheles funestus*, resistant to pyrethroids. *Insect Molecular Biology* **17**, 19-25.

Amichot M, Tarès S, Brun-Barale A*, et al.* (2004) Point mutations associated with insecticide resistance in the *Drosophila* cytochrome P450 *Cyp6a2* enable DDT metabolism. *European Journal of Biochemistry* **271**, 1250-1257.

Andersen SO, Hojrup P, Roepstorff P (1995) Insect cuticular proteins. *Insect Biochemistry and Molecular Biology* **25**, 153-176.

Asante KP, Abdulla S, Agnandji S*, et al.* (2011) Safety and efficacy of the RTS,S/AS01E candidate malaria vaccine given with expanded-programme-on-immunisation vaccines: 19 month follow-up of a randomised, open-label, phase 2 trial. *The Lancet Infectious Diseases* **11**, 741-749.

Awolola TS, Oduola OA, Strode C*, et al.* (2009) Evidence of multiple pyrethroid resistance mechanisms in the malaria vector *Anopheles gambiae* sensu stricto from Nigeria. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **103**, 1139-1145.

Ayala FJ, Coluzzi M (2005) Chromosome speciation: humans, Drosophila, and mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America* **102** 6535-6542.

Baker D, Nolan T, Fischer B*, et al.* (2011) A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, *Anopheles gambiae*. *BMC Genomics* **12**, 296.

Balkew M, Ibrahim M, Koekemoer LL*, et al.* (2010) Insecticide resistance in *Anopheles arabiensis* (Diptera: Culicidae) from villages in central, northern and south west Ethiopia and detection of *kdr* mutation. *Parasites & Vectors* **3**, 40.

Bariami V, Jones CM, Poupardin R, Vontas J, Ranson H (2012) Gene amplification, ABC Ttansporters and cytochrome P450s: unraveling the molecular basis of pyrethroid resistance in the dengue vector, *Aedes aegypti*. *PLoS Neglected Tropical Diseases* **6**, e1692.

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265.

Bass C, Nikou D, Donnelly M*, et al.* (2007) Detection of knockdown resistance (*kdr*) mutations in *Anopheles gambiae*: a comparison of two new high-throughput assays with existing methods. *Malaria Journal* **6**, 111.

Bass C, Nikou D, Vontas J, Williamson MS, Field LM (2010) Development of high-throughput real-time PCR assays for the identification of insensitive acetylcholinesterase (*ace-1*^{R}) in *Anopheles gambiae*. *Pesticide Biochemistry and Physiology* **96**, 80-85.

Black IV WC, Gorrochetegui-Escalante N, Randle NP, Donnelly MJ (2008) The Yin and Yang of linkage disequilibrium: mapping of genes and nucleotides conferring insecticide resistance in insect disease vectors. *Advances in Experimental Medicine and Biology* **627**, 71-83.

Bloomquist JR (1996) Ion channels as targets for insecticides. *Annual Review of Entomology* **41**, 163-190.

Brooke B, Kloke G, Hunt R*, et al.* (2001) Bioassay and biochemical analyses of insecticide resistance in southern African *Anopheles funestus* (Diptera: Culicidae). *Bulletin of Entomological Research* **91**, 265-272.

Brooke B, Koekemoer L (2010) Major effect genes or loose confederations? The development of insecticide resistance in the malaria vector *Anopheles gambiae*. *Parasites & Vectors* **3** 1-8.

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

Caboni P, Sammelson RE, Casida JE (2003) Phenylpyrazole insecticide photochemistry, metabolism, and GABAergic action: ethiprole compared with fipronil. *Journal of Agricultural and Food Chemistry* **51**, 7055-7061.

Caputo B, Nwakanma D, Jawara M*, et al.* (2008) *Anopheles gambiae* complex along The Gambia river, with particular reference to the molecular forms of *An. gambiae s.s*. *Malaria Journal* **7**, 182.

Cassone BJ, Molloy MJ, Cheng C*, et al.* (2011) Divergent transcriptional response to thermal stress by *Anopheles gambiae* larvae carrying alternative arrangements of inversion 2La. *Molecular Ecology* **20**, 2567-2580.

Catteruccia F, Benton JP, Crisanti A (2005) An *Anopheles* transgenic sexing strain for vector control. *Nature Biotechnology* **23**, 1414-1417.

Chanda E, Hemingway J, Kleinschmidt I*, et al.* (2011) Insecticide resistance and the future of malaria control in Zambia. *PLoS ONE* **6**, e24336.

Chandre F, Darrier F, Manga L*, et al.* (1999) Status of pyrethroid resistance in *Anopheles gambie sensu lato*. *Bulletin of the World Health Organization* **77**, 230 - 234.

Chandre F, Darriet F, Duchon S*, et al.* (2000) Modifications of pyrethroid effects associated with *kdr* mutation in *Anopheles gambiae*. *Medical and Veterinary Entomology* **14**, 81-88.

Chouaïbou M, Etang J, Brévault T*, et al.* (2008) Dynamics of insecticide resistance in the malaria vector *Anopheles gambiae s.l.* from an area of extensive cotton cultivation in Northern Cameroon. *Tropical Medicine & International Health* **13**, 476-486.

Christian RN, Strode C, Ranson H*, et al.* (2011) Microarray analysis of a pyrethroid resistant African malaria vector, *Anopheles funestus*, from southern Africa. *Pesticide Biochemistry and Physiology* **99**, 140-147.

Chung H, Bogwitz MR, McCart C*, et al.* (2007) Cis-regulatory elements in the *Accord* retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. *Genetics* **175**, 1071-1077.

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

Cohen MB, Feyereisen R (1995) A cluster of cytochrome P450 genes of the CYP6 family in the house fly. *DNA and cell biology* **14**, 73-82.

Collins F, Besansky N (1994) Vector biology and the control of malaria in Africa. *Science* **264**, 1874-1875.

Collins FH, Mendez MA, Rasmussen MO*, et al.* (1987) A ribosomal RNA gene probe Ddifferentiates mmber secies of the *Anopheles gambiae* complex. *American Journal of Tropical Medicine and Hygiene* **37**, 37-41.

Collins FH, Paskewitz SM (1996) A review of the use of ribosomal DNA (rDNA) to differentiate among cryptic *Anopheles* species. *Insect Molecular Biology* **5**, 1-9.

Coluzzi M, Petrarca V, di Deco MA (1985) Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. *Bolletino di zoologia* **52**, 45-63.

Coluzzi M, Sabatini A (1969) Cytogenetic observations on the salt water species, *Anopheles merus* and *Anopheles melas* of the *gambiae* complex. *Parassitologia* **11**, 117-186.

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

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

Corbel V, Chandre F, Brengues C*, et al.* (2004) Dosage-dependent effects of permethrin-treated nets on the behaviour of *Anopheles gambiae* and the selection of pyrethroid resistance. *Malaria Journal* **3**, 22.

Corbel V, N'Guessan R, Brengues C*, et al.* (2007) Multiple insecticide resistance mechanisms in *Anopheles gambiae* and *Culex quinquefasciatus* from Benin, West Africa. *Acta Tropica* **101**, 207-216.

Cornman RS, Togawa T, Dunn WA*, et al.* (2008) Annotation and analysis of a large cuticular protein family with the R&R Consensus in *Anopheles gambiae*. *BMC Genomics* **9**, 22.

Costantini C, Ayala D, Guelbeogo W*, et al.* (2009) Living at the edge: biogeographic patterns of habitat segregation conform to speciation by niche expansion in *Anopheles gambiae*. *BMC Ecology* **9**, 16.

Coulibaly MB, Lobo NF, Fitzpatrick MC*, et al.* (2007) Segmental duplication implicated in the genesis of inversion 2Rj of *Anopheles gambiae*. *PLoS ONE* **2**, e849.

Craig MH, Kleinschmidt I, Le Sueur D, Sharp BL (2004) Exploring 30 years of malaria case data in KwaZulu-Natal, South Africa: Part II. The impact of non-climatic factors. *Tropical Medicine & International Health* **9**, 1258-1266.

Dabiré KR, Diabaté A, Namountougou M*, et al.* (2009) Distribution of pyrethroid and DDT resistance and the L1014F *kdr* mutation in *Anopheles gambiae s.l.* from Burkina Faso (West Africa). *Transactions of the Royal Society of Tropical Medicine and Hygiene* **103**, 1113-1120.

Daborn PJ, Yen JL, Bogwitz MR*, et al.* (2002) A single P450 allele associated with insecticide resistance in *Drosophila*. *Science* **297**, 2253-2256.

Darvasi A, Soller M (1995) Advanced Intercross Lines, an experimental population for fine genetic mapping. *Genetics* **141**, 1199-1207.

David J-P, Strode C, Vontas J*, et al.* (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**, 4080-4084.

Davidson G (1964) The five mating-types in the *Anopheles gambiae* complex. *Rivista di Malariologia 1964 Dec;43:167-83.*

Davies TGE, Field LM, Usherwood PNR, Williamson MS (2007a) A comparative study of voltage-gated sodium channels in the Insecta: implications for pyrethroid resistance in Anopheline and other Neopteran species. *Insect Molecular Biology* **16**, 361-375.

Davies TGE, Field LM, Usherwood PNR, Williamson MS (2007b) DDT, pyrethrins, pyrethroids and insect sodium channels. *IUBMB Life* **59**, 151-162.

Devonshire AL, Byrne FJ, Moores GD, Williamson MS (1998) Biochemical and molecular characterisation of insecticide-insensitive acetylcholinesterases. *Structure and Function of Cholinesterases and Related Proteins*, 491-496.

Diabate A, Baldet T, Chandre F*, et al.* (2002) The role of agricultural use of insecticides in resistance to pyrethroids in *Anopheles gambiae s.l.* in Burkina Faso. *American Journal of Tropical Medicine and Hygiene* **67**, 617-622.

Diabate A, Baldet T, Chandre F*, et al.* (2004a) First report of a *kdr* mutation in *Anopheles arabiensis* from Burkina Faso, West Africa. *Journal - American Mosquito Control Association* **20**, 195 - 196.

Diabate A, Brengues C, Baldet T*, et al.* (2004b) The spread of the Leu-Phe *kdr* mutation through *Anopheles gambiae* complex in Burkina Faso:

genetic introgression and *de novo* phenomena. *Tropical Medicine & International Health* **9**, 1267-1273.

Diabaté A, Dabire RK, Millogo N, Lehmann T (2007) Evaluating the effect of postmating isolation between molecular forms of *Anopheles gambiae* (Diptera: Culicidae). *Journal of Medical Entomology* **44**, 60-64.

Ding Y, Hawkes N, Meredith J*, et al.* (2005) Characterization of the promoters of Epsilon glutathione transferases in the mosquito *Anopheles gambiae* and their response to oxidative stress. *Biochemical Journal* **387**, 879-888.

Ding Y, Ortelli F, Rossiter L, Hemingway J, Ranson H (2003) The *Anopheles gambiae* glutathione transferase supergene family: annotation, phylogeny and expression profiles. *BMC Genomics* **4**, 35.

Djegbe I, Boussari O, Sidick A*, et al.* (2011) Dynamics of insecticide resistance in malaria vectors in Benin: first evidence of the presence of L1014S *kdr* mutation in *Anopheles gambiae* from West Africa. *Malaria Journal* **10**, 261.

Djogbénou L, Chandre F, Berthomieu A*, et al.* (2008) Evidence of introgression of the *ace-1*^{R} mutation and of the *ace-1* duplication in West African *Anopheles gambiae s. s*. *PLoS ONE* **3**, e2172.

Djogbénou L, Labbe P, Chandre F, Pasteur N, Weill M (2009) *Ace-1* duplication in *Anopheles gambiae*: a challenge for malaria control. *Malaria Journal* **8**, 70.

Djogbénou L, Pasteur N, Akogbéto M, Weill M, Chandre F (2011) Insecticide resistance in the *Anopheles gambiae* complex in Benin: a nationwide survey. *Medical and Veterinary Entomology* **25**, 256–267.

Djogbénou L, Pasteur N, Bio-Bangana S*, et al.* (2010) Malaria vectors in the republic of Benin: distribution of species and molecular forms of the *Anopheles gambiae* complex. *Acta Tropica* **114**, 116-122.

Djouaka R (2008) Study of mechanisms of pyrethroid resistance developed by *Anopheles gambiae* in the south western Nigeria and Southern Benin: mapping of resistance, characterization of *Anopheles* breeding sites and analysis of the implication of agricultural pesticides and oil spillage in the selection of resistance. *PhD thesis* Department of Zoology, Cell Biology and Genetics Unit, University of Ibadan.

Djouaka RF, Bakare AA, Coulibaly ON*, et al.* (2008) Expression of the cytochrome P450s, *CYP6P3* and *CYP6M2* are significantly elevated in multiple pyrethroid resistant populations of *Anopheles gambiae s.s.* from Southern Benin and Nigeria. *BMC Genomics* **9**, 538.

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics* **3**, 43-52.

Donnelly MJ, Corbel V, Weetman D*, et al.* (2009) Does *kdr* genotype predict insecticide-resistance phenotype in mosquitoes? *Trends in Parasitology* **25**, 213-219.

Du W, Awolola TS, Howell P*, et al.* (2005) Independent mutations in the Rdl locus confer dieldrin resistance to *Anopheles gambiae* and *An. arabiensis*. *Insect Molecular Biology* **14**, 179-183.

Dworkin I, Palsson A, Gibson G (2005) Replication of an Egfr-wing shape association in a wild-caught cohort of *Drosophila melanogaster*. *Genetics* **169**, 2115-2125.

Elissa N, Mouchet J, Riviere F, Meunier JY, Yao K (1993) Resistance of *Anopheles gambiae* to pyrethroids in Cote d'Ivoire. *Annales de la Societe Belge de Medecine Tropicale* **73**, 291 - 294.

Elliott R, Ramakrishna V (1956) Insecticide resistance in *Anopheles gambiae* Giles. *Nature* **177**, 532-533.

Enayati AA, Ranson H, Hemingway J (2005) Insect glutathione transferases and insecticide resistance. *Insect Molecular Biology* **14**, 3-8.

Fanello C, Petrarca V, Torre Ad*, et al.* (2003) The pyrethroid knock-down resistance gene in the *Anopheles gambiae* complex in Mali and further indication of incipient speciation within *An. gambiae s.s. Insect Molecular Biology* **12**, 241-245.

Fanello C, Santolamazza F, Della Torre A (2002) Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. *Medical and Veterinary Entomology* **16**, 461-464.

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

Feyereisen R (1999) Insect P450 enzymes. *Annual Review of Entomology* **44**, 507-533.

Fire A, Xu S, Montgomery MK*, et al.* (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806-811.

Fisher RA, Yates F (1949) *Statistical tables for biological, agricultural and medical research.* Oliver and Boyd, Edinburgh

Flaxman AD, Fullman N, Otten MW, Jr.*, et al.* (2010) Rapid scaling up of insecticide-treated bed net coverage in Africa and its relationship with development assistance for health: A systematic synthesis of supply, distribution, and household survey data. *PLoS Medicine* **7**, e1000328.

Fukuto TR (1990) Mechanism of action of organophosphorus and carbamate insecticides. *Environmental Health Persepectives* **87**, 245-254.

Gentile G, Slotman M, Ketmaier V, Powell J, Caccone A (2001) Attempts to molecularly distinguish cryptic taxa in *Anopheles gambiae s.s*. *Insect Molecular Biology* **10**, 25-32.

Gillies M, Coetzee M (1987) A supplement to the anophelinae of Africa south of the Sahara (Afrotropical region). *Publications of the South African Institute for Medical Research*, 55.

Glunt KD, Thomas MB, Read AF (2011) The effects of age, exposure history and malaria infection on the susceptibility of *Anopheles*

mosquitoes to low concentrations of pyrethroid. *PLoS ONE* **6**, e24968.

Gould F (1984) Mixed function oxidases and herbivore polyphagy: the devil's advocate position. *Ecological Entomology* **9**, 29-34.

Habig WH, Pabst MJ, Jakoby WB (1974) Glutathione S-Transferases. *Journal of Biological Chemistry* **249**, 7130-7139.

Hall LM, Spierer P (1986) The *Ace* locus of *Drosophila melanogaster*: structural gene for acetylcholinesterase with an unusual 5' leader. *The EMBO Journal* **5**, 2949-1954.

Hargreaves K, Koekemoer LL, Brooke BD*, et al.* (2000) *Anopheles funestus* resistant to pyrethroid insecticides in South Africa. *Medical and Veterinary Entomology* **14**, 181-189.

Harris AF, Nimmo D, McKemey AR*, et al.* (2011) Field performance of engineered male mosquitoes. *Nature Biotechnology*.

Harris AF, Rajatileka S, Ranson H (2010) Pyrethroid resistance in *Aedes aegypti* from grand cayman. *The American Journal of Tropical Medicine and Hygiene* **83**, 277-284.

He N, Botelho JMC, McNall RJ*, et al.* (2007) Proteomic analysis of cast cuticles from *Anopheles gambiae* by tandem mass spectrometry. *Insect Biochemistry and Molecular Biology* **37**, 135-146.

Hedstrom L (2002) Serine protease mechanism and specificity. *Chemical Reviews* **102**, 4501-4524.

Hemingway J (1989) A note on simple biochemical methods for resistance detection and their field application in Sri Lanka. *Pesticide Science* **27**, 281-285.

Hemingway J, Penilla RP, Rodriguez AD*, et al.* (1997) Resistance management strategies in malaria vector mosquito control. A large-scale field trial in Southern Mexico. *Pesticide Science* **51**, 375-382.

Hemingway J, Ranson H (2000) Insecticide resistance in insect vectors of human disease. *Annual Review of Entomology* **45**, 371-391.

Hittinger CT, Johnston M, Tossberg JT, Rokas A (2010) Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. *Proceedings of the National Academy of Sciences* **107**, 1476-1481.

Holt RA, Subramanian GM, Halpern A*, et al.* (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**, 129-149.

Hoyer RF, Plapp JR. FW (1968) Insecticide resistance in the house fly: Identification of a gene that confers resistance to organotin insecticides and acts as an intensifier of parathion resistance. *Journal of Economic Entomology* **61**, 1269-1276.

Huang DW, Sherman BT, Lempicki RA (2009a) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**, 1-13.

Huang DW, Sherman BT, Lempicki RA (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44-57.

Ingles PJ, Adams PM, Knipple DC, Soderlund DM (1996) Characterization of voltage-sensitive sodium channel gene coding sequences from

insecticide-susceptible and knockdown-resistant house fly strains. *Insect Biochemistry and Molecular Biology* **26**, 319-326.

Jaenisch R (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics* **33**, 245-254.

Jones C, Sanou A, Guelbeogo WM*, et al.* (2012a) Aging partially restores the efficacy of malaria vector control in insecticide-resistant populations of *Anopheles gambiae s.l.* from Burkina Faso. *MalariaJournal*, 11:24.

Jones CM, Liyanapathirana M, Agossa FR*, et al.* (2012b) Footprints of positive selection associated with a mutation (N1575Y) in the voltage-gated sodium channel of *Anopheles gambiae*. *Proceedings of the National Academy of Sciences* **109**, 6614-6619.

Kappe SHI, Vaughan AM, Boddey JA, Cowman AF (2010) That was then but this is now: malaria research in the time of an eradication agenda. *Science* **328**, 862-866.

Kerah-Hinzoumbe C, Peka M, Antonio-Nkondjio C*, et al.* (2009) Malaria vectors and transmission dynamics in Goulmoun, a rural city in south-western Chad. *BMC Infectious Diseases* **9**, 71.

Kerah-Hinzoumbe C, Peka M, Nwane P*, et al.* (2008) Insecticide resistance in *Anopheles gambiae* from south-western Chad, Central Africa. *Malaria Journal* **7**, 192.

Killeen GF, McKenzie FE, Foy BD, Bøgh C, Beier JC (2001) The availability of potential hosts as a determinant of feeding behaviours and malaria transmission by African mosquito populations. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **95**, 469-476.

Killeen GF, Smith TA, Ferguson HM*, et al.* (2007) Preventing childhood malaria in africa by protecting adults from mosquitoes with insecticide-treated nets. *PLoS Med* **4**, e229.

Kirst M, Myburg AA, De Leon JP*, et al.* (2004) Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus. *Plant physiology* **135**, 2368-2378.

Kiszewski A, Mellinger A, Spielman A*, et al.* (2004) A global index representing the stability of malaria transmission. *American Journal of Tropical Medicine and Hygiene* **70**, 486-498.

Knipple DC, Doyle KE, Marsella-Herrick PA, Soderlund DM (1994) Tight genetic linkage between the *kdr* insecticide resistance trait and a voltage-sensitive sodium channel gene in the house fly. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 2483-2487.

Kulkarni M, Rowland M, Alifrangis M*, et al.* (2006) Occurrence of the leucine-to-phenylalanine knockdown resistance (*kdr*) mutation in *Anopheles arabiensis* populations in Tanzania, detected by a simplified high-throughput SSOP-ELISA method. *Malaria Journal* **5**, 56.

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185-199.

Lawniczak MKN, Emrich SJ, Holloway AK*, et al.* (2010) Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* **330**, 512-514.

Lehmann T, Diabate A (2008) The molecular forms of *Anopheles gambiae*: a phenotypic perspective. *Infection, Genetics and Evolution* **8**, 737-746.

Lengeler C (2004) Insecticide-treated bed nets and curtains for preventing malaria. *The Cochrane Database of Systematic Reviews 2004*.

Li X, Schuler MA, Berenbaum MR (2007) Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual Review of Entomology* **52**, 231-253.

Livak KJ (1984) Organization and mapping of a sequence on the *Drosophila melanogaste*r X and Y chromosomes that is transcribed during spermatogenesis. *Genetics* **107**, 611-634.

Lobo N, Sangare D, Regier A*, et al.* (2010) Breakpoint structure of the *Anopheles gambiae* 2Rb chromosomal inversion. *Malaria Journal* **9**, 293.

Lockwood JA, Byford RL, Story RN, Sparks TC, Quisenberry SS (1985) Behavioral resistance to the pyrethroids in the horn fly, *Haematobia irritans* (Diptera: Muscidae). *Environmental Entomology* **14**, 873-880.

Loughney K, Kreber R, Ganetzky B (1989) Molecular analysis of the *para* locus, a sodium channel gene in *Drosophila*. *Cell* **58**, 1143-1154.

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, *Aedes aegypti*. *Insect Biochemistry and Molecular Biology* **35**, 861-871.

Lumjuan N, Rajatileka S, Changsom D*, et al.* (2011) The role of the *Aedes aegypti* Epsilon glutathione transferases in conferring resistance to DDT and pyrethroid insecticides. *Insect Biochemistry and Molecular Biology* **41**, 203-209.

Lycett GJ, McLaughlin LA, Ranson H*, et al.* (2006) *Anopheles gambiae* P450 reductase is highly expressed in oenocytes and *in vivo* knockdown increases permethrin susceptibility. *Insect Molecular Biology* **15**, 321-327.

Lynd A, Ranson H, McCall P*, et al.* (2005) A simplified high-throughput method for pyrethroid knock-down resistance (*kdr*) detection in *Anopheles gambiae*. *Malaria Journal* **4**, 16.

Lynd A, Weetman D, Barbosa S*, et al.* (2010) Field, genetic, and modeling approaches show strong positive selection acting upon an insecticide resistance mutation in *Anopheles gambiae s.s. Molecular Biology and Evolution* **27**, 1117-1125.

Magkrioti C, Spyropoulos I, Iconomidou V, Willis J, Hamodrakas S (2004) cuticleDB: a relational database of Arthropod cuticular proteins. *BMC Bioinformatics* **5**, 138.

Marinotti O, Nguyen QK, Calvo E, James AA, Ribeiro JMC (2005) Microarray analysis of genes showing variable expression following a blood meal in *Anopheles gambiae*. *Insect Molecular Biology* **14**, 365-373.

Marsden CD, Lee Y, Nieman CC*, et al.* (2011) Asymmetric introgression between the M and S forms of the malaria vector, *Anopheles gambiae*, maintains divergence despite extensive hybridization. *Molecular Ecology* **20**, 4983-4994.

Martinez-Torres D, F. C, S. WM*, et al.* (1998) Molecular characterization of pyrethroid knockdown resistance (*kdr*) in the major malaria vector *Anopheles gambiae* s.s. *Insect Molecular Biology* **7**, 179-184.

Martinez-Torres D, C C, A B-B*, et al.* (1999a) Voltage-dependent Na+ channels in pyrethroid-resistant *Culex pipiens* L mosquitoes. *Pesticide Science* **55**, 1012-1020.

Martinez-Torres D, Foster SP, Field LM, Devonshire AL, Williamson MS (1999b) A sodium channel point mutation is associated with resistance to DDT and pyrethroid insecticides in the peach-potato aphid, *Myzus persicae* (Sulzer) (Hemiptera: Aphididae). *Insect Molecular Biology* **8**, 339-346.

Mathenge EM, Gimnig JE, Kolczak M*, et al.* (2001) Effect of permethrin-impregnated nets on exiting behavior, blood feeding success, and time of feeding of malaria mosquitoes (*Diptera*: *Culicidae*) in Western Kenya. *Journal of Medical Entomology* **38**, 531-536.

Mathiopoulos KD, della Torre A, Predazzi V, Petrarca V, Coluzzi M (1998) Cloning of inversion breakpoints in the *Anopheles gambiae* complex traces a transposable element at the inversion junction. *Proceedings of the National Academy of Sciences* **95**, 12444-12449.

Milani R (1954) Comportamento mendeliano della resistenza alla azione abbatante del DDT: correlazione tran abbatimento e mortalia in *Musca domestica*. *Rivista Parasitologia* **15**, 513-542.

Mitchell SN (2011) Molecular mechanisms underlying insecticide resistance phenotypes in the major malaria vector *Anopheles gambiae* sensu stricto. *PhD thesis* Liverpool School of Tropical Medicine, University of Liverpool, UK.

Mitchell SN, Stevenson BJ, Müller P*, et al.* (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**, 6147-6152.

Miyazaki M, Ohyama K, Dunlap D, 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* **252**, 61-68.

Morgan JC, Irving H, Okedi LM, Steven A, Wondji CS (2010) Pyrethroid resistance in an *Anopheles funestus* population from Uganda. *PLoS ONE* **5**, e11872.

Morrison TB, Weis JJ, Wittwer CT (1998) Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *BioTechniques* **24**, 954-958, 960, 962.

Mouatcho JC, Munhenga G, Hargreaves K*, et al.* (2009) Pyrethroid resistance in a major African malaria vector *Anopheles arabiensis* from Mamfene, northern KwaZulu-Natal, South Africa. *South African Journal of Science* **105**, 127-131.

Müller P, ChouaÏBou M, Pignatelli P*, et al.* (2008a) Pyrethroid tolerance is associated with elevated expression of antioxidants and agricultural practice in *Anopheles arabiensis* sampled from an area of cotton fields in Northern Cameroon. *Molecular Ecology* **17**, 1145-1155.

Müller P, Donnelly M, Ranson H (2007) Transcription profiling of a recently colonised pyrethroid resistant *Anopheles gambiae* strain from Ghana. *BMC Genomics* **8**, 36.

Müller P, Warr E, Stevenson BJ*, et al.* (2008b) Field-caught permethrin-resistant *Anopheles gambiae* overexpress *CYP6P3*, a P450 that metabolises pyrethroids. *PLoS Genetics* **4**, e1000286.

Munhenga G, Koekemoer LL (2011) Differential expression of cytochrome P450 genes in a laboratory selected *Anopheles arabiensis* colony. *African Journal of Biotechnology* **10**, 12741-12746.

Murray CJL, Rosenfeld LC, Lim SS*, et al.* (2012) Global malaria mortality between 1980 and 2010: a systematic analysis. *The Lancet* **379**, 413-431.

Mutero A, Pralavorio M, Bride JM, Fournier D (1994) Resistance-associated point mutations in insecticide-insensitive acetylcholinesterase. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5922-5926.

N'Guessan R, Darriet F, Guillet P*, et al.* (2003) Resistance to carbosulfan in *Anopheles gambiae* from Ivory Coast, based on reduced sensitivity of acetylcholinesterase. *Medical and Veterinary Entomology* **17**, 19-25.

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**, 199-206.

Nauen R (2007) Insecticide resistance in disease vectors of public health importance. *Pest Management Science* **63**, 628-633.

Neafsey DE, Lawniczak MKN, Park DJ*, et al.* (2010) SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. *Science* **330**, 514-517.

Niare O, Markianos K, Volz J*, et al.* (2002) Genetic loci affecting resistance to human malaria parasites in a West African mosquito vector population. *Science* **298**, 213-216.

Nikou D, Ranson H, Hemingway J (2003) An adult-specific CYP6 P450 gene is overexpressed in a pyrethroid-resistant strain of the malaria vector, *Anopheles gambiae*. *Gene* **318**, 91-102.

O'Reilly AO, Khambay BP, Williamson MS*, et al.* (2006) Modelling insecticide-binding sites in the voltage-gated sodium channel. *The Biochemical journal* **396**, 255-263.

Oetting WS, Armstrong CM, Ronan SM*, et al.* (1998) Multiplexed short tandem repeat polymorphisms of the Weber 8A set of markers using tailed primers and infrared fluorescence detection. *Electrophoresis* **19**, 3079-3083.

Ortelli F, Rossiter LC, Vontas J, Ranson H, Hemingway J (2003) Heterologous expression of four glutathione transferase genes genetically linked to a major insecticide-resistance locus from the

malaria vector *Anopheles gambiae*. *Biochemical Journal* **373**, 957-963.

Osta MA, Christophides GK, Kafatos FC (2004) Effects of mosquito genes on *Plasmodium* development. *Science* **303**, 2030-2032.

Pates H, Curtis C (2005) Mosquito behaviour and vector control. *Annual Review of Entomology* **50**, 53-70.

Pedra JHF, McIntyre LM, Scharf ME, Pittendrigh BR (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**, 7034-7039.

Penilla PR, Rodríguez AD, Hemingway J*, et al.* (1998) Resistance management strategies in malaria vector mosquito control. Baseline data for a large-scale field trial against *Anopheles albimanus* in Mexico. *Medical and Veterinary Entomology* **12**, 217-233.

Plapp FW (1976) Biochemical genetics of insecticide resistance. *Annual Review of Entomology* **21**, 179-197.

Pombi M, Caputo B, Simard F*, et al.* (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**, 309.

Poupardin R, Reynaud S, Strode C*, et al.* (2008) Cross-induction of detoxification genes by environmental xenobiotics and insecticides in the mosquito *Aedes aegypti*: impact on larval tolerance to chemical insecticides. *Insect Biochemistry and Molecular Biology* **38**, 540-551.

Poupardin R, Riaz MA, Jones CM*, et al.* (2012) Do pollutants affect insecticide-driven gene selection in mosquitoes? Experimental evidence from transcriptomics. *Aquatic Toxicology* **114–115**, 49-57.

Pralavorio M, Fournier D (1992) *Drosophila* acetylcholinesterase: characterization of different mutants resistant to insecticides. *Biochemical Genetics* **30**, 77-83.

Puinean AM, Foster SP, Oliphant L*, et al.* (2010) Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae. PLoS Genetics* **6**, e1000999.

Rajatileka S, Burhani J, Ranson H (2010) Mosquito age and susceptibility to insecticides. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **105**, 247-253.

Ranson H, Abdallah H, Badolo A*, et al.* (2009) Insecticide resistance in *Anopheles gambiae*: data from the first year of a multi-country study highlight the extent of the problem. *Malaria Journal* **8**, 299.

Ranson H, Claudianos C, Ortelli F*, et al.* (2002a) Evolution of supergene families associated with insecticide resistance. *Science* **298**, 179-181.

Ranson H, Jensen B, Vulule J*, et al.* (2000a) 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**, 491-497.

Ranson H, Jensen B, Wang X*, et al.* (2000b) Genetic mapping of two loci affecting DDT resistance in the malaria vector *Anopheles gambiae*. *Insect Molecular Biology* **9**, 499-507.

Ranson H, N'Guessan R, Lines J*, et al.* (2011) Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? *Trends in Parasitology* **27**, 91-98.

Ranson H, Nikou D, Hutchinson M*, et al.* (2002b) Molecular analysis of multiple cytochrome P450 genes from the malaria vector, *Anopheles gambiae*. *Insect Molecular Biology* **11**, 409-418.

Ranson H, Paton MG, Jensen B*, et al.* (2004) Genetic mapping of genes conferring permethrin resistance in the malaria vector, *Anopheles gambiae*. *Insect Molecular Biology* **13**, 379-386.

Ranson H, Rossiter L, Ortelli F*, et al.* (2001) Identification of a novel class of insect glutathione S-transferases involved in resistance to DDT in the malaria vector *Anopheles gambiae*. *Biochemical Journal* **359**, 295-304.

Rasmussen RP (2001) *Quantification on the LightCycler* Springer Press, Heidelberg.

Rebers JE, Riddiford LM (1988) Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle genes. *Journal of Molecular Biology* **203**, 411-423.

Rebers JE, Willis JH (2001) A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochemistry and Molecular Biology* **31**, 1083-1093.

Riehle MM, Markianos K, Niare O*, et al.* (2006) Natural malaria infection in *Anopheles gambiae* is regulated by a single genomic control region. *Science* **312**, 577-579.

Ronis MJJ, Hodgson E (1989) Cytochrome P-450 monooxygenases in insects. *Xenobiotica* **19**, 1077-1092.

Rusinov V, Baev V, Minkov IN, Tabler M (2005) MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence. *Nucleic Acids Research* **33**, W696-W700.

Russell RJ, Claudianos C, Campbell PM*, et al.* (2004) Two major classes of target site insensitivity mutations confer resistance to organophosphate and carbamate insecticides. *Pesticide Biochemistry and Physiology* **79**, 84-93.

Saavedra-Rodriguez KL, Strode C, Flores Suarez A*, et al.* (2008) QTL mapping of genome regions controlling permethrin resistance in the mosquito *Aedes aegypti*. *Genetics* **180**, 1137-1152.

Salazar CE, Mills-Hamm D, Kumar V, Collins FH (1993) Sequence of a cDNA from the mosquito *Anopheles gambiae* encoding a homologue of human ribosomal protein S7. *Nucleic Acids Research* **21**, 4147.

Santolamazza F, Calzetta M, Etang J*, et al.* (2008a) Distribution of knock-down resistance mutations in *Anopheles gambiae* molecular forms in west and west-central Africa. *Malaria Journal* **7**, 74.

Santolamazza F, Mancini E, Simard F*, et al.* (2008b) Insertion polymorphisms of *SINE200* retrotransposons within speciation islands of *Anopheles gambiae* molecular forms. *Malaria Journal* **7**, 163.

Sawicki RM (1978) Unusual response of DDT-resistant houseflies to carbinol analogues of DDT. *Nature* **275**, 443-444.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18**, 233-234.

Schulze A, Downward J (2001) Navigating gene expression using microarrays - a technology review. *Nature Cell Biology* **3**, E190-E195.

Scott JA, Brogdon WG, Collins FH (1993) Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *American Journal of Tropical Medicine and Hygiene* **49**, 520-529.

Scott JG (1999) Cytochromes P450 and insecticide resistance. *Insect Biochemistry and Molecular Biology* **29**, 757-777.

Service MW (2004) *Medical Entomology for Students* Cambridge.

Sharakhov IV, White BJ, Sharakhova MV*, et al.* (2006) Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion (2La) in the *Anopheles gambiae* complex. *Proceedings of the National Academy of Sciences* **103**, 6258-6262.

Sharp B, Ridl F, Govender D, Kuklinski J, Kleinschmidt I (2007) Malaria vector control by indoor residual insecticide spraying on the tropical island of Bioko, Equatorial Guinea. *Malaria Journal* **6**, 52.

Simard F, Ayala D, Kamdem G*, et al.* (2009) Ecological niche partitioning between *Anopheles gambiae* molecular forms in Cameroon: the ecological side of speciation. *BMC Ecology* **9**, 17.

Sinka M, Rubio-Palis Y, Manguin S*, et al.* (2010) The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. *Parasites & Vectors* **3**, 72.

Slate J (2005) Quantitative trait locus mapping in natural populations: progress, caveats and future directions. *Molecular Ecology* **14**, 363-379.

Stam R, van Ooijen JW (1996) JoinMap (TM) version 2.0: Software for the calculation of genetic linkage maps. Wageningen : CPRO-DLO.

Stevenson BJ, Bibby J, Pignatelli P*, et al.* (2011) Cytochrome P450 6M2 from the malaria vector *Anopheles gambiae* metabolizes pyrethroids: sequential metabolism of deltamethrin revealed. *Insect Biochemistry and Molecular Biology* **41**, 492-502.

Strode C, Wondji CS, David J-P*, et al.* (2007) Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*. *Insect Biochemistry and Molecular Biology* **38**, 113-123.

Strode KS, F. Ortelli, H. Ranson, (2006) Differential expression of the detoxification genes in the different life stages of the malaria vector *Anopheles gambiae*. *Insect Molecular Biology* **15**, 523-530.

Stump AD, Atieli FK, Vulule JM, Besansky NJ (2004) Dynamics of the pyrethroid knockdown resistance allele in western Kenyan populations of *Anopheles gambiae* in response to insecticide-treated bed net trials. *American Journal of Tropical Medicine and Hygiene* **70**, 591 - 596.

Taylor B (1975) Changes in the feeding behaviour of a malaria vector, *Anopheles farauti* Lav., following use of DDT as a residual spray in

houses in the British Solomon Islands Protectorate. *Transactions of the Royal Entomological Society of London* **127**, 277-292.

The RTS SCTP (2011) First results of phase 3 trial of RTS,S/AS01 malaria vaccine in african children. *New England Journal of Medicine* **365**, 1863-1875.

Tijet N, Helvig C, Feyereisen R (2001) The cytochrome P450 gene superfamily in *Drosophila melanogaster*: Annotation, intron-exon organization and phylogeny. *Gene* **262**, 189-198.

Torre della A, Tu Z, Petrarca V (2005) On the distribution and genetic differentiation of *Anopheles gambiae s.s.* molecular forms. *Insect Biochemistry and Molecular Biology* **35**, 755 - 769.

Torre della A., Fanello C, Akogbeto M*, et al.* (2001) Molecular evidence of incipient speciation within *Anopheles gambiae s.s.* in West Africa. *Insect Molecular Biology* **10**, 9-18.

Touré YT, Petrarca V, Traoré SF*, et al.* (1994) Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae s.s.* in Mali, West Africa. *Genetica* **94**, 213-223.

Tripet F, Touré YT, Taylor CE*, et al.* (2001) DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae*. *Molecular Ecology* **10**, 1725-1732.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS biology* **3**, e285.

USAID-CDC Interagency Working G (2005) President's Malaria Initiative, Strategic Plan.

Valles SM, Dong K, Brenner RJ (2000) Mechanisms responsible for cypermethrin resistance in a strain of german cockroach, *Blattella germanica*. *Pesticide Biochemistry and Physiology* **66**, 195-205.

Vaughan A, Hawkes N, Hemingway J (1997) Co-amplification explains linkage disequilibrium of two mosquito esterase genes in insecticide-resistant *Culex quinquefasciatus*. *Biochemical Journal* **325**, 359-365.

Verhaeghen K, Van Bortel W, Roelants P, Backeljau T, Coosemans M (2006) Detection of the East and West African *kdr* mutation in *Anopheles gambiae* and *Anopheles arabiensis* from Uganda using a new assay based on FRET/Melt Curve analysis. *Malaria Journal* **5**, 16.

Verhaeghen K, Van Bortel W, Trung HD, Sochantha T, Coosemans M (2008) Absence of knockdown resistance suggests metabolic resistance in the main malaria vectors of the Mekong region. *Malaria Journal* **8**, 84.

Vontas J CB, A. C. Koutsos, J.-P. David, F. C. Kafatos, C. Louis, J. Hemingway, G. K. Christophides, H. Ranson, (2005) Gene expression in insecticide resistant and susceptible *Anopheles gambiae* strains constitutively or after insecticide exposure. *Insect Molecular Biology* **14**, 509-521.

Vontas J, David J-P, Nikou D*, et al.* (2007) Transcriptional analysis of insecticide resistance in *Anopheles stephensi* using cross-species microarray hybridization. *Insect Molecular Biology* **16**, 315-324.

Vontas JG, Small GJ, Hemingway J (2001) Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in Nilaparvata lugens. *The Biochemical journal* **357**, 65-72.

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity* **93**, 77-78.

Walker K, Lynch M (2007) Contributions of *Anopheles* larval control to malaria suppression in tropical Africa: review of achievements and potential. *Medical and Veterinary Entomology* **21**, 2-21.

Wang R, Kafatos FC, Zheng L (1999) Microsatellite Markers and Genotyping Procedures for *Anopheles gambiae*. *Parasitology Today* **15**, 33-37.

Wang S, Basten C, Zeng Z (2005) Windows QTL cartographer 2.5. In: *Department of Statistics, North Carolina State University, Raleigh, NC*.

Wang Y, Qiu L, Ranson H*, et al.* (2008) Structure of an insect epsilon class glutathione S-transferase from the malaria vector *Anopheles gambiae* provides an explanation for the high DDT-detoxifying activity. *Journal of Structural Biology* **164**, 228-235.

Weetman D, Wilding CS, Steen K*, et al.* (2010) Association mapping of insecticide resistance in wild *Anopheles gambiae* populations: major variants identified in a low-linkage disequilbrium genome. *PLoS ONE* **5**, e13140.

Weetman D, Wilding CS, Steen K, Pinto J, Donnelly MJ (2012) Gene flow-dependent genomic divergence between *Anopheles gambiae* M and S forms. *Molecular Biology and Evolution* **29**, 279-291.

Weill M (2002) A novel acetylcholinesterase gene in mosquitoes codes for the insecticide target and is non-homologous to the *ace* gene in *Drosophila*. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**, 2007-2016.

Weill M, Chandre F, Brengues C*, et al.* (2000) The *kdr* mutation occurs in the Mopti form of *Anopheles gambiae s.s.* through introgression. *Insect Molecular Biology* **9**, 451-455.

Weill M, Lutfalla G, Mogensen K*, et al.* (2003) Comparative genomics: Insecticide resistance in mosquito vectors. *Nature* **423**, 136-137.

Weill M., Malcolm C, Chandre F*, et al.* (2004) The unique mutation in *ace-1* giving high insecticide resistance is easily detectable in mosquito vectors. *Insect Molecular Biology* **13**, 1-7.

Werck-Reichhart D, Feyereisen R (2000) Cytochromes P450: a success story. *Genome Biology* **1**, reviews: 3003.3001 - 3003.3009.

Wheelock CE, Shan G, Ottea J (2005) Overview of carboxylesterases and their role in the metabolism of insecticides. *Journal of Pesticide Science* **30**, 75-83.

White BJ, Cheng C, Sangaré D*, et al.* (2009) The population genomics of trans-specific inversion polymorphisms in *Anopheles gambiae*. *Genetics* **183**, 275-288.

White BJ, Collins FH, Besansky NJ (2011a) Evolution of *Anopheles gambiae* in relation to humans and malaria. *Annual Review of Ecology, Evolution, and Systematics* **42**, 111-132.

White BJ, Lawniczak MKN, Cheng C*, et al.* (2011b) Adaptive divergence between incipient species of *Anopheles gambiae* increases

resistance to *Plasmodium*. *Proceedings of the National Academy of Sciences* **108**, 244-249.

White BJ, Santalomazza F, Kamau L*, et al.* (2007) Molecular karyotyping of the 2La inversion in *Anopheles gambiae*. *American Journal of Tropical Medicine and Hygiene* **76**, 334-339.

White GB (1974) *Anopheles gambiae* complex and disease transmission in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **68**, 278-298.

White GB (1975) Notes on a catalogue of *Culicidae* of the ethiopian region. *Mosquito Systematics* **7** 303-344.

WHO (1998) Test procedures for insecticide resistance monitoring in malaria vectors, bio-efficacy and persistence of insecticides on treated surfaces: report of the WHO informal consultation. *World Health Organization*.

WHO (2006) Position statement: Indoor residual spraying - Use of indoor residual spraying for scaling up global malaria control and elimination. *World Health Organization*.

WHO (2008) World Malaria Report 2008. *World Health Organization*.

WHO (2011a) Global Insecticide Use for Vector-Borne Disease Control - A 10-Year Assessment (2000-2009). *World Health Organization*.

WHO (2011b) Monitoring of insecticide Resistance in Malaria Vectors. *World Health Organization*.

WHO (2011c) World Malaria Report 2011. *World Health Organization*.

Wilding C, Weetman D, Steen K, Donnelly M (2009) High, clustered, nucleotide diversity in the genome of *Anopheles gambiae* revealed through pooled-template sequencing: implications for high-throughput genotyping protocols. *BMC Genomics* **10**, 1-11.

Williamson MS, Denholm I, A. BC, L. DA (1993) Knockdown resistance (*kdr*) to DDT and pyrethroid insecticides maps to a sodium channel gene locus in the housefly (*Musca domestica*). *Molecular and General Genetics* **240**, 17-22.

Williamson MS, Martinez-Torres D, Hick CA, Devonshire AL (1996) Identification of mutations in the housefly *para*-type sodium channel gene associated with knockdown resistance (*kdr*) to pyrethroid insecticides. *Molecular and General Genetics* **252**, 51-60.

Willis JH (2010) Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochemistry and Molecular Biology* **40**, 189-204.

Winter A, Weir W, Hunt M*, et al.* (2012) Diversity in parasitic nematode genomes: the microRNAs of *Brugia pahangi* and *Haemonchus contortus* are largely novel. *BMC Genomics* **13**, 4.

Wondji CS, Irving H, Morgan J*, et al.* (2009) Two duplicated P450 genes are associated with pyrethroid resistance in Anopheles funestus, a major malaria vector. *Genome Research* **19**, 452-459.

Wondji CS, Morgan J, Coetzee M*, et al.* (2007) Mapping a Quantitative Trait Locus (QTL) conferring pyrethroid resistance in the African malaria vector *Anopheles funestus*. *BMC Genomics* **8**, 34.

Yadouleton A, Asidi A, Djouaka R*, et al.* (2009) Development of vegetable farming: a cause of the emergence of insecticide resistance in

populations of *Anopheles gambiae* in urban areas of Benin. *Malaria Journal* **8**, 103.

Yadouleton A, Padonou G, Asidi A*, et al.* (2010) Insecticide resistance status in *Anopheles gambiae* in southern Benin. *Malaria Journal* **9**, 83.

Yamamoto K, Shigeoka Y, Aso Y*, et al.* (2009) Molecular and biochemical characterization of a Zeta-class glutathione S-transferase of the silkmoth. *Pesticide Biochemistry and Physiology* **94**, 30-35.

Yewhalaw D, Bortel WV, Denis L*, et al.* (2010) First evidence of high knockdown resistance frequency in *Anopheles arabiensis* (Diptera: Culicidae) from Ethiopia. *The American Journal of Tropical Medicine and Hygiene* **83**, 122-125.

Yewhalaw D, Wassie F, Steurbaut W*, et al.* (2011) Multiple insecticide resistance: An impediment to insecticide-based malaria vector control program. *PLoS ONE* **6**, e16066.

Zeidler O (1874) I. Verbindungen von Chloral mit Brom- und Chlorbenzol. *Berichte der deutschen chemischen Gesellschaft* **7**, 1180-1181.

Zeng Z-B (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 10972-10976.

Zeng Z-B, Kao C-H, Basten CJ (1999) Estimating the genetic architecture of quantitative traits. *Genetics Research* **74**, 279-289.

Zheng L, Benedict MQ, Cornel AJ, Collins FH, Kafatos FC (1996) An integrated genetic map of the African human malaria vector mosquito, *Anopheles gambiae*. *Genetics* **143**, 941-952.

Zheng L, Collins F, Kumar V, Kafatos F (1993) A detailed genetic map for the X chromosome of the malaria vector, *Anopheles gambiae*. *Science* **261**, 605-608.

Chapter 9. Appendix

				all					alive					dead		
Family 15	aa	ab	bb	X2		aa	ab	bb	X2		aa	ab	bb	X2	
803	10	25		6.43	**	6	12		2.00		4	13		4.76	**
197	2	25	11	8.05	**	0	13	6	6.37	**	2	12	5	2.26	
757	12	23		3.46	*	5	13		3.56	*	7	10		0.53	
2R_si_5	18	20		0.11		4	13		4.76	**	10	7		0.53	
Cyp6P3	14	20		1.06		10	9		0.05		8	11		0.47	
637 [$]	23	15		1.68		15	4		6.37	**	8	11		0.47	
675	7	20	11	0.95		0	12	7	6.47	**	7	8	4	1.42	
776	21	9		4.80	**	13	2		8.07	****	8	7		0.07	
59	28	9		9.76	****	14	5		4.26	**	14	4		5.56	**
812	7	30	1	14.63	*****	6	13		6.37	**	1	17	1	11.84	****
249	7	20	10	0.73		5	11	2	1.89		2	9	8	3.84	
30C1	21	15		1.00		6	11		1.47		15	4		6.37	**
CPLCG3	22	15		1.32		7	12		1.32		15	3		8.00	****
6M19	21	16		0.68		6	12		2.00		15	4		6.37	**
119	6	18	12	2.00		3	10	5	0.67		3	8	7	2.00	
555	8	19	11	0.47		4	10	5	0.16		4	9	6	0.47	
6H1	5	22	11	2.84		2	11	6	2.16		3	11	5	0.89	
577	8	20	7	0.77		5	10	3	0.67		3	10	4	0.65	
Family 17	aa	ab	bb			aa	ab	bb	X2		aa	ab	bb	X2	
803	15	28	16	0.19		13	21	10	0.50		2	7	6	2.20	
197	21	29		1.28		16	19		0.26		5	10		1.67	
757 [$]	31	25		0.64		19	22		0.22		12	3		5.40	**
cyp6p3	16	47		15.25	*******	11	36		13.30	******	5	11		2.25	
2R_si_5	8	29	17	3.30		7	19	12	1.32		1	10	5	3.00	

770 $	19	30		2.47		13	25		3.79	*	6	5		0.09	
117	17	46		13.35	******	12	35		11.26	*****	5	11		2.25	
kdr	29	31		0.07		23	22		0.02		6	9		0.60	
787 $	47	16		15.25	*******	33	14		7.68	***	14	2		9.00	****
603	42	18		9.60	****	34	10		13.09	******	8	8		0.00	
776	20	24	13	3.14		14	20	8	1.81		6	4	5	3.40	
59	15	39		10.67	****	8	31		13.56		7	8		0.07	
812	19	26	11	2.57		14	18	8	2.20		5	8	3	0.50	
cyp6m2	22	41		5.73	**	18	29		2.57		4	12		4.00	**
249	15	26	7	3.00		11	17	4	3.19		4	9	3	0.38	
30C1	20	33	6	7.47	**	16	24	4	6.91	**	4	9	2	1.13	
cplc3	20	43		8.40	****	17	30		3.60	*	3	13		6.25	**
6M19	20	26	8	5.41	*	17	17	6	6.95	**	3	9	2	1.29	
341	15	22	10	1.26		13	12	9	3.88		2	10	1	3.92	
6H1	16	30	10	1.57		13	18	9	1.20		3	12	1	4.50	
33C1	25	31		0.64		19	21		0.10		6	10		1.00	
6F5	35	21		3.50	*	23	17		0.90		12	4		4.00	**
30L17 $	30	29		0.02		22	21		0.02		8	8		0.00	
SODI	21	36		3.95	**	14	29		5.23	**	7	7		0.00	
811	31	30		0.02		23	22		0.02		8	8		0.00	
577	39	19		6.90	***	28	16		3.27	*	11	3		4.57	**
817	15	25	19	1.92		15	18	11	2.18		0	7	8	8.60	**
46C2	14	29	17	0.37		12	21	11	0.14		2	8	6	2.00	
53 $	33	22		2.20		23	16		1.26		10	6		1.00	
80 $	15	11	12	7.21	**	7	7	9	3.87		8	4	3	6.60	**
Family 25	aa	ab	bb			aa	ab	bb a	X2		aa	ab	bb a	X2	
46 $	12	8	22	20.86	*******	6	3	12	14.14	*****	6	5	10	7.29	**
757	15	27		3.43	*	6	14		3.20	*	9	13		0.73	
799	10	22	10	0.10		6	11	4	0.43		4	11	6	0.43	
2R_si_5	4	30	7	9.24	***	1	15	5	5.38	*	3	15	2	5.10	*

Cyp6P3	19	24		0.58		8	13		1.19		11	11		0.00	
770	4	26	12	5.43	*	0	15	6	7.29	**	4	11	6	0.43	
117	30	11		8.80	****	13	8		1.19		17	3		9.80	****
787	1	23	15	11.31	****	1	14	4	5.21	*	0	9	11	12.30	****
772	2	26	14	9.24	***	0	16	5	8.14	**	2	10	9	4.71	*
675	4	23	15	6.14	**	3	13	5	1.57		1	10	10	7.76	**
776	16	24		1.60		6	13		2.58		10	11		0.05	
59	11	18	13	1.05		3	10	7	1.60		8	8	6	2.00	
812	14	27		4.12	**	7	13		1.80		7	14		2.33	
Cyp6M2	11	19	13	0.77		5	9	7	0.81		6	10	6	0.18	
249	18	23		0.61		9	11		0.20		9	12		0.43	
30C1	27	15		3.43	*	14	6		3.20	*	13	9		0.73	
CPLCG3	6	34		19.60	*******	3	18		10.71	****	3	16		8.89	****
6M19	12	20	10	0.29		4	11	5	0.30		8	9	5	1.55	
119	16	27		2.81	*	7	14		2.33		9	13		0.73	
158	32	9		12.90	******	15	5		5.00	**	17	4		8.05	****
341 $	33	7		16.90	*******	16	4		7.20	***	17	3		9.80	****
6H1	12	18	12	0.86		6	8	7	1.29		6	10	5	0.14	
6F5	13	29		6.10	**	7	13		1.80		6	16		4.55	**
88	11	18	12	0.66		7	8	5	1.20		4	10	7	0.90	
30L17 $	33	9		13.71	******	15	5		5.00	**	18	4		8.91	****
811 $	11	19	12	0.43		7	10	3	1.60		4	9	9	3.00	
817 $	8	15		2.13		3	14		7.12	***	5	1		2.67	
46C2 $	27	16		2.81	*	14	7		2.33		13	9		0.73	

significance levels: *:0.1 **:0.05 ***:0.01 ****:0.005 *****:0.001 ******:0.0005 *******:0.0001

$ those markers were not implemented in the genetic maps

Appendix 9.1: All markers tested for conformity to Mendelian ratio (Hardy-Weinberg equilibrium) with a χ2 goodness-of-fit analysis using the JoinMap single locus analysis module (JMSLA) procedure.

	marker	χ^2	p-value	Sig	df
family 15	803	0.18	0.67		1
	197	0.50	0.48		2
	757	0.36	0.55		1
	2R_si_5	2.55	0.11		1
	cyp6p3	0.27	0.60		1
	2La	0.05	0.82		
	637	4.07	0.04	**	1
	675	5.32	0.02	**	2
	776	3.27	0.07	*	1
	59	0.07	0.79		1
	812	1.94	0.16		2
	249	4.33	0.04	*	2
	30C1	5.18	0.02	**	1
	cplcg3	6.18	0.01	***	1
	6M19	5.67	0.02	**	1
	119	0.23	0.63		2
	555	0.05	0.82		2
	6H1	0.22	0.64		2
	577	0.51	0.48		2
family 17	803	2.51	0.11		2
	197	0.39	0.53		1
	757	3.59	0.06	*	1
	2R_si_5	0.33	0.57		2
	cyp6P3	0.16	0.69		1
	770	0.83	0.36		1
	117	0.08	0.77		1
	kdr	0.36	0.55		1
	2La	0.58	0.45		
	787	1.61	0.20		1
	603	3.42	0.06	*	1
	776	0.13	0.72		2
	59	1.61	0.21		1
	812	0.02	0.90		2
	cyp6m2	0.48	0.49		1
	249	0.54	0.46		2
	30C1	0.46	0.50		2
	cplcg3	0.81	0.37		1
	6M19	0.90	0.34		2
	341	0.03	0.86		2
	6H1	0.01	0.90		2
	33C1	0.29	0.59		1
	6F5	1.15	0.28		1
	30L17	0.00	0.95		1
	SODI	0.74	0.39		1
	811	0.00	0.95		1

	577	0.87	0.35		1
	817	8.76	0.00	**	2
	46C2	1.75	0.19		2
	53	0.02	0.88		?
	80	3.23	0.07		2
family 25	46	0.20	0.65		2
	757	0.29	0.59		1
	799	0.76	0.38		2
	2R_si_5	1.19	0.28		2
	cyp6P3	0.38	0.54		1
	770	0.79	0.37		2
	117	2.35	0.13		1
	2La	1.32	0.25		
	787	3.44	0.06	*	2
	772	0.21	0.65		2
	675	2.51	0.11		2
	776	0.61	0.43		1
	59	1.78	0.18		2
	812	0.01	0.94		1
	cyp6m2	0.20	0.66		2
	249	0.01	0.91		1
	30C1	0.43	0.51		1
	cplcg3	0.01	0.95		1
	6M19	0.73	0.39		2
	119	0.14	0.71		1
	158	0.19	0.67		1
	341	0.16	0.69		1
	6H1	0.19	0.66		2
	6F5	0.14	0.71		1
	88	1.21	0.27		2
	30L17	0.26	0.61		1
	811	3.83	0.05	*	2
	817	4.35	0.04	**	1
	46C2	0.20	0.65		1

significance levels: *:0.1 **:0.05 ***:0.01

$ those markers were not implemented in the genetic maps

Appendix 9.2: A list of $\chi^2$and p-values for genotype-phenotype association in for all markers *An. gambiae* (haploview 4.2).

degrees of freedom	probability [p]							
	0.9	**0.5**	**0.3**	**0.2**	**0.1**	**0.05**	**0.01**	**0.001**
1	0.02	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.21	1.39	2.41	3.22	4.6	5.99	9.21	13.82

Appendix 9.3: Chi square distribution adapted from Fisher (Fisher& Yates 1949).

Functional enrichment	
Seine proteases (fold enrichment 4.6)	
AGAP008296	Trypsin-1
AGAP008295	Trypsin-2
AGAP008294	Trypsin-3
AGAP008292	Trypsin-4
AGAP008291	Trypsin-5
AGAP008290	Trypsin-6
AGAP008293	Trypsin-7
AGAP008183	serine proteinase stubble
cytochrome P450s (fold enrichment 6.2)	
AGAP008203	*cyp6s2*
AGAP008204	*cyp6s1*
AGAP008205	*cyp6r1*
AGAP008206	*cyp6n2*
AGAP008207	*cyp6y2*
AGAP008208	*cyp6y1*
AGAP008209	*cyp6m1*
AGAP008210	*cyp6n1*
AGAP008212	*cyp6m2*
AGAP008213	*cyp6m3*
AGAP008214	*cyp6m4*
AGAP008217	*cyp6z3*
AGAP008218	*cyp6z2*
AGAP008219	*cyp6z1*
AGAP008552	*cyp4h27*
AGAP008553	cytochrome p450
AGAP008356	*cyp4h16*
AGAP008358	*cyp4h17*

Appendix 9.4: List of serine proteases and cytochrome P450s enriched within the boundaries of the QTL identified on chromosome 3R in *An. gambiae*.

				all					alive					dead		
fam 2	**aa**	**ab**	**bb**	**X^2**		**aa**	**ab**	**bb**	**X^2**		**aa**	**ab**	**bb**	**X^2**	
803	24	11	na	4.8	**	12	4	na	4.0	**	12	7	na	1.32	
199	29	19	na	2.1		18	6	na	6.0	**	11	13	na	0.17	
24	29	19	na	2.1		18	6	na	6.0	**	11	13	na	0.17	
757	19	19	na	0.0		6	12	na	2.0		13	7	na	1.8	
799	21	20	na	0.0		9	12	na	0.4		12	8	na	0.8	
2R_si_5	27	21	na	0.8		15	9	na	1.5		12	12	na	0	
cyp6p3	18	30	na	3.0	*	10	14	na	0.7		8	16	na	2.67	
590	22	24	na	0.1		6	16	na	4.6	**	16	8	na	2.67	
cyp304c1	23	25	na	0.1		8	16	na	2.7		15	9	na	1.5	
720	20	28	na	1.3		11	13	na	0.2		9	15	na	1.5	
770 [$]	22	13	na	2.3		12	6	na	2.0		10	7	na	0.53	
786	19	28	na	1.7		9	15	na	1.5		10	13	na	0.39	
ex27	19	29	na	2.1		11	13	na	0.2		8	16	na	2.67	
787	7	21	12	1.4		3	12	6	1.3		4	9	6	0.47	
603 [$]	9	28	4	6.7	**	7	11	3	1.6		2	17	1	9.9	***
675	23	18	na	0.6		12	9	na	0.4		11	9	na	0.2	
776	5	29	11	5.4	*	2	14	6	3.1		3	15	5	2.48	
59	14	18	3	6.9	**	7	8	2	3.0		7	10	1	4.22	
812	28	15	3	32.7	*******	13	10	1	12.7	****	15	5	2	21.9	*******
249	19	21	2	13.8	****	9	11	1	6.1	**	10	10	1	7.76	**
30C1	1	32	12	13.4	****	0	19	5	10.3	***	1	13	7	4.62	*
6M19	4	27	10	5.9	*	0	15	4	8.1	**	4	12	6	0.55	
119	2	33	10	12.6	****	0	19	5	10.3	***	2	14	5	3.19	
811	1	28	8	12.4	****	0	15	3	9.0	**	1	13	5	4.26	
577	18	26	na	1.5		7	14	na	2.3		11	12	na	0.04	
46C3	3	20	8	4.2		2	12	3	3.0		1	8	5	2.57	
46C2	1	24	12	9.8	***	0	13	5	6.3	**	1	11	7	4.26	
cyp4h24	14	30	4	7.2	**	6	15	3	2.3		8	15	1	5.58	*

53	12	27	5	4.5	5	14	3	2.0	7	13	2	3	
80	6	20	6	2.0	4	15	2	4.2	2	5	4	0.82	
H7	24	24	na	0.0	15	9	na	1.5	9	15	na	1.5	
678	19	27	na	1.4	9	15	na	1.5	10	12	na	0.18	

Appendix 9.5: All family 2 markers tested for conformity to Mendelian ratio (Hardy-Weinberg equilibrium) with a χ2 goodness-of-fit analysis using the JoinMap single locus analysis module (JMSLA) procedure.

				all					alive					dead		
fam 3	**aa**	**ab**	**bb**	**X^2**		**aa**	**ab**	**bb**	**X^2**		**aa**	**ab**	**bb**	**X^2**	
803	7	23	9	1.46		1	10	9	6.40	**	6	13	0	6.37	**
199	7	29	9	3.93		0	13	8	7.29	**	7	16	1	5.67	*
24	7	29	11	3.26		0	14	9	8.13	**	7	15	2	3.58	
757	9	20	11	0.20		1	10	8	5.21	*	8	10	3	2.43	
799 $	11	24	na	4.83	**	7	10	na	0.53		4	14	na	5.56	**
26 $	23	23	na	0.00		5	17	na	6.55	**	18	6	na	6.00	**
cyp6p3 $	25	21	na	0.35		16	6	na	4.55	**	9	15	na	1.50	
590	8	27	13	1.79		1	12	11	8.33	**	7	15	2	3.58	
cyp304c1 $	27	18	na	1.80		14	8	na	1.64		13	10	na	0.39	
117	27	17	na	2.27		9	12	na	0.43		18	5	na	7.35	***
786	8	24	8	1.60		4	12	4	0.80		4	12	4	0.80	
787	2	31	8	12.5	****	1	15	5	5.38	*	1	16	3	7.60	**

772	31	16	na	4.79	**	12	12	na	0.00		19	4	na	9.78	****
603	22	15	na	1.32		8	9	na	0.06		14	6	na	3.20	*
675	9	23	7	1.46		9	11	4	0.30		4	12	3	1.42	
776	16	24	4	6.91	**	8	12	2	3.45		8	12	2	3.45	
59	20	19	7	8.74	**	9	11	3	3.17		11	8	4	6.39	**
812	10	27	10	1.04		6	13	4	0.74		4	14	6	1.00	
249	6	22	5	3.73		3	12	4	1.42		3	10	1	3.14	
30C1	15	26	na	2.95	*	8	13	na	1.19		7	13	na	1.80	
6M19	16	20	4	7.20	**	6	12	3	1.29		10	8	1	9.00	**
119	12	25	3	6.55	**	7	12	2	2.81		5	13	1	4.26	
Chymotrypsin	41	7	na	24.1	*******	20	4	na	10.7	****	21	3	na	13.5	******
811	13	30	na	6.72	***	8	15	na	2.13		5	15	na	5.00	**
46C3	12	23	3	5.95	*	6	13	2	2.71		6	10	1	3.47	
46C2	13	18	4	4.66	*	4	10	2	1.50		9	8	2	5.63	*
cyp4h24	17	31	na	4.08	**	6	18	na	6.00	**	11	13	na	0.17	
53	23	18	na	0.61		12	10	na	0.18		11	8	na	0.47	
80	3	12	11	5.08	*	1	8	3	2.00		2	4	8	7.71	**
H7	20	28	na	1.33		6	18	na	6.00	**	14	10	na	0.67	
678	19	28	na	1.72		7	16	na	3.52	*	12	12	na	0.00	

significance levels: *:0.1 **:0.05 ***:0.01 ****:0.005 *****:0.001 ******:0.0005 *******:0.0001

$ those markers were not implemented in the genetic maps

Appendix 9.6: All family 3 markers tested for conformity to Mendelian ratio (Hardy-Weinberg equilibrium) with a χ2 goodness-of-fit analysis using the JoinMap single locus analysis module (JMSLA) procedure.

Functional enrichment of genes within the boundaries of the *An. arabiensis* QTL on 2R	
25 cytochrome P450s (fold enrichment 4.2)	
AGAP002138	*cyp325h1*
AGAP002195	*cyp325f2*
AGAP002196	*cyp325g1*
AGAP002197	*cyp325f1*
AGAP002202	*cyp325e1*
AGAP002203	*cyp325d2*
AGAP002205	*cyp325c2*
AGAP002207	*cyp325c1*
AGAP002209	*cyp325a2*
AGAP002210	*cyp325b1*
AGAP002211	*cyp325a1*
AGAP002416	*cyp4k2*
AGAP002417	*cyp4ar1*
AGAP002418	*cyp4d16*
AGAP002419	*cyp4d22*
AGAP002429	*cyp315a1*
AGAP002555	*cyp325k1*
AGAP002862	*cyp6aa1*
AGAP002864	*cyp6aa2*
AGAP002865	*cyp6p3*
AGAP002866	*cyp6p5*
AGAP002867	*cyp6p4*
AGAP002868	*cyp6p1*
AGAP002869	*cyp6p2*
AGAP002870	*cyp6ad1*

Appendix 9.7: List of cytochrome P450s enriched within the boundaries of the QTL identified on chromosome 2R in *An. arabiensis*.

Description	Accession number	Probe name	FDR p	FC
5 nucleotidase	AGAP007140-RA	CUST_2671_PI422575199	9.03E-04	4.65
protease m1 zinc metalloprotease	AGAP004860-RA	CUST_214_PI422575199	1.74E-03	2.93
chymotrypsin-like protein	AGAP001198-RA	CUST_3526_PI422575199	3.12E-03	6.33
methionine-trna synthetase	AGAP007891-RA	CUST_11187_PI422575199	3.12E-03	10.66
isoform a	AGAP009224-RA	CUST_12451_PI422575199	3.19E-03	3.89
formyl-coenzyme a transferase	AGAP010040-RA	CUST_13266_PI422575199	3.59E-03	2.27
tyrosine aminotransferase	AGAP000327-RA	CUST_8699_PI422575199	3.61E-03	2.58
AGAP001226-PA [Anopheles gambiae str. PEST]	AGAP001226-RA	CUST_3743_PI422575199	3.61E-03	2.75
protease m1 zinc metalloprotease	AGAP001881-RA	CUST_3452_PI422575199	3.61E-03	2.81
26s proteasome non-atpase regulatory subunit 8	AGAP002606-RA	CUST_4439_PI422575199	3.61E-03	3.43
secretory carrier-associated membrane protein	AGAP002922-RA	CUST_4928_PI422575199	3.61E-03	3.28
26s protease regulatory subunit 4	AGAP003216-RA	CUST_5314_PI422575199	3.61E-03	5.49
prefoldin subunit 5	AGAP003416-RA	CUST_5573_PI422575199	3.61E-03	2.09
electron transfer flavoprotein subunit mitochondrial	AGAP004031-RA	CUST_6389_PI422575199	3.61E-03	18.63
protease m1 zinc metalloprotease	AGAP004809-RA	CUST_157_PI422575199	3.61E-03	2.40
rna exonuclease 4	AGAP005050-RA	CUST_423_PI422575199	3.61E-03	5.97
bax inhibitor	AGAP005775-RA	CUST_1242_PI422575199	3.61E-03	2.65
atp-binding cassette sub-family a member	AGAP006380-RA	CUST_1856_PI422575199	3.61E-03	6.09
ichit	AGAP006432-RA	CUST_1913_PI422575199	3.61E-03	9.46
cg30152-like protein	AGAP007363-RA	CUST_2916_PI422575199	3.61E-03	4.40
signal recognition particle 14 kda protein	AGAP008339-RA	CUST_11616_PI422575199	3.61E-03	2.38
isoform b	AGAP009156-RA	CUST_12385_PI422575199	3.61E-03	4.96
microfibril-associated glycoprotein 4	AGAP009556-RA	CUST_12776_PI422575199	3.61E-03	8.40
ctp synthase	AGAP009624-RA	CUST_12846_PI422575199	3.61E-03	3.95
larval cuticle protein lcp-30	AGAP009874-RA	CUST_13099_PI422575199	3.61E-03	2.28
zinc metalloproteinase nas-12	AGAP010764-RA	CUST_9407_PI422575199	3.61E-03	3.55

AGAP011627-PA [Anopheles gambiae str. PEST]	AGAP011627-RA	CUST_10210_PI422575199	3.61E-03	5.43
ornithine decarboxylase	AGAP011808-RA	CUST_10385_PI422575199	3.61E-03	3.29
protein takeout	AGAP012703-RA	CUST_13822_PI422575199	3.61E-03	3.88
AGAP013115-PA [Anopheles gambiae str. PEST]	AGAP013115-RA	CUST_4843_PI422575199	3.61E-03	2.23
glucose dehydrogenase	AGAP003785-RA	CUST_6118_PI422575199	3.68E-03	4.80
ribose-phosphate pyrophosphokinase	AGAP004890-RA	CUST_248_PI422575199	3.68E-03	5.30
serine protease	AGAP004900-RA	CUST_259_PI422575199	3.68E-03	6.70
kazal domain-containing peptide	AGAP007906-RA	CUST_11202_PI422575199	3.68E-03	4.18
notch 2	AGAP008065-RA	CUST_11358_PI422575199	3.68E-03	6.12
dipeptidyl peptidase 4	AGAP008176-RA	CUST_11463_PI422575199	3.68E-03	4.07
inactive dipeptidyl	AGAP008764-RA	CUST_12005_PI422575199	3.68E-03	4.34
GSTE5 - Glutathion S-transferase	AGAP009192-RA	DETOX_625_PI422610884	3.68E-03	2.20
organic cation transporter	AGAP012383-RA	CUST_10939_PI422575199	3.68E-03	2.09
serine protease	AGAP012946-RA	CUST_5873_PI422575199	3.68E-03	2.12
AGAP013147-PA [Anopheles gambiae str. PEST]	AGAP013147-RA	CUST_7999_PI422575199	3.68E-03	3.91
trehalose-6-phosphate synthase 1	AGAP008227-RA	CUST_11510_PI422575199	3.77E-03	2.75
trypsin	AGAP008861-RA	CUST_12093_PI422575199	3.78E-03	31.89
glucose dehydrogenase	AGAP003785-RC	CUST_6121_PI422575199	3.95E-03	4.07
anopheles gambiae pest agap012679-pa	AGAP009909-RA	CUST_13132_PI422575199	3.97E-03	2.37
agcp14332	AGAP001616-RA	CUST_7438_PI422575199	3.97E-03	2.11
kynurenine formamidase	AGAP009433-RA	CUST_12655_PI422575199	3.97E-03	4.90
cuticular protein analogous to peritrophins 1-i	AGAP002052-RA	CUST_3665_PI422575199	3.98E-03	4.36
apolipoprotein d	AGAP002594-RA	CUST_4421_PI422575199	3.98E-03	2.50
ctlma3 protein	AGAP007412-RA	CUST_2497_PI422575199	4.26E-03	2.61

Appendix 9.8: Microarray top-table showing top 50 (ranked by false discovery rate adjusted p-value) up-regulated genes. FDR=adjusted p-value; FC=fold change.

Description	Accession number	ProbeName	FDR p	FC
maltase 1	AGAP008962-RA	CUST_12197_PI422575199	3.61E-03	2.72
folylpolyglutamate mitochondrial	AGAP004679-RA	CUST_5_PI422575199	3.61E-03	2.10
viral a-type inclusion protein repeat containing protein	AGAP013358-RA	CUST_5926_PI422575199	3.61E-03	2.02
AGAP002771-RA	AGAP002771-RA	CUST_4690_PI422575199	3.61E-03	3.31
facilitated trehalose transporter tret1-like	AGAP001236-RA	CUST_3918_PI422575199	3.61E-03	2.37
AGAP009535-PA [Anopheles gambiae str. PEST]	AGAP009535-RA	CUST_12755_PI422575199	3.61E-03	2.88
AGAP009817-PA [Anopheles gambiae str. PEST]	AGAP009817-RA	CUST_13047_PI422575199	3.61E-03	2.02
sterol desaturase	AGAP002769-RA	CUST_4692_PI422575199	3.61E-03	2.51
endomembrane protein emp70	AGAP004882-RA	CUST_239_PI422575199	3.61E-03	2.01
hypothetical conserved protein	AGAP004410-RA	CUST_7021_PI422575199	3.61E-03	4.37
AGAP011104-PA [Anopheles gambiae str. PEST]	AGAP011104-RA	CUST_9719_PI422575199	3.61E-03	2.21
sodium solute symporter	AGAP008359-RA	CUST_11634_PI422575199	3.61E-03	2.60
isoform b	AGAP005005-RA	CUST_371_PI422575199	3.68E-03	2.18
cklf-like marvel transmembrane domain-containing protein 4	AGAP009132-RA	CUST_12363_PI422575199	3.68E-03	6.71
cytochrome p450	AGAP009374-RA	CUST_12597_PI422575199	3.68E-03	3.84
adenosine deaminase	AGAP006906-RA	CUST_2427_PI422575199	3.68E-03	2.79
anopheles gambiae pest agap012543-pa	AGAP012543-RA	CUST_13685_PI422575199	3.68E-03	3.62
folylpolyglutamate mitochondrial	AGAP004679-RB	CUST_4_PI422575199	3.68E-03	2.21
pupal cuticle	AGAP010122-RA	CUST_13347_PI422575199	3.68E-03	2.39
sterol desaturase	AGAP002769-RB	CUST_4691_PI422575199	3.68E-03	2.61
cytochrome p450	AGAP002204-RA	CUST_3915_PI422575199	3.71E-03	4.91
elongation of very long chain fatty acids protein aael008004-like	AGAP008780-RA	CUST_12021_PI422575199	3.77E-03	3.16
glutathione transferase gst1-6	AGAP004380-RA	CUST_6991_PI422575199	3.77E-03	2.23
gaba-a receptor interacting factor-	AGAP008801-RA	CUST_12034_PI422575199	3.77E-03	3.60
zinc finger protein 43	AGAP011256-RA	CUST_9866_PI422575199	3.77E-03	3.00

isoform a	AGAP012127-RA	CUST_10686_PI422575199	3.77E-03	3.42
cytochrome p450	AGAP003066-RA	CUST_5130_PI422575199	3.97E-03	2.67
phosphatidylinositol-binding clathrin assembly protein	AGAP010087-RA	CUST_13313_PI422575199	3.97E-03	2.63
venom allergen	AGAP000356-RA	CUST_8716_PI422575199	3.97E-03	5.79
isoform b	AGAP004179-RC	CUST_6654_PI422575199	4.17E-03	2.36
matrix metalloproteinase	AGAP006904-RA	CUST_2423_PI422575199	4.26E-03	5.20
matrix metalloproteinase	AGAP006904-RC	CUST_2425_PI422575199	4.26E-03	5.16
short-chain dehydrogenase	AGAP011852-RA	CUST_10429_PI422575199	4.26E-03	4.21
isoform b	AGAP011702-RA	CUST_10284_PI422575199	4.39E-03	5.07
matrix metalloproteinase	AGAP006904-RB	CUST_2424_PI422575199	4.42E-03	5.17
AGAP005614-RA	AGAP005614-RA	CUST_1068_PI422575199	4.52E-03	6.51
uncharacterized protein kiaa1797	AGAP005343-RA	CUST_775_PI422575199	4.52E-03	3.26
3-alpha-(or 20-beta)-hydroxysteroid dehydrogenase	AGAP005645-RA	CUST_1104_PI422575199	4.52E-03	4.24
isoform a	AGAP004553-RA	CUST_7147_PI422575199	4.52E-03	2.07
odorant binding protein (agap012321-pa)	AGAP012321-RA	CUST_10876_PI422575199	4.74E-03	3.56
lethal malignant brain tumor	AGAP003268-RA	CUST_5384_PI422575199	4.75E-03	2.69
cytochrome p450	AGAP002209-RA	CUST_3897_PI422575199	4.75E-03	2.15
high-affinity cgmp-specific -cyclic phosphodiesterase	AGAP002927-RB	CUST_7632_PI422575199	4.91E-03	3.08
AGAP005610-PA [Anopheles gambiae str. PEST]	AGAP005610-RA	CUST_1063_PI422575199	4.97E-03	3.47
membrane glycoprotein lig-1	AGAP007037-RA	CUST_2563_PI422575199	5.10E-03	4.67
itg-containing peptide	AGAP008993-RA	CUST_12228_PI422575199	5.14E-03	2.61
chromatin assembly factor p180-	AGAP001568-RA	CUST_7398_PI422575199	5.21E-03	6.44
isoform b	AGAP007503-RB	CUST_3049_PI422575199	5.21E-03	3.40
atp-dependent rna helicase vasa	AGAP008578-RA	CUST_11837_PI422575199	5.21E-03	7.08
CYP6P4 - Cytochrome P450 monooxygenase	AGAP002867-RA	DETOX_463_PI422610884	5.21E-03	5.12

Appendix 9.9: Microarray top-table showing top 50 (ranked by false discovery rate adjusted p-value) down-regulated genes. FDR=adjusted p-value; FC=fold change.

Annotation Cluster 1	**Enrichment Score: 3.4**						
Category	**Term**	**Count**	**%**	**PValue**	**Fold Enrichment**	**Benjamini**	**FDR**
GOTERM_MF_FAT	GO:0070011~peptidase activity, acting on L-amino acid peptides	23	17.8	1.1E-07	3.50	1.34E-05	1.28E-04
GOTERM_MF_FAT	GO:0008233~peptidase activity	23	17.8	2.7E-07	3.33	1.63E-05	3.09E-04
GOTERM_BP_FAT	GO:0006508~proteolysis	23	17.8	1.2E-06	2.98	1.67E-04	0.001
GOTERM_MF_FAT	GO:0008236~serine-type peptidase activity	14	10.9	9.1E-05	3.55	3.65E-03	0.104
GOTERM_MF_FAT	GO:0017171~serine hydrolase activity	14	10.9	9.1E-05	3.55	3.65E-03	0.104
SMART	SM00020:Tryp_SPc	10	7.8	3.7E-04	4.12	1.01E-02	0.310
INTERPRO	IPR001314:Peptidase S1A, chymotrypsin	10	7.8	4.8E-04	4.28	4.28E-02	0.594
INTERPRO	IPR018114:Peptidase S1/S6, chymotrypsin/Hap, active site	9	7.0	8.4E-04	4.46	4.92E-02	1.026
INTERPRO	IPR001254:Peptidase S1 and S6, chymotrypsin/Hap	10	7.8	2.5E-03	3.39	1.06E-01	3.019
GOTERM_MF_FAT	GO:0004252~serine-type endopeptidase activity	10	7.8	8.9E-03	2.73	1.44E-01	9.803
SP_PIR_KEYWORDS	Protease	8	6.2	1.1E-02	3.24	3.85E-01	10.00
GOTERM_MF_FAT	GO:0004175~endopeptidase activity	11	8.5	2.9E-02	2.12	3.00E-01	28.73
SP_PIR_KEYWORDS	Serine protease	4	3.1	1.4E-01	3.05	7.30E-01	75.74
SP_PIR_KEYWORDS	hydrolase	6	4.7	4.2E-01	1.39	9.29E-01	99.42
Annotation Cluster 2	**Enrichment Score: 2.5**						
Category	**Term**	**Count**	**%**	**PValue**	**Fold Enrichment**	**Benjamini**	**FDR**
INTERPRO	IPR003146:Proteinase inhibitor, carboxypeptidase propeptide	4	3.1	7.4E-05	44.09	1.32E-02	0.09
GOTERM_MF_FAT	GO:0004180~carboxypeptidase activity	5	3.9	2.0E-04	16.28	6.02E-03	0.23
GOTERM_MF_FAT	GO:0008237~metallopeptidase activity	8	6.2	4.5E-04	5.60	1.09E-02	0.52
GOTERM_MF_FAT	GO:0008238~exopeptidase activity	5	3.9	3.7E-03	7.64	7.20E-02	4.17
SMART	SM00631:Zn_pept	3	2.3	1.1E-02	17.85	1.41E-01	8.90

INTERPRO	IPR000834:Peptidase M14, carboxypeptidase A	3	2.3	1.7E-02	14.70	4.03E-01	19.02
GOTERM_MF_FAT	GO:0004181~metallocarboxypeptidase activity	3	2.3	2.0E-02	13.22	2.68E-01	21.08
GOTERM_MF_FAT	GO:0008235~metalloexopeptidase activity	3	2.3	5.5E-02	7.75	4.63E-01	47.84
Annotation Cluster 3	**Enrichment Score: 0.9**						
Category	**Term**	**Count**	**%**	**PValue**	**Fold Enrichment**	**Benjamini**	**FDR**
GOTERM_MF_FAT	GO:0009055~electron carrier activity	7	5.4	2.4E-02	3.07	2.78E-01	24.32
SP_PIR_KEYWORDS	metal-binding	10	7.8	5.0E-02	2.07	6.66E-01	37.79
INTERPRO	IPR017972:Cytochrome P450, conserved site	4	3.1	5.7E-02	4.52	7.81E-01	51.51
INTERPRO	IPR017973:Cytochrome P450, C-terminal region	4	3.1	5.9E-02	4.46	7.46E-01	52.63
INTERPRO	IPR001128:Cytochrome P450	4	3.1	7.8E-02	3.96	8.05E-01	63.29
SP_PIR_KEYWORDS	heme	4	3.1	8.5E-02	3.83	7.19E-01	56.14
SP_PIR_KEYWORDS	iron	4	3.1	1.5E-01	2.97	6.85E-01	77.71
SP_PIR_KEYWORDS	oxidoreductase	5	3.9	1.5E-01	2.40	6.39E-01	78.60
COG_ONTOLOGY	Secondary metabolites biosynthesis, transport, and catabolism	4	3.1	1.7E-01	2.62	5.64E-01	65.41
GOTERM_BP_FAT	GO:0055114~oxidation reduction	7	5.4	1.8E-01	1.81	1.00E+00	89.73
INTERPRO	IPR002401:Cytochrome P450, E-class, group I	3	2.3	1.9E-01	3.67	9.80E-01	92.93
SP_PIR_KEYWORDS	Monooxygenase	3	2.3	2.0E-01	3.61	6.96E-01	87.28
GOTERM_MF_FAT	GO:0020037~heme binding	4	3.1	2.4E-01	2.32	9.07E-01	95.74
GOTERM_MF_FAT	GO:0046906~tetrapyrrole binding	4	3.1	2.4E-01	2.32	9.07E-01	95.74
GOTERM_MF_FAT	GO:0005506~iron ion binding	5	3.9	2.5E-01	1.93	9.02E-01	96.37

Appendix 9.10: Results list from DAVID functional clustering analysis for the 144 significantly up-regulated entities.

Annotation Cluster 1	**Enrichment Score: 2.9**						
Category	**Term**	**Count**	**%**	**PValue**	**Fold Enrichment**	**Benjamini**	**FDR**
GOTERM_MF_FAT	GO:0020037~heme binding	9	7.89	1.78E-05	7.44	1.97E-03	2.01E-02
GOTERM_MF_FAT	GO:0046906~tetrapyrrole binding	9	7.89	1.78E-05	7.44	1.97E-03	2.01E-02
INTERPRO	IPR001128:Cytochrome P450	8	7.02	2.51E-05	9.04	4.33E-03	3.07E-02
GOTERM_MF_FAT	GO:0005506~iron ion binding	10	8.77	5.60E-05	5.43	3.11E-03	6.34E-02
INTERPRO	IPR002401:Cytochrome P450, E-class, group I	7	6.14	6.47E-05	9.94	5.58E-03	7.91E-02
SP_PIR_KEYWORDS	Monooxygenase	6	5.26	5.04E-04	9.02	2.05E-02	4.63E-01
INTERPRO	IPR017972:Cytochrome P450, conserved site	6	5.26	8.90E-04	7.91	5.00E-02	1.08
GOTERM_MF_FAT	GO:0009055~electron carrier activity	8	7.02	9.00E-04	4.92	3.28E-02	1.01
INTERPRO	IPR017973:Cytochrome P450, C-terminal region	6	5.26	9.39E-04	7.82	3.98E-02	1.14
COG_ONTOLOGY	Secondary metabolites biosynthesis, transport, and catabolism	8	7.02	1.05E-03	4.32	0.01	0.60
SP_PIR_KEYWORDS	heme	6	5.26	1.45E-03	7.14	0.03	1.33
SP_PIR_KEYWORDS	oxidoreductase	8	7.02	1.45E-03	4.71	0.02	1.33
GOTERM_BP_FAT	GO:0055114~oxidation reduction	9	7.89	3.75E-03	3.33	0.61	4.77
SP_PIR_KEYWORDS	iron	6	5.26	4.49E-03	5.51	0.05	4.06
SP_PIR_KEYWORDS	metal-binding	9	7.89	5.22E-02	2.18	0.36	38.92
GOTERM_MF_FAT	GO:0043169~cation binding	19	16.67	5.34E-02	1.48	0.78	46.25
GOTERM_MF_FAT	GO:0043167~ion binding	19	16.67	5.34E-02	1.48	0.78	46.25
GOTERM_MF_FAT	GO:0046914~transition metal ion binding	15	13.16	9.07E-02	1.50	0.78	65.91
GOTERM_MF_FAT	GO:0046872~metal ion binding	17	14.91	1.22E-01	1.38	0.84	77.10

Appendix 9.11: Results list from DAVID functional clustering analysis for the 137 significantly down-regulated entities.