Molecular and evolutionary genetics of *Anopheles gambiae s.l*, a malaria vector in Africa

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

By

Samuel Kweku Dadzie

APRIL 2008

DECLARATION

THIS WORK HAS NOT PREVIOUSLY BEEN ACCEPTED IN SUBSTANCE FOR ANY DEGREE AND IS NOT BEING CURRENTLY SUBMITTED IN CANDIDATURE FOR ANY DEGREE

Signed

…… ……..(Candidate)

Date ……29/07/08……

STATEMENT 1

THIS THESIS IS THE RESULT OF MY OWN INVESTIGATION, EXCEPT WHERE OTHERWISE STATED. OTHER SOURCES ARE ACKNOWLEDGED AND BIBIOGRAPHY IS APPENDED

Signed…… ……(Candidate)

Date……29/07/08……

STATEMENT 2

I HEREBY GIVE MY CONSENT FOR MY THESIS, IF ACCEPTED, TO BE AVAILABLE FOR PHOTOCOPYING AND FOR INTER-LIBRARY LOAN, AND FOR THE TITLE AND SUMMARY TO BE MADE AVAILABLE TO OUTSIDE ORGANIZATIONS

Signed …… ……(Candidate)

Date ……29/07/08……

ACKNOWLEGEMENTS

First and foremost, I would like to extend my profound gratitude to my primary supervisor, Dr Martin James Donnelly for his ideas, vision and direction in guiding me to the successful completion of my PhD programme. I also thank Professor Harold Townson, my second supervisor for his guidance and encouragement. I also wish to thank Dr. Themba Mzilahowa, Derick Charlwood, Dr João Pinto and Dr Kwang Shik Choi for providing samples for some of the localities studied. Many thanks to all my colleagues at Liverpool School of Tropical Medicine for valuable advice and assistance, especially Mr Ken Sherlock for laboratory assistance and Amy Lynd for providing me with mosquito samples for the melanism study. I am also indebted to Mrs Gwen Finnegan for her assistance and support throughout my stay at the school and for helping me put the thesis in good shape.

Many thanks also to the Director of Noguchi Institute for Medical Research, Ghana, Professor David Ofori-Adjei; Head of Parasitology Unit, Professor Michael Wilson and Mr Maxwell Appawu for giving me the opportunity to pursue this PhD programme and to all my work mates at the institute, I say a big thank you for your moral support.

I wish to acknowledge the contribution of my family, my wife Rose and children Kojo, Adorkoh and Kwabena to the successful completion of my PhD programme.
And never enough thanks to the Ghana Government for providing me with funding under the Ghana Educational Trust Fund and the Liverpool School of Tropical Medicine for some financial assistance during the course of my programme.

Finally I will give thanks to the almighty God who gave me strength and sustained me through very difficult times of my PhD programme.

Molecular and evolutionary genetics of *Anopheles gambiae s.l.*, a malaria vector in Africa

Samuel Kweku Dadzie

GENERAL ABSTRACT

One of the most important areas of research regarding adaptation of malaria vectors that has received very little attention is the effect of selection on the population structure of *An. arabiensis* and the ecological significance of melanisation on the ability of *An. gambiae* to survive in different habitats. Most microsatellite-based studies of population structure have selected loci that are located within inversions many of which may be subject to selection. The structure of the population will therefore depend on the nature of the selection on the selected loci and the effect of other nearby loci. Locus 33C1 within chromosome 3Ra inversion of *An. gambiae* showed abberation in the genetic differentiation of *An. arabiensis* population sampled from the north-south transect of East Africa (Donnelly and Townson 2000).

The aim of the first part of this thesis which constituted Chapter 3 is to determine if there is any signature of positive selection on some candidate loci located within the 3Ra inversion and investigate its effect on nucleotide variation of *An. arabiensis* populations. In line with this objective, we obtained 20 sequences each of 4 Epsilon class *Glutathione S-transferases* (GSTs) i.e. (GSTε1, 441bp (n=20), GSTε2, 800bp (n=20), GSTε6, 735bp (n=20) and GSTε8, 796bp (n=20) from samples of *An. arabiensis* captured from Sudan, Ethiopia, Malawi and Tanzania. Glutathione s-transferases (GSTs) are detoxication enzymes that are involved in the metabolism, detoxication and excretion of a large number of endogenous and exogenous compounds including DDT from the cells of organisms. Using evolutionary models, we showed that 3 out of the 4 loci deviated from neutral expectations. This is indicative of selective constraints on these loci. Pairwise estimates of population differentiation, F_{ST} values were far in excess of those observed from microsatellite-based studies of the same samples with specimen from Ethiopia much more genetically differentiated from the rest (*e.g.* range of F_{ST} values for pairwise comparisons involving Ethiopia 0.571 to 0.800 other comparisons - 0.25 to 0. Inference from phylogenetic analysis indicates that gene duplication events in GST gene family occurred prior to *An. gambiae s.l., An. arabiensis* and *An. stephensi* split.

In Chapter 4, we analysed 15 sequences of a nearby locus, *Dopa decarboxylase* (DDC) from *An. arabiensis* captured in Sudan. The DNA sequence was highly conserved at this locus and phylogenetic analyses of the sequences of *An. arabiensis* at the DDC locus produced a monophyletic genealogy. Fay and Wu *H* test on the coding region of the gene was significant at (H=0.094, $P<0.05$), indicative of the effect of selection. We hypothesize that the effect of selection on DDC in *An. arabiensis* and the occurrence of its orthologues in *Aedes* and *Drosophila* species suggest that DDC may be a single pleiotropic quantitative gene that is responsible for the generation of adaptive phenotypes.

The third part of my thesis reported in Chapter 5 is the first of a study to determine the interplay between phenotypic plasticity and some life-history traits of *An. gambiae.* To achieve this objective, we induced melanised and non-melanised phenotypes of *An. gambiae* larvae by rearing inbred larvae in different colour containers. A total of 800 larvae of laboratory colony of KISUMU strain and 400 *An. gambiae* collected from Ghana were used for the induction experiment. More than 70% (corrected for mortality) of larvae reared in the dark background developed darker pigmentation whereas those reared in the white containers became paler relative to the parental larval pigmentation, an indication of the occurrence of bi-directional response. Darker melanic larvae had faster developmental time (14 days) than the non-melanic forms (16 days). There were differences in the survivorship of the melanic and non-melanic forms of larvae and this was consistent with a trade-off hypothesis. The melanization rate measured as average grey values on a scale of 0 (black) - 225 (white) was 64.3 $\pm$ 7.5 for melanic and 106 $\pm$ 8.6 for the non-melanic forms and these were significantly different ($P<0.001$). The results of the study indicate a trade-off between melanin production and some life-history traits; darker individuals had faster development time but low survival rate. Our results suggest that the development of melanin in *An. gambaie* larvae was a phenotypic response modulated by the expression of DDC and phenooxidase (PO) genes.

The results of this PhD thesis provide baseline information that will be key to understanding the evolutionary forces that control the generation of adaptive phenotypes within populations of *An. gambiae s.l.* The study also shares some information on the role of *phenoloxidase* (PO) and *Dopa decaboxylase* genes in the phenotypic response of *An. gambiae* to different colour environments.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

An.	Anopheles
ITNs	Insecticide Treated Nets
M	Molar
Kb	Kilobasepair(s)
MgCl2	Magnesium Chloride
U	Unit
Mg	Magnesium
mL	Milliliter (s)
µl	Microliter (s)
DNA	Deoxyribonucleic acid
G	gram
^{o}C	Degrees centigrade
dNTP	Deoxyribonucleic triphosphate
DDT	Dichlorodiphenyltetrachloroethane
mM	Millimolar
PCR	Polymerase Chain reaction
rpm	Revolutions per minute
TE	Tris EDTA
Tris	2-amino-2(hydroxymethyl)-propane-1,3-diol
UV	Ultraviolet
WHO	World Health Organization
mtDNA	mitochondrial DNA
rDNA	ribosomal Deoxyribonucleic acid
KCl	Potassium chloride
HCl	Hydrochloric Acid
TAE	Tris, Acetic Acid and EDTA in solution

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

1.1 INTRODUCTION

Anopheles gambiae Giles *sensu stricto, Anopheles arabiensis* Patton *and Anopheles funestus* Giles are the most widespread (Figure 1) and important malaria vectors in Africa (White, 1974). Present anti-vector malaria strategies are based on reducing the contact between human host and vectors by the use of insecticides for indoor spraying, bednets, and environmental management. One of the many reasons why control of these vectors has proved difficult is that the species exhibit extreme versatility in their tolerance of a wide variety of micro and macro environmental conditions, as evidenced by their broad geographic distributions and the rapidity with which they evolve insecticide resistance (Coetzee *et al.,* 2000; David *et al.,* 2005). Current work is centred at the development of molecular-level techniques for the genetic control of malaria vectors (Rai, 1996; Collins *et al.,* 2000). However, malaria control strategies based on the genetic manipulation of vectors will require extensive knowledge on vector population genetics.

This thesis examined the rate of selection and local adaptation on two selectively important traits; melanism and insecticide detoxification in *Anopheles gambiae s.l.* Selection on these traits may result from diverse evolutionary pressures including varying physical environments, predator and insecticide evasion as well as sexual selection (Reznick and Ghalambor 2001). It has been demonstrated that genes that are involved in the generation of divergent adaptive traits could have family lineages and evolutionary histories that are different from those that are not (Wilder *et al.,* 2004). This first part of this thesis analysed the pattern of genetic variation at the *Glutathione s-transferases* (GSTs) and *Dopa decaboxylase* (DDC) genes. These genes respectively encode a number of alleles that are involved in the development of insecticide resistance and pigmentation patterns in insects (Prapanthadara *et al.,* 1995; True *et al.,* 1999; Gompel and Carroll 2003). Infact, one of the GST genes, GSTε2 and has been implicated in the detoxication of DDT in *An. gambiae* (Ranson *et al.,* 2001). The second part of the thesis further examined the effect of melanism on the fitness of melanised phenotypes of *An. gambiae* against non-melanised types. The importance of melanism in the evolution of *Drosophila* and other *Lepidoptera*

has been well studied (McMillan *et al.,* 2002; Wittkopp *et al.,* 2003b; Talloen *et al.,* 2004; Brisson *et al.,* 2005). However, the adaptive significance this trait in *An. gambiae* is not known.

The general focus of the present study was to use sequences from candidate genes, GSTs and DDC to investigate the evolutionary forces shaping *An. gambiae s.l.* species genetic variability. The study further investigated the evolutionary and molecular basics of the development of melanic phenotypes in populations of *An. gambiae*.

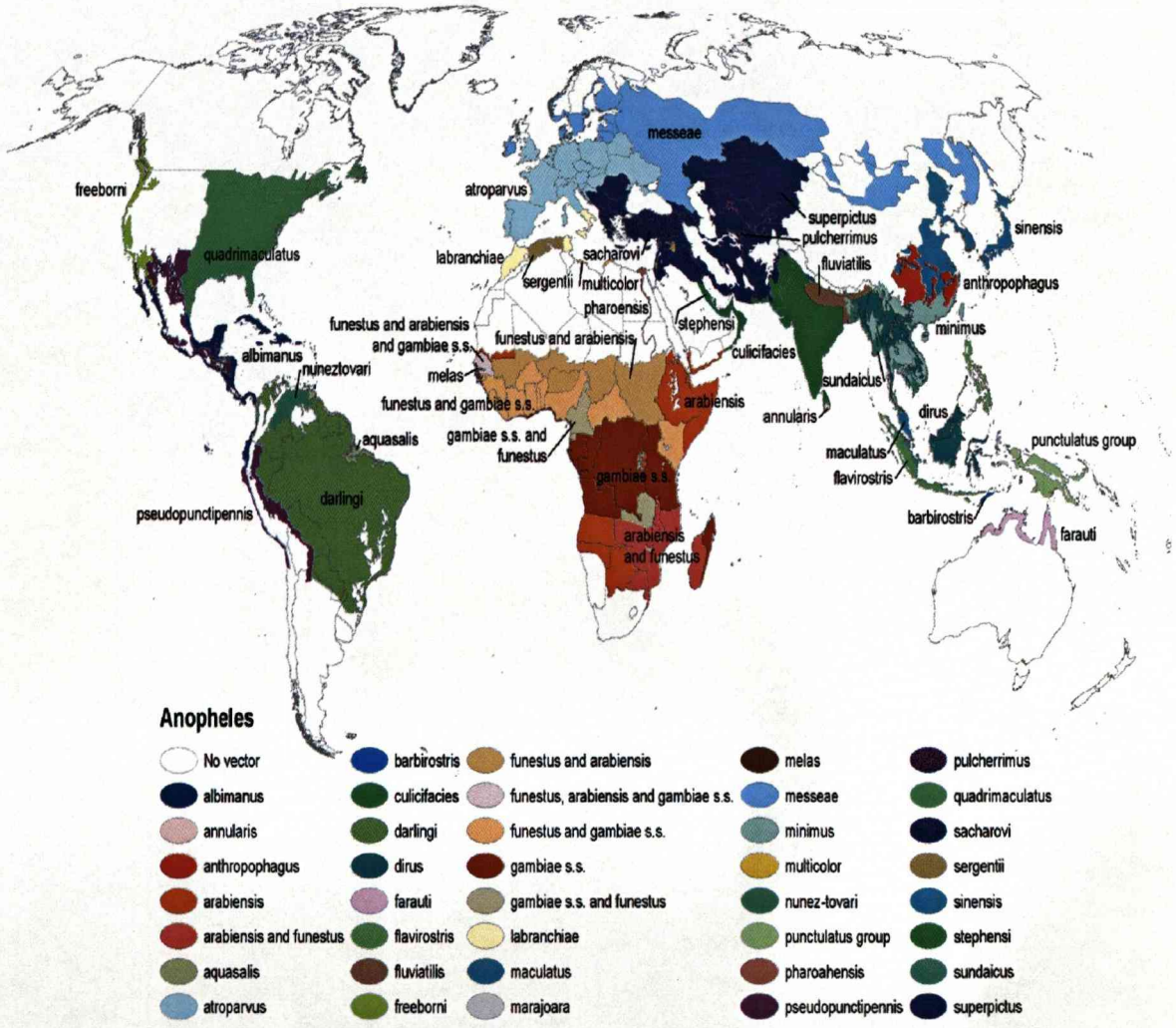


Figure 1: Global distribution (Robinson projection) of dominant or potentially important malaria vectors (Source: Kiszewksi *et al.*, 2004).

1.2 Rationale of study

This thesis focused on understanding how genetic variation is generated and maintained in natural populations of *Anopheles* mosquitoes. The specific interest is the role of natural selection in shaping adaptive genetic and phenotypic variations in populations of *Anopheles gambiae*. The long-term goals of this research are two-fold. First, I hope to provide insight into how natural selection affects DNA nucleotide sequence variation in *Anopheles* mosquitoes and and secondly to determine the rate of phenotypic plasticity on a trait such as the evolution of melanism.

In this study, we selected GSTs and DDC as candidate genes to help us understand more about their role in generating variation within populations of malaria vectors. GSTs are a major family of detoxification enzymes found in most organisms, which are involved in the metabolism, detoxification and excretion of a large number of endogenous and exogenous compounds from the cell. Insect GSTs belonging to the Epsilon class are of particular interest because of their potential to cause resistances to all the major families of insecticide especially organophosphates and DDT (Huang, 1998; Ranson *et al.,* 2001).

The *Dopa decarboxylase* gene (DDC) is involved in diverse physiological events in insects. The product of this enzyme's activity, dopamine, acts as an intermediate substrate for various tanning and melanization reactions, as well as a neurotransmitter (Eveleth and Marsh, 1986; Scholnick *et al.,* 1986). The DDC has been found to affect variation in longevity in *Drosophila* (De Luca *et al.,* 2003) and in mosquitoes the DDC is very important in the biochemical pathway of tanning and melanization reactions (Figure 2). Melanism is very common in several species of insects and melanic forms of Tiger swallowtails, *Papilio glaucus;* Two-spot ladybirds, *Adalia bipunctata; Drosophila elegans* and peppered moth, *Biston betularia* has been observed in nature (True, 2003). However, its adaptive value in natural populations has received very little attention. The present study was designed to answer some important questions regarding the adaptation of *Anopheles* vectors to colour changes in nature. Firstly, I investigated the nature of selection likely to be acting on the *Dopa decarboxylase* gene involved in melanisation pathway and

secondly, I provided some insight into the role of phenotypic plasticity in the evolution of melanism.

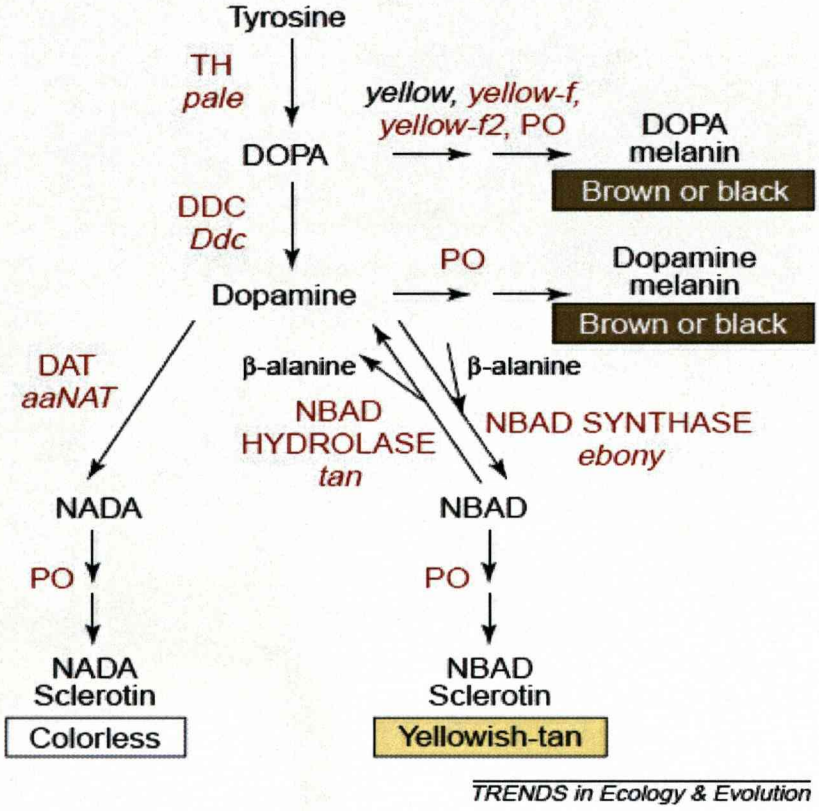


Figure 2: Candidate genes for melanism in insect melanin biosynthesis (Source: True, 2003).

Enzymes are indicated in red type. Genes in *Drosophila melanogaster* known to encode these enzymes or corresponding to known pathway steps are in italics. Pathways shown with two arrows comprise several enzymatic steps and are still not characterized fully. Final pigmentation states are given in boxes indicating their color at the end of each pathway. Abbreviations: DAT, dopamine acetyltransferase (also called arylalkylamine-N-acetyltransferase, encoded by aaNAT genes); DDC, DOPA decarboxylase; DOPA, dihydroxyphenylalanine; NADA, N-acetyl dopamine; NBAD

1.2.1 Effect of natural selection on GST and DDC nucleotide variation

Previous studies using neutral microsatellite markers have estimated level of genetic differentiation, (F*st* values) for *An. gambiae* s.s. and *An. arabiensis* from East and West Africa to be consistently high at locus, 33C1 on chromosome 3R (Lehmann *et*

al., 1997, 2003; Donnelly and Townson 2000, 2001; Kayondo *et al.,* 2005). A clinal change in modal allele classes in *An. arabiensis* populations from north–south was also observed at this locus (Donnelly and Townson 2000). Authors proposed that the pattern of allele frequencies observed at the locus could be due to selective foces i.e natural selection or population demographic history such as population expansion, migration and genetic drift (Donnelly and Townson 2000). In *An. arabiensis,* locus 33C1 falls within the 'a' inversion on chromosome 3R. Inversions are informative places to investigate signatures of selection (Kreitman and Wayne 1994; Depaulis *et al.,* 1999). This is because recombination is much reduced within chromosomal inversions and this facilitates the coexistence of favourable adaptive genes that are individually adapted to local conditions (Kirkpatrick and Baton 2006). It is known that the frequency of 2La polymorphic inversion in *An. gambiae s.l.* varies seasonally and clinally in response to aridity tolerance (White *et al.,* 2007). Inversions per say do not confer selective advantage but genes located within the inversions which are free from recombination. The study seeks to determine if putative candidate genes located within the 3Ra inversion showed marked signature of selective pressure at locus 33C1. The effect of selective pressure can be detected within DNA sequence by comparing the observed variation to the distribution derived from neutral theory (Tajima, 1989). One challenge to this population genetics-based signature is how to determine whether a signature is due to selection or demographic events such as recombination, population bottlenecks, population expansion etc. To circumvent this problem, the study uses a multi-locus approach. This is because whereas selection is locus-specific demographic events will affect the entire genome.

1.2.2 Effect of melanisation on life-history traits in *An. gambiae*

Melanism, defined as the appearance of mostly dark or dark forms of an organism is one of the most conspicuous evidence of biodiversity in insects and melanic forms of *An. gambiae, An. albimanus, An. quadrimaculatus, Culex* species have been observed (Benedict *et al.,* 1987, Besansky *et al.,* 1997a; Benedict and Chang 1996). Also, *An. daudi* a closely related melanic form to *An. gambiae s.l.* was found in Sudan. Melanic forms of adult *An. gambiae* have also been observed in nature in Sudan (Aboud, 2003).

The production of melanin is known to be detrimental in the sense that it affects some life-history traits such as developmental time, hatching and fecundity rates as well as pupal weight of the phenotypes. A study on the effect of melanization on the juvenile stages of the European map butterfly showed that melanized 5th instars grow more slowly than the pale ones (Windig, 1999), and a similar study investigating the interaction between melanization and drought-stress environment in Satyrine butterflies, *Pararge aegenia* showed that darker melanic individuals had slower development and lower survival rates than the pale forms (Talloen, 2004). Although melanic phenotypes have been observed in nature, very little information exists on their origin and adaptive significance in natural populations. This is the first of a study, which investigated the effect of melanisation on the some life-history traits of *An. gambiae s.l.*

1.2.3 Objectives

1. To determine the role of natural selection in shaping nucleotide variation at four Glutathione s-transferases (GST) genes within sample populations of *An. arabiensis* from Ethiopia, Sudan, Malawi and Tanzania.

2. To determine role of natural selection in shaping nucleotide variation at the *Dopa decaboxylase* (DDC) locus, which is important in the melanisation pathway.

3. To determine the response of larval samples of *An. gambiae* to different rearing coloured environments.

4. To determine the fitness cost of the induced melanised phenotypes and investigate the role of *Dopa decaboxylase* and *Prophenooxidase* (PO) genes in the phenotypic expression of melanism during the induction.

1.2.4 Research Question 1

The first part of this thesis is a follow-up of previous work by Donnelly and Townson 2000). In that study, microsatellite based analysis of samples of *An. arabiensis* from the same locality showed that a locus, 33C1, within the 3Ra inversion of chromosome was an outlier (i.e the F*st* values were far in excess at this locus compared to seven others). The question is whether this aberrant behaviour observed at this locus may be indicative of differences in selective pressure on locus 33C1 or nearby loci within sample population of *An. arabiensis*. To answer this question, we applied some evolutionary models on sequences of four GST genes obtained from specimen of *An arabiensis* captured from Ethiopia, Sudan, Malawi and Tanzania along north- south transect of Eastern Africa and as well as study the molecular evolution of a nearby locus, DDC from samples obtained from Sudan.

1.2.5 Research Question 2

The first aim is to determine whether the development of melanism is a plastic or genetic response to different environmental or colour cues. To answer the question, we first determine whether melanism can be induced in a population of *An. gambiae.* The second aim is to determine whether the fitness of the melanic phenotypes induced in the population will be affected in anyway and if so what gene are differentially expressed in the melanised phenotypes. To test this we measured the degree of melanisation of the larvae and determined the relationship between their developmental time, pupal weight and survival rates etc. We also determined the role of two genes involved in the melanism pathway, DDC and PO in the expression of the phenotypes.

1.3 LITERATURE REVIEW

1.3.1 Malaria: Disease burden

Malaria remains a disease with a serious public health importance in many areas in the world and approximately 40% of the world's population, mostly those living in the world's poorest countries are at risk of the disease (Figure 3). Globally, it is responsible for 350–500 million clinical cases annually and between one and three million deaths mostly children are attributable to the disease (Breman, 2001).

In Africa, malaria is already estimated to kill between 1 and 2.7 million people every year (Breman 2001) and recent estimates indicated that around 60% of the cases of clinical malaria and over 80% of the deaths occur in Africa south of the Sahara (World Health Organization, 2003). At the core of the magnitude of malaria in Africa is the environment, which is highly conducive to malaria transmission. The most important reason for the persistence of malaria in Africa is the presence of the vectors *An. gambiae* and *An. funestus*. The complexes of *An. gambiae s.l.* and *An. funestus* are about ubiquitous in Africa and they create the world's most efficient vectorial systems for malaria transmission. Of all the factors related to malaria transmission, the number (density), human biting habits, and longevity of anopheline mosquito vectors are the most important. The dominant parasite in Africa is *Plasmodium falciparum*, which is the most pathogenic of all the four human plasmodia and responsible for almost all mortality due to malaria (World Health Organization, 2003).

The socio-economic impact of malaria on African countries cannot be over-emphasized. The disease is estimated to be responsible for an estimated average annual reduction of 1.3% in economic growth for those countries with the highest burden and in Africa the economic cost is about 0.6-1% of Gross Domestic Product (GDP), although recent reports indicates that the disease burden on national income is likely to be higher (Sachs, 2001).

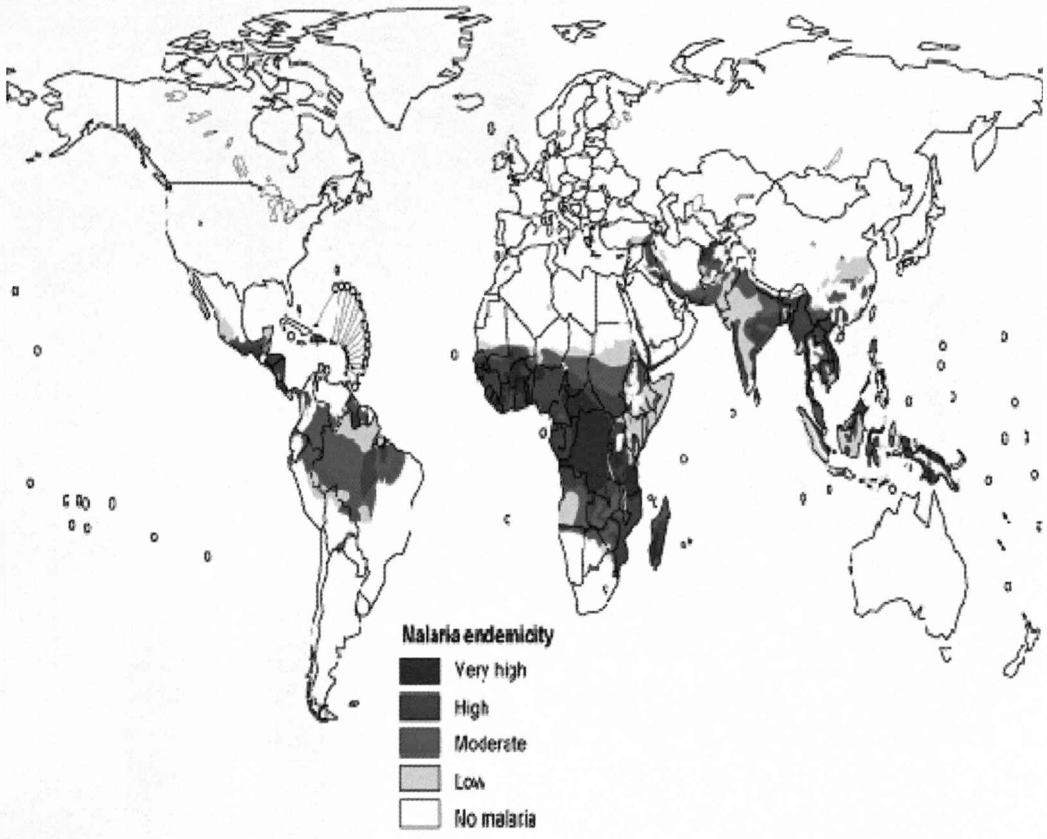


Figure 3: Global distribution of malaria transmission risk with the degree of severity measured by malaria endemicity (Source: World Health Organization, 2003)

1.3.2 Malaria Control Strategies

In malaria-endemic countries the goal of malaria control is to reduce as much as possible the health impact of malaria on a population, using the resources available, and taking into account other health priorities. The main strategies in the control of malaria disease include 1) case management of patients suffering from malaria 2) prevention of infection through vector control and 3) prevention of disease by the use of anti-malarial drugs and vaccines. However, the greatest challenge to malaria control efforts has been the development of resistance to the drugs and insecticides by the parasite and vectors respectively (Nuwaha, 2001).

1.3.2.1 Indoor Residual spraying

Studies have shown that many malaria vectors are endophilic, resting inside houses after taking a blood meal and therefore are particularly susceptible to control through indoor residual spraying (IRS). This method was the primary malaria control tool used during the Global Malaria Eradication Campaign (1955-1969) with great success in some countries such as United States, Europe, Chile and South Asia (Hays, 2000). A curtailment of commitment and resources for malaria control and research characterizes the decades following the global eradication period, as malaria became a disease of the poor nations situated primarily in tropical areas (Shiffman *et al.,* 2002). As a result of the cost of IRS, the negative publicity due to the failure of the Malaria Eradication Campaign, and environmental concerns about residual insecticides, IRS programmes were largely disbanded other than in a few countries with resources to continue them. In fact, previous consensus has been that IRS cannot be used as the main tool for malaria control on long-term basis in tropical Africa (Zahar, 1984). However, the recent success of IRS in reducing malaria cases in South Africa by more than 80% has revived interest in this malaria prevention tool. In fact, IRS is still being used as part of malaria control strategy and some recent large field trials in mainland tropical Africa have produced better results than any of the more recent trials that employed treated nets (Curtis and Mnzava, 2000). In Zanzibar for example and perhaps certain parts of West Africa, DDT resistance in *An. gambiae* now precludes effective use of DDT. In South Africa, however, reversion to DDT spraying from pyrethroid during 2001 has relieved the increasing malaria problem associated with pyrethroid resistance in *An. funestus* (Hargreaves *et*

al., 2000). The effectiveness of indoor residual spraying (IRS) as a vector control measure against malaria have been examined by many studies (Curtis and Mnzava 2000; Pinto *et al.,* 2000; Romi *et al.,* 2002; Sharp *et al.,* 2002; Sharma *et al.,* 2005 ; Sharp *et al.,*2007). Kolaczinski *et al.* (2007) suggested that in high transmission settings, IRS must be implemented indefinitely and at high quality to achieve control objectives.

Recent body of evidence suggest that IRS is a more cost-effective intervention against malaria than other intervention strategies (Goodman *et al.,* 1999, 2001; Guyatt *et al.,* 2002; Conteh *et al.,* 2004). Infact, suggestions have been made that IRS may be both more effective and cheaper than ITNs in communities subjected to low, seasonal risks of infection (Guyatt *et al.,* 2002). A recent study demostrated the effectiveness of IRS by dramatically reducing *An. funestus*, *An. gambiae* and *An. melas* populations, followed by a substantial decrease in the transmission index (Sharp *et al.,* 2007). DDT continues to be effective in many indoor spraying programmes and some advocates are currently promoting its reintroduction into the WHO malaria control strategy (Schapira, 2004, 2006) .

1.3.2.2 Use of Insecticide-treated Bednets

The use of insecticide-treated nets (ITN) represents a new dimension in the use of physical barriers and chemicals in malaria control. Their efficacy in reducing man-vector contact, malaria morbidity and mortality has been demonstrated in various epidemiological situations (Alonso *et al.,* 1991; D'Alessandro *et al.,* 1995; Binka *et al.,* 1996; Lengeler and Snow 1996 ; Nevill *et al.,* 1996; Maxwell *et al.,* 2002; Hawley *et al.,* 2003) Recent studies in Burkina Faso suggested that children who were protected from malaria by Insecticide Treated Materials (ITMs) acquired functional immunity more rapidly than did the control children and are able to clear drug-resistant parasites better than unprotected children (Toure *et al.,* 1998b). Bednets and curtains treated with insecticides such as permethrin have been shown in Mali to have both insecticidal and repellent effects, and may reduce the number of anopheline mosquitoes within houses by more than 95 percent (Toure *et al.,* 1998b). Bifenthrin-impregnated bednets have also been found to have considerable personal protection against *An. funestus* and metabolic pyrethroid resistant *An. gambiae* populations (Chouaibou *et al.,* 2006).

Despite the benefits of ITNs, many workers have argued that the use of this strategy might selectively encourage the proliferation of anopheline mosquitoes that bite outside the home or it could shift the malaria mortality currently observed among young children to older children and young adults. However, this assertion was rejected by Lindblade *et al.* (2004) who showed in a study in Kenya that benefits of insecticide-treated bednets in infants were sustained for up to 6 years and there was no evidence that bednet use from birth increases all-cause mortality in older children in areas of intense perennial transmission of malaria. Another issue which has attracted attention recently is the treatment concentration that provides maximum protection against mosquito bites and the effect this has on the evolution of *kdr* (knock down resistance) mutations. A recent study showed that nets treated with high permethrin concentrations provided better blood feeding prevention against pyrethroid-resistant *An. gambiae* than did lower concentrations and that permethrin-treated nets seem unlikely to select for pyrethroid resistance in areas where the *kdr* mutation is rare and present mainly in heterozygous form (Corbel *et al.,* 2004). There has been a report of reduced efficacy of treated bednets as a result of the intensive use of pyrethroids in bednet impregnation and for agricultural purposes (Vulule *et al.,* 1994). In Benin, a major loss of efficacy associated with pyrethroid resistance was observed in *An. gambiae*, in that only 19% of mosquitoes in the ITN hut and only 22% in the IRS hut were killed, a clear evidence of pyrethroids' failing to control an *An. gambiae* population that contains *kdr* resistance at high levels (N'Guessan *et al.,* 2007)..

1.3.2.3 Modification of vector populations (Transgenesis)

The used of transgenesis in vector control involves the concept of genetically manipulating mosquitoes to make them inefficient vectors. This include identifying gene products that block development of the parasite in the mosquito and introducing these genes into the mosquito germ line to affect their ability to transmit diseases or by reducing reproductive ability of female insects by introducing sterile males or repressible female specific lethal genes in wild populations.

The first demonstration of the genetic transformation of an insect was by (Rubin and Spradling 1982). They showed that the *P*- transposable element could be used to

introduce foreign genes into the germ line of *Drosophila*. Thereafter, some scientist have demonstrated the *P*-element transformation in *Aedes triseriatus* (McGrane *et al.,* 1988), *Ae. aegypti* (Morris *et al.,* 1989) and quite recently in *An. gambiae* (Grossman *et al.,* 2001). Other transformation mechanisms such as the use of transposon-based vectors and promoters have been developed recently (Handler, 2002). Moreira *et al.* (2000) demonstrated that both *An. gambiae* and *An. aegypti* carboxypeptidases promoters (Edwards *et al.,* 1997, 2000) could drive strong blood-inducible transgene expression in transgenic *Ae. aegypti* mid-gut. However, the challenge that many scientist are likely to face will be how to effectively release these transgenic mosquitoes into wild populations. There has been a concern on the environmental impact of the strategy (Moreira *et al.,* 2002) or the fitness of the genetically transformed mosquitoes in their ability to compete with the untransformed ones (Scott *et al.,* 2002). Quite recently a study found that nearly all aspects of development and reproduction of transgenic mosquitoes was severely impaired when compared to non-engineered mosquitoes of the same type (Irvin *et al.,* 2004). However, another study indicated that transgenic mosquitoes lived longer and produced more eggs compared to wild-type mosquitoes (Marrelli *et al.,* 2007) with some malaria-resistant mosquitoes having a fitness advantage when feeding on *Plasmodium*-infected blood (Marrelli *et al.,* 2007). In the light of these challenges, much progress has been made in the area of mosquito transgenesis, although much work remains to be done.

Sterile insect technique (SIT) is another method which has been used successfully to control tsetseflies on the island of Zanzibar in Tanzania (Kabayo 2002). It relies on mass rearing, sterilization and release of large numbers of sterile males over a target. The sterile males will mate with the females and produce sterile offspring therefore reducing the reproductive potential of the wild population. The advantage of the method is that it is highly species-specific and has no adverse effect on non-target species. However, the use of radiation in the method has not encouraged its continous use in vector control.

Another genetic sytem known as RIDL ^{TM}(Release of Insects with a Dominant Lethal) has been used in the control of *Drosophila* species (Alphey and Andreasen 2002) and it is recently been used in *Aedes aegypti* populations. It involved a system in which mass reared insects carry a repressible female specific lethal gene that

enables females to produce only males in the population. In comparison with the SIT, this method does not use radiation and the males are normally fit and able to mate and produce viable male offspring.

1.3.2.4 Larval control

This strategy of vector control takes advantage of the fact that *An. gambiae* complex occupy temporary aquatic discrete habitats (Gillies and De Meillon 1968).
Controlling aquatic stages of malaria vectors depends upon finding where and when they occur and then planning the appropriate intervention measures (Killeen *et al.,* 2002). Some of this measures include the application of 1) oils to the water surface, suffocating the larvae and pupae 2) biological control agents include toxins from the bacterium *Bacillus thuringiensis* var. *israelensis* (Bti) 3) insect growth regulators such as methroprene to retard the development of larvae 4) mosquito fish (*Gambusia affinis*) to control mosquitoes in larger bodies of water and 5) other potential biological control agents, such as fungi (e.g. *Laegenidium giganteum*) or mermithid nematodes. Many larval control strategies have proven successful in many countries including Brazil (Soper *et al.,* 1943), Egypt (Shousha *et al.,* 1948) and quite recently in Djibouti (Louis and Albert 1988) and Ethiopia (Fletcher *et al.,* 1992), Mauritius (Gopaul, 1995) and Cameroon (Barbazan *et al.,* 1998). However, successful implementation of the program depends on how well larval habitats are identified (Killeen *et al.,* 2002) and with the advent of the Geographical Information System (GIS) and molecular techniques, it is now possible to map and characterize the distribution of larval habitats prior to the start of larval control programme (Pinto *et al.,* 2000; Carter *et al.,* 2000a). A recent review of large scale trials of *Anopheline* larval control methods in Africa within the past 15 years suggest that targeting larvae, particularly in human-made habitats, can significantly reduce malaria transmission in appropriate settings (Lynch, 2007). Compared to other malaria control interventions such as ITNs etc, larval control is more practical and when well organized can be integrated successfully with other control measures (Killeen *et al.,* 2002).

1.3.2.5 Vaccines

Over the years, research has shown that the development of an effective cost-effective malaria vaccine represents one of the most important approaches that would be integrated into currently available malaria control strategies. There are many lines of evidence that suggest prophylactic malaria vaccine for humans is feasible (Sabchareon *et al.,* 1991; Egan *et al.,* 1993;Trape *et al.,* 1994; Hoffman *et al.,* 2002). The traditional approach to develop malaria vaccines has focused on targeting one of the different stages of parasite development, i.e. the pre-erythrocytic, the asexual (intra-erythrocytic) or the sexual stage (Girard *et al.,* 2007). In line with this approach, different types of vaccines are currently under consideration. This include 1) pre-erythrocytic vaccines including circumsporozoite protein vaccines and DNA vaccines and live recombinant vaccines 2) asexual blood-stage vaccines based on gene mutations as MSP-1, MSP-2 and RESA combination, MSP-1 and AMA-1, MSP-3, GLURP and SERA and 3) transmission-blocking vaccines which are aimed at inducing antibodies against the sexual stage antigens in order to prevent the development of infectious sporozoites in the salivary glands of the *Anopheles* mosquitoes (Carter *et al.,* 2000b). Although no malaria vaccine has yet gone into routine use, considerable progress has been made in the development of malaria vaccines during the past 20 years. Field trials involving many of these candidate vaccines have taken place in many areas including Africa (Girard *et al.,* 2007) with promising but in some cases limited success. The challenge facing malaria vaccine development is enormous because of the genetic diversity of the parasite antigens and even with the most optimistic scheme of unlimited resources, it will be many years from now before vaccine formulations with sufficient efficacy is obtained. However, current trends in vaccine development provides some optimism. Currently, a consortium of researchers are in the final step of phase III (step before licencing) of RTS,S vaccine. It is envisaged the vaccine could give protection from malaria by 30% and reduce severe malaria by 50% (Maher, 2008) and it could be in use by the year 2011.

1.3.2.6 Intermittent preventive therapy

Intermittent preventive treatment (IPT), involves the administration of a full course of an anti-malarial treatment to a population at risk at specified time points regardless of whether or not they are known to be infected (Greenwood, 2006). The aim is to decrease the frequency of malarial illness while allowing the development of natural immunity (Munday, 2007). The target group for this strategy are pregnant women, infants and older children who are at a high risk of being infected with malaria.
Non-immune pregnant women risk both acute and severe clinical disease, resulting in up to 60% fetal loss and over 10% maternal deaths, including 50% mortality for severe disease. In pregnant women, IPTp is able to reduce placental infection, maternal anemia, and low birth weight (LBW) in HIV-negative women. Intermittent treatment of pregnant women with sulfadoxinepyrimethamine (SP) has been shown to reduce the risk of maternal anaemia, placental parasitaemia, and low birth weight, and is now being integrated into the malaria control programmes of a number of African countries.

New research has lead to a new treatment regime known as intermittent preventive treatment in infants (IPTi). This involves giving a treatment course of antimalarial drugs, regardless of the level of parasitaemia at intervals over the first year of life. In Tanzania, treatment of infants with SP at 2, 3, and 9 months of age, at the time of routine immunization, reduces episodes of clinical malaria by 60%, and episodes of severe anaemia by 50% (Schellenberg *et al.,* 2001, 2005). Also in northern Ghana, where malaria transmission is intense and highly seasonal, IPTi with SP gave 25% protection against clinical attacks of malaria and 35% protection against hospital admissions with anaemia during the first year of life and no protection during the second (Chandramohan *et al.,* 2005). Recent studies of IPT in children have focused on its impact on malaria. In Mali and Senegal, two treatments with SP reduced the annual incidence of clinical attacks of malaria in children aged 6 months to 10 years by 40% (Dicko *et al.,* 2004; Cisse *et al.,* 2006). IPT has promise as a method of malaria control in older children living in an area with a short transmission season. However, a number of important issues need to be resolved with regard to this new treatment regime. These issues include 1) the interaction of IPTs with other malaria control programmes e.g. ITNs 2) delivery and sustainability 3) choice of drug and

drug resistance 4) impairment of the acquisition of natural immunity and 4) the cost of drugs. The challenge is enormous but IPT, employed as part of an integrated malaria control programme, has the potential to make an immediate impact on the many preventable deaths from malaria that occurs in areas where the transmission of malaria is seasonal.

1.3.3 *Anopheles gambiae* complex and disease transmission in Africa

Global distribution of dominant or potentially important malaria vectors is shown in figure 1. Although there are about 400 species of *Anopheles,* only 60 of them transmit malaria under natural conditions, and only 30 are of major importance (Bruce-Chwatt, 1985). Of these, the *An. gambiae* complex and *An. funestus* are the most efficient vectors for *P. falciparum* transmission in Africa: the highest rates of sporozoite development are in *An. gambiae*, the most widespread throughout tropical Africa. *An. gambiae* belongs to a complex of morphologically similar but genetically different species. The species differ in their behaviour and vectorial roles in malaria transmission. Hitherto, the species was initially considered to be a single species.

Presently, seven formally named species with different vectorial capacities are widely recognised as sibling species of the *An. gambiae* complex and they can be distinguished by cytotaxonomic and molecular methods (Coluzzi *et al.,* 2002; Scott *et al.,* 2003).

An. gambiae sensu stricto is the most widely distributed and the most efficient and important vector in Africa (White, 1974; Gilles and Coetzee 1987). This exceptional ability to transmit malaria is due to its high anthropophilic and endophilic behavior (Gilles and Coetzee 1987). The genetic structure of this species is complex, with several recognized ''chromosomal forms'' that differ ecologically. It has also been found that there exist cryptic taxa (M and S forms) within *gambiae s.s.* (Powell *et al.,* 1999). *An. arabiensis* although more zoophilic in many areas, is a major vector of malaria particularly in arid or montane areas such as Ethiopia, Sudan, Malawi and Tanzania (White, 1974). In many areas in the absence of *An. gambiae* it can be a major vector. In areas of sympatry, *An. arabiensis* is more abundant than *An. gambiae* in the dry season (Coluzzi *et al.,* 1979). *An. gambiae* sensu stricto and

An. arabiensis, are subdivided into discrete subpopulations, each carrying a unique set of polymorphic chromosomal inversions, some of which have been associated with differences in seasonality of breeding, adaptation to natural vs. human-disturbed habitats, microhabitat selection, and host preferences (Coluzzi, 1979; Rishikesh *et al.*,1985).

1.3.4 Population genetics of *An. gambiae* populations

An. gambiae and *An. arabiensis* are widespread in many areas in Africa (White, 1974). The numerous fixed and polymorphic chromosomal inversions seem to be one of the driving forces promoting ecological partitioning in *An. gambiae s.s.* (della Torre *et al.,* 2002). Polymorphic chromosomal inversions are common in *An. gambiae s.s* and *An. arabiensis* and these inversions are mostly located on chromosome 2*R* (Petrarca *et al.,* 1983; Bryan *et al.,* 1987). *An. gambiae s.s.* is polymorphic for two arrangements, 2La and 2L+^{a} and it is believed that the 2L+^{a} arose from the 2La (Ayala and Coluzzi 2005; Sharakhov *et al.,* 2006). The frequency of 2La varies clinally and seasonally in a pattern suggesting response to selection for aridity tolerance (White *et al.,* 2007). A molecular technique is now available for karyotyping the 2La inversion in *An. gambiae* (White *et al.,* 2007). Based on the distribution of *2R* inversions, numerous chromosomal forms have been defined (Toure *et al.,* 1998a). The chromosomal forms identified based on inversion karyotypes include the Forest and Savanna forms (Coluzzi *et al.,* 1985) as well as Bamako, Bissau, Mopti and Savanna also identified based on different inversion karyotype frequencies. These chromosomal forms are not reproductively isolated throughout West Africa but rather represent adaptations to particular habitats (Touré *et al.,* 1998a; della Torre *et al.,* 2001). The FOREST form can normally be found in the forest areas whereas the MOPTI form occurs in Guinea, Mali and Burkina Faso (Coluzzi *et al.,* 1985; Toure *et al.,* 1994). The Savanna form is the most widespread throughout the sub-saharan Africa whereas the BISSAU form is restricted to Gambia and Senegal and BAMAKO form to Mali and northern Guinea. Chromosomal inversion patterns of *An. gambiae* in sourthern Sudan showed characteristics of intergrading Savanna/Forest populations similar to those observed in West Africa (Petrarca *et al.,* 2000). On the other hand, *An. arabiensis* is polymorphic for inversion systems (2Ra, 2Rb, 2Rd1, 3Ra) in West Africa (Petrarca *et al.,* 2000).

Marked differences exist between and within chromosomes in their ability to introgress from *An. gambiae* into *An. arabiensis* (Slotman *et al.,* 2005). Earlier on, a study had suggested local specific unidirectional introgression of mitochondria from *An. arabiensis* into *An. gambiae* (Donnelly *et al.,* 2004). Studies of the genetic structure of *An. gambiae* have reached conflicting results. Using mitochondrial DNA (Donnelly and Townson 2000; Besansky *et al.,* 1997), microsatellite markers (Lehmann *et al.,* 1996; Kamau *et al.,* 1999; Onyabe and Conn 2001), allozymes (Lehmann *et al.,* 1996), some workers have suggested that there is extensive gene flow across Africa in *An. gambiae* at least 5000 km apart. Other microsatellite-based studies had postulated that there exists limited gene flow between *An. gambiae* populations (Donnelly *et al.,* 2001; Wondji *et al.,* 2002). Levels of population differentiation were lower in *An. arabiensis* than *An. gambiae* in East Africa and Kenya (Donnelly and Townson 2000; Lehmann *et al.,* 2003). However, it is important to take into consideration the location of the microsatellite marker under study since its been reported that the extent of genetic differentiation in *An. arabiensis* populations from three ecological zones in Kenya varied significantly with respect to microsatellite markers' location relative to chromosomal inversions (Temu and Yan 2005).

An. gambiae s.s. exist as two distinct molecular forms, named M and S and their classification was based on fixed differences in the *X*-linked rDNA (Favia *et al.,* 1997). However, studies have found that barriers to gene flow do not show throughout the entire range of distribution of M and S forms (Toure *et al.,* 1998a; Taylor *et al.,* 2001; Gentile *et al.,* 2002). Studies of gene flow have found genetic differentiation between the molecular forms to be linked to the rDNA (Wondji *et al.,* 2002; Lehmann *et al.,* 2003) although some other workers could not detect any differentiation between the two cryptic species in Kenya (Gentile *et al.,* 2001; Besansky *et al.,* 1997; della Torre *et al.,* 2001). The Great Rift Valley has been postulated to serve as barrier limiting gene flow within populations in East Africa (Lehmann *et al.,* 1996, 1999).

1.3.5 Insecticide resistance in mosquitoes

Insecticide resistance can be defined as ''the development of an ability in some individuals of a given organism to tolerate doses of a toxicant which will prove lethal to a majority of individuals in a normal population of the same organism'' (World Health Organization, 1957). Insecticide application remains the most important component of mosquito vector control strategy (Zaim *et al.,* 2000; Najera, 2001; Hemingway *et al.,* 2002). Apart from other classes of insecticides, e.g organochlorines and carbamates, pyrethroids account for 25% of the world insecticide market (Hemingway, 2004) and the most widely used for the treatment of mosquito nets (Najera, 2001). However, with the continuous use of insecticides for mosquito control and agricultural purposes, insecticide resistance has developed in over 100 mosquito species (Hemingway and Ranson 2000) including mosquito vectors (Curtis *et al.,* 1998; Hargreaves *et al.,* 2000; McAbee *et al.,* 2004; Enayati *et al.,* 2005).

1.3.5.1 Mechanism of insecticide resistance

Insecticide resistance in organisms may manifest either biochemically or behaviourally or both. For the purpose of this study, we will focus on biochemical resistance. The two major forms of biochemical resistance are target-site insensitivity and metabolism resistance (Hemingway and Ranson 2000).

1.3.5.1.1 Target site insensitivity of sodium channels

Knock down resistance (kdr) mutation An. gambiae

Sodium channels are special gates within the membrane of insects that separates nerve cells from their surrounding environments (Xu *et al.,* 2006). These channels have been the important target for both pyrethroid insecticides and DDT action (Narahashi, 1996). In the development of resistance to pyrethroids and DDT, insects are able to structurally modify the channels and reduce their sensitivity to insecticides. The term 'knock down resistance' (*kdr*) refers to resistance to pyrethroids and DDT as a result of reduced sensitivity of the sodium channel. *Kdr* mutation *in An. gambiae s.l.* is the substitution of leucine to phenylalanine (Leu to Phe) or leucine to serine in the domain II segment 6 of the sodium channel (Martinez-Torres *et al.,*1998; Hemingway *et al.,* 2004). The *kdr* mutation in

mosquitoes has received much interest because of the use of pyrethroids in malaria control.

Kdr was first reported in *An. gambiae* from Africa in 1998 (Martinez-Torres, 1998). Two mutations (L1014F and L1014S) have been detected in pyrethroid resistant *An. gambiae s.s.* (Ranson *et al.,* 2000b) and *An. arabiensis* (Diabate *et al.,* 2002; Stump *et al.,* 2004). The predominant *kdr* mutation in *An. gambiae* in several West African countries is the L1014F, termed *kdr*-west (*kdr*-w) whereas the L1014S or *kdr*-east (*kdr*-e) predominates in eastern Africa (della Torre *et al.,* 2001; Fanello *et al.,* 2003b; Diabate *et al.,* 2004; Yawson *et al.,* 2004; Stump *et al.,* 2004; Awolola *et al.,* 2005). However, both the East and West African *kdr* mutations have been found to co-occur in samples from Gabon (Pinto *et al.,* 2006).

The *kdr* mutation has more recently been found in both M and S-forms in sympatry in Benin (Fanello *et al.,* 2000), Burkina Faso (Diabate *et al.,* 2004) , Ghana (Yawson *et al.,* 2004) and in Cameroon (Etang *et al.,* 2006). In general the frequency of the *kdr* allele in M-form is much lower than that of S-form. Yawson *et al.,* (2004) observed very high frequency (98–100%) of the *kdr* mutation within the S form but reached a maximum of only 3.38% in the M form in one population at an irrigation scheme in the Ghanaian coastal savannah zone. Recently, *kdr* was detected for the first time within the S molecular form of the Bamako chromosomal form (Tripet *et al.,* 2007). Weill *et al.* (2000) postulated that the *kdr* allele is present in both M and S-forms as a result of introgression from the S-form to M-form based on sequence analysis of the Intron 1 sequence upstream of the *kdr* mutation. A single specimen of *An. arabiensis* carrying *kdr* was discovered in Burkina Faso and Kenya (Diabaté *et al.,* 2004; Stump *et al.,* 2004) and it was postulated that the *kdr* may have introgressed into *An. arabiensis* from *An. gambiae* (Slotman *et al.,* 2005). However, the *kdr* mutation observed in *An. arabiensis* in Burkina Faso was a new and independent mutation event (Diabaté *et al.,* 2004). The impact of *kdr* on the efficacy of ITNs has come under discussion recently. Recent studies suggest that the efficacy of ITN may be drastically reduced in the presense of *kdr* (N'Guessan *et al.,* 2007) although studies have shown that mosquito mortality from ITNs remains high and personal protection from biting remains for free flying wild mosquitoes (Darriet *et al.,* 1998, 2000; Chandre *et al.,* 1999).

1.3.5.1.2 Metabolic resistance in *An. gambiae*

Metabolic detoxification of xenobiotics including insecticides by enzymes is common in insects (Scott *et al.,* 1991) and increased metabolic detoxification is one of the most common mechanisms of insecticide resistance (Hemingway and Karunaratne 1998; Hemingway, 2000). Several detoxifying enzymes have been implicated in the development of metabolic resistance. These include 1) insect cytochrome P450s, 2) insect glutathione S-transferases (GSTs) and 3) esterases or carboxylesterases. There are 111 P450, 31 GST and 51 carboxylesterase in the genome of *An. gambiae* (Collins and Haseltine 2000). Although each of these enzyme families is encoded by supergene families, most of the individual genes that are up-regulated or amplified in insecticide resistant individuals are yet to be identified (David *et al.,* 2005).

Cytochrome P450s are a superfamily of haemoproteins that are responsible for the oxidative metabolism of a wide variety of xenobiotic compounds and endogenous compounds. Insect cytochrome P450s have been implicated in the detoxification of insecticides and plant toxins (Scott *et al.,* 1998; Feyereisen, 1999). There are approximately 100 different P450 enzymes predicted in *An. gambiae* and over 25 classes of insect P450s have been identified (Ranson *et al.,* 2002a). P450 enzymes are known to confer high levels of resistance (Scott and Georghiou 1986) as well as being cross-resistant to other unrelated compounds (Scott *et al.,* 1991). Several studies have shown elevated levels of cytochrome P450 monooxylase and esterase activities in pyrethroid resistant mosquitoes (Vulule *et al.,* 1999; Enayati *et al.,* 2003; Hemingway, 2004). Other P450s such as *CYP6Z1* was overexpressed in pyrethroid resistant *An. gambiae* (Nikou *et al.,* 2003).

Carboxylesterases structurally belong to a superfamily of α/β-fold proteins, which consist of alternate α-helix and β-sheets connected by loops with a varying length (Oakeshott *et al.,* 1999). These enzymes hydrolyze chemicals containing such functional groups as a carboxylic acid ester, amide, and thioester (Wheelock *et al.,* 2005). Carboxylesterase activity is widely distributed in mammalian tissues, with the highest levels present in liver microsomes (Wheelock *et al.,* 2005).

1.3.6 *Glutathione s-transferases in insects*

Glutathione s-transferases (GSTs) are a major family of detoxication enzymes found in most organisms. They are soluble dimeric proteins that are ubiquitous in nature and are involved in the metabolism, detoxication and excretion of a large number of endogenous and exogenous compounds from the cell. A typical GST-catalysed reaction involves the transfer of the tripeptide glutathione to an electrophilic substrate. To facilitate this reaction, each GST monomer has two distinct binding sites, a G site, which binds glutathione, and an H site which binds the substrate. A typical structure of a GST gene is shown in Figure 4. Variability in the structure of the H site, largely accounts for the wide range of substrate specificities of the GSTs (Sheeban, 2001). Over 40 GST genes have been detected in the genomes of higher eukaryotes and these have been classified into at least 13 different classes based on their amino acid sequence identities, immunological properties and substrate specificities. Some classes are found across multiple eukaryote phyla, for example the Zeta and Omega classes (Board *et al.,* 2000), whereas others, such as the insect-specific Delta and Epsilon classes are more restricted in their distribution (Ranson, 2001).

Epsilon class GSTs belonging to insect specific GSTs are of particular interest because of their potential to cause resistances to all the major families of insecticide especially organophosphates and DDT (Huang, 1998). Many studies have associated GSTs with the detoxification of insecticides (Fournier *et al.,* 1992; Tang and Tu 1994; Ranson *et al.,* 1997; Brogdon and McAllister 1998; Hemingway, 2000; Hemingway and Ranson 2000; Lumjuan *et al.,* 2007). Several GSTs have been found to be overexpressed in *An. gambiae* (Ding *et al.,* 2003) and their role in the dehydrochlorinase activity of DDT has been studied (Ranson, 2001). A study implicated one member of the GST cluster, GSTe2 in the development of DDT resistance on the basis that this gene was over-transcribed in a DDT-resistant strain of *An. gambiae* and the recombinant protein from GSTe2 possessed DDT dehydrogenase activity (Ranson, 2001). This gene belong to the Glutathione S-Transferase (GST) group and occur on the right arm of chromosome 3 within the polymorphic chromosomal inversion 3Ra (subdivision locus 33B). Earlier biochemical studies had indicated that DDT resistance is associated with both

qualitative and quantitative changes in multiple GST enzymes (Prapanthadara and Ketterman 1993). Muller *et al.* (2007) recently profiled samples of *An. gambiae* from Ghana, West Africa and found that genes up-regulated in these samples differed from those East African strain of pyrethroid resistant strains. This finding implied that metabolic resistance may have multiple origin in *An. gambiae.*

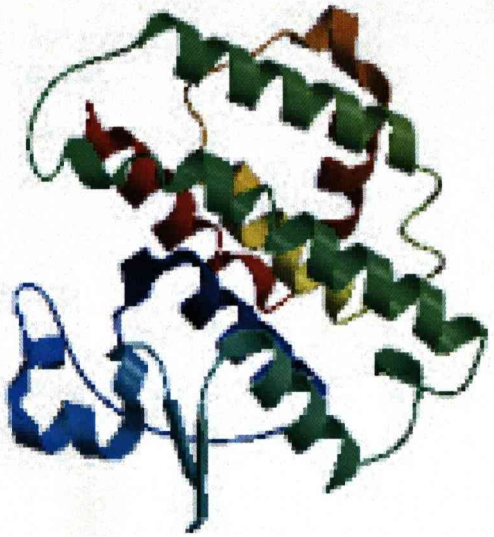


Figure 4: 3D structure of Glutathione s-transferase E2 from *An. arabiensis*

N-terminus domain is coloured green and the C-terminal is coloured red, GST is the substrate binding site is coloured light blue (Modelled from one GSTe2 sequence described in Chapter 3).

1.3.7 Melanism in insects

Melanism (the occurrence dark or black form of a species) has intrigued evolutionary biologists for over a century (Majerus, 1998). Numerous hypotheses have been proposed to explain the origin and maintenance of melanic phenotypes. Melanism has been particularly well studied in insects as it plays an important role in ecology and defence against parasites (Siva-Jothy, 2000), mate signalling (Ellers and Boggs 2003) and thermoregulation (Jong *et al.,* 1996); Ottenheim *et al.,* 1999). Melanism in larval *Lepidoptera* has been particularly well studied, and previous work has shown that the expression of melanism may be affected by a range of environmental factors, including temperature (Gunn, 1998; Hazel, 2002; Solensky and Larkin 2003), light (Faure, 1943), humidity (Goulson, 1994), nutrition (McGraw, 2007) and population density (Kazimirova, 1992; Goulson and Cory 1995; Gunn, 1998; Hagen *et al.,* 2006).

Melanism is often thought to occur as a result of genetic mutation, but can it now known that it can result from other stimuli, such as exposure to abnormal temperature changes which transiently alter gene transcription or translation. In larval *Lepidoptera* and other soft-bodied insects, melanism has been shown to be a dynamic trait both within and between larval instars (Chapman, 1998) and its expression changes with the development of the larvae. The relationship between melanization and life history traits has been studied in several species of lepidoptera including *Inachis io* and *Araschnia levana* (Windig, 1999). Studies have also linked short juvenile development time and large adult size to fitness (Nylin and Gotthard 1998), and variation in melanization is also linked to these traits (Majerus, 1998).

1.3.7.1 *Melanism in An. gambiae complex*

Numerous phenotypic mutants have been described in anophelines (Kitzmiller, 1976; Seawright *et al.,* 1985). A red stripe phenotype, which was apparent in larvae and pupae of *An. arabiensis,* was described earlier by Mason (1967) but he was uncertain of the pattern of inheritance of the trait. A similar phenotype called red stripe has been observed and analyzed in *An. albimanus* (Nakashima *et al.,* 1975) and *An. quadrimaculatus* (Mitchell and Seawright 1984). Recessive lethals of anopheline mosquitoes have been observed (Seawright and Benedict 1985) but they were

characterized by various deformities of setae, eye, and body shape. However, among this class are *An. albimanus* mutants, *bubblehead* (Seawright and Benedict 1985) and *curled* (Seawright *et al.,* 1985) which had no deformities because they were expressed in hemizygous males (XY), and such mutations can be maintained through viable heterozygous sibling females (XX).

Among the rarely detected mutations are those that modify normal physiological responses to environmental conditions but otherwise have no visible effect. Homochromy (larval colour change as a result of rearing background-colour) in *Anopheles* is one response in which such mutants are easily detected, since the colour-change is dramatic and simple to induce (Fuzeau-Braesch, 1972). When larvae of some organisms are reared on an illuminated dark background, larvae darken significantly and this effect has been described extensively in various mosquito species (Benedict and Seawright 1987). However, it was interesting to note that larvae (mutants) that did not respond to the background colour were eye-colour mutants (Benedict and Chang 1996; Benedict and Seawright 1987). It was therefore suggested that the colour change is stimulated by the perception of the rearing background colour by the larval ocelli and the response is affected through the neurophysiological pathway. Therefore, it is reasonable to expect that numerous mutations that interfere with transduction of the appropriate signals or pigment synthesis and transport might interfere with the response (Benedict and Seawright 1987).

1.3.8 *Dopa decaboxylase* gene

Dopa decarboxylase gene (DDC) contributes to diverse physiological events in insects. The product of this enzyme's activity, dopamine, acts as an intermediate substrate for various tanning and melanization reactions (Eveleth *et al.,* 1986; Scholnick *et al.,* 1986). The DDC gene has been characterized in many insects (Hirsh and Davidson 1981; Hiruma *et al.,* 1995; Ferdig *et al.,* 2000; Noguchi *et al.,* 2003). In mosquitoes the biochemical pathway of tanning and melanization reactions has been well characterized. The tanning process is initiated by the action of phenol oxidase (PO) on the substrate tyrosine to produce 3, 4-dihydroxyphenylalanine that subsequently converted by DDC to dopamine (Li and Christensen 1993). DDC

activity in tanning of the egg chorion of *Ae. aegypti* was initiated by the ingestion of a blood meal (Li and Christensen 1993). During oviposition, the chorion blackens to form a protective outer layer that allows the eggs to withstand an indeterminate desiccation period required in this species life cycle. Also differential tissue-specific regulation of DDC has been studied in *D. melanogaster* and *Manduca sexta* by experimental promoter analysis (Scholnick *et al.,* 1986; Konrad and Marsh 1987; Hodgetts *et al.,* 1995; Hiruma *et al.,* 1995) and several regulatory elements required for tissue-specific expression have been identified (Bray and Kafatos 1991; Scholnick *et al.,* 1986). In each case, the regulation of this enzyme seems to occur at the level of transcription. The *Drosophila* DDC gene produces two different transcripts, one in the epidermis and another in the central nervous system (CNS), which can be accounted for by an alternative splicing mechanism in which all four exons are involved in neuronal transcription, but the second exon is spliced out of the epidermal message (Eveleth *et al.,* 1986; Morgan *et al.,* 1986). Catecholamine metabolism in mosquitoes also plays a role in the melanotic encapsulation defence response against parasites.

1.3.9 Population genetic signatures of selection

One of the main effects of selection is to modify levels of variability within and between species (Nielsen, 2005). Selection leaves signatures that can be detected within the genome. For example, a selective sweep illustrated in Figure 5 tends to drastically reduce variation and the reduction affects level of population subdivision especially if the sweep is new in the population (Charlesworth, 1997; Slatkin and Wiehe 1998). When a locus shows a high level of genetic differentiation compared with other loci, it may be interpreted as evidence of balancing selection. Akey *et al.* (2002) used estimates of differentiation, *Fst* to identify regions of increased population subdivision. Selection can also affect the distribution of alleles within populations. Selection against deleterious mutations will increase the fraction of mutations segregating at low frequency in the population whilst a positive selection will tend to increase the frequency in a sample of mutations segregating at high frequency (Nielsen, 2005).

The statistical tests of neutrality are designed to detect selection within populations or between species. The basics of these statistical test emanates from the neutral theory of molecular evolution (Kimura, 1968) that stipulates that sequence differences between species reflect neutral polymorphisms that have drifted to fixation in one or the other species. Adaptive evolution focuses largely on cases where the data departs from neutral model. Such departures from neutrality may be detected through both interspecific and intraspecific DNA sequence comparison. A number of statistical tests have been proposed to detect departures from neutrality using intraspecific polymorphism data. Many of the common population genetic methods for detecting selection are based on comparing variation within and between species or comparing the rate of polymorphisms to divergence for multiple loci (Hudson *et al.*, 1987).

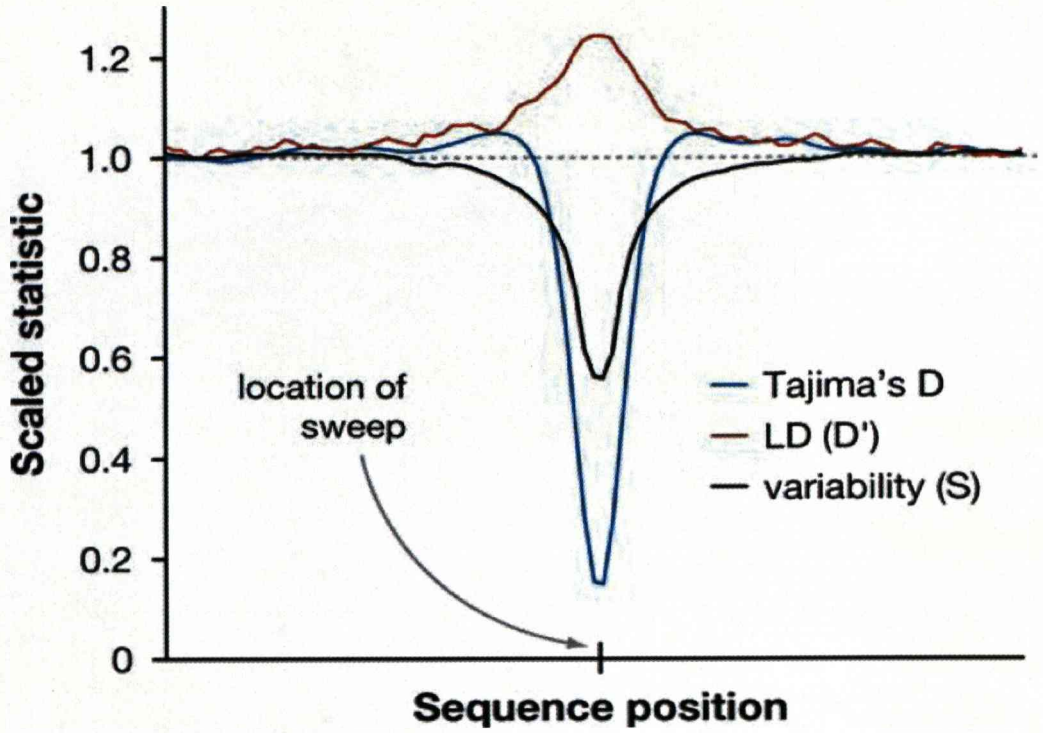


Figure 5: The effect of a selective sweep on genetic variation.

The figure is based on averaging over 100 simulations of a strong selective sweep. It illustrates how the number of variable sites (variability S) is reduced, Linkage disequilibrium (LD) is increased, and the frequency spectrum as measured by Tajima's D is skewed in the region around the sweep. All statistics are calculated in a sliding window along the sequence after the advantageous allele has reached frequency 1 in the population. All statistics are scaled so that the expected value under neutrality equals one (Source: Nielsen, 2005).

1.3.9.1 *Tajima's test of neutrality*

The test of Tajima uses intra-population data and compares the observed frequencies of variants at polymorphic sites to the frequencies expected under the neutral theory (Tajima, 1989). It compares the difference between the two estimators of the parameter θ, the number of segregating sites in a sample, S and the average pairwise difference, π, in the number of nucleotides. The expected difference between the two estimators should be zero under the neutral model of molecular evolution. Positive and negative values of the test correspond to departures from equilibrium neutral

expectations in the direction of having the frequency spectrum skewed towards many intermediate-frequency polymorphisms or too many low-frequency polymorphisms respectively. Significantly positive values of Tajimas D test are consistent with balancing selection for two or more alleles, but can also indicate the presence of population sub-division. Significantly negative values of Tajimas D test are consistent with a recent selective sweep or purifying selection of a linked mutation, population bottleneck and/or, a recent expansion in population size (Slatkin and Hudson 1991). However, the Tajimas test has a limited power to detect selection because the test statistics is sensitive to sample size, S and timing of selection events (Simonsen *et al.,* 1995).

1.3.9.2 *The Hudson, Kreitman and Aguade (HKA)*

The Hudson, Kreitman and Aguade (HKA) test uses intraspecific and interspecific nucleotide data to test the null hypothesis that the ratio of polymorphism to divergence is equivalent among loci (Hudson *et al.,* 1987), The HKA test rejects a neutral model if the ratio of polymorphism to divergence differs significantly among independent loci, provided that one genetic locus evolves according to predictions of the neutral theory. A ratio of polymorphism to divergence that is significantly lower neutral expectation indicates that directional selection has recently fixed a new adaptive mutation within species. When the observed polymorphism levels exceed neutral expectations, then balancing selection has maintained alleles within populations for many generations. This test may not always detect directional selection events if the adaptive increase in allele frequency is fairly recent (Hudson *et al.* 1992).

1.3.9.3 *McDonald and Kreitman test*

McDonald and Kreitman (1991) proposed a test of neutral protein evolution that compares the ratio of variability in replacement and synonymous sites, but do so for both within-species polymorphism and between-species divergence. The test estimates from interspecific DNA data the ratio of nonsynonymous substitution rate (Ka) to that of synonymous substitution (Ks). Ka/Ks=1 is expected for genes evolving neutrally, where selection neither favors nor disfavors changes in the amino acid sequence. Ka/Ks <1 is the commonly observed situation and suggests negative

(purifying) selection acting to remove amino acid replacements. Ka/Ks >1 indicates positive selection and is rarely observed. Directional selection is detected if the nonsynonymous/synonymous ratio is greater for fixed differences than for polymorphism, suggestion that too many amino acid replacements have accumulated between species relative to the interspecific number of synonymous changes.

1.3.9.4 *Fu and Li test of neutrality*

Fu and Li (1993) derived a test of selective neutrality that examines the number of mutations that occur on external versus internal branches of a geneology. The test uses the within-and between-species data to estimate the number of internal and external mutations. The test statistic D is used to determine if there is an excess or deficiency of external mutations. A significantly negative value of D indicates directional selection that is due to an excess of external mutations, while a significant positive value of D denotes balancing selection that is due to a deficiency of external mutations.

1.4 EXPECTED OUTCOME OF THE STUDY

The present study on the molecular level generated information on the evolutionary forces shaping patterns of nucleotide variation within natural populations of *An. arabiensis* and also provided insight into the concept of phenotypic plasticity and the generation of adaptive phenotypes in populations of *An. gambiae,* both of which are malaria vectors in Africa. *Anopheles gambiae* Giles genome sequencing project is almost completed. This is a necessary tool towards developing more effective strategies in reducing not only malaria but other vector borne diseases transmitted by the mosquito. To be able to meet this challenges, one needs baseline information on genome structure, gene function and environmental effects on genetic expression. The information generated in this study will be useful in understanding the various molecular and evolutionary mechanisms underlying adaptation of *An. gambiae s.l.* to heteregenous environments and help formulate new genetic tools or enhance already existing malaria control strategies.

CHAPTER 2 MATERIALS AND METHODS

2.1 MATERIALS AND METHODS I

The role of natural selection in shaping nucleotide polymorphisms within GST and DDC loci of *An. arabiensis* populations

2.1.1 Collection of *An. arabiensis* populations

The mosquito samples used in this study were collected from 4 locations in Africa (Figure 6) and have been described in detail elsewhere (Donnelly and Townson 2000). For this study, we analysed mosquito samples from Wad Awad, Sudan ; Harosha, Ethiopia ; Mkali, Malawi and Ifakara, Tanazania. Species collection location and date were as follows: Mkali, Malawi (August 1997), Wad Awad, Sudan (1997), Harosha, Ethiopia (August 1997) and Ifakara, Tanzania (1999). For interspecific analyses, *An. stephensi* samples from Bahewahay, Pakistan (2002) were used.

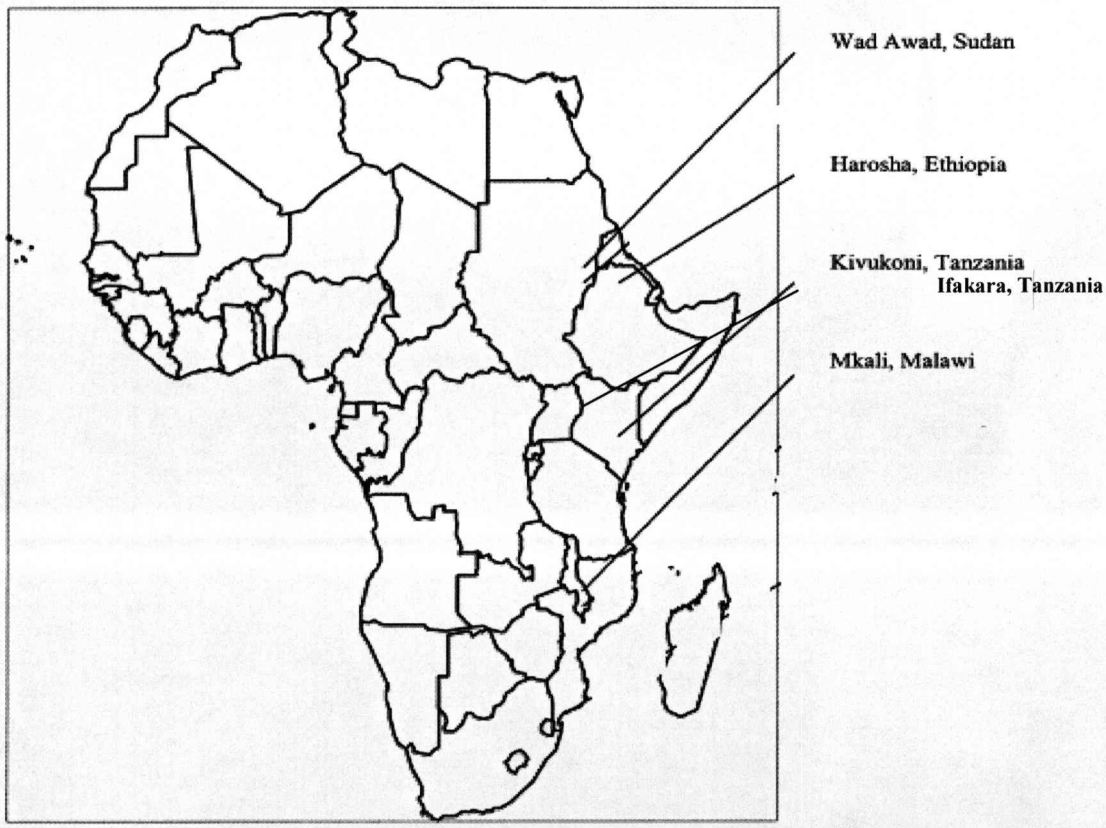


Figure 6: Map showing the mosquito collection sites (Modified from Donnelly and Townson, 2000)

2.1.2 Mosquito collection sites

2.1.2.1 *Ethiopia*

Mosquitoes were collected using CDC light-trap methods by Asefaw Getachew on the 16th and 17th August 1997 from a village in Harosha (12° 19'00''N 39°39'00''E) at an altitude of 1450m above sea level on the Ethiopian Plateau (Figure 6). Ethiopia has population of 74,777, 981 million with a land area of 1,127,127km^2. Harosha is one of the most populated places in Ethiopia.

2.1.2.2 *Sudan*

Specimens from Sudan was collected from Wad Awad (13°53'N 34°34'E) 140km East of Khartoum the capital of Sudan (Figure 6). The collections were done by Derek Charlwood using Pyrethrum Spray collection method. Sudan is the largest country in Africa, comprising more than 8% of the entire continent. The total population is estimated to be 30.3 million inhabitants, of whom 75% live in rural areas. In Sudan, there are 7.5 million cases and 35,000 malaria deaths every year.

2.1.2.3 *Malawi*

Specimen from Malawi used for this study were collected during the period from July to September 1997 from a village, Mkali (14°21'S 35°18'E) at an altitude of 500m above sea level (Figure 6). Some specimens were also provided by Dr Themba Mzilahowa previously of the Liverpool School of Tropical Medicine, UK. Malawi is situated in southeastern Africa (Figure 6) and it is surrounded by Lake Malawi which is the third-largest lake in Africa, making about 20% of Malawi's land area. The country has a population of 12,884,000 (July 2005 estimate) with a total area of 118,484 km^2. Malawi's climate is subtropical. The rainy season runs from November through April. There is little or no rainfall throughout most of the country from May to October. It is hot and humid from October to April along the lake and in the Lower Shire Valley. Malaria is the most frequent cause of morbidity and mortality in children under five years of age, and over 40% of deaths in children under two years.

2.1.2.4 *Tanzania*

Specimens from Ifakara (8° 18' 0 S, 36° 25' 0 E), Tanzania were aspirated from rooms between the 18th and 21st July 1995 by Derek Charlwood. Tanzania is located in East Africa. It borders the Indian Ocean and located between Kenya and Mozambique (figure 6). It has a population of 39,384,223 and covers a land area of 945,087 km^2. Its climate varies from tropical along coast to temperate in the highlands.

2.1.3 DNA extraction methods

2.1.3.1 *Ballinger–Crabtree Method (Ballinger–Crabtree et al., 1992)*

Individual mosquitoes were homogenized in a 1.5 ml micro centrifuge tube in 270µl of lysis buffer and 30µl of 10% SDS and 5µl of 20mg/ml proteinase K. The suspension was then incubated overnight at 50°C and thereafter 305ul of phenol: isoamyl:chloroform (24:1:1) solution was added and mixed by spinning on a wheel at 20 rpm for 20 minutes. After centrifugation at 13,000 rpm for 10 minutes, equal volume of phenol: chloroform (24:1) was added to the supernatant, mixed and centrifuged again as previous. The supernatant was removed and 0.2 volume of 10M ammonium acetate added to precipitate pellets of DNA. The solution was vortexed and centrifuged at 13,000 rpm at 4°C for 30 minutes. The supernatant is discarded and the pellets were washed first in 500 µl ice-cold absolute alcohol. The suspension was then centrifuged again and the supernatant discarded. The process was repeated with 70% alcohol and the pellets were finally air-dried and resuspended in 200µl of distilled water. The DNA was stored at -20°C.

2.1.3.2 *Livak method (Livak, 1984)*

Individual mosquitoes were homogenized in 100µl of Livak grinding buffer in 1.5 micro-centrifuge tubes and incubated at 65°C for 30 minutes. 14µl of 8M K^+ acetate solution was added and left on ice for 30 minutes. Samples were centrifuged at 14,000 rpm at 4°C for 20-25 minutes and 200µl of ice cold 100% ethanol was added to the supernatant. The samples were again centrifuged at 14,000 rpm for 20 minutes at 4°C and the supernatant was discarded. The above process was repeated with 70%

ethanol and the pellets were left at room temperature to dry before resuspending in 100µl of distilled water.

	GENE				
Place of collection	***GSTε1***	***GSTε2***	***GSTε6***	***GSTε8***	**DDC**
Ethiopia					
Harosha	5	5	5	5	-
Sudan					
Wad awad	5	5	5	5	-
Shambat/Hagusuf	-	-	-	-	15
Malawi					
Mkali	5	5	5	5	-
Tanzania					
Ifakara	5	5	5	5	-
Total	20	20	20	20	15

Table 1: Sample collection sites and number of *An. arabiensis* samples analysed for each gene surveyed

(GSTε1, Glutathione s-transferases E1:GSTε2,Glutathione s-transferases E2: GSTε6,Glutathione s-transferases E6:GSTε8,Glutathione s-transferases E8)

2.1.4 PCR identification of *An. arabiensis*

Members of *An. gambiae* complex were identified by PCR (Scott *et al.,* 1993) using species-specific primers. PCR was done in a total reaction mix of 25µl containing 1X PCR buffer, 0.5mM each of dNTPs, 0.5µM each of primer and 0.5U taq DNA polymerase. Amplification conditions were: One cycle at 94°C for 5 minutes followed by 30 cycles of at 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 1 minute. An additional extension of 72°C for 10 minutes was included for 1 cycle at the end of the reaction

2.1.5 Design of primers for GST and DDC amplification in *An. arabiensis*

Putative GST genes from *An.gambiae* were identified online (www.ncbi.nlm.nih.gov/blast/mmtrace.html) and primers were designed for each gene using primer 3 software (Rozen and Skaletsky 1999). Details of the primers are given in the Appendix. Four epsilon class GSTs were selected; GSTε1, GSTε2, GSTε6 and GSTε8. GSTε2 is known to be over- expressed in DDT resistant *An. gambiae (*Ding *et al.,* 2003*).*

2.1.6 PCR amplification of GST and DDC genes in *An. arabiensis*

Amplification of GST genes was performed using technique as previously described by Ranson *et al.,* 2001 with slight modifications. The PCR conditions were as follows: 25 µl reaction containing, 1X reaction buffer (500mM KCl, 100mM Tris-HCl pH 8.3), 1.5 mM $MgCl_2$, 0.5mM each of dNTP, 0.5µM each of primer, 0.5U taq DNA polymerase. Amplification conditions were: One cycle at 94°C for 15 minutes followed by 30 cycles of at 94°C for 20 seconds, 52°C for 30 seconds, and 72°C for 20 seconds. An additional extension of 72°C for 10 minutes was included for 1 cycle at the end of the reaction.

PCR amplification of the Dopa gene (DDC) was performed in a 25µl reaction mix containing, 1X reaction buffer (500mM KCl, 100mM Tris-HCl, pH 8.3), 1.5 mM $MgCl_2$, 10mM each of dNTP, 0.5μM each of primer, 0.5U taq DNA polymerase. Amplification conditions were: One cycle at 95°C for 10 minutes followed by 29 cycles of at 95°C for 30 seconds, 58°C for 30 seconds, and 72°C for 1min. An.

additional extension of 72°C for 10 minutes was included for 1 cycle at the end of the reaction. The sequences of the primers used for the PCR are shown in the appendix.

2.1.7 Gel electrophoresis

Gel electrophoresis was used to separate and visualize DNA fragments after amplification with PCR. Between (5μl and 10μl) of the genomic PCR products were electrophoresed in 2% agarose (proportion 1grams agarose + 50mls of TAE), stained with ethidium bromide and visualised on a UV-light spectrophotometer.

2.1.8 Subcloning and sequencing of GST and DDC genes in *An. arabiensis*

The resultant PCR products of Epsilon GSTs and DDC were ligated into pGEM-T Easy vector (Promega) overnight at 4°C in a 10µl reaction mix. The 10µl ligation reaction mix contained 1µl (strong band) or 2µl (weak band) of PCR product, 5µl of 2X Rapid ligation buffer, 1µl of T4 DNA ligase (3 weiss units/µl), 0.5µl of pGEM-T Easy Vector (50ng) and distilled water.

The resultant ligation product was used to transform *Escherichia coli* competent cells (JM109). This was done by adding 25µl of competent cells to 1.5µl of the ligated product on ice. The reaction was left on ice for 25minutes and then heat-shocked at 42°C for 45sec. 450µl of SOC medium was then added and incubated at 37°C for 1.5 hours whilst shaking at 250 rpm. 50µl of the transformed reaction was then spread on Agar plates inoculated with 50 ug/ml of ampicillin and coated with 10mM isopropyl-beta-D-thiogalactopyranoside (IPTG) and 2% 5-bromo-4 chloro-3-indol-β-D-galactopyranoside (X-gal). The reaction was then incubated at 37°C overnight. White colonies were then screened by PCR using M13 forward and reverse primers (Stratagene) to determine colonies containing the full-length cDNA. The colonies containing the desired inserts were picked and grown overnight in 2ml LB inoculated with 2µl of 50ug/ml ampicillin at 37°C with shaking at 250rpm. The plasmids were purified using QIAprep spin miniprep kits (Qiagen) and then the plasmid concentration was determined using Nanodrop® (Nanodrop Technologies) according to manufacturer's instruction. Sequencing was performed after electrophoresis using an ABI 3100 sequencing machine at the Cardiff DNA sequencing service.

2.1.9 Sequence analysis

Sequences were multiple aligned using the Clustal W program within BioEdit software, version 5.0.9.1 (http://www.mbio.ncsu.edu/bioedit/page2.html) and then adjusted manually. Most interlocus, intraspecific and interspecific statistics were calculated using DnaSP, version 3.51 (Rozas and Rozas 1999; Rozas *et al.*, 2003). DNA polymorphism $\theta = 4Neu$ was estimated from the mean number of pairwise differences between sequences л, and from the number of segregating sites (θ_k). Phylogenetic analysis was performed with genetic distances using the neighbour-joining (Saitou and Nei, 1987) implemented in the MEGA version 3 (Kumar *et al.*, 2004).

2.2 MATERIALS AND METHODS II

Ecological implications of intraspecific differences in larval melanization within populations of *Anopheles gambiae s.l.*

2.2.1 Melanism induction in *An. gambiae s.l.*

2.2.1.1 *Rearing and handling of larvae*

The melanin induction experiment on *An. gambiae* was performed in the insectary of the Liverpool School of Tropical Medicine. Adult female susceptible inbred *An. gambiae* mosquitoes were bloodfed and females allowed to lay eggs in individual larval rearing trays in the insectary. The larvae from each individual adult mosquito constituted an iso-female line. This was to provide progeny that have a common genetic background prior to the induction. The larvae were fed every two days on Tetra Min Fish food.

Four days after hatching, when the larvae were sufficiently robust to withstand handling, up to 30 larvae from each iso-female line were transferred into one of three test containers (1000ml Polypropylene Beakers, VWR International: Catalogue No 212-9306). The three treatment groups were classified into DARK (those larvae reared in containers with dark background), WHITE (those larvae reared in containers with white background) and PARENTAL (those larvae reared in containers that had no covering). Each container contained 200 ml of distilled water. The outside surfaces of the transparent, colourless plastic containers described earlier

were covered with either white or black adhesive materials to provide the white and dark rearing background (Figure 7). The larvae were reared concurrently in the white, dark and normal containers under fluorescent lighting and the larvae were observed daily to assess larval pigmentation changes (Figure 22).

Initial trials showed there was a 1°C difference in temperature of the rearing water in the black and white containers so the experiment was run using a water bath as a holding tank for the rearing containers. This was to maintain an approximate temperature of 28.0°C and remove any comfounding effect of temperature on the development of larvae.

An attempt was made to maintain a constant pH in the containers by using distilled water for rearing the larvae and changing the rearing water every two days prevent any build up of mould in the containers. A detailed plan of the experimental set up is shown in Figure 8.

The larvae were examined every day under a dissection microscope (MEIJI) for signs of colour change. When larvae have reached the fourth instars, individuals are in each treatment regime were examined under a stereomicroscope (MEIJI) against a white backgrounds.

In order to investigate the genetic basis of the observed phenotypic, melanised pupae were allowed to emerge into adults. They were then fed and put in egg laying tubes for laying. The eggs laid were hatched into 1st instar larvae and reared in normal larval trays till they develop into 4th instar larvae. Individual larvae were then examined under the microscope for any visible changes in larval pigmentation. We obtained insufficient data from this aspect of the study and therefore we did not include it in the current analysis.

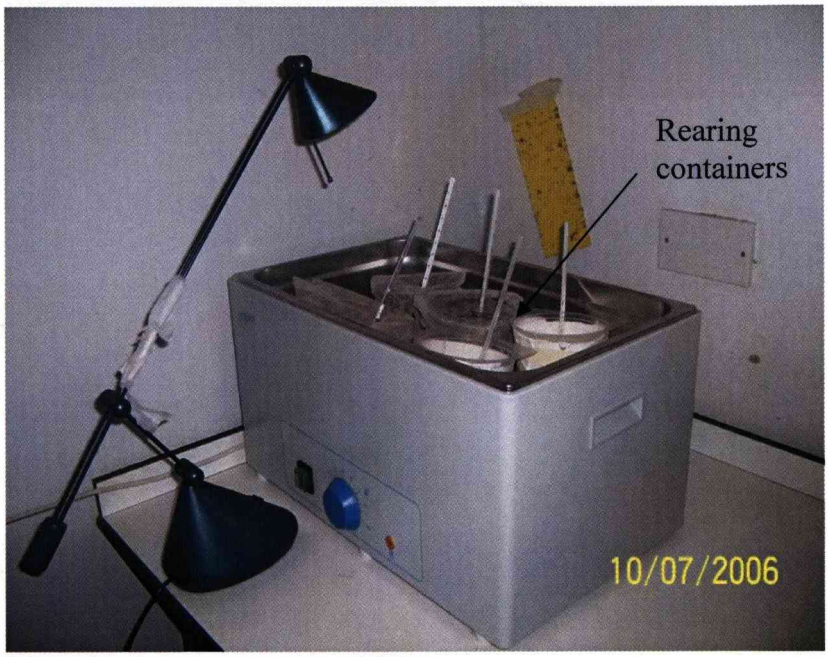


Figure 7: Picture of the experimental set up used for the selection of melanic, non-melanic and parental forms of *An. gambiae s.l.*

(The rearing containers were mounted inside a water bath filled with 2 liters of distilled water to control for temperature differences between the containers).
Rearing containers i.e DARK, WHITE and TRANSPARENT indicated by the arrow.

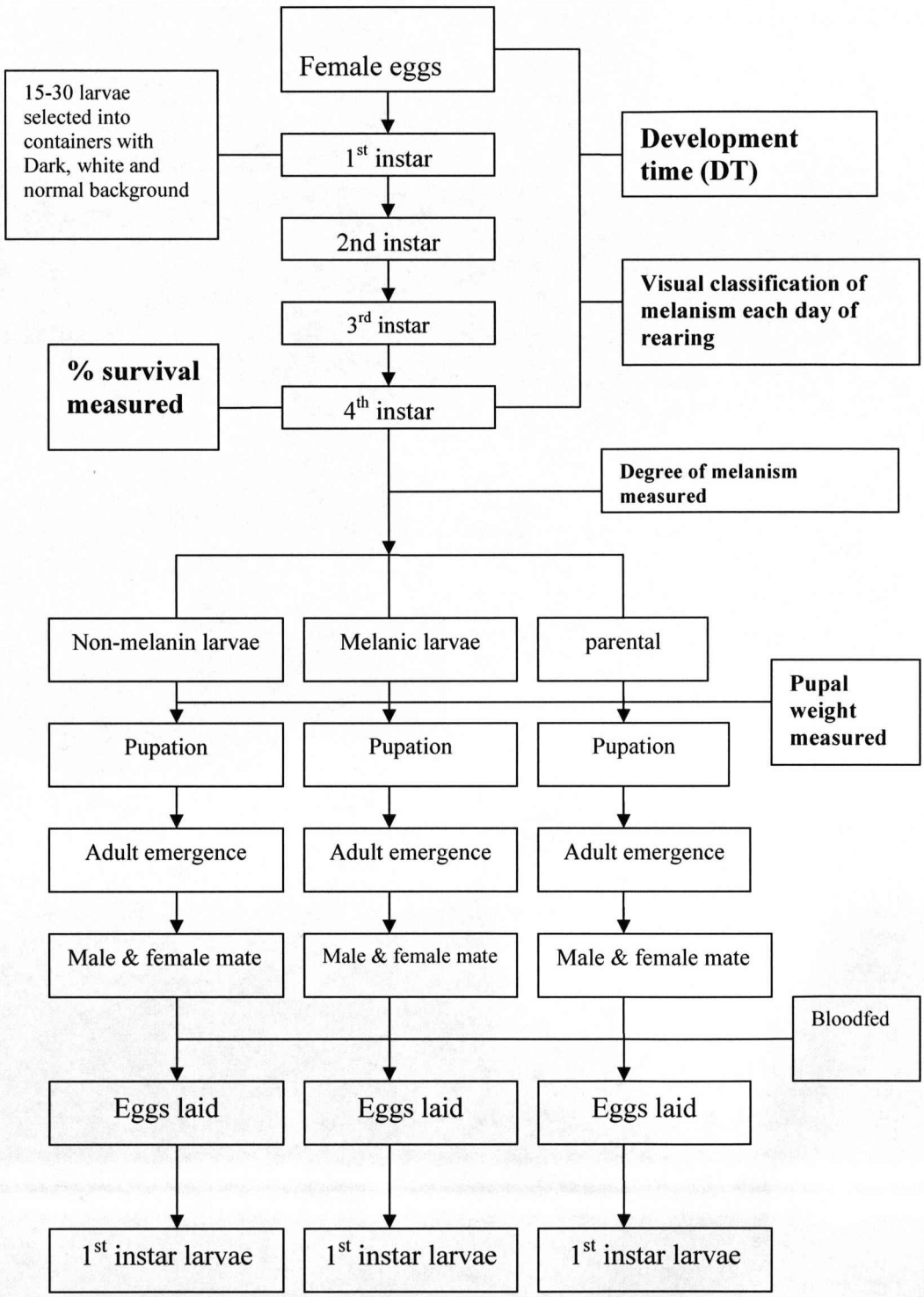


Figure 8: Flow chart of the experimental design used for the study of the effect of melanisation on the fitness of *An. gambiae s.l.* larvae.

2.2.2 Measurement of melanization

2.2.2.1 *Quantification of degree of melanism in larvae*

Photographs of all 4th instar larvae were taken against a white background using a Nikon Coolpix 4500 digital camera mounted on a MEIJI dissection microscope with an annular ring light source. Variation in larvae colour was measured on the head, thorax and abdomen (Figure 9). All the images were imported into Able Image Analyzer® software (Mu Labs Imaging) and the degree of larval melanisation was quantified on an average grey scale (scaled from 0, i.e. black, to 225. i.e white) of the areas shown in figure 9. Based on an initial subsample of 15 individual larvae, melanisation measurements were highly repeatable and the average grey value were averaged over the three measurements taken from each segment (Head/thorax, abdomen and final segment) of larval body. Initial estimated of average grey values measured on the head and thorax separately did not show any significant differences between the two segments. Subsequent estimates were therefore made on both the head and thorax together (Figure 9). The last segment (tail) was measured separately because of the attachments of numerous setae and the siphon which the larvae uses for gaseous exchange.

2.2.2.2 *Visual classification*

Fourth instars larvae were visually classified as DARK (appearance of black pigment above the dorsal thoracic and abdominal integument and/or setae and head capsule) and PALE (no signs of dark scales on the abdominal integuments or paleness of the cuticle). Figure 22 shows pictures of larvae visually classified into dark, pale and parental forms. All larvae were visually scored as melanic using the criteria described above and the frequency of melanic and pale forms estimated in percentage.

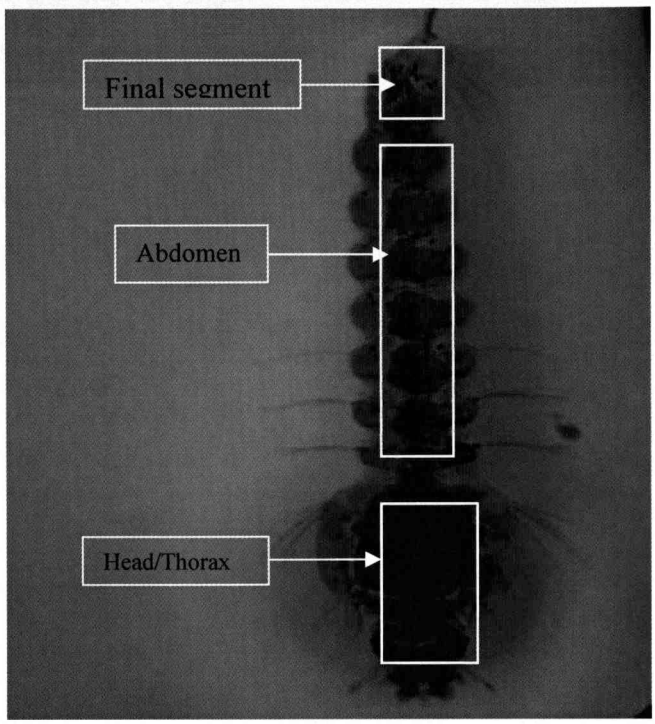


Figure 9: Measurement of melanization in *An. gambiae* larvae.

Dorsal view of larvae showing the regions of the body (Head/Thorax, abdomen and final segment) where measurements were made. Total area where average grey value were estimated is indicated by the rectangle

2.2.3 Measurement of life history traits

Larval development time of larvae was estimated as the time (days) from oviposition to entering the pupal stage. **Larval survival** was measured as percentage of larvae that survive up to on the average day 14 (black background rearing trays) and 17 (white background containers). **Pupal weight** was estimated as the average weight of 10 randomly selected pupae from each rearing trays.

2.2.4 Gene expression analysis

2.2.4.1 *Total RNA extraction and cDNA synthesis*

Total RNA was extracted from a pool of 10 larvae or pupae using the TRI reagent (SIGMA), according to the manufacturer's instructions. After the extraction, the total RNA was treated with DNase to remove any genomic DNA. The mRNA was then reverse transcribed into cDNA using superscript II (GIBCO BRL) and an oligo (dT) adapter primer (5'-GACTCGAGTCGACATCGA$(dT)_{17}$-3'). The PCR conditions for amplifying DDC and PO cDNAs were determined empirically for each gene. The DDC and PO standard plasmid contained 249-bp and 146-bp cDNA fragments respectively and this was amplified using primers for DDC and PO. Products of the expected size were subcloned into pGEM-T Easy vectors (Promega) and used as templates for sequencing. Sequencing reactions were performed using Beckman chemistry and the resultant products analysed on a Beckman CEQ800 capillary sequencer.

2.2.4.2 *Quantitative PCR*

Total RNA was reverse transcribed into cDNA as described above. Plasmids containing the gene of interest were diluted to produce seven standard templates at concentrations ranging from 1 ng/µl to 1 fg/µl. The incorporation of the fluorescent dye SYBR Green during PCR amplification of these templates was detected using a DNA Engine Opticon (MJ Research). For each experiment two replicates of each of the seven plasmid templates and two replicates of three cDNA samples from each form were used. A control plasmid containing a partial fragment of the S7 ribosomal protein gene was used to standardise the initial cDNA concentration in each sample.

CHAPTER 3 THE ROLE OF NATURAL SELECTION IN SHAPING NUCLEOTIDE VARIATION IN THE EPSILON CLASS GLUTATHIONE-S-TRANSFERASES GENE FAMILY IN MALARIA VECTOR, ANOPHELES ARABIENSIS (THEOBALD)

3.1 ABSTRACT

This chapter follows on from the observation of aberrant patterns of genetic differentiation among *An. arabiensis* samples at microsatellite locus 33C1 on the right arm of chromsome 3. This microsatellite locus showed a clinal change in modal allele classes in samples from East Africa. It was suggested that the pattern observed among the samples at locus 33C1 may be indicative of a low rate of mutation at the locus as a result of mutational constraints. This locus (33C1) is adjacent (~10-20kb) to a cluster of Glutathione-S-Transferase genes. Members of this gene family metabolise insecticides and other toxins and are therefore likely to be subject to strong selection. In this chapter we describe DNA sequence variation within and among four populations of *Anopheles arabiensis* at four Glutathione s-transferase genes (*GSTε1*, *GSTε2, GSTε6* and *GSTε8)*. We used the sequence data to investigate whether patterns of intraspecific and interspecific variation (used *An. stephensi* as outgroup) in *An. arabiensis* species will indicate any evidence of positive selection on GST loci surveyed. We detected marked sequence conservation of GSTs within the Cellia group (subgenus of anopheles Meigen) i.e *An. gambiae Giles* and *An.stephensi.* Based on the analysis, we postulate that the divergence of the GST epsilon family may predate the *An. gambiae* and *An. stephensi* split. GSTε2 locus showed a relatively high (5.9%) percentage sequence divergence than the rest of the loci. We detected evidence of a weak selection at three of the four loci and also show evidence of restricted gene flow between Ethiopia and the rest of the populations. Based on our findings, we infer that gene duplication events are important within the GST gene family in generating and maintaining adaptive phenotypes in natural populations of *An. arabiensis*.

3.2 INTRODUCTION

Across the genome of *An. gambiae*, the effect of random genetic drift on loci is similar whereas natural selection creates a marked difference in variation from purely neutral loci (Bonin *et al.,* 2006). These loci (outliers) affected directly by selection, are scattered throughout the genome and may be detected by their divergence from empirical neutral expectations (Luikart *et al.,* 2003). In contrast to what happens in the rest of the genome, these outliers have an atypical behaviour sometimes involving having an excess or a deficit of rare alleles in a given population or an aberrant pattern of genetic variability within and between populations eg. a higher differentiation between populations (Luikart *et al.,* 2003). In the context of local adaptation studies, such loci deserve particular attention because selection may be the underlying cause of their atypical behavior either because they are direct targets of selection or because they are genetically linked to a selected locus (Bonin *et al.,* 2006).

Therefore, scanning the patterns of DNA polymorphisms at the genomic level enables one to evaluate the amount of neutral genetic diversity and to identify "outlier" loci, i.e., loci that behave differently from the rest of the genome. So far, genome scans have been used to track outlier loci under selection especially in situations where individual adaptive traits segregate into two contrasting phenotypes (Bonin *et al.,* 2006). However, some local adaptations may arise often along environmental gradients and create an adaptive phenotypic continuum instead of discrete phenotypes.

Recently, inbreeding coefficients (F_{st}) have been used to detect outlier loci and use the distribution of estimates of F_{st} from individual genetic loci to detect the effects of natural selection (Storz and Nachman 2003; Storz and Dubach 2004; Storz, 2005; Beaumont, 2005; Guinand *et al.,* 2006).

A microsatellite based study of genetic differentiation between samples of *An. arabiensis* showed that locus 33C1, within 3Ra inversion was an outlier (Donnelly and Townson 2000) . This is because interpopulation values of F*st* measured at the locus were far in excess with a characteristic clinal change in modal allele classes in samples from the north and south of the East African region. The authors suggested

that because the locus is located within an inversion, the clinal change observed could reflect the increasing isolation with distance between populations studied or it could be a result of a clinal change in the inversion frequency, resulting from selective pressure upon genes within the inversion ((Donnelly and Townson 2000). The present study hypothesizes that the observed pattern and distribution of allele frequencies at microsatellite locus 33C1 may largely reflect differences in selective pressures exerted on this or nearby loci in the sample populations. Given that locus 33C1 is intronic and microsatellites are predicted to be selectively neutral it is likely that neighbouring loci are the ones subject to selection. Locus 33C1 is within the polymorphic 3Ra inversion on the right arm of chromosome together with, amongst other loci, a cluster of Glutathione s-Transferases (*GST*s) genes and the *Dopa decarboxylase* gene. Details on GST genes can be found in Chapter 1 whereas *dopa decarboxylase* locus is discussed in greater detail in Chapter 4. Genetic mapping showed that chromosome 3 division 33B adjacent to locus 33C1 contains a major locus controlling DDT resistance (Ranson *et al.,* 2000).

Interestingly, insecticides have been used intensively to control mosquito populations over the last 50 years and many species of insects have developed resistance to several families of insecticides. Previous study found that adult *An. arabiensis* Patton from Sudan were resistant to malathion and phenthoate but susceptible to all other organophosphates and the authors of the study suggested that a carboxylesterase enzyme may be responsible for malathion resistance in this strain (Hemingway, 1983). However, surveys of insecticide resistance carried out recently in Sudan on *An. arabiensis* colonies showed that the species exhibits the most widespread resistance in terms of its response to different classes of insecticides than any member of the *An. gambiae* complex including permethrin (Matambo *et al.,* 2007). The first report of *kdr* mutation was detected in samples of *An. arabiensis* from Burkina Faso (Diabate *et al.,* 2004). A study by Balkew *et al.* (2003) also established 43% DDT resistance in populations of *An. arabiensis* from Eastern Ethiopia. As part of the study by Matamboa *et al.* (2007), elevated glutathione s-transferase was detected in DDT-resistant strain of *An. arabiensis* as well as *kdr* (Leu-Phe mutation in the sodium channel gene). Of interest also is the fact that, whilst Kulkarni *et al.* (2006) identified the West African leucine-phenylalanine *kdr* mutation in two *An.arabiensis* heterozygous individuals from Tanzania and in

Uganda, Verhaeghen *et al.* (2006) detected the presence of both East-*kdr* and West-*kdr* mutations in *An. arabiensis* in the same geographical region. These findings may be a good evidence for the direct change in DDT resistance status of *An. arabiensis* and the wide geographic distribution of some of these *kdr* mutations. This pattern can also be explained by the balance between migration and selection at the population level. However, we know little about several important properties of this beneficial *kdr* mutations, including their mutational origin, their phenotypic effects and the frequency and rapidity with which they become fixed in a population. One signature of the spread of beneficial mutations is the reduction of heterozygosity at linked sites.

To characterize the signature of natural selection at the loci and to understand the levels and patterns of sequence variation at this locus, we surveyed variation in DNA sequences at four GST loci (GSTε1, GSTε2, GSTε6 and GSTε8) in four populations of *An. arabiensis* from East Africa. Comparing loci within the same gene family that have or have not been implicated in conferring resistance to insecticides is a powerful for approach for identifying the role of selection in shaping patterns of genetic variation in *An. arabiensis* populations. The ability to identify targets of selection using this approach will be critical in understanding the evolution of *An. arabiensis,* one of the major vectors of malaria. Perhaps, a more compelling reason for understanding the genetic mechanisms i.e. natural selection, genetic drift etc. responsible for adaptations of *An. arabiensis* at the individual level is that it can allow for prediction of future response to selection and rates of response., the combination of which could be important in improving our understanding of evolutionary principles underlying a range of biological functions (*eg* immune function, longevity *etc*) that have impacts on public-health.

3.3 MATERIALS AND METHODS

Sample collection sites

The mosquito samples i.e *An. arabiensis* used in this study were collected from Ethiopia, Sudan, Malawi and Tanzania in Africa. *An. stephensi* were collected from Bahewahay, Pakistan in South Asia. Samples from Ethiopia, Sudan, Tanzania and

Malawi used for the current study have been used in a previous study (Donnelly and Townson, 2000).

Sample collection, DNA extraction, species identification and locus selection are given in details in Chapter 2.

3.4 SEQUENCE ANALYSIS

Sequences were aligned using the Clustal W program within BioEdit software, version 5.0.9.1 (http://www.mbio.ncsu.edu/bioedit/page2.html) and then adjusted manually. Interlocus, intraspecific, interspecific analyses were calculated using DnaSP, version 3.51 (Rozas *et al.,* 2003) and relationships between haplotypes were illustrated using neighbor-joining trees (Saitou and Nei 1987) built by running MEGA version 2.1 (Kumar *et al.,* 1994) on Kimura's two-parameter distance (Kimura, 1980). The recombination parameter C = 2Nc, where N is the effective population size and c is the recombination rate per generation between the most distant sites was estimated using the methods of (Kaplan and Hudson 1985) and (Hudson,1987). The method was based on R_M or the minimum number of recombination events in the sample and the estimates of R_M were used to estimate C by coalescent simulations.

The decay of linkage disequilibrium with physical distance was estimated using nonlinear regression of linkage disequilibrium between polymorphic sites vrs the distance in base pairs between sites (Remington *et al.,* 2001). This analysis was done within populations at each locus. Linkage disequilibrium was scored between pairs of polymorphic sites using the squared allele frequency correlations, r^2 (Weir, 1979). An algorithm used for computing the statistic from DNA sequence data and for estimating its confidence intervals are implemented in DNAsP software. Details of other analytical methods are given in Chapter 2.

The *An. arabiensis* and *An. stephensi* DNA sequences used in this study were all obtained by sequencing products from amplified fragments of GST genes. GSTε2 of *An. stephensi* was obtained from the GeneBank (Accession No: AY573189). Data for *An. gambiae* were also obtained from GeneBank with corresponding accession numbers AY063776, AF316636, AY070256 and AY070257 for *GSTε1, GSTε2, GSTε6 and GSTε8* respectively.

3.5 RESULTS

3.5.1 Phylogenetic relationship between the GSTs of An. arabiensis and An. gambiae and An. stephensi

Figure 10 shows a phylogenetic tree illustrating the relationship between *An. arabiensis*, *An. gambiae* and *An. stephensi* GST sequences, based on a CLUSTAL W Kimura 2 parameter distance with 1000 bootstrap alignment. GST sequences of *An. gambiae* and *An. stephensi* were retrieved from the *An. gambiae* and *An. stephensi* genome database. The tree showed marked conservation of GSTs within the Cellia group and indicated that the divergence of the GST epsilon family predates the *An. gambiae* and *An. stephensi* split. The tree also showed a substantial divergence between GSTε2 in *An. stephensi* and two close relatives (*An.arabiensis* and *An. gambiae*) in 90% of 1000 replicates. Levels of divergence for GSTε2 (5.9%) far exceeded that for the other 3 genes, suggesting that this locus may have been subject to selection in one or both of the taxa.

Figures 11 and 12 shows phylogenetic trees illustrating the relationships between *An. arabiensis and An. gambiae* sequences. On the basis of the phylogeny, all GST haplotypes within *An. arabiensis* were relatively undifferentiated from *An. gambiae* although the distance between loci were all over 75% of 1000 bootstrap replicates (Figures 11 and 12). With the exception of GST haplotypes from Ethiopia which showed some degree of geographical isolation, all the GST haplotypes from Sudan, Malawi and Tanzania were to some extent geographically undifferentiated. Previous classifications have designated GSTs as being members of the same class if their amino acid sequences are more than 40% identical. By this criterion, GSTε8 of the *An. arabiensis* GSTs would be classified as not belonging to Epsilon class GSTs. This suggests that the present classification of insect GSTs into only two classes might need re-evaluating.

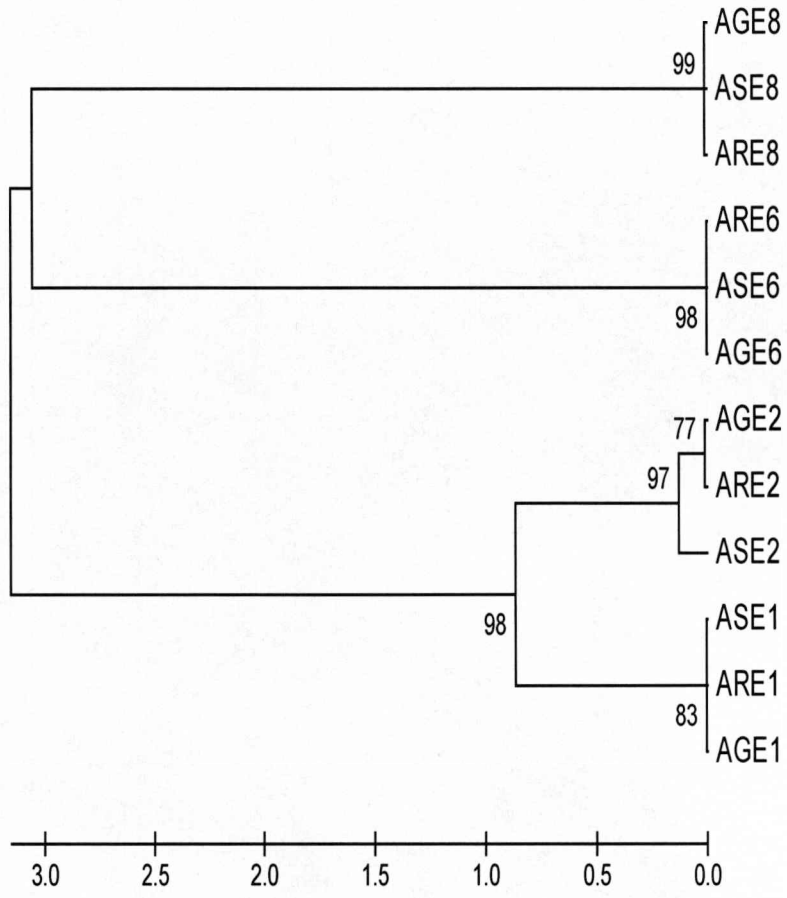


Figure 10: Neighbor-joining tree of the coding region of *GSTs* of *An. arabiensis*, *An. gambiae* and *An. stephensi* based on Kimura's two-parameter distance.

The number of bootstrap probability values based on 1000 replicates. The trees are based on total length of 298 bp. The *An. gambiae* sequences for *GSTε1*, *GSTε2* and *GSTε8* were obtained from the Genebank (AR=*An. arabiensis*, AG= *An.gambiae*, AS= *An. stephensi).*

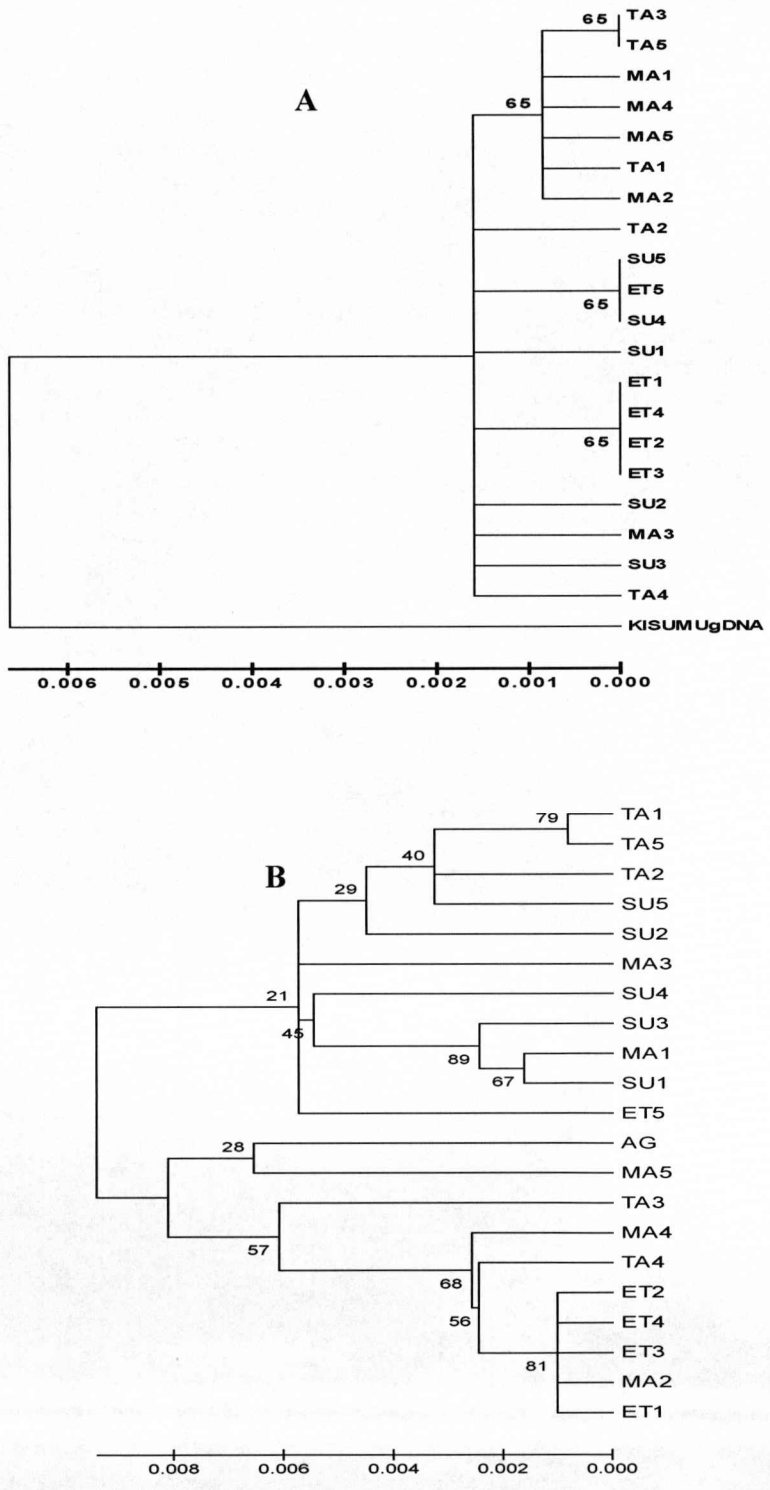
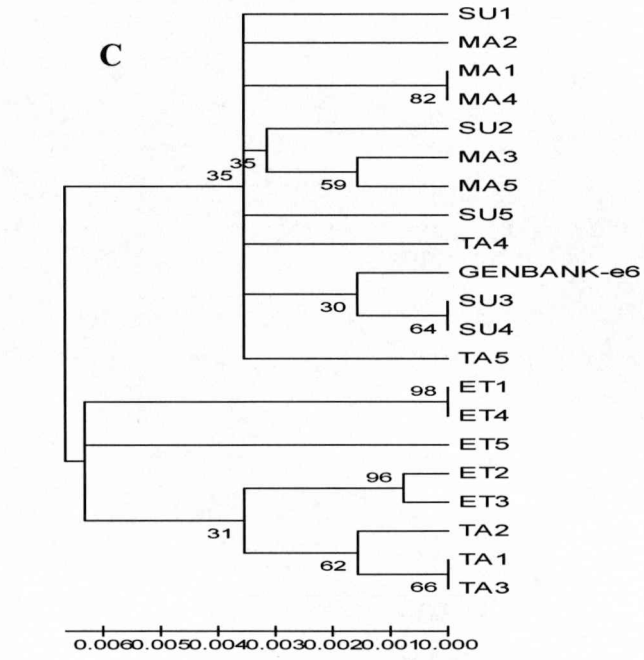


Figure 11: Neighbour-joining tree illustrating the relationship between the exonic regions of *An. arabiensis* and *An. gambiae* epsilon class A) GSTε1 and B) GSTε2 based on Kimura's two-parameter distance with bootstrap probability values based on 1000 replicates.

The trees are based on amplified GST fragments (GSTε1, 298bp; GSTε2, 500bp ; GSTε6, 644bp ; GSTε8, 651bp). The *An. gambiae* sequences for *GSTε1* and *GSTε2* were obtained from the Genebank (SU=Sudan, MA= Malawi, ET= Ethiopia, TA=Tanzania, *An.gambiae*=KISUMUgDNA, AG).



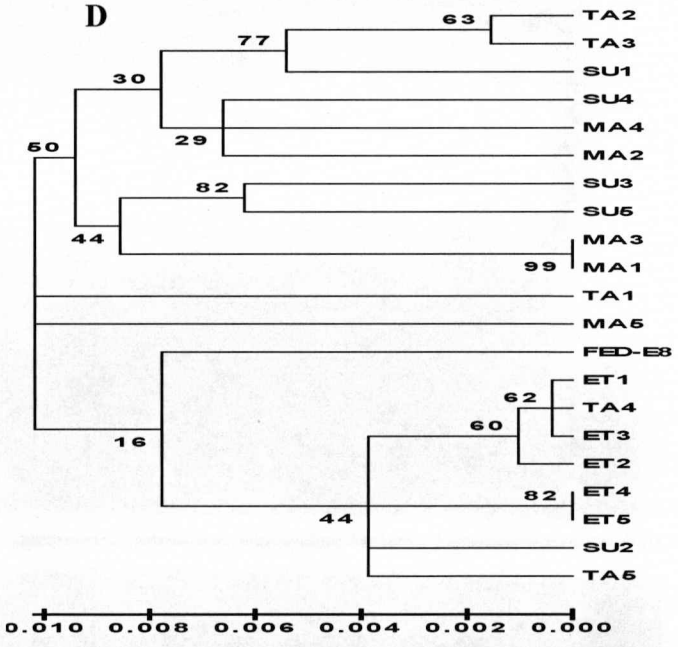


Figure 12: Neighbour-joining tree illustrating the relationship between the exonic region of *An. arabiensis* and *An. gambiae* epsilon class C) GSTε6 and D) GSTε8. based on Kimura's two-parameter distance with bootstrap probability values based on 1000 replicates.

The trees are based on amplified GST fragments. The *An. gambiae* sequences for *GSTε6* and *GSTε8* were obtained from the Genebank (SU=Sudan, MA= Malawi, ET= Ethiopia, TA=Tanzania, *An. gambiae*=GENBANK, AG

3.5.2 Tajima's D and Fu and Li's D test of selection

We tested whether the observed pattern of nucleotide variation is compatible with that expected under neutrality. We applied several tests that compare different estimates of θ either using only intraspecific data (Tajima,1989) or using intraspecific data and sequence information of another species (the outgroup) to determine the polarity of mutations (Fu and Li 1993). The HKA test (Hudson *et al.*,1987) was conducted to assess whether levels of polymorphism and divergence were correlated. We applied the tests first to the exonic region to determine any signature of selection on the coding region of the four loci. To further investigate the effect of selection on the genomic regions of the genes, we applied the test to the whole region of the genes (Table 2). Also because population structuring can affect the tests, the analysis was repeated on individual populations with a concomitant reduction in power (Table 3.). Negative values of Tajima D were obtained for all the loci i.e. GSTε1, GSTε2, GSTε6 and GSTε8 and none of the values approached significance (Table 2). Tajima's *D*, a statistic that measures the difference between two estimators of $4N\mu$ (π and Θ), is expected to be zero under a neutral model with constant population size (Tajima, 1989). Although, we obtained negative values of D and Fu & Li D, they were not significantly different at all GST loci of *An. arabiensis* and the pattern was the same for the corresponding estimates of Fu & Li D test (Table 2).

We repeated some of these analyses within populations (Table 3). The tests revealed positive values for Fu & Li* test for Ethiopia at GSTε8 locus and in 2 out of four loci in Tanzania populations (Table 3). Positive but not significant ($P>0.05$) values of Tajima's D values at the GSTε8 locus were estimated for the Ethiopia and Malawi populations. The MKA (Mcdonald and Kreitman 1991) test for loci

GSTε6 and GSTε8 in the Ethiopian sample showed P-values approaching significance ($p<0.05$). The MKA test on Ethiopian population could not be computed for locus GSTε2 because there were no synonymous and nonsynonymous polymorphic sites between the sequences.

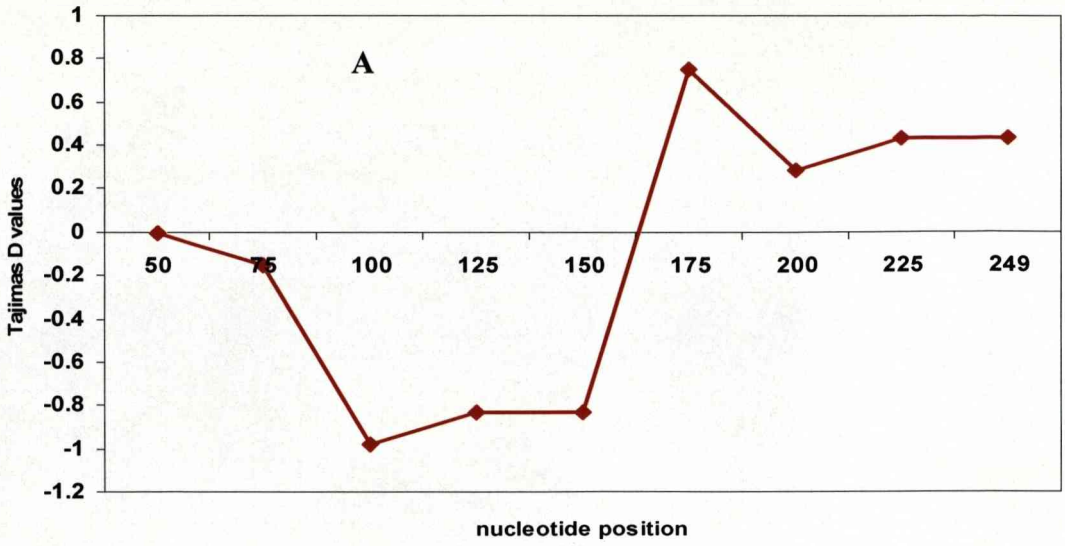
Locus /region	*S*	Nonsynonymous/ Synonymous mutations	π	Tajimas D	Fu & Li's D*
GSTε1(441bp) N=20					
Coding	6	-	0.005	-0.409 NS	-0.154 NS
overall	10	2 / 4	0.005	-0.878 NS	-0.966 NS
GSTε2 (800bp) N=20					
Coding	37	16 / 21	0.015	-0.535 NS	-1.232 NS
overall	53	16 / 21	0.017	-0.385 NS	-0.754 NS
GSTε6(788bp) N=20					
Coding	38	-	0.012	-1.375 NS	-1.448 NS
overall	46	-	0.012	-1.245 NS	-1.297 NS
GSTε8(796bp) N=20					
Coding	65	-	0.018	-1.521 NS	-1.384 NS
overall	83	-	0.022	-1.244 NS	-1.859 NS

Table 2: Summary of nucleotide polymorphism and tests of selection in the coding and noncoding regions of GST loci in *An. arabiensis*

Tajimas D and Fu and Li's D* test used the total number of mutations, rather than the number of segregation sites to estimate theta because the latter does take into account several instances of multiple mutations at the same sites. NS indicates a nonsignificantly negative value of a given index ($P<0.05$). S, number of segregation sites (number of mutations π, average pair wise nucleotide diversity; D, Tajimas D value)

3.5.3 Sliding-window analysis of coding region of GST loci

Because selection may be localized within the gene and the tests of selection are not sufficiently sensitive to detect, we applied a sliding window analysis to the data (with a window size of 100bp and a slide of 25bp). Although, whole estimates of Tajima's D values did not reveal any significant deviations from neutrality, the sliding window plot of the coding regions of the four loci revealed areas of significantly ($P<0.10$) low Tajima D. This values deviations from neutral expectations at the GSTε2, GSTε6 and GSTε8 loci (Figure 13 and 14). We also applied the Fu and Li D* sliding window analysis and the pattern was qualitatively the same. Overall, the sliding window analysis suggests that polymorphism distribution in three GST genes; GSTε2, GSTε6 and GSTε8 significantly deviate from the expectations of neutrality at specific regions within the genes (indicated by the circles in Figures 13 and 14).



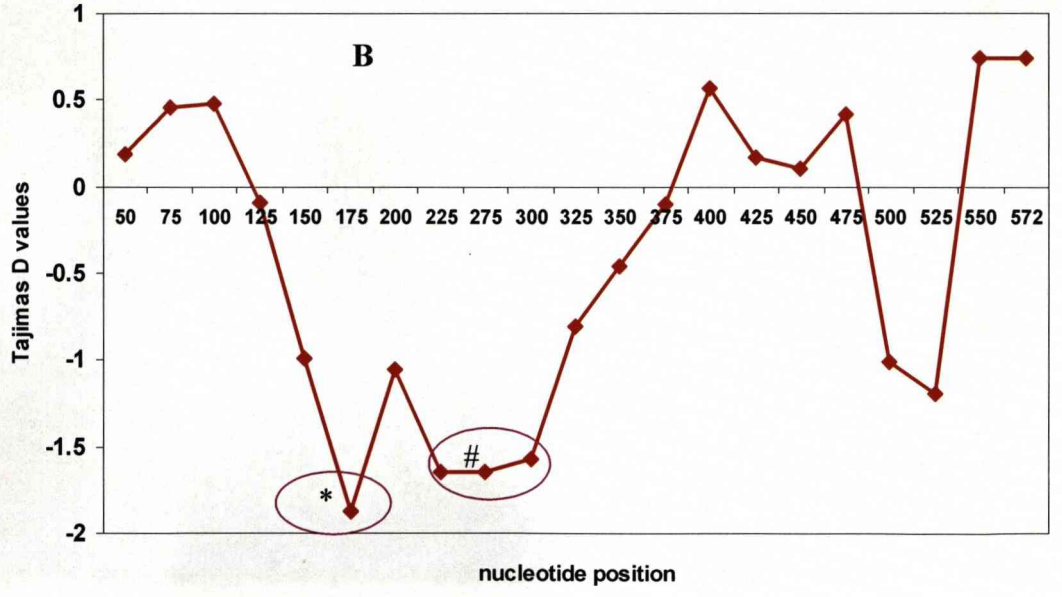


Figure 13: Sliding window plot of the protein coding region of A) GSTε1 and B) GSTε2 sequenced from *An. arabiensis* populations

Region indicated by the # are significant at P<0.105 respectively

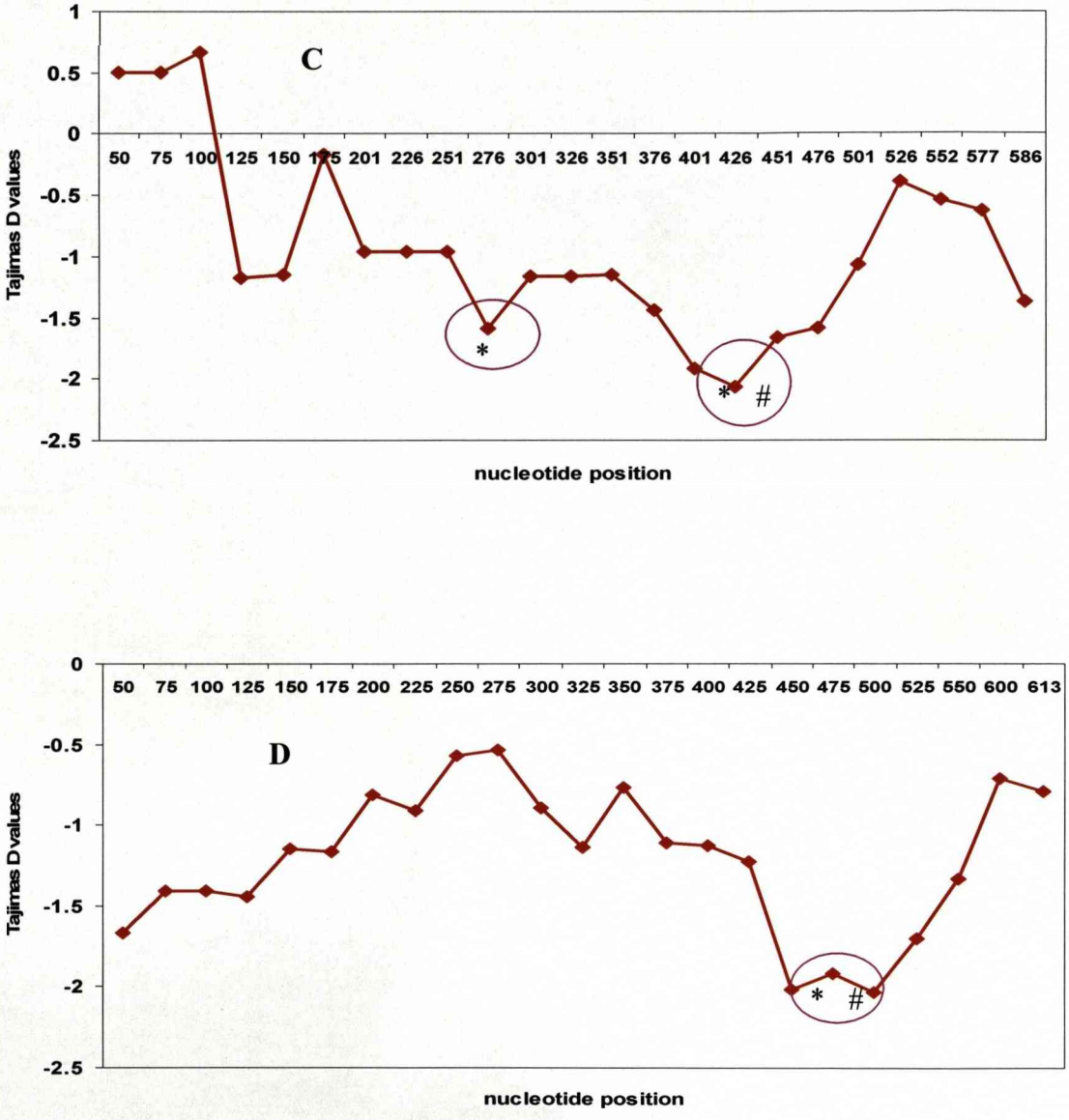


Figure 14: Sliding window plot of the protein coding region of C) GSTε6 and D) GSTε8 sequenced from *An. arabiensis* populations

Region indicated by the # and * are significant at P<0.10, P<0.05 respectively

Locus/population		N	*S*	π	$D(T)$	Fu & Li*	HKA P-value	MK (P-value)	Average pairwise sequence Divergence (%)
GSTε1 (N=20)									1.0
	Ethiopia	5	2	0.003	-0.973	-0.973	ID	1.00	
	Malawi	5	2	0.003	-0.973	-0.973	0.37	1.00	
	Sudan	5	2	0.003	0.243	0.243	0.58	1.00	
	Tanzania	5	2	0.004	1.459	1.459	0.60	0.33	
GSTε2 (N=20)									5.9
	Ethiopia	5	9	0.006	-1.184	-1.184	ID	ID	
	Malawi	5	23	0.018	-0.027	-0.027	0.37	0.76	
	Sudan	5	12	0.010	0.301	0.301	0.58	0.53	
	Tanzania	5	16	0.013	0.496	0.496	0.60	1.00	
GSTε6 (N=20)									2.0
	Ethiopia	5	15	0.012	-0.267	-0.076	0.70	0.065	
	Malawi	5	9	0.007	-0.197	-0.197	0.59	0.400	
	Sudan	5	13	0.008	-0.977	-0.977	0.35	0.099	
	Tanzania	5	9	0.006	-0.526	-0.526	0.80	1.000	
GSTε8 (N=20)									1.6
T	Ethiopia	5	4	0.003	0.957	0.957	0.15	0.119	
	Malawi	5	30	0.022	0.155	0.207	0.63	1.000	
	Sudan	5	31	0.021	-0.644	-0.644	0.75	0.319	
	Tanzania	5	19	0.013	-0.421	-0.258	0.72	1.000	

Table 3: Summary statistics of intra population and inter-species variation at the coding region of GST loci in *An. arabiensis* populations from Ethiopia, Sudan, Malawi and Tanzania

Tajimas D test and Fu and Li's D and F tests used the total number of mutations, rather than the number of segregating sites were used to estimate theta because the latter does take into account several instances of multiple mutations at the same sites. * indicates a significantly positive value of a given index ($P<0.05$). S, number of segregation sites (number of mutations) ; Π, average pairwise nucleotide diversity; $D(T)$, Tajimas D value; D(F&L), Fu and Li's D* value; HKA was done considering all substitutions; MK (Mcdonald and Kreitman,1991) was done considering for both synonymous and non-synonymous substitutions. *An. stephensi* was used as outgroup for the interspecies comparison, ID = statistics could not be determined

3.5.4 Interspecific analysis of nucleotide polymorphism between *An. arabiensis* populations

3.5.4.1 *Genetic differentiation and gene flow between An. arabiensis populations from Ethiopia, Sudan, Tanzania and Malawi*

The locus-by-locus pairwise estimates of population differentiation, F*st* (Hudson *et al.,* 1992) are shown in Table 4. The estimates showed that pairwise F_{ST} comparisons involving Ethiopian population (ET/SU, ET/MW and ET/TZ) for GSTε1 and GSTε8 were significantly greater that zero ($P<0.05$; permutation test; range (0.306 - 0.600) Table 4. This pattern of genetic differentiation between populations at all the GST loci is reflected in the neighbour-joining trees (Figures 11 and 12). The F_{ST} values estimated in this study are far in excess of those observed from microsatellite and mitochondrial (mtDNA) sequence based studies of the same populations, Table 4. This may be highly suggestive of differential selection pressures acting upon one or all of these genes in Ethiopian populations of *An. arabiensis*. It may also reflect the effect of homoplasy on microsatellite markers and how estimates of differentiation are constrained by heterozygosity (Estoup *et al.,* 2002).

We computed the average level of gene flow, assuming the island model of population structure (Wright, 1946). Estimates of gene flow, Nm within and between *An. arabiensis* across the four study sites at GSTε1, GSTε2, GSTε6 and GSTε8 were Nm= 0.15, 0.36, 0.21 and 0.53 respectively (Table 5).

Population comparison	loci					
	GSTε1	*GSTε2*	*GSTε6*	*GSTε8*	microsatellites	ND5 mtDNA
ET / SU	0.467*	0.599**	0.392*	0.388*	0.0175^{o}	0.01425^{+}
ET / MW	0.600*	0.251	0.323**	0.388*	0.0459$^{\#}$/0.0383^{o}	0.10884^{+}
ET / TZ	0.533*	0.348*	0.136	0.306*	0.0378^{o}	-
SU / MA	0.350	0.081	0.357	0.084	0.0292^{o}	0.07172^{+}
SU / TZ	0.250	0.169*	0.309	0.016	0.0212^{o}	-
MA / TZ	0.028	0.038	0.239	0.144	0.0313^{o}	-

Table 4: Estimates of genetic differentiation (F_{st}) between *An. arabiensis* populations from Ethiopia, Sudan, Tanzania and Malawi

*P<0.05, **P<0.005, Fst estimates in Donnelly & Townson 2000^{o}, Donnelly *et al.,* 2004^{+}, Donnelly *et al.,* 2001$^{\#}$, ET=Ethiopia, SU=Sudan, MW=Malawi, TZ=Tanzania

Populations	*GSTε1*	*GSTε2*	*GSTε6*	*GSTε8*
	Nm	Nm	Nm	Nm
ET-SU	0.06	0.22	0.32	0.48
ET-MA	0.14	0.59	0.33	0.48
ET-TZ	0.10	0.25	0.50	0.52
SU-MA	0.50	1.21	0.32	1.10
SU-TZ	0.50	1.04	0.24	2.21
MA-TZ	0.83	1.05	0.35	0.96
Average	0.15	0.36	0.21	0.53

Table 5: Gene flow estimated (Nm) between *An. arabiensis* population captured from Ethiopia, Sudan, Malawi and Tanzania

Nm (Nei, 1982) values were estimated from sequence data information, (ET-Ethiopia, SU-Sudan, MA-Malawi, TZ-Tanzania)

3.5.4.2 Estimates of linkage disequilibrium and recombination rates

To obtain a high power of detection, recombination rates and linkage disequilibrium was estimated from the whole set of sequences for the four *GST* loci. Using the four gamete test of Hudson and Kaplan (1985), all loci except *GSTε1* showed marked recombination indicated by the percentage of pairs of sites where four gametes were found i.e. *GSTε2* (*11.2%*), *GSTε6* (*7%*) and *GSTε8* (*3%*). The minimum recombination events, Rm (Hudson and Kaplan 1985) needed to explain these data were 10, 5 and 5 for *GSTε2*, *GSTε6* and *GSTε8* respectively with corresponding population recombination parameters.

Within *GSTε1*, 1 out of 45 pairwise comparisons (2.2%) between nonsingleton pairs of polymorphisms show statistically significant linkage disequilibrium ($p<0.05$) by the Fisher exact test; although after a Bonferroni correction for multiple comparisons, this was not significant. *GSTε2* and *GSTε6* showed 4.1% and 1.5% significant associations respectively but no association after Bonferroni correction. For *GSTε8*, 99 out of 3081 pairwise comparisons (32.1.6%) between showed statistically significant linkage disequilibrium by the Fisher exact test; after the Bonferroni correction for multiple comparisons, 4 (0.1%) were still significant.

Figures 15 and 16 show the nonlinear relationship between linkage disequilibrium (LD) and nucleotide distance in the specimen analysed. With the exception of *GSTε1* and *GSTε2*, all the other loci showed that LD decays with distance (Figure 16C-D). It should be noted that estimates of within-population rate of decay of LD are subject to much larger standard errors, due to the smaller number of sites that were polymorphic within populations. Despite the rapid decline of LD, several sites in *GSTε2* and *GSTε8* showed linkage disequilibrium over distances that approach the length of the sequenced region.

Locus	R_M	4*Nc*
GSTε1	0	0.0260
GSTε2	10	0.1629
GSTε6	5	0.1530
GSTε8	5	0.1644

Table 6: Estimates of recombination parameter for all the four GST loci sequenced from samples of *An. arabiensis* populations

Estimate of the population recombination parameter, C=4*N*c (Hudson, 1987) from the minimum number of recombination events, R_M (Hudson and Kaplan 1985). N=the effective population size and c=mutation rate

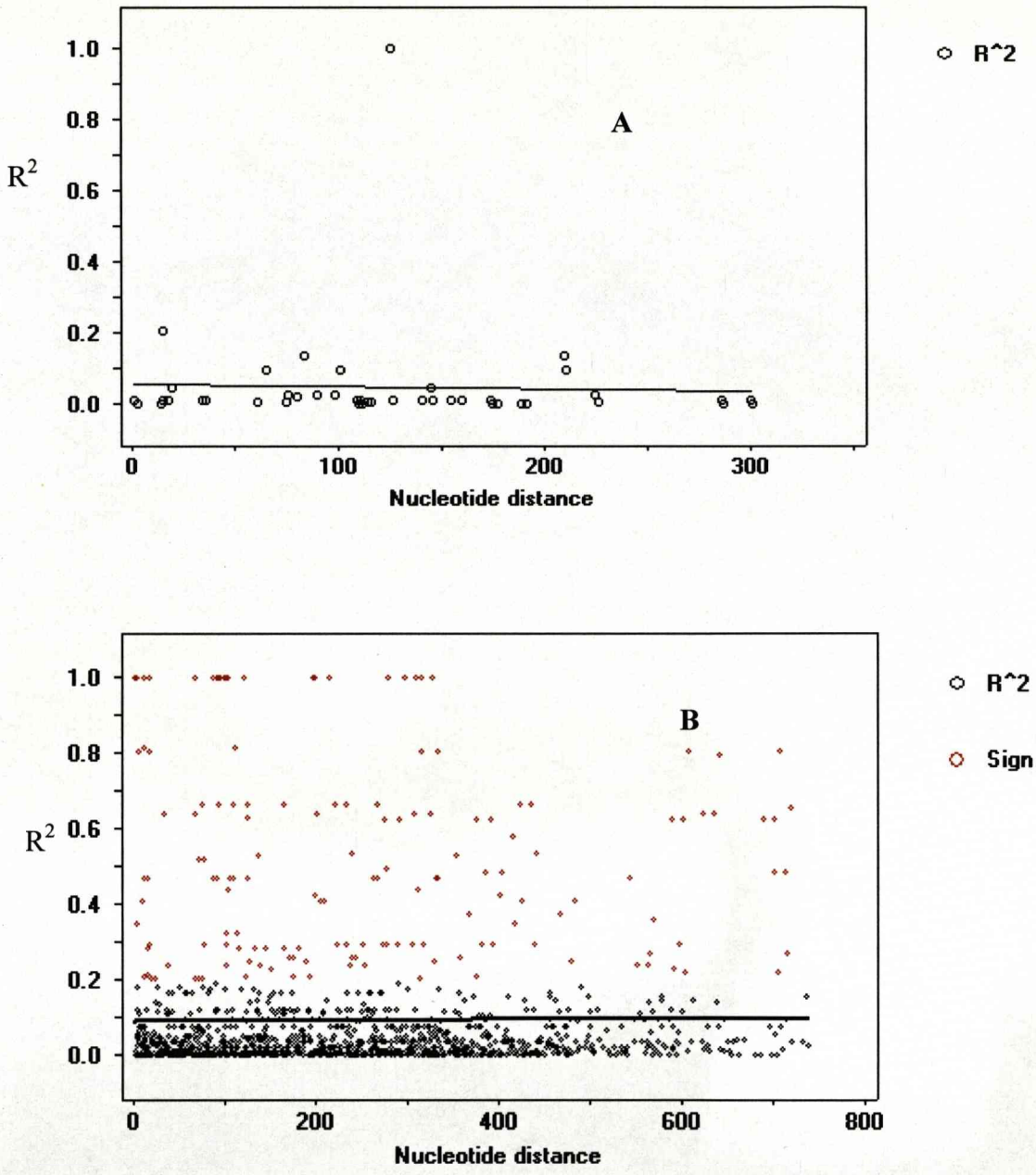


Figure 15: Plots showing the squared correlations of allele frequencies (r^2) as a function of physical distance between sites for two GST genes (A) GSTε1 and (B) GSTε2 in *An. arabiensis.*

Thin lines depict within-population decline in linkage disequilibrim. The red dots denote areas of significant linkage disequilibrium.

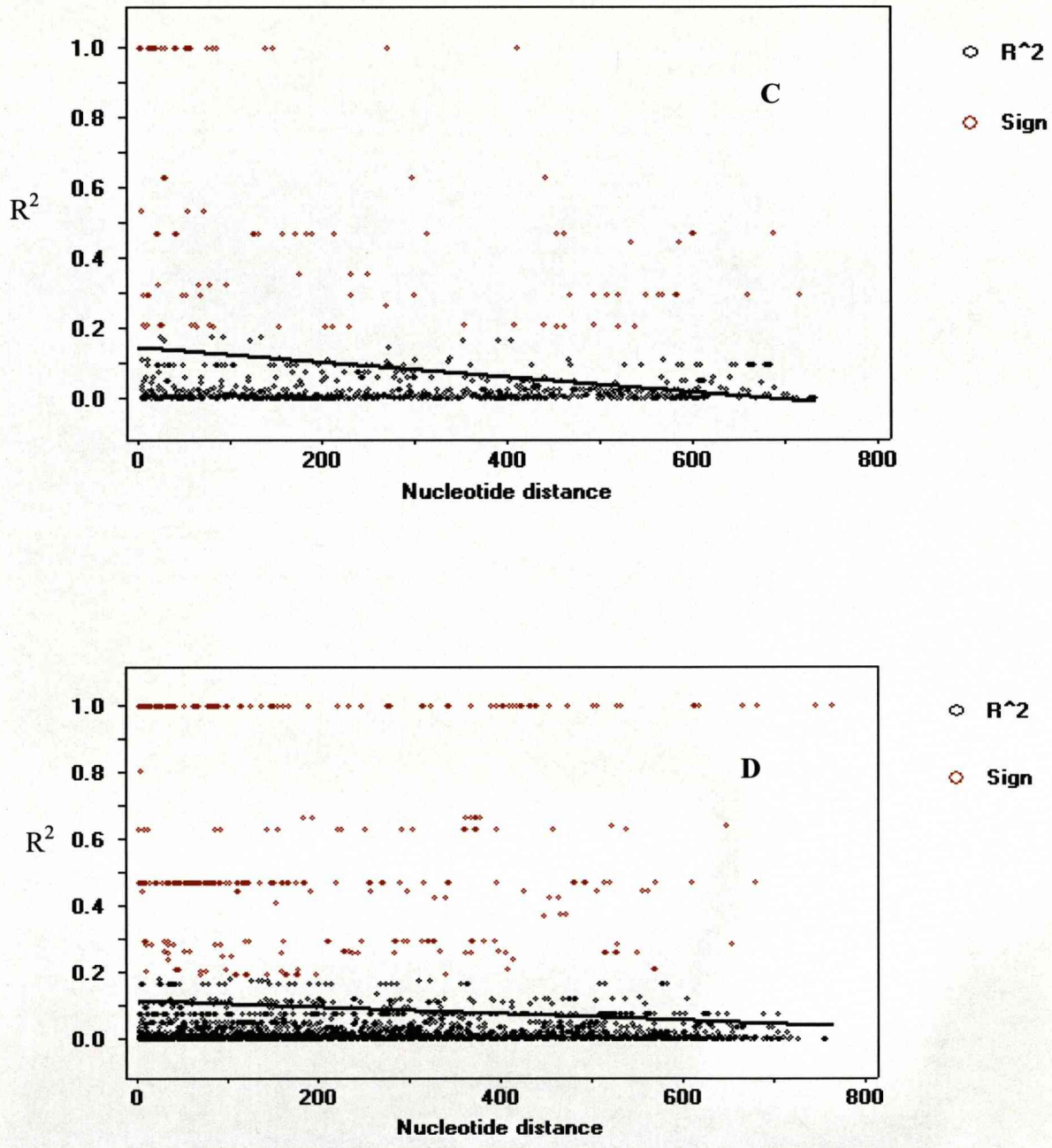


Figure 16: Plots showing the squared correlations of allele frequencies (r^2) as a function of physical distance between sites for two GST genes, C) GSTε6 and (D) GSTε8 in *An. arabiensis.*

Thin lines depict within-population decline in linkage disequilibrim. The red dots denote areas of significant linkage disequilibrium.

3.6 DISCUSSION

3.6.1 Phylogenetic lineage between GSTs within the Cellia group

The present study provides evidence of marked conservation within the GST sequences in *Anopheles* from within the Cellia group *An. gambiae, An. arabiensis* and *An. stephensi*. This implies that gene duplication events that resulted in the cluster of GST genes on chromosome 3R are likely to predate the division of Cellia from other anopheline sub-groupings i.e apparent radiation occurred before Cellia split about 90 – 106 million years ago (Krzywinski *et al.,* 2001; Della Torre *et al.,* 2002). Using the white gene as a marker, Krzywinski *et al.* (2001) suggested that there is a close relationship between members of the *Anopheles* subgenera Cellia and our results to some extent is in agreement with that assession. Although, we did not produce sequence data on the *Anopheles* subgenus for comparison during this particular study, it will be interesting to extend the study to *Anopheles* to confirm our findings. The study detected no obvious associations between haplotypes and geographic locations as evident in the phylogenetic analysis.

Surprisingly, with the exception of Ethiopia samples which seem to be grouped into clusters at all loci, samples of *An. arabiensis* from Sudan, Tanzania and Malawi showed little differentiation. Using other markers such as mitochondrial DNA (mtDNA), similar patterns have been observed in *An. arabiensis* in East Africa (Besansky *et al.,* 1997; Donnelly and Townson 2004) and in *An. gambiae* and these pattern were thought to reflect similarities in ecological zones and the absence of topographic barriers to gene flow (Lehmann *et al.,* 2003). Some workers had detected extensive mtDNA haplotype sharing between *An. gambiae* and *An. arabiensis* across Africa (Besansky *et al.,* 1997; Donnelly *et al.,* 2004).

This study showed over 98% sequence similarity between *An. gambiae* and *An. arabiensis* in the coding region of all four loci. With the exception of locus GSTε2 which showed 86% sequence identity between *An. gambiae* and *An. stephensi,* all the other three loci, GSTε1, GSTε6, GSTε8 showed between 97-98% identities. It is now known that GSTs share sequence and structural similarities with several stress-related

proteins in a wide range of organisms and this allows GSTs to be used to identify patterns of divergence in many organisms (Rossjohn *et al.,* 1996).

Insect GSTs have been shown to be orthologous to the Sigma GST class found in a diverse range of species from nematodes to mammals (Agianian *et al.,* 2003). Epsilon GST classes are also known to have expanded independently in *D. melanogaster, Ae. eagypti* and *An. gambiae*, suggesting that the enzymes play important roles in the adaptation of the species to their specific environments (Ranson *et al.,* 20002b; Lumjuan *et al.,* 2007).

GSTε8 locus shares very little similarity with the other three GST loci. The classification of GSTε8 within the Epsilon cluster has previously been described ambiguous because it has less than 40% identity with other members of the Epsilon class, the cut off widely used to classify *GST*s (Hayes and Pulford 1995). However, protein products of this gene was found to react with anti-sera raised against Epsilon GST (Ding *et al.,* 2003) and also a BLAST search using putative translation of GSTε8 as query identified *P. xyloslla* Epsilon GSTs as the characterized protein (Altschul *et al.,* 1997). This study found that GSTε8 in *An. arabiensis* had between 6-13% identities to other *GST*'s in the Epsilon class, much less than what was reported by Hayes and Pulford 1995 and Ortelli *et al.,* 2003. Using this criterion (cut off of 40% identity), we postulate that GSTε8 may belong to other class of GSTs. More research on the classification of GSTε8 is needed in line with the evidence from this study.

3.6.2 Intraspecific nucleotide diversity of GSTs within An. arabiensis populations

Intraspecific analyses revealed a significant non-neutral pattern of diversity at three of the loci. The nucleotide variability across three of the loci i.e. GSTε2, GSTε6 and *GSTε8* was relatively high compared to that of locus GSTε1. Highly negative estimates of Tajimas D and Fu and Li's D* for GSTε2, GSTε6 and GSTε8 loci were observed but they failed to reach significance. The relatively low numbers of individuals sampled would have reduced the power of the two tests to detect significance. However there were regions within each gene (GSTε2, GSTε6 and GSTε8) which were associated with non-synonymous changes and this was revealed by less conservative sliding window plots of Tajima's D statistic. More non-synonymous mutations were found in 3 out of the 4 populations in the domain of the

GSTε2 and GSTε8 loci and sequence divergence estimated for all the loci showed GSTε2 to be highly divergent, 5.9% compared to the other loci (Table 3). Elevated GST activity has been associated with resistance to all the major classes of insecticides (Prapanthadara and Ketterman 1993; Wheelock *et al.,* 2005; Vontas *et al.,* 2001; Enayati *et al.,* 2001). GSTε2 specifically has been found to be over expressed in DDT resistant mosquitoes as it encodes an enzyme that is effective at catalyzing the dehydrogenation of DDT insecticide (Ranson *et al.,* 2000a; Ortelli *et al.,* 2003; David *et al.,* 2005).

The observed nucleotide variability within GSTs may be adaptive in the context of insecticide usage and the data strongly suggest that the variations observed within the GSTε2, GSTε6 and GSTε8 loci may be shaped by purifying selection which is consistent with most metabolic enzymes (Harwood, 1988). The signature of positive selection seen only in Ethiopia population suggests that GSTε6 may possess specific alleles that have conferred local adaptation to some environmental pressure. The significance of the Great Rift Valley as barrier to gene flow in this case cannot be overlooked as postulated by Lehmann *et al.* (1999). Also there was a 12 base pair nucleotide insertion in the intronic region of GSTε2 and this insertion was only found in the Ethiopia samples. Ranson *et al.* (2003) had previously shown that there is considerable conservation of intron positions within the epsilon and delta class of *GST*s and therefore further research is needed to determine the evolutionary significance of the insertion in these samples.

3.6.3 Interpopulation variation and differentiation within and between species of An. arabiensis

Inter-population analysis with all the GST loci showed different patterns of nucleotide variation and differentiation especially between Ethiopia and the rest of the populations. The relatively high F_{ST} values for all loci especially for the GSTε2 are in contradiction with neutral expectations, suggesting diversifying and/or local selection on the between populations. The F_{ST} values estimated for *An. arabiensis* are much greater than what was estimated by Donnelly and Townson (2000). The authors of explained that the large between-locus variation in values observed in their study may have been caused by selection upon some of the loci; possibly a result of linkage effects with nearby coding regions or chromosomal inversions. Based on the non-

uniform pattern of F_{ST} estimates across GST loci, we can infer that the GST gene family may be under some selective constraint in *An. arabiensis.*

3.6.4 Recombination and linkage disequilibrium at GST loci

The study detected that recombination is frequent within the 3Ra inversion of chromsome 3 of *An. arabiensis.* This was evident in the recombination rates C= (4*N*c) estimated for 3 of the GST loci (Range: 4*N*c= 0.1530 -1644) Table 6. This pattern was also reflected in the minimum number of recombination events/generation, R_M estimated for the three loci (Table 6). A plot of linkage disequilibrium (LD) against nucleotide distance (Figures 15 and 16) showed a reduction in LD with nucleotide distance especially within loci GSTε6 and GSTε8. Another mtDNA based analysis and also detected high linkage disequilibrium within population of *An. arabiensis* in Kenya (Donnelly *et al.,* 2001). It has been suggested that there could be marked heterogeneity in recombination rates across the genome of *An. gambiae* (Black *et al.,* 2008). Interestingly we detected recombination within GST genes located within 3Ra inversion and we postulate that recombination, together with selection may be responsible for the pattern of nucleotide variation observed at the GST loci in this study. The role of chromosomal inversions in determining the pattern of nucleotide polymorphisms in the genome is also worth noting. Inversions have been shown to have effects on recombination rates (Andolfatto, 2001) and hence modulate nucleotide variability in a complex way. (Navarro *et al.,* 1997) postulated that inversions strongly influence nucleotide polymorphisms by reducing and redistributing recombination in chromosomal variants. Inversions are common in *An. gambiae,* they are thought to confer selective advantages under different environmental conditions (Toure *et al.,* 1994). However, despite the potential impact of inversions on nucleotide variability; they have been overlooked in nucleotide variation surveys (Andolfatto, 2001). The frequency of *An. gambiae* 3Ra inversion in Ethiopia is high, 0-35% (Petrarca and Beier 1992; Abose, 1998) whilst that in Sudan is 13% (Petrarca, 2000). Further research is needed to elucidate the effect of inversions on nucleotide variability in *An. gambiae* since this may help explain the pattern of population differentiation and recombination rates observed in this study.

Linkage disequilibrium is important for detecting valid genotype-phenotype associations i.e linkage mapping depends on the markers under study being in disequilibrium with the genes that actually conditions the expression of the phenotypes (Black *et al.,* 2008). Recent work has shown that inversions may facilitate speciation by creating linkage disequilibrium between these genes (Noor *et al.,* 2001; Rieseberg, 2001; Navarro and Barton 2003). Linkage mapping studies for insecticide reistance in *An. gambiae* is currently being pursued and the extensive disequilibrium of segregating sites observed at the GST loci will make them good candidates for Linkage mapping. GSTs are known for detoxifying insecticides (Ranson *et al.,* 2000a). However, marked recombination observed at the loci may reduce LD across and these needs to be considered in future when doing linkage mapping studies.

3.6.5 Gene flow estimates across An. arabiensis populations

There was variation in the values of gene flow, Nm (Nei, 1982) averaged across all the loci. Gene flow estimates between Sudan, Malawi and Tanzania *An. arabiensis* populations were relatively high especially at the GSTε2 locus (Nm > ~1.0) compared to the gene flow values from Ethiopia populations, (Nm < ~1.0) for each of the four loci (Table 5). This is consistent with findings from the phylogenetic analysis and genetic differentiation, i.e. F*st* estimates. This pattern of genetic differentiation could be explained by the fact that Ethiopia population may have an independent origin and hence may be geographically isolated from the rest of the population.

In comparison, Lehmann *et al.* (1996) obtained much significantly higher values of gene flow between populations of *An. gambiae* from East and West Africa separated by a distance of approximately 6000 km (*Nm* 7.7 and 3.4 for *Fst* and *Rst* respectively). Using microsatellite and mitochondrial DNA based analysis, another study found that the Great Rift Valley in East Africa is an important gene flow barrier for *An. gambiae* (Kamau *et al.,* 1998, 1999). Again, microsatellite analysis of samples from nine localities along a 4500 km transect from Sudan to Mozambique revealed highly significant differences in genotype frequencies between populations separated by more than 200 km with extensive barriers to gene flow in this region (Donnelly and Townson 2000). Both *An. gambiae* and *An. arabiensis* are known to have recently expanded its range into new habitats (Donnelly *et al.,* 2001). Based on our findings,

we postulate that the Ethiopian population of *An. arabiensis* may be undergoing local adaptation in response to selection and the local selection may be responsible for the pattern of differentiation observed. The selection could be due to insecticide pressure as result of their continuous use for malaria control as well as for agricultural purposes or some other source of selective pressure.

CHAPTER FOUR MOLECULAR EVOLUTION OF A DOPA DECABOXYLASE GENE IN POPULATION OF *AN. ARABIENSIS* FROM SUDAN

4.1 ABSTRACT

For insects, several environmental factors including temperature influence an individual's ability to adapt to various environments. *An. arabiensis* is a cosmopolitan species that has had great success in adapting and colonizing dry environments. This adaptation may have resulted in complex pattern of genetic variation. Identifying genes controlling adaptive genotypes in natural population of *An. arabiensis* is crucial for understanding the evolutionary history of the species. In this study, we examine DNA sequence variation among specimens of *An. arabiensis* from Sudan at the *Dopa decaboxylase* (DDC) locus. This locus plays an integral role in pigmentation patterning and other physiological processes in insects. Interspecies analysis of polymorphisms showed an excess of high-frequency derived variants within the coding region of the sequence. Phylogenetic analyses sequences of *An. arabiensis* at the DDC locus produced gene genealogies with topologies that mirror one lineage. However, when compared to the sequences of *Drosophila* and *Aedes* species, three distinct lineages were observed with the *Aedes* lineage sharing extensive ancestral polymorphism with *An. arabiensis*. Fay and Wu *H* test on the coding region of the gene was significant at ($P=0.094$, $P<0.05$). The conservation of the DDC gene over evolutionary time scale in *An. arabiensis, Drosophila* and *Aedes* highlights the importance of the gene in the evolution of the three groups of insects. We postulate that positive selection within the coding region of the DDC gene is maintaining advantageous mutations in the gene so as to maintain its multiple functional polymorphisms in the adaptation of *An. arabiensis* populations in Sudan.

4.2 INTRODUCTION

Understanding the genetic architecture of quantitative traits in organisms begins by identifying the genes regulating the traits in natural population. Throughout its long evolutionary history, the *Dopa decarboxylase gene* (DDC) has acquired a variety of functions in insects (Hodgetts and O'Keefe 2006) including cuticular sclerotization in insects as well as melanic encapsulation in malaria parasites (Beerntsen *et al.,* 2000). *Dopa decarboxylase* converts dopa to dopamine (Figure 2) and also catalyzes the production of the neural transmitter dopamine and serotonin (Hiruma and Riddiford 1985; Hiruma *et al.,* 1985). Its key role in behaviour and development of insects and its potential utility in studies of molecular systematics in Noctuoid moths were latter highlighted by Fang *et al.,* 2000. (Refer chapter 1 and 5 for detailed information on DDC). Subsequent workers have characterized the DDC gene in many insects (Hirsh and Davidson 1981; Hiruma *et al.,* 1995; Ferdig *et al.,* 2000; Noguchi *et al.,* 2003). The most important use of this gene is its role in the production of melanin and defence in parasites and insects (Paskewitz *et al.,* 1998, 1999; Nappi and Christensen 2005; Huang *et al.,* 2005). Also many DDC mutations have been shown to affect the acquisition of learned responses in *Drosophila* (Tempel *et al.,*1984) and morphological defects of the cuticle and/or catecholamine-related abnormalities in *Drosophila melanogaster* (Wright, 1996). A study has also established that there exist a high degree of sequence conservation within DDC-coding regions and this has allowed comparisons from various insects, facilitating a number of recent studies on insect systematics (Hodgetts and O'Keefe 2006) and melanism has been known to be the target for parallel evolution in many animal species (Pearse and Pogson 2000).

The present chapter is a follow-up of our findings in Chapter 3. In chapter 3, we found that GST genes use duplication events to acquire diverse phenotypes in *An. arabiensis* populations. To understand the pattern of genetic variation at this locus and identify genes within the genome that are responsible for generating diverse phenotypes, we selected the DDC gene. The DDC is of interest to us for two reasons 1) it is located near microsatellite locus 33C1 within the 3Ra inversion on chromosome 3 in proximity to the 4 Epsilon class GSTs studied in Chapter 3, and 2) its role in diverse immunological, physiological as well as biochemical mechanisms in many species of

insects including the development of insect melanism (True, 2003). A number of workers have also observed that *An. gambiae* or closely related species in Sudan showed marked increase in melanism (Aboud, 2003). Analysis of sequence data showed a clustering of mtDNA haplotypes corresponding to melanic and normal forms and the authors postulated that the melanic forms may have adapted to survive severe drought and heat in the Sudan (Aboud, 2003). However, there is very little information on the molecular basis of the evolution of these phenotypes within populations. It is known that genes that encode for divergent adaptive traits may have genealogies that contrast with those from loci that are not functionally involved in differentiation (Wilder *et al.,* 2004). Findings from studies cited above indicates that the DDC may be involved in the evolution of adaptive phenotypes in populations of insects and any selective constraints on this gene is likely to affect the pattern of nucleotide variation observed at other nearby loci. In the context of adaptation in insects, melanization is of interest for two reasons. First, the ecological niche of *An. arabiensis* is postulated to be largely defined by the climatic conditions (species prefer dry arid habitats) and therefore mechanisms such as the tanning of the cuticle may be important for the adaptation of *An. arabiensis* to its environment. Secondly, *An. arabiensis* is a vector of the malaria parasite and studies has shown that menalic encapsulation of the melanise ookinetes and early oocysts which involves the DDC locuss is a defensive mechanism against the parasite. This means that the DDC locus may be under some selective constrains and therefore may be the obvious target of natural selection.

An. arabiensis is known to predominate in arid savanna and montane areas (White, 1974; Lindsay *et al.,* 1998; Hargreaves *et al.,* 2000) and it is a major vector of malaria particularly in areas such as Ethiopia, Sudan, Malawi and Tanzania (White, 1974).
The present study examined the molecular evolution of *Dopa decarboxylase*, in the context of identifying the type of evolutionary forces shaping nucleotide variation at the locus. We therefore used DDC sequences of *An. arabiensis* from Sudan to gain insight into the evolution of the species in nature.

4.3 MATERIALS AND METHODS

Source of samples and sequencing of DDC *gene*

Mosquito samples were collected from Sudan using Pyrethrum spray technique and were kindly provided by Drs Kwang Shik Choi and Derek Charlwood (See Chapter 2 for detailed information on the samples). DNA extraction followed the technique of Ballinger–Crabtree *et al.*(1992) (see Chapter 2). Species identification followed Scott *et al.,* (1993) i.e. see Chapter 2. Details of PCR amplification and sequencing of DDC gene are detailed in Chapter 2.

4.4 SEQUENCE ANALYSIS

Sequences were multiple aligned using the Clustal W program within BioEdit software, version 5.0.9.1 (http://www.mbio.ncsu.edu/bioedit/page2.html). Intraspecific and interspecific statistics were calculated using DnaSP, version 3.51 (Rozas *et al.,* 2003). Nucleotide variation was estimated as nucleotide diversity, π (Nei, 1987) for all sites and subsequently for the coding region. The genetic divergence between species was estimated as the average number of pairwise differences between species, *K* (average proportion of nucleotide differences between species), according to the methods of Nei, 1987, by computing the average of all comparisons between sequences of *An. arabiensis* and *Ae. aegypti* (outgroup). We examined the frequency spectrum of polymorphisms for deviations from neutral equilibrium expectations using Tajimas D (TD), Fu & Li F and Fay & Wu's H (FWH). FWH compares the estimator π against θ_H which is weighted toward high frequency derived alleles. Significant deviations from neutral expectations for both TD and FWH were assessed using 1000 coalescent simulations conditioned on the observed value of θ from the data. Based on the observed recombination events in our dataset (Table 6), we assumed an intermediate rate of recombination (R=10) for the simulation. A distance tree was constructed using the neighbour-joining algorithm as implemented in MEGA version 3.51 (Kumar *et al.,* 1994). Thousand bootstrap replicates were performed and distances were estimated according to the two-parameter model of Kimura (1980). Sequences from *An. gambiae*, *Ae. aegypti* and two *Drosophila* species included in the analysis were obtained from the GeneBank (*Anopheles gambiae*: AF063021, *Aedes aegypti*: AAU27581, AY064102.1,

AY064101.1; *Drosophila immigrans*: AF293738, *Drosophila sordidula:*AF324979). Phylogenetic analysis were performed on the sequences using the MEGA version 3 software (Kumar *et al.*,2004).

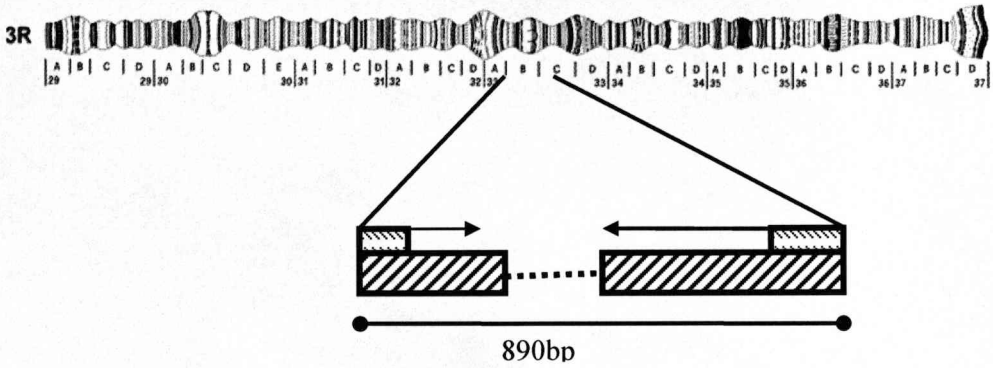


Figure 17: Diagram of chromosome 3R showing the DDC region sequenced in *An. arabiensis* from Sudan

Exons are represented by ▨ ntron ..... Arrows indicate the direction of forward and reverse primers and ▨ indicated the position of primers

4.5 RESULTS

4.5.1 Nucleotide variation at DDC locus

We sequenced 890bp of the DDC gene (Figure 17) from 15 samples of *An. arabiensis* from Sudan. In the nucleotide survey of the 890bp region of DDC, we detected a total of 35 polymorphic sites within the entire sequence out of which 19 were found within the coding region of the DDC locus (Table 7). 12 of the 19 sites were synonymous and 7 nonsynonymous (i.e. encoding amino acid substitution). The rest were found in the intronic region of the gene. Overall total nucleotide diversity, π = 0.0120 was

estimated for the DDC gene with the coding region contributing about 66% (π = 0.0083) of the diversity (Table 7). Polymorphic sites were not uniformly distributed throughout the coding region of the gene; a sliding window plot of genetic diversity showed several peaks of polymorphism in the samples (Figure 18). Minimum number of recombination events in the sequence was 5 per species (Rm=5) for the DDC locus Table 7.

4.5.2 Intraspecific test of selection at DDC locus

We applied tests of selection on both the coding and non-coding regions of the DDC locus to assess whether the observed allele frequency deviates from neutral expectation. Although, no test was statistically significant at 95% confidence interval, the frequency spectrum of polymorphic sites across the sequenced region was skewed towards rare variants, as revealed by the negative Tajima's test (D=-0.1883), Table 7. The Fu and Li F* statistics provided concordant results. However, a positive Tajima's D (D=0.0296) was observed within the coding region of the DDC gene. The synonymous and non-synonymous substitutions at the DDC locus were markedly different in the entire sequence (Table 7).

Sliding-window analysis of Tajima's D values is shown in Figure 18. The figure showed several peaks indicating that the distribution of allele frequency varied across the entire coding region of DDC gene, indicated by the numerous peaks observed (Figure 18). The frequency distribution of derived variants in the DDC was estimated by the FWH (Fu and Way's H statistics) using coalescent simulations that incorporated recombination equal to 10 per gene. The values of H, using *Ae. aegypti* as outgroup revealed a skew towards an excess of high-frequency variants within the coding region relative to neutral equilibrium model, indicated by the near significance of H (H=0.094, P<0.05).

Region	*S*	S +NS	π	TD	Fu & Li F*	FWH *H*-value	R_M
Coding N=707bp	19	12+7	0.0083	0.0296 (NS)	0.2694 (NS)	0.094	5
Total N=890bp	35	14+8	0.0120	-0.1883 (NS)	-0.26011 (NS)	0.135	5

Table 7: Diversity statistics for the DDC locus in *An. arabiensis* populations captured from Sudan

S= the number of segregating sites, S=synonymous substitutions, NS=nonsynonymous substitution, π = the average number of pairwise differences between sequences, TD= Tajimas (1989) D value, Fu & Li F*, FWH=Fay and Wu's (2000) *H* value, R_M=the observed minimum number of recombination events. Significance was determined at 95%

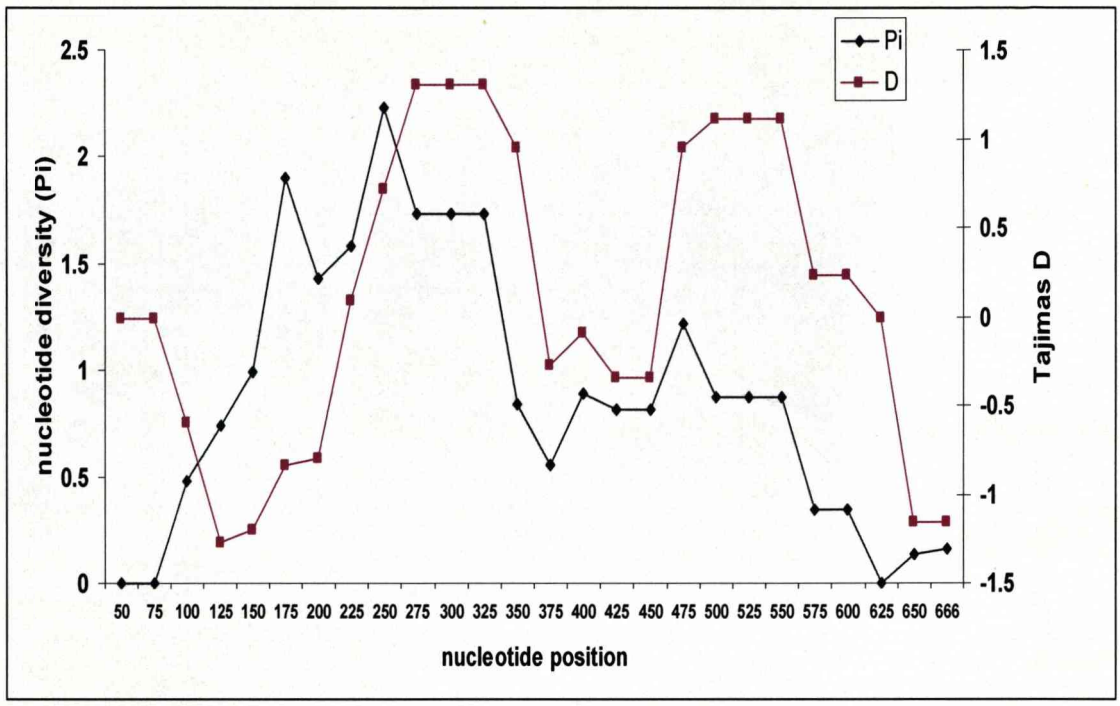


Figure 18: Sliding window analysis of the genetic diversity A) Pi and B) Tajimas D value D within the coding region of the DDC locus in *An. arabiensis* populations

Window length of 100bp was used with 25bp step length. The nucleotide values (Pi) values were multiplied by 100

4.5.3 Nucleotide polymorphism and divergence

The silent polymorphism and divergence at the coding resgion of the DDC locus is shown in Figure 19. The overall silent polymorphism in *An. arabiensis* was 0.02622, and divergence from *Ae. aegypti* was 0.7529. Nucleotide diversity at nonsynonymous (replacement) sites was estimated as 0.00270, and divergence between *An. arabiensis* and *Ae. aegypti* (outgroup) at nonsynonymous sites was estimated as 0.08055. Figure 19 shows that silent divergence was much higher than silent polymorphism (values multiplied by 100) across the coding region.

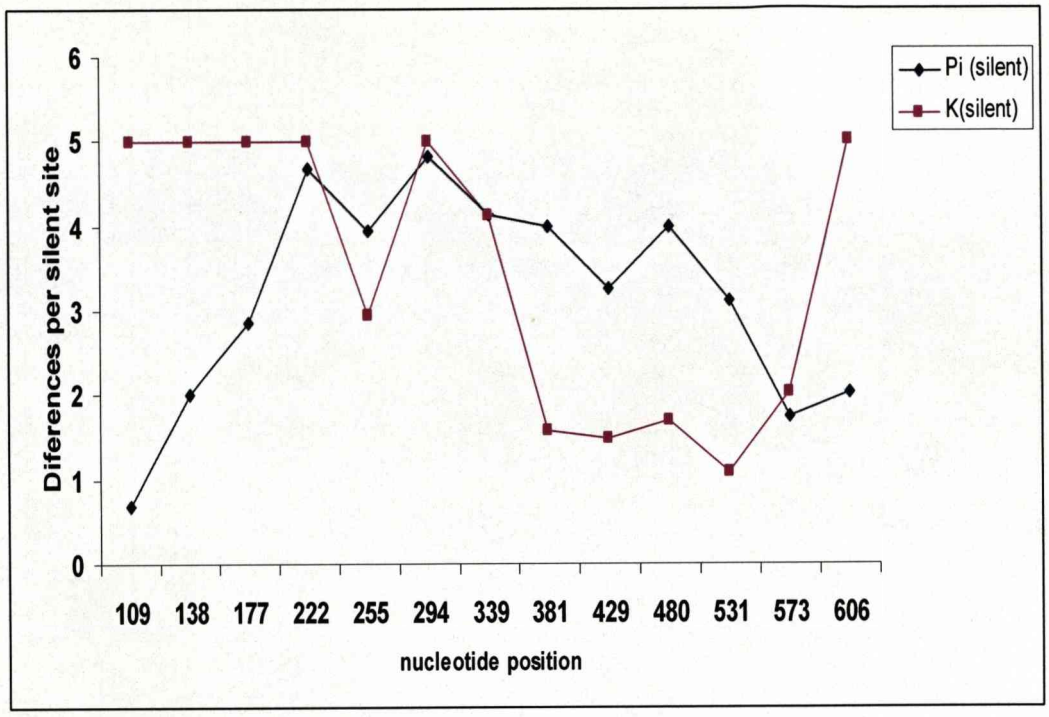


Figure 19: Sliding-window plot of silent polymorhism, *Pi* (silent) and silent divergence, K (silent) between *An. arabiensis* and *Ae. aegypti*

Pi (polymorphism) and and *K* (Divergence) was measured as differences per silent site on the y-axis. Window length of 100bp was used for the analysis. Silent polymorphism values were multiplied by 100. *Ae. aegypti* was used as an outgroup

4.5.4 Phylogenetic analysis

Phylogenetic analysis of the DDC revealed three major clades as shown in Figure 20. One lineage consists of all sequences from *An. arabiensis* and *An. gambiae* (with 99% bootstrap support-values not shown in diagram). The second lineage contains all *Ae. aegypti* individuals (99% bootstrap support) and the third contains sequences from *Drosophila* species also with 99% bootstrap support. The data did not allow resolution of the two species individuals from the *An. gambiae* complex.

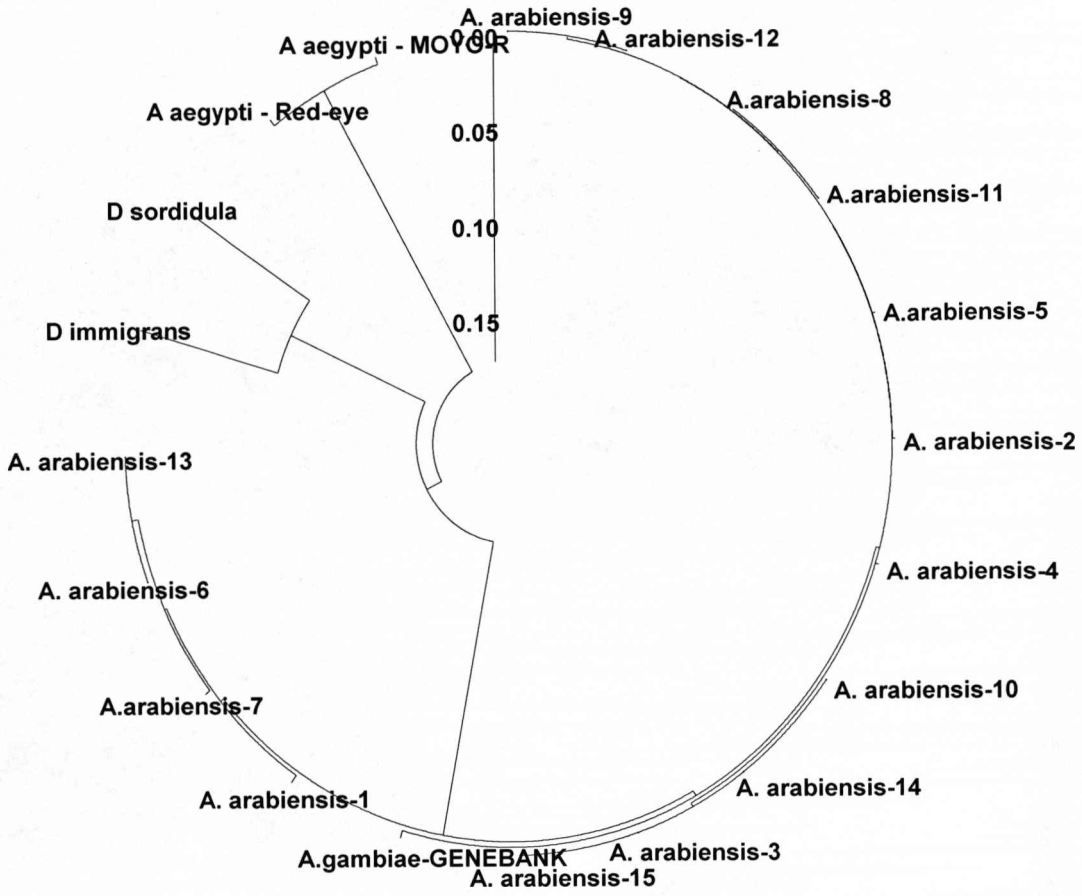


Figure 20: Neighbour joining tree of DDC at the coding region of *An. arabiensis*, *An. gambiae*, *Drosophila* species and *Ae. aegypti*

Bootstrap support for major nodes indicated (percentage of 1000 bootstrap). Abbrevations are as follows D=*Drosophila*, *Ae. aegypti* = *Aedes aegypti*, *An. arabiensis* = *Anopheles arabiensis*, Genebank = (*An. gambiae Dopa decarboxylase isoform)*

4.6 DISCUSSION

Colouration is one of the most variable characters among animals and is a rich source of models of phenotypic evolution (Wittkopp *et al.,* 2003a). Also of interest are loci that underlie the phenotypic divergence between species. *Dopa decarboxylase* contributes to diverse physiological events in insects including cuticular sclerotization (Hiruma and Riddiford 1985), defence and survival mechanisms (Gorman *et al.,*

1997; Paskewitz, 1999; Wilson *et al.,* 2001; Nappi and Christensen 2005). In *Anopheles* mosquitoes, the DDC gene is located in the middle of a large cluster of other functionally related genes e.g GSTs on chromosome 3. We observed a high degree of sequence conservation within DDC coding region within populations of *An. arabiensis* from Sudan. This pattern has been observed within DDC sequence of *Drosophila* (Wang *et al.,* 1996) and a study by Joron *et al.* (2006) showed that a conserved 'supergene' (group of neighbouring genes on a chromosome which are inherited together because of close genetic linkage and are functionally related in an evolutionary sense, although they are rarely co-regulated genetically) locus controls colour pattern in *Heliconius* butterflies. It has been shown that the conservation of genes over evolutionaty time-scales is important for many biological processes that affect homologous traits in many species (Mackay, 2001).

In the present study, analysis of nucleotide variation at the DDC locus revealed an excess of high-frequency variants within the DDC as evident by the near significant *H* value of Fay and Wu's (2000) test. An excess of high-frequency derived alleles have been observed in *D. melanogaster* DDC locus (Tatarenkov and Ayala 2007). This pattern of evolution may have lead to a high rate of synonymous substitutions within the species. More than 50 DDC mutations have been isolated in *Drosophila*, many of which are recessive lethals (Wright, 1996) indicating the importance of the gene in the evolution of the species. The gene has been used recently to study insect systematics (Fang *et al.,* 2000; Hodgetts and O'Keefe 2006). The DDC therefore provides a unique opportunity to examine the evolution of not only pigmentation but other functions such as parasite encapsulation, defence etc. in natural populations of *An. arabiensis* and their adaptive value in habitat selection.

We estimated a minimum recombination events/generation of 5 (R_M=5) in 707bp of DNA sequence. Another study (Goto *et al.,* 2004) observed 7 recombination events in 580bp of *Drosophila* sequence. Although we cannot extrapolate the recombination events observed in *An. arabiensis* to *Drosophila* species because the two species may not be under the same selective pressure or evolutionaty constrains, we can highlight the fact that recombination events are common in both species.

Based on the pattern of phylogenetic differentiation, the DDC delineate three divergent lineages between *Anopheles*, *Aedes* and *Drosophila* species but shows one lineage within the *An. arabiensis* species. The DDC is known to be involved in the cuticular pigmentation in insects such as Tobacco hornworm, *Manduca sexta* (Hiruma *et al.,* 1985) and in colonies of the New Zealand black coral, *Antipathes fiordensis* (Holl *et al.,* 1992).

Neutrality tests on the coding region of DDC showed that both Tajima's *D* and Fu and Li's test as well as Fay and Wu's *H* test had values that deviated from what is expected under neutral expectation. In most populations, positive D or Fu and Li values are often interpreted as evidence for balancing selection, population subdivision, or decrease in population size many of which are factors that result in a relative overabundance of derived variants with intermediate frequencies. The *H* test, derived by Fay and Wu (2000) showed that in the presence of recombination, positive selection may be acting on the locus and the effect may be contributing to relative excess of intermediate- and high-frequency variants. However, an excess of derived alleles at high frequency, according to the *H*-test, may also be consistent with the effect of hitchhiking caused by directional selection. This hitchhiking may be occurring with other functionally important genes such as GSTs (described in Chapter 3) clustered the DDC locus (distance of about 3MB). Tatarenkov and Ayala (2007) observed hitchhiking effect on DDC in *D. melanogaster* and discussed that the complicated pattern of variation and pattern observed at the locus may be a result of an unusually high density of functionally important genes located around the locus. However, this further investigations is needed to ascertain the effect of these neighbouring genes on the overall pattern of nucleotide variation at the DDC locus. Sliding window analysis of genetic diversity Pi and Tajimas D values showed a fluctuating symmetry with both peaks coinciding with each other. We observed a low intraspecific but high interspecific divergence within the DDC locus of *An. arabiensis* as evident in the relatively high (8%) nonsynonymous site divergence between *An. arabiensis* and *Ae. aegypti*.

Another evolutionary scenario could be that, after several million years of evolution of the DDC gene, there may be an increase in the selective constraints on this locus. The nature of this constrains is not very clear. It may be as a result of environmental

changes or it could be due to its functional role in many biological processes. Wittkopp *et al.* (2003b) recently studied the DDC locus in *D. americana* and *D. novamexicana* and found that there was no association between DDC and the observed melanisation pattern. The authors postulated that the lack of association between DDC and melanisation could be as a result of the DDC locus being functionally constrained due to its upstream role as a regulator of melanin precursors and also because it acts pleiotropically in many different biological pathways. For example, it is known that the DDC is alternatively spliced and produces two primary transcripts, one which is expressed in the central nervous system and may be necessary for neural transmission and a second, which is necessary for melanization, in hypodermal tissue (Morgan *et al.*,1986). The 4th exon, which Wittkopp *et al.* (2003b) surveyed was included in both of these transcripts and therefore that region of DDC may be constrained due to its role in several biological processes. This highlights the importance of the DDC in many biological mechanisms and one has to be cautious when linking the DDC locus to the expression of phenotypes or phenotypic patterns.

It has also been postulated that genes in evolutionary conserved signalling pathways affecting metabolism, nutritional control and stress response regulate longevity in novel organisms (Hopkins and Kramer 1992). The DDC has been shown to affect variation in *Drosophila* longevity (De Luca *et al.,* 2003) and also polymorphisms at DDC have been found to be associated with naturally occurring genetic variation in locomotor behaviour in *D. melanogaster* (Jordan *et al.,* 2006). Given the highly pleotropic effect of DDC on many phenotypes , the DDC can be excellent candidates for the generation and maintenance adaptive phenotypes in natural populations of *An. arabiensis,* although we recommend that further studies using much larger samples size to confirm this assertion.

Vectorial capacity is also dependent on the continuous transmission of pathogens including malaria parasites by *Anopheles* mosquitoes (Beerntsen *et al.,* 2000). Earlier studies have revealed the importance of melanisation response of *An. gambiae* to intrathoracically inoculated CM-Sephadex beads (Gorman and Paskewitz 1997; Gorman *et al.,* 1997; Paskewitz and Riehle 1994). As part of the defence mechanisms against malaria parasites, *An. gambiae* is able to encapsulate and melanise ookinetes

and early oocysts of the malaria parasite and kill them. The deposition of melanin during the encapsulation process is commonly initiated by the haemocytes and in some cases by melanogenic enzymes (phenoloxidases) circulating in the plasma (Nappi and Christensen 2005). This means that in terms of mosquito-parasite (in this case *An. arabiensis-Plasmodium*) interaction, the DDC known to be involved in this process (Figure 21) may be under intense mutational contrains due to the presence of malaria parasites.

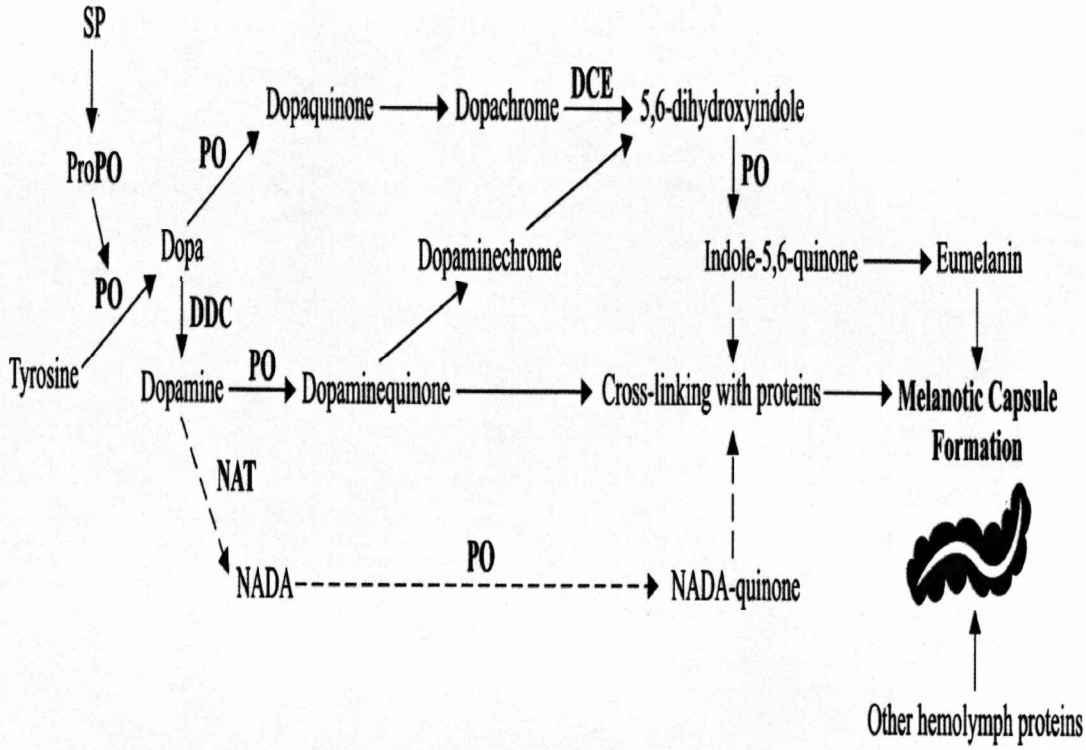


Figure 21: Proposed melanic encapsulation pathway of mosquitoes.

In the proposed melanotic encapsulation pathway of mosquitoes, a serine protease (SP) proteolytically cleaves an inactive prophenoloxidase (ProPO) to form an active phenoloxidase (PO). Tyrosine, the initial substrate, then is hydroxylated by the activated PO to form dopa, a key branchpoint substrate. Next, dopa is oxidized by PO to form dopaquinone, which then forms a dopachrome intermediate. A dopachrome conversion enzyme (DCE) converts this intermediate to 5,6-dihydroxyindole, which subsequently is oxidized by PO to form indole-5,6-quinone. This latter compound forms eumelanin or cross-links with proteins to eventually produce a melanized capsule. Dopa also can be decarboxylated by dopa decarboxylase (DDC) to form dopamine, which forms another branch point. A PO-based oxidation of dopamine produces dopaminequinone, which can cross-link with proteins or form melanin via the indole pathway. Dopamine also may be acetylated by *N*-acetyltransferase (NAT) to form *N*-acetyldopamine (NADA). PO oxidation of NADA then produces NADA-quinone, which cross-links with other proteins to form a melanotic capsule. Solid arrows designate likely major pathways, and dashed arrows denote probable minor pathways. (Source: Beerntsen *et al.*, 2000).

An. arabiensis is an important vector of malaria in many areas in Africa and malaria parasite challenge may put selective pressure on the DDC gene. These mutational constrains may generate nucleotide polymorphisms within the DDC locus. We postulate that positive selection within the coding region of the DDC gene is maintaining advantageous mutations in the gene so as to maintain its pleotropic functional roles in the adaptation of *An. arabiensis* populations in Sudan.

In conclusion, based on our result and the review of other studies, we can suggest that the DDC is a very important gene in the genome of *An. arabiensis* involved in many biological processes. This was evident in its pleotropic effect on many quantitative traits, whilst maintaining its functional role in many species after several million years of evolutionary time-scale. The DDC may be involved in generating diverse phenotypes through a single gene (epistasis) as way of adapting populations of *An. arabiensis* to heterogeneous environments. Further work is needed to examine the influence of the observed nucleotide changes on pigmentation phenotypes and the role of DDC in the generation of adaptive phenotypic traits in the *An. arabiensis*.

5.0 CHAPTER FIVE ECOLOGICAL IMPLICATIONS OF INTRASPECIFIC DIFFERENCES IN LARVAL MELANIZATION WITHIN POPULATIONS OF *ANOPHELES GAMBIAE S.L.*

5.1 ABSTRACT

This chapter details the effect of a phenotypic response in the form of larval pigmentation on the life-history traits of the malaria vector, *An. gambiae.* Larval melanism, i.e appearance of black pigment above the dorsal thoracic and abdominal integument, setae and head capsule) was induced by rearing concurrently larvae from iso-female lines in containers with dark and white backgrounds. A total of 900 larvae of the KISUMU laboratory strain and 400 progeny from females collected from Ghana were used for the induction experiment. There were no significant differences in the phenotypic response of the KISUMU and the Ghana samples so the two samples were pooled for all analysis. More than 70% (corrected for mortality) of larvae reared in the dark background developed darker pigmentation. The melanization rate measured as average grey values on a scale of 0 (black) - 225 (white) was 64.3 $\pm$ 7.5 for melanic and 106 $\pm$ 8.6 for the non-melanic forms and these were significantly different ($P<0.001$). A higher proportion of larvae survived in white containers (N=456/510, 89.4%) than in Dark containers (N=372/510, 72.9%, $P<0.05$) although the difference was not significant. Larvae reared in dark containers produced significantly larger pupae, 5.3 $\pm$ 2.2mg (mean $\pm$ S.E) than larvae reared in white, 3.0 $\pm$ 1.5mg ($P<0.001$). There was considerable variation in juvenile melanization rates and these correlated with rearing habitat. Melanization in *An. gambiae* influenced developmental time and pupal weight. Darker melanic larvae had faster developmental time (13.9 days) than the non-melanic larvae (15.6 days). There were differences although not significant in the survivorship of the melanic and non-melanic forms of larvae and this was consistent with a trade-off hypothesis. We demonstrated a bi-directional response of *An. gambiae* larvae to different rearing habitats and showed that phenotypic response was modulated by the elevated activity of DDC and PO in the melanised larvae relative to the parental during the induction of melanism. The effect of the phenotypic response on the fitness of the larvae is discussed.

5.2 INTRODUCTION

5.2.1 Phenotypic plasticity

Organisms live in complex environments that change over a variety of scales in space and time and they have evolved mechanisms that enable them to adapt to this changing variable environment. In some cases, the variability in environment stimulates individual organism to produce different phenotypes, a phenomenon referred to as phenotypic plasticity (Agrawal, 2001; Garland and Kelly 2006). This change in phenotype may take the form of changes in morphology, behavioural or physiological (Price,2006). For example, phenotypic plasticity induced morphological change in a planktonic crustacean, *Daphnia galeata* (Oda *et al.,* 2007), behavioural change in tits, *Parus* species on a Sweddish Island (Alatalo and Moreno 1987) as well physiological change (influence of colour preference by carotenoids in diet) as observed in the house finch, *Carpodacus mexicanus* (Hill, 1994). This changes may be important for the survival, reproduction and persistence of the organisms in their environment.

Phenotypic plasticity has been observed to be common in nature (Harvell 1990; Karban and Baldwin 1997; West-Eberhard, 2003). Previously, phenotypic plasticity within populations was thought to result solely from genetic differences among individuals. However, it is now known that in certain cases, phenotypes are not fixed but can be influenced by the environment in which an organism lives (Schlichting, 1986; Tollrian and Harvell 1999; Pigliucci, 2001; West-Eberhard, 2003). However, some questions still remain over the basis and significance of phenotypic plasticity in the life–history of organisms. Since it has been established that plasticity does not necessarily reflect genetic variation, the question is what is the link between natural selection and plasticity? Secondly, since it will be expected that alternative phenotypes within populations must have greater fitness in different environmental conditions, what will be the link between phenotypic plasticity and fitness measured by life-history traits? The life history of an organism in terms of evolutionary biology refers to how the various mechanisms that enables an organism to develop, reproduce and die in its habitat is shaped by natural selection (Roff, 1992). Examples of some life history traits include developmetal time, generation time, fecundity rate, mortality

rate etc. It has been shown that an organisms life-history defines its finess because positive effects of plasticity on life-history enhances adaptation to environment (Stearne, 1992). For example, male body size can be important in male-male competition for females (Goldsmith and Alcork 1993) as well as for female choice of mates (Gilburn and Day 1994). Studies have also expounded the theory that trade-off in fitness between alternative phenotypes produced in response to different environments exist (Tufto, 2000; Sultan, 2000). Several studies in *Lepidoptera* and *Drosophila* species have linked some life history traits, such as short juvenile development time and large adult size as being highly relevant to fitness (Nylin and Gotthard 1998).

For example, it was shown that larval melanization in European map butterfly, *Araschnia levana* influences plasticity of development time in the 5th instar and adult size so that dark larvae have short 5th instar periods but instars 1–4 take longer relative to pale larvae and the difference is especially large in long day (Windig, 1999). In that same study, adult wing size from dark larvae were larger, especially in short day, resulting in a significant daylength by colour interaction (Windig, 1999). Also a study by Bochdanovits and De jong (2003) showed that under conditions of food deprivation two temperate and two tropical populations of *Drosophila melanogater* reared at high and low temperatures produced different adult body sizes coinciding with different chances of reaching adult stage. These changes in life-history traits in *Araschnia levana* and *Drosophila melanogater* may have important consequences for the survival and fitness of individuals in all stages.

Therefore increasing our knowledge of phenotypic plasticity for fitness components is important to gaining more insight into observed ecological differences among species (Ford and Seigel 1989).

5.2.2 Melanization

The study of pigmentation is one of the traits that is most amenable to exploring the connection between genotype and phenotype (Hoeskstra, 2006). Pigmentation is one of the most variable traits in the genus *Drosophila* (Wittkopp *et al.*, 2003b) and many studies have used *D. melanogaster* tools to help explain the genetic and developmental mechanisms involved in pigment patterns.

The present study is a follow-up from Chapter 4. The results from Chapter 4 indicated that the *Dopa decaboxylase* gene (DDC), a pigmentation gene important in the cuticular pigmentation in insects (refer to Figure 2 in Chapter 1) showed very little differentiation between specimens of *An. arabiensis* from Sudan, suggesting a selective constraint on the locus. We hypothesize that this pigmentation gene may be important in generating adaptive phenotypic traits (melanic patterns) in populations of *An. gambiae*.

In order to further understand the ecological implications of phenotypic plasticity in nature, we used a pigmentation system known as melanism which is defined as the appearance of wholly or mostly dark forms of an organism, to investigate the relationship between melanization in juvenile *An. gambiae* species and some life-history traits such as developmental time, survival rates and pupal weight.

Melanism plays important roles in insect ecology, including defence against parasites and predators (Siva-Jothy, 2000) e.g. winter moth larvae (*Operophtera brumata*) and in the peppeth moth (*Biston betularia*) ; mate signaling (Wiernasz, 1989; Ellers and Boggs 2002) e.g. populations of eastern mosquitofish (*Gambusia holbrooki*), and thermoregulation (Ottenheim *et al.,* 1999) e.g. butterfly, *Parnassius phoebus*. Melanism is common in many different animal groups (Majerus *et al.,* 1998). Studies has hypothesized that melanin is a complex polymer, and its synthesis may be hindered if ambient conditions limit the resource budget (Talloen *et al.,* 2004). Other studies suggest that both faster growth and larger size may be traded off against one another or against potentially costly traits such as melanin production (Blois, 1978). For example, a study on the effect of melanization on the juvenile stages of the European map butterflies showed that melanized 5th instars grow more slowly than the pale ones (Windig, 1998). A recent study investigated interaction between melanization and drought-stress environment in Satyrine butterflies, *Pararge aegenia* and also indicated that darker larvae showed slower development and lower survival compared to pale larvae (Talloen *et al.,* 2004).

5.2.3 Melanization in *An. gambiae*

Melanised forms of *An. gambiae* have been observed in nature and sampled in the desert scrub areas of Sudan (Aboud, 2003) and a dark species of *An. daudi* thought to be closely related species to *An. gambiae s.l.* has also been observed in the same area. The ability of *Anopheles* to change colour (homochromy) in response to the background colour of rearing trays was first demostrated by Fuzeau-Braesch (1972). This prompted other studies in laboratory colonies of *Anopheles* mosquitoes; in this case *An. albimanus* and *An. quadrimaculatus* (Benedict *et al.,* 1996) demonstrated homochromy but *Culex* and *Aedes* species did not (Benedict *et al.,* 1987). However, the ecological significance of this observed phenomenon and its effect on the fitness components of the species was never investigated in these mosquitoes.

To gain more insight into the phenomenon of phenotypic plasticity and its effect on life history traits in populations of *An. gambiae*, I formulated the following hypotheses:

1) As a result of the cost involved in the production of melanin, phenotypic response in the form of larval pigmentation (melanization) induced by environmental colour cues will have effect on the survival, development time and pupal weight of *An. gambiae.*

2) Larval pigmentation (melanization) in the juvenile stages of *An. gambiae* larvae is a plastic response to environmentally induced stimulus as a way of an adaptation to heterogeneous habitats. To investigate this, we induced homochromy in inbred lines of *An. gambiae* larvae and looked for signatures of bi-directional selection.

3) We expect to find that populations collected from different localities (location effect) will respond differently to the same environmental stimulus with resultant effect on their life history traits. To investigate this hypothesis, we used laboratory colonies (KISUMU strain) of *An. gambiae* and as well as field samples from northern Ghana.

4) Larval pigmentation is modulated by genes that are involved in the melanin production pathway. To investigate this hypothesis, we did gene expression analysis the melanised as well as non-melanised larvae using two candidate genes, *Dopa decaboxylase* (DDC) and *Phenoloxidase* (PO) important in the melanin production pathway to ascertain their role in larval pigmentation.

The present study is the first of a kind that investigated the effect of melanization on the life-history traits of juvenile stages of *An. gambiae*, an important malaria vector in Africa.

5.3 MATERIALS AND METHODS

5.3.1 Study species

5.3.1.1 *Laboratory colonies of KIS strain An. arabiensis*

Laboratory colonies of the Kisumu strain of *An. gambiae* were used to set up iso-female lines for the induction experiment. The Kisumu strain (S-form of *An. gambiae*) originated from Kisumu, Western Kenya, and is susceptible to permethrin (Vulule *et al.,* 1994). They were established in the insectary of the Liverpool School of Tropical Medicine, UK and provided to me by Amy Lynd.

5.3.1.2 *Field colonies of An. gambiae*

Additional *An. gambiae s.l.* adults were collected resting indoor from northern Ghana in June 2006 by Mr Victor Asoala using indoor resting collecting techniques. The gravid females were kept in cages in the insectary of Navrongo Health Research Centre in Ghana and allowed to lay eggs. The eggs were then transported to Liverpool School of Tropical Medicine by courier service and immediately placed in distilled water. Larvae were reared through the first generation of adults and iso-female lines used for the induction experiments. The iso-female lines were used in the experiment to provide a common genetic background to the population under study.

Information on the procedure for laboratory induction experiments and larval melanization measurements as well the molecular analysis detailed in Chapter 2.

5.4 STATISTICAL ANALYSIS

Simple t-test was used to estimate significance of the differences in the experimental variables. Multifactor Analysis of variance (ANOVA) was used in EPIIFO statistical software to evaluate the influence of larval melanization on the life history characters. GraphPad Prism, version 5.01 software (www.graphpad.com) was used to plot graphs of the life-history traits. A comparative analysis of the effect of larval pigmentation on some life history traits of laboratory (Kisumu samples) and field (Ghana samples) did not show location effect (See results). We therefore pooled the data from Ghana

samples and Kisumu strain for our analysis to increase the power of the analysis. We used the Delta Delta (Ct) method (Livak and Schmittgen 2001) to relatively quantify and determine the relative gene expression profiling. In this method, the threshold cycles of the gene of interest and a reference or housekeeping gene are subtracted to yield a Delta Ct value for each RNA sample. The difference between the Delta Ct values for an experimental RNA sample and control RNA sample (Delta Delta Ct) is directly related to the relative amount of product in the two samples if and only if the PCR amplification efficiencies of the gene of interest and the reference gene are similar.

5.5 RESULTS

5.5.1 Rearing environmental conditions and phenotypic response of larvae collected from different locations

A total of 1300 larvae of *An. gambiae* were used for the induction experiment. These comprised, 69.2%, (900) of KISUMU strain and 30.3%, (400) of field samples from Ghana. Visual classification of dark and pale larvae is based on the appearance of larvae as shown in Figure 22. In the study, over 70% of larvae in the dark rearing environment were scored as dark morphs. Table 1 shows the temperature, pH and other life-history traits measured for KISUMU samples and field (Ghana) samples. Both KISUMU and Ghana samples showed similar phenotypic response to the colour stimulus in the three rearing environments (Table 8). There was no significant difference ($P>0.05$) between the rearing water temperature and pH of the rearing containers. Average temperature in the dark container for KISUMU and Ghana samples were 28.1 ± 0.1 and 28.2 ± 0.1 respectively whilst that for the white container averaged 27.8 ± 0.5 and 28.1 ± 0.06 respectively. No dark larvae were observed in the rearing containers with the white background (Table 8).

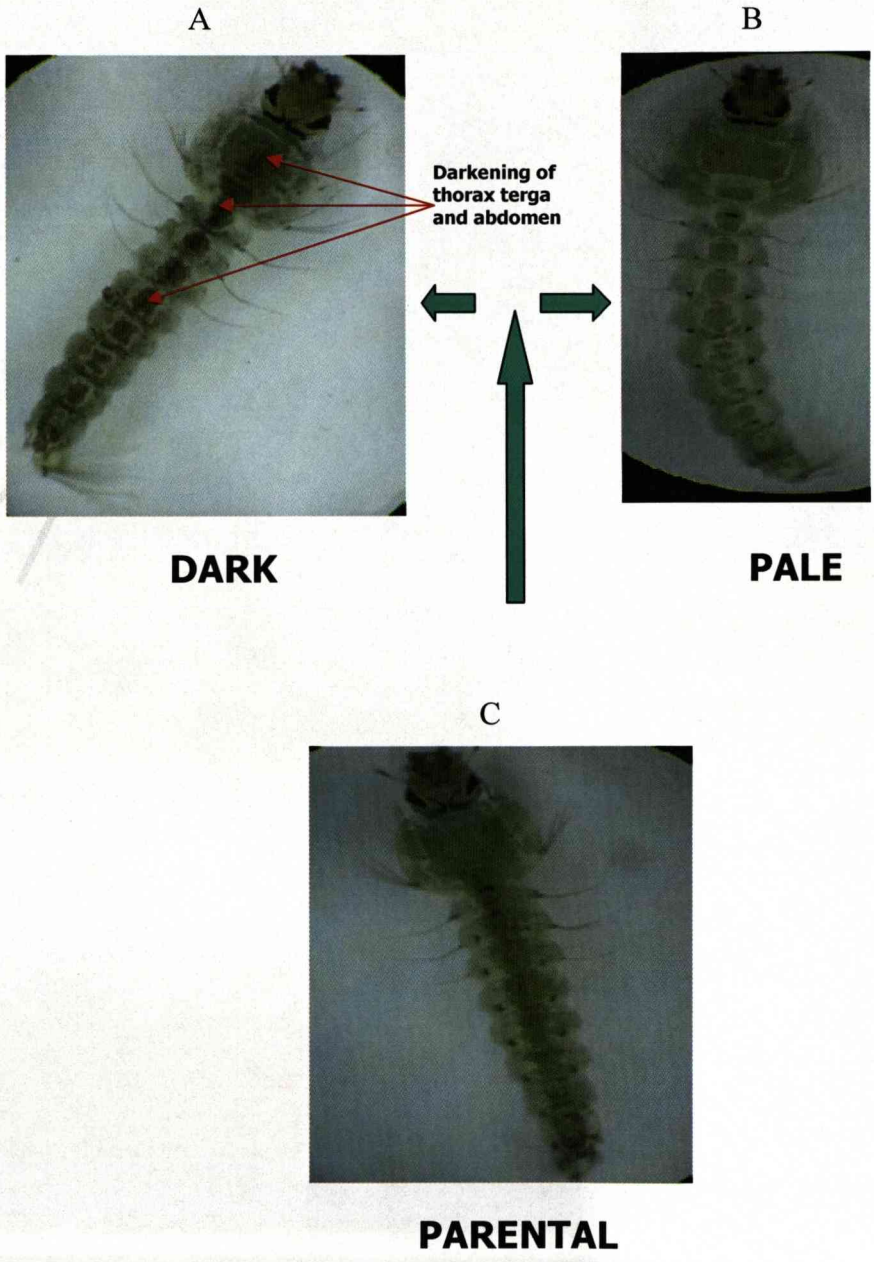


Figure 22: Observed colour variation of *An. gambiae* larvae induced in different rearing colour environments

A) dark larvae reared in dark container, B) pale larvae reared in white containers and C) parental larvae reared in transparent larval bowls (no colour background)

Variable	DARK			WHITE		
	KISUMU strain (n=360)	Ghana samples (n=150)	Test of location effect (t-test) P value	KISUMU strain (n=360)	Ghana samples (n=150)	Test of location effect (t-test) P value
Water temperature	28.1 ±0.1	28.2 ±0.1	0.294 (NS)	27.8 ±0.5	28.05 ±0.06	0.158 (NS)
pH	7.2 ± 0.02	7.3± 0.01	0.147 (NS)	7.2 ±0.04	7.2 ±0.02	0.373 (NS)
% survival rate	74.7 ± 17.3	71.2±14.2	0.197 (NS)	90.9 ±7.6	87.6 ±4.3	0.097 (NS)
Larval development time (days)	13.6 ± 1.5	14.2 ± 1.6	0.187 (NS)	16.4 ±1 .1	14.8 ± 0.8	0.345 (NS)
Pupal weight (mg)	5.1 ± 2.9	5.5 ± 1.5	0.313 (NS)	2.8 ±1.6	3.2 ±1.1	0.621 (NS)
Melanism score (%)	76.2 ± 8.6	73.5 ± 7.7	0.064 (NS)	0.00	0.00	-
Degree of melanization (av.grey values)	63.5 ±9.0	65.1 ± 5.3	0.212 (NS)	110 ± 10.8	102 ± 6.4	0.390 (NS)

Table 8: Temperature, pH and other life history traits estimated for laboratory samples (KISUMU strain) and Ghana samples of *An. gambiae* reared in DARK and WHITE background containers.

Student t- test was used and significance determined at 95% confidence interval, P>0.05, i.e NS=test not significant

NB: A total of 180 KISUMU and 100 larvae from Ghana were used as controls (Parentals)

5.5.2 Effect of rearing environment on larval pigmentation of An. gambiae

In all experiments, there was a significant effect of colour cues on the phenotypic responses of *An. gambiae* larvae (Figure 22). Figure 23 shows the effect of rearing environment on larval pigmentation. There was a significant difference ($P<0.003$) in the degree of melanization between the dark and pale larvae. Larval pigmentation was most observable at the 4th instar stage of development. The results showed a bi-directional selection in that larvae in the dark container became more melanised, whereas those reared in the white containers became much paler than parentals (Figure 23). The differences observed among treatments indicates that background colour induced a plastic response which in turn affect some life-history traits such as developmental time, pupal weight and to a less extent the survival rate of the larvae (Figure 24A-C). There was variation in the pattern of larval pigmentation as shown in Figure 25. There was a significant difference ($P<0.05$) between the average grey values estimated from the Head/thorax and the rest of the body. The Head/thorax region was the most melanised part of the larval body.

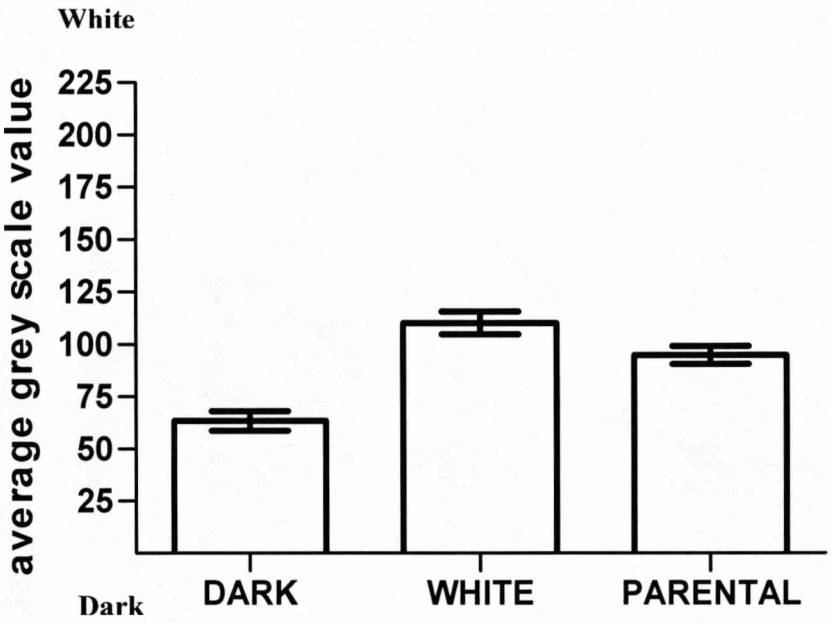


Figure 23: The effect of rearing environment on *An. gambiae* larval pigmentation measured as average grey value on a scale of 0-225. (0=black and 225=white)

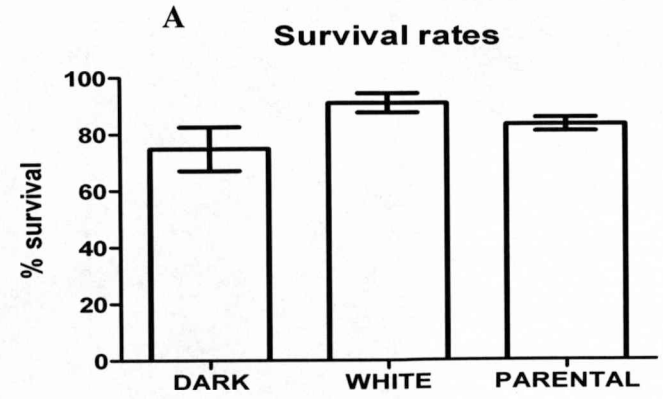
X-axis represents rearing environment categories of DARK, WHITE and TRANSPARENT containers

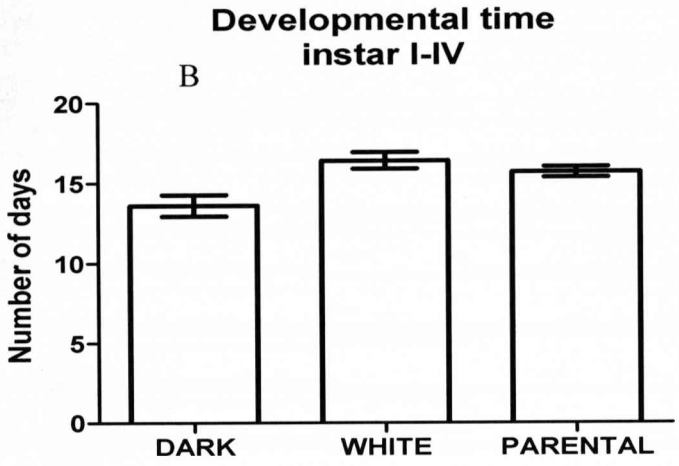
5.5.3 Effect of larval melanization on developmental time, survival rates and pupal weight of An. gambiae

The effect of larval pigmentation on the life history traits of *An. gambiae* is shown in Figures 24A-C. The results of our study showed that dark larvae reared in dark containers produced significantly larger pupae, 5.3 $\pm$ 2.2mg than larvae reared in white, 3.0 $\pm$ 1.5 mg ($P<0.001$). However, the dark larvae had a lower survival rates compared to the pale larvae although the values were not statistically significant (Figure 24A). Dark larvae in the dark containers developed much faster than the pale larvae ones in the white containers (Figure 24B) and difference between the mean developmental times was significant ($P<0.05$).

There was variation in the life-history traits estimated for larvae reared in different environments. Larvae from dark containers also had much faster average development time (13.9 days) than larvae in the white containers (15.6 days). Results of the ANOVA test (Table 9) showed that the rearing environment did affect developmental

time, DT (F=10.89, P<0.011), pupal weight (F=7.19, P<0.028), larval melanization (F=44.12, P<0006) and to a less extent the survival rates (F=3.68, P<0.091).





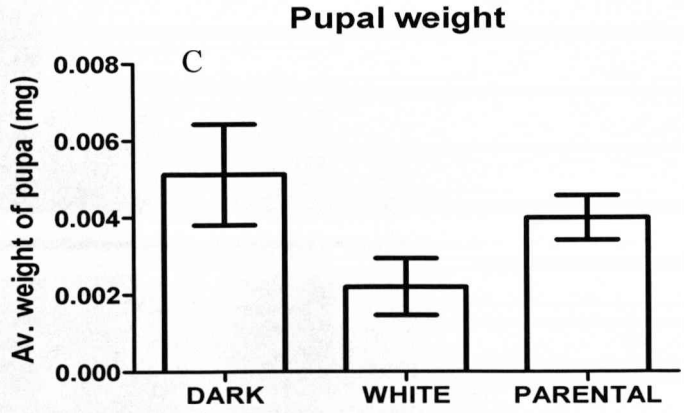


Figure 24: Graphs showing the effect of *An. gambiae* larval coloration on A) survival rate B) developmental time and C) pupal weight.

X-axis represents rearing larval categories. Error bars indicate level of significance

Trait	Source	Sum of squares	df	F	P value
Developmental time (Days)	**between**	19.6	1	10.89	<0.011
	error	14.4	8		
	total	34.0	9		
Pupal weight (mg)	**between**	26.9	1	7.19	<0.028
	error	29.9	8		
	total	56.8	9		
Melanization (Degree of darkening)	**between**	4364	1	44.12	<0.0006
	error	593.4	6		
	total	4957	7		
Survival rate (%)	**between**	655.6	1	3.68	0.091
	error	1425	8		
	total	20.81	9		

Table 9: Results of ANOVA testing the effects of rearing environment (i.e. dark, white or normal) on larval developmental time, pupal weight, degree of melanization and survival rates of *An. gambiae*.

F statistics (F) fixed effects, (df) degree of freedom

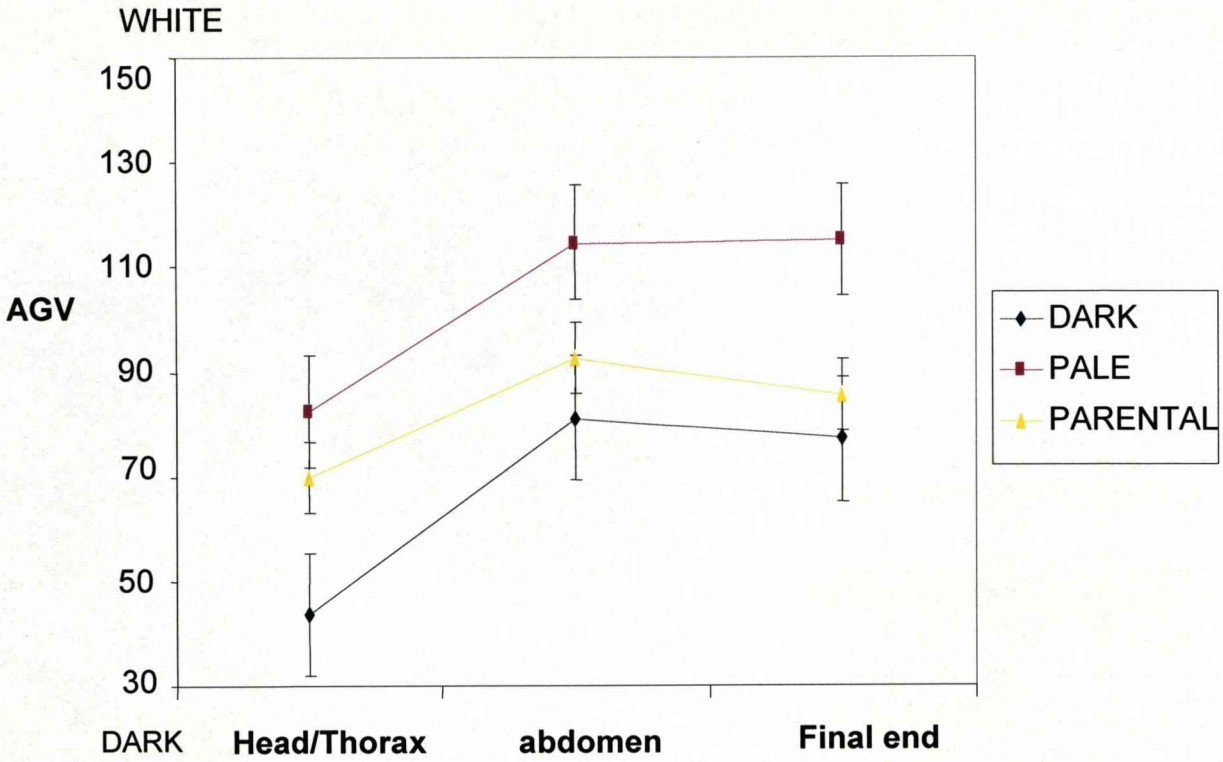


Figure 25: Plasticity for larval melanisation in population of *An. gambiae* measured as AGV (average grey value) on the y-axis.

Measurement was made on a scale of 0-225, dark –white). Error bars show standard error

5.5.4 Relative gene expression profiling of DDC and PO

In order to determine the potential role of DDC and PO in pigment synthesis of the induced larvae, we examined the DDC and PO activity in melanised (dark), non-melanised (pale) and parental larvae (control) Figure 26. The average threshold cycle (Δt) between the melanised and non-melanised larvae was significantly different ($P<0.05$). The pattern observed showed that there was 33% increase of DDC activity in the melanised larvae compared to the parental with a corresponding decrease of about 4% in the non-melanised larvae. However, the PO activity increased by about 20% in the melanised and 13% in non-melanised larvae relative to the parental (control) larvae.

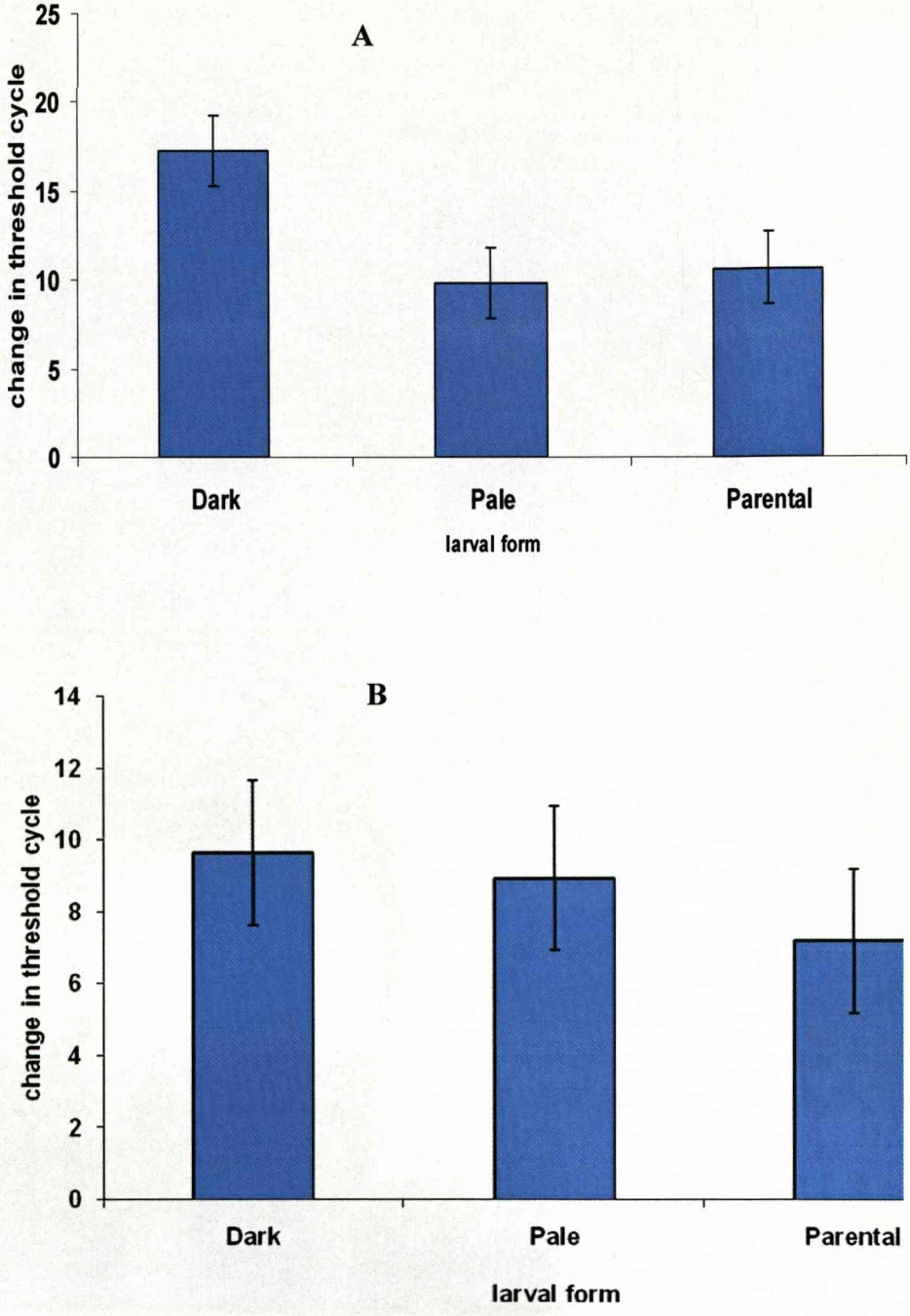


Figure 26: RT-PCR analysis of A) DDC and B) PO showing average Delta C_T values in the melanised (dark), non-melanised (pale) and parental (control) forms of *An. gambiae* .

Error bars are standard error from the mean at 95% CI (confidence interval). Change in threshold cycle (Δt) of Delta C_T is shown on the y-axis

5.6 DISCUSSION

The ecological distribution of a species is often influenced by the individual pattern of response to environment for traits that contribute to fitness (Sultan 2001). For example, at the population level, species may be excluded from certain environmental conditions in which individuals produce few or poor-quality offprings, and hence fail to establish populations (Futuyma and Moreno 1988). Phenotypic plasticity is now recognized as the major source of variation in nature (Ford and Seigel 1989; Travis 1994). However, little is known about the nature of species differences in plasticity for fitness-related traits (Lönn *et al.,* 19198; Roskam and Brakefield 1999). Issues about the evolution of adaptive phenotypic plasticity, the underlying genetic mechanisms and the way plasticity is affected by natural selection in different environments remains a subject of much debate (Via *et al.,* 1995). For example, in terms of adaptation to environment, individuals can become more adapted to different parts of the environment or more so individuals can evolve plastic phenotypes that that are adapted to the extreme ends of the environments (Scheiner and Callahan 1999). This was illustrated in a study by Buskirk *et al.* (1997) where it was shown that when confronted with a dragonfly predator *Anax*, *Pseudacris triseriata* tadpoles adjusted their tail shape but not body shape in the direction of selection imposed by the predator. This indicates that phenotypic plasticity of some morphological traits can evolved under intermittent selection imposed by the inducer (predator, temperature etc.). How this selection will affect fitness components such as survival and growth is not clear. Therefore increasing our understanding of phenotypic plasticity for fitness components will provide the framework for gaining more insight into ecological differences among species at the population level.

The results of the present study showed that larval pigmentation can be induced in *An. gambiae* larvae as phenotypic response to different colour environments. This phenomenon i.e the ability of organisms to alter its morphology in different environments has been demonstrated in other animals. Examples include the predator-induced morphological defences in *Daphnia pulex* (Schneider and Berrigan 1998); induced melanization in the juvenile stages of the European map butterfly (*Araschnia levana*) and the peacock butterfly (*Inachis io)* (Windig, 1999); coat colour polymorphisms in rock pocket mice, *Chaetodipus intermedius* (Majerus and Mundy

2003) and quite recently in some species of *Anopheles* mosquitoes (Benedict *et al.*,1987, 1996a). It is known that when *Anopheles* is reared on an illuminated dark background, larvae darken significantly and *Anopheles* background-colour can stimulated larval colour change (Benedict *et al.*, 1996). *An. gambiae* and its closely related species in has been shown to have marked increase in melanism in Sudan (Aboud, 2003). However, none of previous investigators had tested the effect of induced melanisation on life-history traits in populations of *An. gambiae*.

The study also showed that different environmental colour cues can induce a bi-directional effect on *An. gambiae* population (Figure 23). Over 70% of dark and pale larvae were selected in dark and white background containers respectively. That is, larvae reared in a white rearing container resulted in paler larvae and larvae reared in black rearing container resulted in darker larvae compared to the parental larvae. This observation can be viewed in the context of adaptive evolution in heterogeneous environments which is very important in contemporary adaptation in natural populations (Reznick and Ghalambor 2001). Extreme environments have not only been shown to disrupt normal development of organism but also induce large phenotypic changes in new directions, and these changes simultaneously exert strong phenotypic selection that favours changes in these directions (Bradshaw and Hardwick 1989; Jablonka *et al.*, 1995). However, the ecological significance of these phenotypic changes remains to be investigated. Other environmental factors including temperature (Brady and Jones 1994; Bayoh and Lindsay, 2003, 2004 ; Lyimo *et al.*, 1992 ; Stillwell *et al.*, 2007), food availability and photoperiod (Kooi and Brakefield 1996; Reilly *et al.*, 2006) and stress (Badyaev, 2005) have been known to induce variation in populations of organism including *An. gambiae* (Bayoh and Lindsay, 2003, 2004; Lyimo *et al.*, 2002). Also, Lehmann *et al.* (2006) recently demonstrated that there is a genetic component to the variation observed in the developmental time, adult size and longevity of starved wild populations of *An. gambiae*. However, we did not think that there was any genetic component in the colour variation observed in this study neither do we think that any of these factors listed above affected the pattern of pigmentation observed because these environmental factors were kept constant during the induction experiment.

We hypothesize that since melanin is a complex polymer, its synthesis may be constrained if ambient conditions limit the resource budget, and in this case one will expect a trade-off between melanisation and its fitness components. In this study, larvae in the dark environment produced larger pupae than those in the white environment at a relatively short developmental time. However, dark larvae had a much lower survival rates than pale.

Although, one will expect that if the theory 'bigger is better' is true, then melanised larvae should have a much better chance of adapting to the dark rearing environment. However, this study indicated the reverse i.e smaller pale non-melanised larvae induced in white containers had high survival rates than the large darker individuals reared in black containers. Furthermore the with the exception of pupal weight, estimates of life-history traits of parental larvae reared in normal containers and those reared in the white containers were not significantly different (Figure 24).

There appears to be a trade-off between melanin production and life history traits such that dark individuals had faster development time but had low survival rates. This finding is in contrast to what has been found in other species such as European map butterflies (Windig, 1998) and recently in Satyrine butterflies (Talloen, 2004) where melanised larvae grew slower than the non-melanized ones. However, several studies in *Lepidoptera* and *Drosophila* species have shown that short juvenile development time and large adult size are highly relevant to fitness (Nylin and Gotthard 1998), and so is variation in melanization (Majerus, 1998). Therefore, in the present study, if the dark colour induced in the larvae is at least initially is a predator avoidance behaviour, then one would expect that naturally, they will have a much higher survival rates than the pale or parental larvae whose colour makes them vulnerable to potential predators. However, in this study, larvae reared in the white containers also produced pale larvae in responce to the rearing white colour background. This presupposes that with the exception of larvae reared in transparent containers (controls), the colour change of larvae to the dark and pale forms may be an adaptation in response to the background colour for reason of either camouflaging (crypsis) themselves against predators or some other reasons which needs further investigation. However, the low survival rates of dark larvae are surprising in this study. The possible explanation is that since melanin production is costly, more energy in the form of food is used in its production

during the course of the development of the larvae and by the 4th stages this competition for food leads to the death of many larvae in the dark environment. This hypothesis needs to be further investigated in future studies. In adult moths of *Ephestia kuhniella,* melanised strains are more active relative to pale strain due to the presence of dopamine, a precursor of melanin and also a neurotransmitter.

Studies in Lepidoptera have suggested that both faster growth and larger size may be traded off against one another or against costly traits such as melanin production, which requires protein (Blois, 1978). Significant amount of dark pigmentation was observed on the head/thorax compared to the other segments of the body of the 4th instar larvae reared in containers with dark background. The adaptive significance of this pattern needs further elucidation.

Fast growth rate of dark 4th instars, relative to pale larvae observed in *An. gambiae* may be the consequence of their coloration. It is well known that dark insects such as butterflies heat up fast and have high growth rates (Dennis, 1993). Although temperature has been shown to have effect on larval development of *An. gambiae* (Bayoh and Lindsay 2004; Lyimo *et al.,* 1992), in this particular study the temperature was not significantly different in the rearing environments and therefore the effect of temperature could be discounted here. The study detected a increase of both DDC and PO Delta C_T values in the melanised larvae compared to the parental larvae. Delta C_T values correlates with the relative amount of RNA produced in the sample during the RT-PCR analysis. These two genes are important in the melanization pathway (Figure 2) and the DDC has been shown to be expressed during colour pattern formation in melanic tiger swallowtail butterflies (Koch *et al.,* 1998) and in the tanning of the egg chorion of *Ae. aegypti* (Li and Li 2006). This is the first of a study that has showed the direct link between elevated DDC and PO activity and formation of melanised phenotypes of *An. gambiae* under different environmental colour cues. However, considering the fact that our analysis involved a low sample size, we recommend a much detailed study in this area of research to ascertain the veracity of our finding.

*An. gambiae s.s.*is the most widely distributed and the most efficient and important vector of malaria in Africa (Gilles and Coetzee 1987). This species is most versatile

and most adaptable to many heterogeneous environments and considering the fact that the species breeds in small temporary pools which easily dries out, the need to maximize resources for fast growth and protection against predators is paramount. We hypothesize that unlike other organisms which have more permanent breeding sites where predation may not be high, the phenotypic plasticity (melanization of larvae) demostrated in response to colour cues by *An. gambiae* may be a costly but necessary adaptation to developmental constrains such as predation. The study concludes that the generation of adaptive melanised phenotypes is important for the survival of *An. gambiae* in nature and that may explain the selective constrains on DDC gene observed in Chapter 4. Melanism has been shown to be a dynamic trait both within and between larval instars and because the expression changes with the development of the larvae it means that melanism is likely to be a highly dynamic and plastic trait, and this has potential implications not only for quantifying melanism in individual insects, but also for understanding the adaptive value of melanism and the selection pressures associated with it.

CHAPTER 6 GENERAL DISCUSSION AND CONCLUSIONS

6.1. Natural selection and maintenance of nucleotide polymorphisms within Glutathione s-transferases in *An. arabiensis*

One of the important aims of evolutionary biology is to understand the various forces that control how populations and species evolve. In terms of molecular evolution, this involves understanding the contribution of genetic drift and natural selection to patterns of DNA variation (Kimura and Takahata 1983; Gillespie, 1991). Of interest to the evolutionary biologist in terms of adaptive evolution is the role of natural selection in shaping nucleotide polymorphisms of genes within the genome. Scanning the genome of *An. arabiensis* to identify genes that are important for fitness and adaptation is therefore crucial for understanding mechanisms involved in adaptive population divergence as well as the evolutionary history of the species. One of the traditional methods of investigating this is to use intra and interspecific comparisons of DNA sequence variation within and between species and determine 'outlier' loci (loci effected directly by positive selection). These loci (outliers) are scattered throughout the genome are responsible for deviant level of variation relative to the rest of the genome and in most cases diverge from empirical neutral expectations (Luikart *et al.,* 2003).

In this study we used a multi-locus approach to investigate locus-specific rather than genome-wide effects of DNA variation on the evolution of the species. This is because, it is only locus-specific effects that help to identify genes that contribute to fitness and hence the adaptation of the species.

6.1.1. Intraspecific variation between An. arabiensis populations

Intra-specific comparison of *An. arabiensis* samples provided marked evidence of sequence conservation of GSTs within the Cellia group with the exception of the GSTε2 locus which showed high percentage divergence. This conservation was evident in the fact that both *An. gambiae* and *An. arabiensis* share over 98% sequence identity in the coding region of all four GST loci. With the exception of locus GSTε2 which showed 86% sequence identity between *An. gambiae* and *An. stephensi,* all the

other three loci, GSTε1, GSTε6, GSTε8 showed between 97-98% identities. The sequence conservation observed implies that gene duplication events that resulted in the cluster of GST genes on chromosome 3R are likely to predate the division of Cellia from other anopheline sub-groupings i.e apparent radiation occurred before Cellia split (Besansky and Fahey 1997; Krzywinski *et al.,* 2001; Besansky *et al.,* 2003; Turner *et al.,* 2005). GSTs have been implicated in insecticide resistance in many species of insects (Fournier *et al.,* 1992; Tang and Tu 1994; Collins *et al.,* 2000; Hemingway 2000; Hemingway and Ranson, 2000; Lumjuan *et al.,* 2005). Similar findings were observed in a study by Low *et al.* (2007) who showed that GST orthologs and 24 duplication events occurred within genus *Drosophila* and subgenera *Sophophora* which include *D. melanogaster* species group. Also Lumjuan *et al.* (2007) confirmed that the GST supergene family is conserved between *Ae. aegypti* and *An. gambiae* and that several clear orthologs could be identified between the two but not in *Drosophila*. This means that over several million years since *An. arabiensis* and other members of the Cellia group diverge, GST gene cluster have undergone duplication events and still has slowly diverging orthologs. Gene duplication involved duplication of a region of DNA that contains a gene; it may occur as an error in during homologous recombination, a retrotransposition event, or duplication of an entire chromosome (Zhang, 2003). Gene duplication is known to be fundamental to the creation of substantial heterogeneity among the genomes of organisms (Ohta 2003; Moore and Purugganan 2003). This is because it is known that the two genes that exist after a gene duplication event (paralog genes) usually code for proteins with a different function and/or structure or loose their function. GST gene family is made up of a cluster of genes many of which are involved in detoxification of endogenous and xenobiotic substances including insecticides (Huang, 1998). This means that many of this genes may be under enomours selective pressure and mutational constraints. We infer from our findings that the occurrence of duplication events prior to the split of the Cellia (including *An. arabiensis*) may be necessary in the context of adaptation of this group of mosquitoes to heterogenous environments.

Also of interest is the high sequence divergence observed at the GSTε2 locus. However, although GSTε1 locus (region sequenced was shorter in length in this study) showed lower nucleotide diversity in the coding region (π=0.005) in our study, that for GSTε2 was much higher (π=0.015) than expected. The pattern of nucleotide

variation observed at this (i.e GSTε2) locus can be explained against the background that some authors have specifically implicated GSTε2 in the detoxifixation of insecticides including DDT in *An. gambiae* and *Ae. aegypti* (Ranson *et al.,* 2001; Ding *et al.,* 2003; Lumjuan *et al.,* 2005) and therefore this gene may be under some insecticide selective pressure as a result of its role in the generation of adaptive traits in the context of insecticide resistance. Generally, based on our findings and the review from other authors, one can deduce that the GST gene cluster may have special functions and these unique functions or temporal expression patterns may impose significant selective constraints on the cluster and this selective pressure may be important in shaping nucleotide polymorphisms in populations of *An. arabiensis.*

The study also found that GSTε8 locus in *An. arabiensis* had between 6-13% identities to other GSTs in the Epsilon class. This percentage identity was less than what Ortelli *et al.,* 2003 found in *An. gambiae* and much less than the 40% identity criterion (Hayes and Pulford 1995) required for the inclusion of GSTs into the Epsilon class. Although protein products of GSTε8 reacted with anti-sera raised against Epsilon class GSTs (Ding *et al.,* 2003), in line with the evidence from our study, we recommend that the classification of this particular gene as a member of the Epsilon class requires further investigation.

Another issue of concern in the classification of GSTs is the occurrence of pseudogenes (non-functional genes) within the GST clusters. Ding *et al.,* (2003) detected three unclassified GSTs in *An. gambiae* and clear orthologues of each of these unclassified genes were also found in *Ae. aegypti* mosquitoes (Lumjuan *et al.,* 2007). Again Lumjuan *et al.* (2007) did not detect transcript for GSTε1 in *Ae. aegypti* and therefore classified GSTε1 as a non-transcribed pseudogene. This means that pseudogenes do occur within the GST cluster. Considering the fact that GST orthologues are common within the Cellia, one has to be careful in distinguishing pseudogenes from functional one. With the exception of GSTε8, the striking parallel or homology observed in the GST sequences of *An. arabiensis, An. gambiae* and *An. stephensi* suggest that these genes may be undergoing parallel evolution in nature (independent evolution of similar traits starting from a similar ancestor due to similar environment or other evolutionary pressures). This phenomenon of parallel evolution is frequently seen in instances of insecticide resistance (Low *et al.,* 2007) and

therefore its importance in the evolution of GSTs, some of which has been known to detoxify insecticides cannot be overemphasized.

The study also detected evidence of weak selection at the GST loci as evident by the departure from expectations of Tajima's values under the standard neutral model. This departure was significant at some regions (Figures 13 & 14) of GSTε1, GSTε6 and GSTε8 genes when a detailed scan of the entire gene was done using the sliding window-plot. However, one has to be cautious in interpreting this data because we hypothesize that the lack of significance of Tajima's values may be due to the reduction in the power of the test as a result of the low sample sizes used in the analysis. Further studies using much larger sample may be needed to confirm our finding. This effect may be more apparent in the GSTε1 because we unable to obtain the full length of the gene.

The rejection of the neutral model in *An. arabiensis* populations may reflect the combined action of natural selection and demographic forces shaping genome variability patterns of *An. arabiensis* populations. The idea of demographic forces affecting the pattern of nucleotide variability of the species seems to play on well since the species has been shown recently to inhabit novel habitats (Donnelly and Townson 2000). However, we can rule out the effect of demographic forces including population expansion on the observed nucleotide variation because the effects of demographic forces are expected to be uniform across the genome whereas selection is locus-specific. Therefore the pattern of nucleotide variation observed at the three loci (i.e. GSTε2, GSTε6 and GSTε8) may be suggestive of the action of selection on the loci.

We also estimated recombination rates for GST genes and again based on the four gamete test (Hudson and Kaplan 1985), GSTε2 showed twice the number of recombination events/generation (R_M=10) in comparison with that of GSTε6 (R_M=5) and GSTε8 (R_M=5). This means that more recombination events have occurred between sites within the GSTε2 gene in their past history. In human, multiple recombination events have been known to cause heterogeneity of thalassemia haplotypes (Fodde *et al.,* 1991). As discussed earlier, GSTε2 has been specifically implicated in the detoxification of DDT and other insecticides. Based on this premix,

we hypothesize that the multiple recombination events in the presence of selection i.e. albeit insecticide pressure may had led to the great degree of genetic divergence observed at the locus. If this is generally true, this finding may reflect the fact that a high rate of substitution is required at locus GSTε2 in order to generate adaptive phenotypes within populations of *An. arabiensis* in the context of the development of insecticide resistance phenotypes. Also we suggest that the non-uniformity of recombination rates estimated for the GST loci across the genome indicates that these genes evolve independently.

The pattern of linkage disequilibrium observed at the GST loci in this study is quite interesting. This is because it is known that how fast a pair of alleles at a given locus approach linkage equilibrium is primarily a function of the recombination rate between the two loci and recombination is known to reduce linkage disequilibrium. Based on this premix, one will expect a much reduced non-random associations between the GST loci in the presence of recombination. However, this was not the case in our findings. We detected evidence of linkage disequilibrium within the GST loci and GSTε8 locus much more showed significant linkage disequilibrium after Bonferroni correction in the presence of recombination. Although, it may be difficult to assign any one reason for the observed pattern, but we can attempt to provide some explanations based on the physical location of these loci within the genome and the effect of selection. *An. gambiae* epsilon class GSTs are located within inversion 3Ra of chromosome 3. Chromosomal inversions have been found to be associated with disequilibrium among loci and they have the potential to lock up co-adapted genes by suppression recombination (Hoffmann *et al.,* 2004). Therefore the observed pattern is to be expected for genes closely linked to inversion breakpoints (Andolfatto *et al.,* 1999; Andolfatto and Kreitman 2000). Inversions are frequent within *An. gambiae* genome and their frequency varies in the sample collection sites. Ethiopia has the highest *An. gambiae* 3Ra inversion frequency of 0-35% (Abose *et al.,* 1998) and as discussed in the next section (interpopulation comparisons), Ethiopia populations of *An. arabiensis* are highly differentiated from the rest of the samples. Also in Sudan, the degree of polymorphism for chromosomal paracentric inversion 3Ra was high compared to that of *An. gambiae* (Petrarca *et al.,* 2000). It is known that inversions are able to maintain favourable groups of alleles at loci within them (Coluzzi, 1982) and therefore they are expected to influence recombination and linkage disequilibrium

and hence the general pattern of DNA variation. It is not known how far inversion frequencies may have influenced the pattern of nucleotide variation observed in our sampling sites. Hitherto, the role of inversions in DNA variation has been overlooked in many cases and we recommend that this needs to be considered in interpreting DNA sequence variation.

The non-random association of alleles within the GST loci will be important for Linkage disequlibrium (LD) mapping since LD mapping relies on the markers being in disequilibrium to the genes that aids in the expression of the phenotypes (Black *et al.,* 2008) and for a pest species such as *An. arabiensis,* linkage disequilibrium within GSTs will facilitate their being used for genotype-phenotype mapping. Linkage mapping using single nucleotide polymorphisms (SNPs) clustered around detoxification enzymes as cytochrome P450s and GSTs is currently underway and information from our study will help formulate new approaches to LD mapping.

6.1.2. Interpopulation comparison of An. arabiensis

The study showed that the populations of *An. arabiensis* from Ethiopia were highly differentiated from the rest of the samples. This was evident from the significant genetic differentiation, F*st* estimates observed in comparison with samples from Sudan, Malawi and Tanzania. This pattern of differentiation was consistent with findings of other microsatellite-based and mtDNA studies (Lehmann *et al.,* 1997; Donnelly and Townson 2000; Kayondo *et al.,* 2005). As proposed by Donnelly and Townson (2000), we will expect a locus under locus-specific selection to have high F*st* values i.e values that will exceed the expected variance (Figure 27).

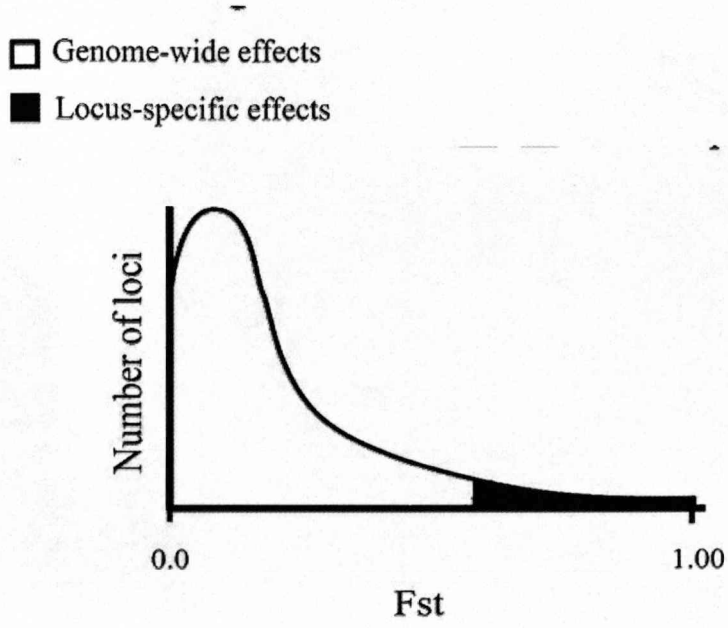


Figure 27: Hypothetical sampling distributions of F_{st} among loci distributed throughout a genome. *F*st follows a Poisson (random) distribution owing largely to genome-wide effects. However, *F*st at the locus that is under selection will exceed the expected variance because it is subject to locus specific selection and will fall within the black region of the sampling distribution (Source: Modified from Black *et al.,* 2001)

The study also showed extensive gene flow between populations of *An. arabiensis.* Microsatellite-based analysis also showed high level of gene flow estimated for *An. arabiensis* in Tanzania and Ethiopia (Donnelly *et al.,* 2004). This pattern is most worrying because in terms of control of the vector, it may facilitate the spread of insecticide resistant genes which if under selection will spread rapidly.

The isolation of Ethiopian populations from the rest may be due to the effect of Great Rift Valley which restricted migration and hence gene flow between *An. arabiensis* populations. Ethiopia occupies most of the Horn of Africa and it shares frontiers with Sudan, Kenya, Somalia, and Djibouti. The major physiographic features are a massive highland complex of mountains and plateaus divided by the Great Rift Valley and surrounded by lowlands along the periphery. Much of the Ethiopian landmass is part of the East African Rift Plateau. Some microsatellite and mitochondrial DNA based

studies has implicated the Great Rift Valley in East Africa as an important barrier to gene flow in *An. gambiae* (Kamau *et al.,* 1998, 1999) and in *An. arabiensis* populations (Donnelly and Townson 2000). The Great Rift Valley has also been shown to affect the population structure of *An. arabiensis* populations in Ethiopia (Nyanjom *et al.,*2003). The specimens used for this analysis were collected from different geographical areas, i.e. Sudan, Ethiopia, Tanzania and Malawi where the pressure of insecticide usage may be different. We propose that selection at the GST loci due to insecticide pressure will lead to alleles being fixed in the different populations and therefore create an excess of heterozygotes in the population. We might therefore expect estimates of F*st* to be far greater than zero (Figure 27). However, one must be careful in interpreting the high F*st* estimates because the high values may reflect the choice and location of loci within the genome.

Over several million years of the evolution of GST genes, apparent sequence conservation within the Cellia group may be indicative of the importance of this gene cluster in adaptive response to selective pressure in heterogeneous environments. Epsilon class multigene family represent an example of genes that may have provided important insight into the recent population history that microsatellites and mtDNA variation have not been able to reveal and it is hoped that future studies will take advantage of this insight whenever possible.

In conclusion, based on our findings we postulate the gene duplication events which occurred within the Cellia group of mosquitoes over the evolutionary time- scale may predate the split of the subgenus such that the nucleotide sequence were much conserved within the subgenus. We also demonstrated that GSTe2 may be important in the generation of adaptive phenotypes in natural populations of *An. arabiensis* with reference to its role in the detoxification of DDT and other insecticides.

6.2 Evolution of *Dopa decaboxylase* gene and its role in the generation of divergent phenotypes

In Chapter 4 of this thesis, based on phylogenetic analysis, the DDC locus showed marked sequence conservation between *An. arabiensis*, *Aedes* and *Drosophila* species. This pattern showed that the locus is constrained with regard to the pattern of amino acid substitutions. The low recombination rates and the strong non-random

association of the locus to other nearby genes within the 3Ra inversion of chromosome 3 may reflect the action of selection on the DDC and the subsequent hitchhiking effect on advantageous substitution occurring within the GST loci. A high degree of sequence conservation has allowed comparisons of the DDC-coding regions from various insects, facilitating a number of recent studies on insect systematics (Hodgetts and O'Keefe 2006).

A study of nucleotide variation of DDC in *Drosophila* species showed that the three species groups studied, *melanogaster*, *obscura*, and *willistoni* were each monophyletic and all three combined form a monophyletic group, which corresponds to the subgenus Sophophora (Tatarenkov *et al.,* 2001). Linkage disequilibrium analysis shows presence of two groups of haplotypes in the populations, each of which is fairly diverged, suggesting epistasis or inversion polymorphism. Also the McDonald-Kreitman test indicates a deficit of fixed amino acid differences between *D. melanogaster* and *D. simulans*, which may be due to negative selection. An excess of derived alleles at high frequency, significant according to the *H*-test, is consistent with the effect of hitchhiking (Tatarenkov and Ayala 2007). De Luca *et al.* (2003) applied the test of selection on DDC in samples of *Drosophila* but did not detect any significant departure from neutrality. However, he observed considerable linkage disequilibrium between polymorphic sites throughout the region.

A study demonstrated sufficient conservation between DDC sequences of *D. virilis* and *D. melanogaster* (Bray and Hirsh 1986) and others have shown evidence of duplication events within the DDC of *Drosophila* (Eveleth and Marsh 1986). The same authors showed for the first time the structural homology of DDC and other genes in *Drosophila* and suggested that the functional relationship between the DDC and other genes may be that they both share common evolutionary histories.

It has also been shown that DDC occurs as a cluster of genes which are functionally related and it was postulated that the clustering of the genes is necessary to ensure that their regulation could be coordinated (Wright, 1996). It was shown that even with the proximal and distal clusters of DDC separated by a large segment of the chromosome in *D. virilise* and *D. pseudoobscura,* the clusters were identical in the two species and even more importantly in *An. gambiae*. In line with our findings, we suggest that

some functional rationale may underlie the clustering of these genes in *An. arabiensis* genome.

We therefore propose two models to explain the pattern of nucleotide variation observed at the DDC locus: 1) parallel evolution may be maintaining the functional role of DDC in all the species surveyed, 2) positive selection may be acting on the DDC in *An. arabiensis* and this facilitates the generation of adaptive phenotypes in nature. Further studies using much larger sample size is recommended to assess the veracity of the explanation. We suggest that given the pleotropic effect (genetic effect of a single gene on multiple phenotypic traits) of DDC on many phenotypes, the DDC can be excellent candidates for the maintenance of natural variation in *An. arabiensis* populations.

6.3 Effect of melanization on life-history traits of *An. gambiae*

6.3.1 Relationship between melanisation and life-history traits

To understand the evolutionary and ecological significance of phenotypic plasticity in nature, it is important to know how phenotypic traits are generated. In chapter 5, we induced phenotypic plastic response in the 1st instars of *An. gambiae* by rearing larvae in different colour environments and generating melanic and non-melanic phenotypes. We observed a bi-directional response in the response of the larvae to the rearing containers. This means that dark containers induced melanised larvae whereas white ones induced pale larvae relative to the parental colouration.

The present study showed that larvae reared in dark containers produced bigger larvae but had a low survival rate and short developmental time whereas smaller pale larvae produced from white containers had a much higher survival rate and a much longer developmental time. This finding contradicts what Windig (1999) discovered in a *Lepidoptera, Araschnia levana.* In that study, it was demostrated that melanized 5th instars of *A. levana* grow more slowly in early instars. However, one has to be cautious in comparing *A. levana* and *An. gambiae* since each has its own unique habitats and therefore each may be under different selective pressures.

As discussed in Chapter 5, temperature has been shown to affect the development of *An. gambiae* (Bayoh and Lindsay 2003, 2004; Impoinvil *et al.,* 2007). However, in

this particular study because water temperature was constant and larvae were fed the same amount of food, it means that other factors other than temperature and food may be responsible for the fast growth and large size of larvae in the dark containers. We can however hypothesize that the fast developmental time of larvae reared in dark containers in comparison to the pale larvae observed in *An. gambiae* may be the consequence of their coloration. Further research is needed to fully understand the mechanisms involved. The study established that melanisation of larvae did affect the life-history traits of juvenile *An. gambiae* populations and this effect has also been demonstrated in other insect species as in *Inachis io* and *Araschnia levana* (Windig 1999) and in *Helicoverpa armigera* (Ma *et al.,* 2007).

6.3.2 Phenotypic plasticity of An. gambiae larval colour trait

It has been assumed that phenotypic plasticity acts as a constraint on evolutionary change, however, it has become increasing clear that phenotypic plasticity actually represents a fundamental component of evolutionary change (Thompson, 1991). It is also now clear that genotypes that perform best in one environment usually perform less well than other genotypes in a different environment, and therefore a plastic response is not necessarily an adaptation to the environmental. This is because a response to environmental variation is only adaptive if it represents a mechanism by which relative fitness is maintained in the face of environmental variation (Thompson, 1991).

In this study we were able to demonstrate bi-directional selection of dark and pale *An. gambiae* larvae in the laboratory using the background colour of rearing containers as the environmental cue. The mechanism underlying this colour change is not clear, however, in a similar experiment by (Benedict and Seawright 1987), it was suggested that the noticeable colour change was stimulated by perception of the background colour by the larval ocelli which in turn send signals to the neurophysiological pathway for the responce.

During the induction, we used inbred lines of *An. gambiae*. This was to ensure that all the larvae used for the induction had a uniform genetic background prior to their introduction into the rearing containers. The appearance of dark larvae in the dark containers may therefore signal a plastic rather genetic response to colour stimuli. It

can also be suggested that the elevated activity of DDC and PO in the melanised phenotypes during the induction process may be an indication that these two genes are involved in the melanin synthesis pathway and therefore are important in the development of the plastic response observed in the study. Other genes not included in the current study, such as the *ebony* (Wittkopp *et al.,* 2003a;Takahashi *et al.,* 2007) and *tan* (True *et al.,* 2005) have been implicated in the development of pigments in insects. It is known that like most traits, pigmentation is controlled by regulatory genes, such as transcription factors, which control the expression of other genes although different regulatory genes might control their expression in different body regions (Wittkopp *et al.,* 2003b). The question is whether it is the same genes that controlled the observed regional variation in larval body pigmentation. Further studies will be needed to investigate whether it is the same DDC and PO genes that are expressed in the different areas (head, thorax and abdomen) of the *An. gambiae* larvae.

6.3.3 Trade-offs and cost

Although it is known that natural selection favors certain kinds of plastic response in populations, there may be some measurable cost to maintaining plasticity, as well as limits to the ability of an organism to be adaptively plastic (DeWitt *et al.*, 1998).

In this study, the importance of the generation of melanised phenotypes of *An. gambiae* was tested against its effect on some life-history traits such as developmental time, survival rates and pupal weight. We detected that melanised larvae had shorter developmental time and much larger pupal weight when compared to the pale and parental individuals (Figure 24). However, the survival rates of the menalised larvae was lower, though not significant than the pale and parental larvae. Previous studies on the European map butterflies (Windig, 1999) and Satyrine butterflies (Talloen *et al.,* 2004) found that melanised larvae grew slower than the non-melanized ones. Also studies in *Lepidoptera* and *Drosophila* species have linked short juvenile development time and large adult size (Nylin and Gotthard 1998) and variation in melanization (Majerus, 1998) to be important for the fitness of the species.
Therefore, the observed behaviour (melanised larvae with short developmental time and low survival rate) in *An. gambiae* may be a trade-off between being able to

develop mechanisms (camouflage) to avoid being preyed upon and its ability to survive in its habitat. Since melanin production is known to be costly, it means that the larvae in the dark rearing containers may be competing for the available food resources to produce melanism for use as camouflage and as a result those larvae that that do not get adequate food resources die before the pupation stage. The behaviour may therefore be viewed in the context of adaptation to different heterogeneous environments i.e different organisms will develop adaptation in relation to the peculiarity of its breeding habitats. In case of *An. gambiae*, the darkening of the cuticle and the subsequent fast development time may be a predator avoidance behaviour. Further investigation on the non-additive effects of predators on plasticity will need to be conducted in further to confirm this accession. Further studies will also be necessary to ascertain the effect the colour change in larvae will have on the development of adult individuals.

The study concludes that the generation of adaptive melanised visible phenotypes although costly is important for the survival of *An. gambiae* in nature especially in avoiding potential predators, and that may explain the selective constrains on DDC gene observed in Chapter 4. However, the question that will continue to challenge and inform future research is that if plasticity is adaptive, at what stage and how does natural selection favour or maintain the phenotypic phenotypes? Issues about costs of plasticity are also difficult to detect (Dewitt *et al.,* 1999; Karan *et al.,* 2000), although more recently they have been detected in plants (Agrawal e*t al.,* 2002), ; juvenile treefrogs, *Oecologia* (Relyea and Hoverman 2003) and in the common frog, *Rana temporaria* populations (Merila *et al.,* 2004). However, we hypothesize that the cost of melanin production in darker individual larvae observed in the study, although adaptive in the context of predator avoidance strategy may be responsible for the low survival rates estimated. Other authors demonstrated that compared to the pale individuals, the development of darker individuals was slower and less stable as estimated by the level of fluctuating asymmetry of the Satyrine butterfly, *Pararge aege* (Talloen *et al.,* 2004). However, the trade-off hypothesis is not supported by some authors. A study investigated the effect of melanisation on antibacterial immune response of *An. gambiae* (Lambrechts *et al.,* 2004) and determined that genetic association between the melanization response and an antibacterial response in well-fed and undernourished larvae were positively genetically correlated (Lambrechts *et*

al., 2004). This meant that melanised larvae were able to clear injected bacteria irrespective of whether they were well fed or undernourished and hence the trade-off hypothesis was not supported in this case.

We demonstrated in this study that DDC and PO genes are up-regulated in the development of melanised phenotypes of *An. gambiae*. Some authors (Koch *et al.,* 1998) also demonstrated the regulation of DDC expression during colour formation in wild-type and melanic swallowtails. In consonance with our study objectives, we selected only DDC and PO for our gene expression analysis, therefore we do not rule out other genes that may be involved in the development of melanised phenotypes. For example, some authors have shown the *yellow* (Wittkopp *et al.,* 2003b) and *tan* (True *et al.,* 2005) genes are also involved at some stages of the melanisation pathway. The *yellow* gene converts DOPA to DOPA melanin whislt the *tan* gene converts Dopamine to N-acetyl dopamine (NBAD) in the biochemical pathway of tanning and melanisation (Figure 2).

Information on the effect of melanization and its adaptive significance in natural populations of *An. gambiae* will require multiple lines of evidence on the role of trade-offs, phenotypic plasticity and its effect on life-history traits in the development of both the larvae and adults. We also recommend further studies into the impact of the melanic trait on the adult life of *An. gambiae*. It will be a very useful and important area of research to pursue in future in line of the evidence discussed in this study.

6.4 Conclusions and implications for malaria control

Our study has provided insight into three different areas that have bearing on the ability of *An. gambiae* to adapt to changing environments.

a) The first finding is that *An. arabiensis* uses GSTs to acquire diverse phenotypes through gene duplication events. This is important for maintaining the function of these genes within the species as well as between different members of the Cellia group (*An. gambiae* and *An. stephensi*). We also showed that GSTε2 may be important in the generation of adaptive phenotypes within the Cellia group.

b) The second finding is that unlike GSTs, the DDC cluster acquires diverse phenotypes through a single gene. This gene shows much pleotropic effect i.e. plays different roles in the survival and adaptation of *An. arabiensis* to heterogeneous environments.

c) The third finding of our study showed it is possible to induce bi-directional selection of *An. gambiae* melanic phenotypes. We further showed that the induction is modulated by the expression of genes, in this case *Phenoloxidase* (PO) and *Dopa decaboxylase* (DDC) and that the survival and hence longevity (fitness) of juvenile *An. gambiae* is affected by the characteristics of the rearing environments.

An. gambiae s.s. and *An. arabiensis* are important vectors of malaria in Africa and several mosquito-related and environmental factors influence the vectorial competence of these two species. Generally the study provides some information necessary to understand how nucleotide polymorphisms are maintained by genes under selection and also shows how the local ecology (environment) affects the fitness of the malaria vectors. The information generated from this study can be used a basis for future studies and also for formulation of future malaria control interventions such as the use of transgenic mosquitoes for malaria control in order to formulate strategies for successful vector control.

6.5 Recommended future studies

Based on the findings of our study, we wish to recommend the following areas of interest for future research. More information on these areas will provide a more comprehensive picture of the mechanisms involved in the expression of phenotypic plasticity and also allow one to gain more insight into strategies that enables *An. gambiae* to adapt to different habitats. We will recommend the following for future studies:

a) A study into the role of other genes such as e.g. *ebony*, *yellow*, *tan* etc in the expression on *An. gambiae* melanised phenotypes during the induction process. This may be done by inducing melanised phenotypes of *An. gambiae*

larvae and using RT-PCR analysis to investigate the expression profile of *ebony*, *yellow, tan* and any other genes known to be involved in the melanisation pathway.

b) Determine the effect of melanisation on the life-history traits of adult mosquitoes. This study may be done my selecting melanised and non-melanised forms of *An. gambiae* larvae (as in this study) and raising the larvae through several generations with a view to finding out how the change in larval colour will affect the colour, survival, fecundity rates, hatching rate and the general morphology of adults.

c) Predator-induced morphological and behavioural change has been observed in many organisms and it is known that melanin production is also linked to the protection from adverse effect of UV-light. It will be interesting to investigate if the introduction of a predator during the induction process will affect the number of larvae being melanised and also what effect UV-light will have on the induction response when used instead of the normal florescent light.

d) Do RNA interference studies by introducing RNAi (homologous double stranded RNA to specifically target gene's product of DDC and PO) and investigate its effect on phenotypic response. The Specific RNAi pathway proteins will "cleave" to the target mRNA, breaking it down into smaller portions that can no longer be translated into protein. The absence of observable colour change in *An. gambiae* will show that DDC and PO are important for the expression of melanin in larvae.

REFERENCES

Aboud, M.A.E.M (2003) Ecology and population structure of *Anopheles arabiensis* in Sudan. PhD thesis, The University of Liverpool, Liverpool, United Kingdom

Abose, T., Yeebiyo, Y., Olana, D., Alamirew, D., Beyene, Y., Regassa, L. and Mengesha, A. (1998) Re-orientation and definition of the role of malaria vector control in Ethiopia. *World Health Organization/Malaria/*, **98,**1085

Agianian, B., Tucker, P.A., Schouten, A., Leonard, K., Bullard, B. and Gros, P. (2003) Structure of a *Drosophila* sigma class *glutathione S-transferase* reveals a novel active site topography suited for lipid peroxidation products. *Journal of Molecular Biology,* **326,** 151-65.

Agrawal, A. (2001) Phenotypic plasticity in the interactions and evolution of species. *Science,* **294,** 321-326.

Agrawal, A.A., Conner, J.K., Johnson, M.T. and Wallsgrove, R. (2002) Ecological genetics of an induced plant defense against herbivores: additive genetic variance and costs of phenotypic plasticity. *Evolution International Journal of Organic Evolution,* **56,** 2206-2213

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

Alatalo, R.V. and Moreno, J. (1987) Body size, interspecific interactions, and use of foraging sites in tits (Paridae). *Ecology,* **68**, 1773–1777

Alonso, P.L., Lindsay, S.W., Armstrong, J.R., Conteh, M., Hill, A.G., David, P.H., Fegan,G., de Francisco, A., Hall, A.J. and Shenton. F.C. (1991). The effect of insecticide-treated bed nets on mortality of Gambian children. *Lancet,* **337,** 1499-1502.

Alphey, L. and Andreasen, M. (2002) Dominant lethality and insect population control. *Molecular and Biochemical Parasitology*, **121**, 173–178.

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

Ananthakrishnan, T. N. and Whitman, D. (2005) Insect Phenotypic Plasticity:diversity of Responses. *Entomology,* : 222

Andolfatto, P. (2001) Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans. Molecular Biology Evolution,* **18,** 279-290.

Andolfatto, P. and Kreitman, M. (2000) Molecular variation at the In(2L)t proximal breakpoint site in natural populations of *Drosophila melanogaster* and *Drosophila simulans. Genetics*, **154,** 1681-1691.

Andolfatto, P., Wall, J.D. and Kreitman, M. (1999) Unusual haplotype structure at the proximal breakpoint of In(2L)t in a natural population of *Drosophila melanogaster*. *Genetics*, **153,** 1297-1311.

Awolola, T.S., Oyewole, I.O., Amajoh, C.N., Idowu, E.T., Ajayi, M.B., Oduola A., Manafa, O.U., Ibrahim, K., Koekemoer, L.L. and Coetzee, M. (2005) Distribution of the molecular forms of *Anopheles gambiae* and pyrethroid knock down resistance gene in Nigeria. *Acta Tropica*, **95,** 204-209.

Ayala, F.J. and Coluzzi, M. (2005) Chromosome speciation: humans, *Drosophila*, and mosquitoes. *Proceedings of the National Academy of Science,* USA, **102,** Suppl 1: 6535-6542.

Balkew, M., Gebre-Michael, T. and Hailu, A. (2003) Insecticide susceptibility level of *Anopheles arabiensis* in two agrodevelopment localities in eastern Ethiopia. *Parassitologia,* **45,** 1-3.

Badyaev A. (2005) Stress-induced variation in evolution : from behavioural plasticity to genetic assimilation. *Proceedings of the Royal Society of London*, **272**, 877-886

Ballinger-Crabtree, M.E., Black, W.C. and Miller, B.R. (1992) Use of genetic polymorphisms detected by the polymerase chain reaction (RAPD-PCR) for differentiation and identification of *Aedes aegypti* subspecies and populations. *American Journal of Tropical Medicine and Hygiene,* **42,** 893-901.

Baird, J.K. (1998) Age-dependent characteristics of protection v. susceptibility to *Plasmodium falciparum. Annals of Tropical Medicine and Parasitology,* **92,** 367-390.

Barbazan, P., Baldet, T., Darriet, F., Escaffre, H., Djoda, D.H. and Hougard, J.M. (1998) Impact of treatments with *Bacillus sphaericus* on *Anopheles* populations and the transmission of malaria in Maroua, a large city in a savannah region of Cameroon. *Journal of American Mosquito Control Association,* **14,** 33-39.

Bayoh, M.N. and Lindsay, S.W. (2004) Temperature-related duration of aquatic stages of the Afrotropical malaria vector mosquito *Anopheles gambiae* in the laboratory. *Medical and Veterinary Entomology,* **18,** 174-179.

Bayoh M.N. and Lindsay S.W. (2003) Effect of temperature on the development of the aquatic stages of *Anopheles gambiae* sensu stricto (Diptera: Culicidae). *Bulletin of Entomological Research*, **93**, 375-381

Beaumont, M.A. (2005) Adaptation and speciation: what can F(st) tell us? *Trends in Ecology and Evolution*, **20,** 435-440.

Beerntsen, B.T., James, A.A. and Christensen, B.M. (2000) Genetics of mosquito vector competence. *Microbiology and Molecular Biology Review,* **64,** 115-137.

Benedict, M.Q., Besansky, N.J, Chang, H, Mukabayire, O. and Collins, F.H. (1996a) Mutations in the *Anopheles gambiae* pink-eye and white genes define distinct, tightly linked eye-color loci. *Journal of Hereditary,* **87,** 48-53

Benedict, M.Q. and Chang, H. (1996) Rapid isolation of *anopheline* mosquito eye-colour mutants based on larval colour change. *Medical and Veterinary entomology*, **10,** 93-96

Benedict, M.Q. and Seawright, J.A. (1987) Changes in pigmentation in mosquitoes (Diptera: Culicidae) in response to the color of the environment. *Annals of the Entomological Society of America,* **80,** 55-61.

Besansky, N. J. and Fahey, G.T. (1997) Utility of the white gene in estimating phylogenetic relationships among mosquitoes (Diptera: Culicidae). *Molecular Biology Evolution*, **14,** 442-454.

Besansky, N.J., Lehmann, T., Fahey, G.T., Fontenille, D., Braack, L.E., Hawley, W.A. and Collins, F.H. (1997) Patterns of mitochondrial variation within and between African malaria vectors, *Anopheles gambiae* and *An. arabiensis*, suggest extensive gene flow. *Genetics,* **147**, 1817-1828.

Besansky, N.J., Krzywinski, J., Lehmann, T., Simard, F., Kern, M., Mukabayire, O., Fontenille, D., Toure, Y. and Sagnon, N. (2003) Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. *Proceedings of the National Academy of Sciences,* USA, **100,** 10818-10823.

Besansky, N.J., Mukabayire, O., Benedict, M.Q., Rafferty, C.S., Hamm, D.M. and McNitt, L. (1997a). The *Anopheles gambiae* tryptophan oxygenase gene expressed from a baculovirus promoter complements *Drosophila melanogaster* vermilion. *Insect Biochemistry and Molecular Biology,* **27,** 803-805.

Binka, F.N., Kubaje, A., Adjuik, M., Williams, L.A., Lengeler, C., Maude, G.H., Armah, G. E., Kajihara, B., Adiamah, J.H. and Smith, P.G. (1996) Impact of permethrin impregnated bednets on child mortality in Kassena-Nankana district, Ghana: a randomized controlled trial. *Tropical Medicine and International Health,* **1,** 147-154.

Black, W.C., Baer, C.F., Antolin, M.F. and DuTeau, N.M. (2001) Population genomics: genome-wide sampling of insect populations. *Annual Review of Entomology,* **46,** 441-469.

Black, W.C., Gorrochetegui-Escalante, N., Randle, N.P. and Donnelly, M.J. (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,*71-83.

Blois, M.S. (1978) The melanins: their synthesis and structure. In. *Photochemical and Photobiological Reviews.* Smith, K.C. (ed.), **3,** pp. 115-134, Plenum Press, New York.

Board, P.G., Coggan, M., Chelvanayagam, G., Easteal, S., Jermiin, L.S., Schulte, G.K., Danley, D.E., Hoth, L.R., Griffor, M.C., Kamath, A.V., Rosner, M.H., Chrunyk, B. A., Perregaux, D.E., Gabel, C.A., Geoghegan, K.F. and Pandit, J. (2000) Identification, characterization, and crystal structure of the Omega class *glutathione transferases*. *Journal of Biological Chemistry,* **275,** 24798-24806.

Bochdanovits, Z. and De Jong, G. (2003) Temperature dependent larval resource allocation shaping adult body size in *Drosophila melanogaster. Journal of evolutionary biology,* **16**, 1159–1167

Bonin, A., Taberlet, P., Miaud, C. and Pompanon, F. (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution,* **23,** 773-783.

Braby, M.F. and Jones, R.E. (1994) Effect of Temperature and Hostplants on Survival, Development and Body-Size in 3 Tropical Satyrine Butterflies From North-Eastern Australia. *Australian Journal of Zoology*, **42,** 195 – 21

Bradshaw, A.D. and Hardwick, K. (1989) Evolution and stress—genotypic and phenotypic components. *Biological Journal of the Linnean Societ,.* **37**:137–155.

Bray, S.J. and Hirsh, J. (1986) The *Drosophila virilis dopa decarboxylase* gene is developmentally regulated when integrated into *Drosophila melanogaster. The European Molecular Biology Organization Journal,* **5,** 2305-2311.

Bray, S.J. and Kafatos, F.C. (1991) Developmental function of Elf-1: an essential transcription factor during embryogenesis in *Drosophila. Genes and Development,* **5,**1672–1683

Breman, J.G. (2001) The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *American Journal of Tropical Medicine and Hygiene,* **64,** 1-11.

Brogdon, W.G. and McAllister, J.C. (1998) Insecticide resistance and vector control. *Emerging Infectious Diseases Journal,* **4,** 605-613.

Bruce-Chwatt, L.J. (1985) John Hull Grundy lecture. Mosquitoes, malaria and war; then and now. *Journal of Royal Army Medical Corps,* **131,** 85-99.

Bryan, J.H., Petrarca, V., Di Deco, MA. and Coluzzi, M. (1987) Adult behaviour of members of the *Anopheles gambiae* complex in the Gambia with special reference to *An. melas* and its chromosomal variants. *Parassitologia,* **29,** 221-249.

Brisson, J.A., De Toni, D.C., Duncan, I. and Templeton, A.R. (2005) Abdominal pigmentation variation in *Drosophila polymorpha*: geographic variation in the trait, and underlying phylogeography. *Evolution International Journal of Organic Evolution,* **59,** 1046-1059.

Buskirk, J., Van McCollum, S.A and Werner, E.E. (1997) Natural Selection for Environmentally Induced Phenotypes. *Evolution*, **51**, 1983-1992

Carter, R., Mendis, K.N. and Roberts, D. (2000a) Spatial targeting of interventions against malaria. *Bullentin of World Health Organization,* **78,** 1401-1411.

Carter, R., Mendis, K.N., Miller, L.H., Molineaux, L. and Saul, A. (2000b) Malaria transmission-blocking vaccines--how can their development be supported? *Nature Medicine,* **6,** 241-244.

Chandramohan, D., Owusu-Agyei, S., Carneiro, I., Awine, T., Amponsa-Achiano, K., Mensah, N., Jaffar, S., Baiden, R., Hodgson, A., Binka, F. and Greenwood, B. (2005) Cluster randomised trial of intermittent preventive treatment for malaria in infants in area of high, seasonal transmission in Ghana. *British Medical Journal,* **331,** 727-733.

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

Chapman, R.F. (1998) *The Insects. Structure and Function.* (4th edn) Cambridge: Cambridge University Press

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

Chouaibou, M., Simard, F., Chandre, F., Etang, J., Darriet, F. and Hougard, J.M. (2006) Efficacy of bifenthrin-impregnated bednets against *Anopheles funestus* and pyrethroid-resistant *Anopheles gambiae* in North Cameroon. *Malaria Journal*, **5,** 77.

Cisse, B., Sokhna, C., Boulanger, D., Milet, J., Ba el, H., Richardson, K., Hallett, R., Sutherland, C., Simondon, K., Simondon, F., Alexander, N., Gaye, O., Targett, G., Lines, J., Greenwood, B. and Trape. J.F. (2006) Seasonal intermittent preventive treatment with artesunate and sulfadoxine-pyrimethamine for prevention of malaria in Senegalese children: a randomised, placebo-controlled, double-blind trial. *Lancet,* **367,** 659-667.

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

Collins, F.H., Kamau, L., Ranson, H.A. and Vulule, J.M. (2000) Molecular entomology and prospects for malaria control. *Bullentin of World Health Organization,* **78,** 1412-1423.

Collins, F.S. and Haseltine, W.A. (2000) Of genes and genomes: what lies between the base pairs. *Journal of Investigative Medicine,* **48**, 295-301.

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

Coluzzi, M., Petrarca, V. and di Deco, M.A. (1985). Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. *Bollettino di Zoologia*, **52**: 45–63.

Coluzzi, M (1982) Spatial distribution of chromosomal inversions and speciation in *Anopheline* mosquitoes. *Progress in Clinical Biological Research*, **96**:143-53.

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

Conteh, L., Sharp, B.L., Streat, E., Barreto, A. and Konar, S. (2004) The cost and cost-effectiveness of malaria vector control by residual insecticide house-spraying in southern Mozambique: a rural and urban analysis. *Tropical Medicine and International Health,* **9,** 125-132.

Corbel, V., Chandre, F., Brengues, C., Akogbeto, M., Lardeux, F., Hougard, J.M. and Guillet, P. (2004) Dosage-dependent effects of permethrin-treated nets on the behaviour of *Anopheles gambiae* and the selection of pyrethroid resistance. *Malaria Journal,* **3,** 22.

Curtis, C.F. and Mnzava, A.E. (2000) Comparison of house spraying and insecticide-treated nets for malaria control. *Bullentin of World Health Organization,* **78,** 1389-1400.

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

D'Alessandro, U., Olaleye, B.O., McGuire, W., Thomson, M.C., Langerock, P., Bennett, S. and Greenwood, B.M. (1995) A comparison of the efficacy of insecticide-treated and untreated bed nets in preventing malaria in Gambian children. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **89,** 596-598.

Darriet, F., Guillet, P., N'Guessan, R., Doannio, J.M., Koffi, A., Konan, L.Y. and Carnevale, P. (1998) Impact of resistance of *Anopheles gambiae* s.s. to permethrin and deltamethrin on the efficacy of impregnated mosquito nets. *Médecine tropicale,* **58**, 349-354.

Darriet, F., N'Guessan, R. and Carnevale, P. (2000) Evaluation in test huts of the protective effects of untreated mosquito nets against *Anopheles gambiae* s.s. bites. *Sante,* **10,** 413-417.

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

della Torre, A., Fanello, C., Akogbeto, M., Dossou-yovo, J., Favia, G., Petrarca, V. and Coluzzi, M. (2001) Molecular evidence of incipient speciation within *Anopheles gambiae* s.s. in West Africa. *Insect Molecular Biology*, **10,** 9-18.

della Torre, A., Costantini, C., Besansky, N.J., Caccone, A., Petrarca, V., Powell, J.R. and Coluzzi M. (2002) Speciation within *Anopheles gambiae*--the glass is half full. *Science,* **298,** 115-117.

De Luca, M., Roshina, N.V., Geiger-Thornsberry, G.L., Lyman, R.F., Pasyukova, E.G. and Mackay, T.F. (2003) *Dopa decarboxylase* (Ddc) affects variation in *Drosophila* longevity. *Nature Genetics,* **34,** 429-433.

Dennis, R.L.H. (1993). *Butterflies and Climate Change*. Manchester University Press, Manchester

Depaulis, F., Brazier, L. and Veuille, M. (1999) Selective sweep at the *Drosophila melanogaster* Suppressor of Hairless locus and its association with the In(2L)t inversion polymorphism. *Genetics,* **152,** 1017-1024.

Dewitt, T.J., Sih, A. and Hucko, J.A. (1999) Trait compensation and cospecialization in a freshwater snail: size, shape and antipredator behaviour. *Animal Behaviour,* **58,** 397-407.

DeWitt, T.J. (1998) Costs and limits of phenotypic plasticity: tests with predator-induced morphology and life history in a freshwater snail, *Journal of Evolutionary Biology,* **11,** 465–480.

Diabate, A., Baldet, T., Chandre, F., Guiguemde, R.T., Brengues, C., Guillet, P., Hemingway, J. and Hougard, J.M. (2002) First report of the kdr mutation in *Anopheles gambiae* M form from Burkina Faso, West Africa. *Parassitologia,* **44,** 157-158.

Diabate, A., Baldet, T., Chandre, F., Dabire, K.R., Simard, F., Ouedraogo, J.B., Guillet, P. and Hougard, J.M. (2004) First report of a kdr mutation in *Anopheles arabiensis* from Burkina Faso, West Africa. *Journal of Americal Mosquito Control Association,* **20,** 195-196.

Diabate, A., Brengues, C., Baldet, T., Dabire, K.R., Hougard, J.M., Akogbeto, M., Kengne, P., Simard, F., Guillet, P., Hemingway, J. and Chandre, F. (2004) The spread of the Leu-Phe kdr mutation through *Anopheles gambiae* complex in Burkina Faso: genetic introgression and de novo phenomena. *Tropical Medicine and International Health,* **9,** 1267-1273.

Dicko, A., Klion, A.D., Thera, M.A., Sagara, I., Yalcouye, D., Niambele, M.B., Sogoba, M., Dolo, G., Dao, A., Diallo, D.A., Doumbo, O.K. and Miller, L.H. (2004) The etiology of severe anemia in a village and a periurban area in Mali. *Blood,* **104,** 1198-1200.

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

Donnelly, M.J. and Townson, H. (2000) Evidence for extensive genetic differentiation among populations of the malaria vector A*nopheles arabiensis* in Eastern Africa. *Insect Molecular Biology,* **9**, 357-367.

Donnelly, M.J. Licht, M.C. and Lehmann, T. (2001) Evidence for recent population expansion in the evolutionary history of the malaria vectors *Anopheles arabiensis* and *Anopheles gambiae. Molecular Biology and Evolution,* **18,** 1353-1364.

Donnelly, M.J., Pinto, J., Girod, R., Besansky N.J. and Lehmann, T. (2004) Revisiting the role of introgression vs shared ancestral polymorphisms as key processes shaping genetic diversity in the recently separated sibling species of the *Anopheles gambiae* complex. *Heredity,* **92,** 61-68.

Edwards, M.J., Lemos, F.J., Donnelly-Doman, M. and Jacobs-Lorena, M. (1997) Rapid induction by a blood meal of a carboxypeptidase gene in the gut of the mosquito *Anopheles gambiae*. *Insect Biochemistry and Molecular Biology*, **27,** 1063-1072.

Edwards, M.J., Moskalyk, L.A., Donelly-Doman, M., Vlaskova, M., Noriega, F.G., Walker, V.K. and Jacobs-Lorena, M. (2000) Characterization of a carboxypeptidase A gene from the mosquito, *Aedes aegypti*. *Insect Molecular Biology*, **9,** 33-38.

Egan, J.E., Hoffman, S.L., Haynes, J.D., Sadoff, J.C., Schneider, I., Grau, G.E., Hollingdale, M.R., Ballou, W.R. and Gordon, D.M. (1993) Humoral immune responses in volunteers immunized with irradiated *Plasmodium falciparum* sporozoites. *American Journal of Tropical Medicine and Hygiene,* **49**, 166-173.

Ellers, J. and Boggs, C.L. (2002) The evolution of wing color in Colias butterflies: heritability, sex linkage, and population divergence. *Evolution International Journal of Organic Evolution,* **56,** 836-840.

Ellers, J. and Boggs, C.L. (2003) The evolution of wing color: male mate choice opposes adaptive wing color divergence in Colias butterflies. *Evolution International Journal of Organic Evolution,* **57,** 1100-1106.

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

Enayati, A.A., Vatandoost, H., Ladonni, H., Townson, H. and Hemingway, J. (2003) Molecular evidence for a kdr-like pyrethroid resistance mechanism in the malaria vector mosquito *Anopheles stephensi*. *Medical and Veterinary Entomolology,* **17**, 138-144.

Enayati, A.A., Vontas, J.G., Small, G.J., McCarroll, L. and Hemingway, J. (2001) Quantification of pyrethroid insecticides from treated bednets using a mosquito recombinant glutathione S-transferase. *Medical and Veterinary Entomolology,* **15,** 58-63.

Estoup A., Jarne, P. and Cornuet, J.M. (2000) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology,* **11**, 1591-1604.

Etang, J., Fondjo, E., Chandre, F., Morlais, I., Brengues, C., Nwane, P., Chouaibou, M., Ndjemai, H. and Simard, F. (2006) First report of knockdown mutations in the malaria vector *Anopheles gambiae* from Cameroon. *American Journal of Tropical Medicine and Hygiene*, **74**,795-797

Eveleth, D.D. and Marsh, J.L. (1986) Evidence for evolutionary duplication of genes in the *dopa decarboxylase* region of *Drosophila*. *Genetics,* **114,** 469-483.

Eveleth, D.D., Gietz, R.D., Spencer, C.A., Nargang F.E., Hodgetts, R.B. and Marsh J.L. (1986) Sequence and structure of the *dopa decarboxylase* gene of *Drosophila*: evidence for novel splicing variant. *European Molecular Biology Organization Journal,* **5,**2663–2672

Fanello, C., Akogbeto, M. and Della Torre, A. (2000) Distribution of the pyrethroid knockdown resistance gene (kdr) in *Anopheles gambiae* s.l. from Benin. *Transaction of the Royal Society of Tropical Medicine and Hygiene*, **94**, 132.

Fanello, C., Petrarca, V., della Torre, A., Santolamazza, F., Dolo, G., Coulibaly, M., Alloueche, A., Curtis, C.F., Toure, Y.T. and Coluzzi, M. (2003b) 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.

Fang, Q.Q., Mitchell, A., Regiera, J.C., Mitter, C., Friedlander, T.P. and Poole, R.W. (1997) Phylogenetic Utility of the Nuclear Gene *Dopa Decarboxylase* in Noctuoid Moths (Insecta: Lepidoptera: Noctuoidea). *Molecular Phylogenetics and Evolution*, **15**, 473-486

Faure, J.C. (1943) Phase variation in the armyworm *Laphygma exempta* (Walk.) *Union of South Africa, Department of Agriculture and Forestry Scientific Bulletin,* **17**, pp234.

Favia, G., della Torre, A., Bagayoko, M., Lanfrancotti, A., Sagnon, N., Toure,Y.T. and Coluzzi, M. (1997) Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation. *Insect Molecular Biology,* **6,** 377-383.

Fay, J.C. and Wu, C.I. (2000) Hitchhiking under positive Darwinian selection. *Genetics,* **155,** 1405-1413.

Ferdig, M.T., Taft, A.S., Smartt, C.T., Lowenberger, C.A., Li, J. and Christensen, B.M. (2000) *Aedes aegypti* dopa decarboxylase gene structure and function. *Insect Molecular Biology,* **9,** 231–239

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

Fletcher, M., Teklehaimanot, A. and Yemane, G. (1992) Control of mosquito larvae in the port city of Assab by an indigenous larvivorous fish, *Aphanius dispar*. *Acta Tropica*, **52,** 155-166.

Fodde, R., Harteveld, C.L., Losekoot, M., Giordano, P.C., Khan, P.M., Nayudu, N.V. and Bernini, L.F. (1991) Multiple recombination events are responsible for the heterogeneity of alpha(+)-thalassemia haplotypes among the forest tribes of Andhra Pradesh, India. *Annals of Human Genetics,* **55,** 43-50.

Fournier, D., Bride, J.M., Poirie, M., Berge, J.B. and Plapp, Jr. F. W. (1992) Insect *glutathione S-transferases*. Biochemical characteristics of the major forms from houseflies susceptible and resistant to insecticides. *Journal of Biological Chemistry*, **267,** 1840-1845.

Ford, N.B. and Seigel, R.A. (1989) Relationships among body size, clutch size, and egg size in three species of oviparous snakes *Herpetologica,* **45,** 75-83

Fu, Y.X. and Li, W.H. (1993) Statistical tests of neutrality of mutations. *Genetics,* **133,** 693-709.

Futuyma, D.J. and Moreno, G. (1998) The Evolution of Ecological Specialization. *Annual Review of Ecology and Systematics,* **19**, 207-233

Fuzeau-Braesch, S. (1972) Pigments and color changes. *Annual Review of Entomology,* **17,**403-424

Garland, T. Jr and Kelly, S.A. (2006) Phenotypic plasticity and experimental evolution *Journal of Experimental Biology,* **209,** 2344-2361

Gentile, G., Della Torre, A.B., Maegga, J., Powell, R. and Caccone, A. (2002). Genetic differentiation in the African malaria vector, *Anopheles gambiae s.s.*, and the problem of taxonomic status. *Genetics,* **161,** 1561-1578.

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

Gilburn, A.S. and Day, T.H. (1994). Evolution of female choice in seaweed flies: Fisherian and good genes mechanisms operate in different populations. *Proceedings of the Royal Society of Medicine,* **255,** 159– 165

Gillespie, J. H. (1991) The Burden of Genetic Load. *Science,* **254,** 1049.

Gillies, M.T. and Coetzee, M. (1987) *Supplement to the Anophelinae of Africa South of the Sahara.* Publication no. 55. Johannesburg, South Africa: South African Institute of Medical Research.

Gilles, M.T. and De Meillon, G. (1968) *The Anophelinae of Africa South of Sahara,* Ethiopian Zoogeographical Region, 2nd edn, No.54. The South African Institute for Medical Research, Johannesburg, 343pp

Girard, M.P., Reed, Z.H., Friede, M. and Kieny, M.P. (2007) A review of human vaccine research and development: malaria. *Vaccine,* **25,** 1567-1580.

Goldsmith, S.K. and Alcock, J. (1993) The mating chances of small males of the cerambycid beetle *Trachyderes mandibularis* differ in different environments (Coleoptera, *Cerambycidae*). *Journal of Insect Behaviour,* **6,** 351– 360

Gompel, N. and Carroll, S.B. (2003) Genetic mechanisms and constraints governing the evolution of correlated traits in *drosophilid* flies. *Nature,* **424,** 931-935.

Goodman, C.A., Coleman, P.G. and Mills, A.J. (1999) Cost-effectiveness of malaria control in sub-Saharan Africa. *Lancet,* **354,** 378-385.

Goodman, C. A., A. E. Mnzava, S. S. Dlamini, B. L. Sharp, D. J. Mthembu, and J. K. Gumede. 2001. Comparison of the cost and cost-effectiveness of insecticide-treated bednets and residual house-spraying in KwaZulu-Natal, South Africa. *Tropical Medicine and International Health* **6**: 280-95.

Goulson, D. (1994) Determination of larval melanization in the moth, *Mamestra brassicae*, and the role of melanin in thermoregulation. *Heredity*, **73,** 471–479.

Goulson, D. and Cory, J.S. (1995) Responses of *Mamestra brassicae* (Lepidoptera, Noctuidae) to crowding – interactions with disease resistance, color phase and growth. *Oecologia*, **104,** 416–423.

Gopaul, R. (1995) Entomological surveillance in Mauritius. *Sante,* **5,** 401-405.

Gorman, M.J. and Paskewitz, S.M. (1997) A genetic study of a melanization response to Sephadex beads in *Plasmodium*-refractory and-susceptible strains of *Anopheles gambiae*. *American Journal of Tropical Medicine and Hygiene,* **56,** 446-451.

Gorman, M.J., Severson, D.W., Cornel, A.J., Collins, F.H. and Paskewitz, S.M. (1997) Mapping a quantitative trait locus involved in melanotic encapsulation of foreign bodies in the malaria vector, *Anopheles gambiae*. *Genetics,* **146**, 965-971.

Goto, H., Inomata, N., Szmidt, A.E. and Yamazaki,T. (2004) Nucleotide variation of seven genes in *Drosophila kikkawai*. *Genes and Genetic Systems,* **79,** 19-26.

Greenwood, B. (2006) Review: Intermittent preventive treatment - a new approach to the prevention of malaria in children in areas with seasonal malaria transmission. *Tropical Medicine and International Health,* **11,** 983-991.

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

Gunn, A.(1998) The determination of larval phase coloration in the African armyworm, *Spodoptera exempta* and its consequences for thermoregulation and protection from UV light, *Entomologia Experimentalis et Applicata.* **86,**. 125–133.

Guinand, B., Scribner, K.T., Page, K.S., Filcek, K., Main, L. and Burnham-Curtis, M.K. (2006) Effects of coancestry on accuracy of individual assignments to population of origin: examples using Great Lakes lake trout (*Salvelinus namaycush*). *Genetica,* **127,** 329-340.

Guyatt, H.L., Corlett, S.K., Robinson, T.P., Ochola, S.A. and Snow, R.W. (2002) Malaria prevention in highland Kenya: indoor residual house-spraying vs. insecticide-treated bednets. *Tropical Medicine and International Health,* **7,** 298-303.

Hagen, S.B., Sørlibråten, O., Ims R.A. and Yoccoz, N.G. (2006) Density-dependent melanism in winter Moth larvae (*Lepidoptera:Geometridae*): A countermeasure against parasitoids? *Environmental Entomology,* **35,** 1249-1253

Handler, A.M. (2002) Use of the piggyBac transposon for germ-line transformation of insects. *Insect Biochemistry and Molecular Biology,* **32,** 1211-1220.

Hargreaves, K., Koekemoer, L.L., Brooke, B.D., Hunt, R.H., Mthembu, J. and Coetzee, M. (2000) *Anopheles funestus* resistant to pyrethroid insecticides in South Africa. *Medical and Veterinary Entomolology,* **14,** 181-189.

Harvell, C.D. (1990) The ecology and evolution of inducible defenses. *Quarterly Review of Biology,* **65,** 323-340.

Harwood, J.L. (1988) Fatty acid metabolism. *Annual Review of Plant Physiology,* **39**, 101-138

Hawley, W.A., Phillips-Howard, P.A., ter Kuile, F.O., Terlouw, D.J., Vulule, J.M., Ombok, M., Nahlen, B.L., Gimnig, J.E., Kariuki, S.K., Kolczak, M.S. and Hightower, A. W. (2003) Community-wide effects of permethrin-treated bed nets on child mortality and malaria morbidity in western Kenya. *American Journal of Tropical Medicine and Hygiene,* **68,** 121-127.

Hazel, W.N. (2002) The environmental and genetic control of seasonal polyphenism in larval color and its adaptive significance in a swallowtail butterfly. *Evolution International Journal of Organic Evolution,* **56,** 342-348.

Hayes, J.D. and Pulford, D.J. (1995) The glutathione S-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Critical Reviews in Biochemistry and Molecular Biology,* **30,** 445-600.

Hays, C.W. (2000) The United States Army and malaria control in World War II. *Parassitologia,* **42,** 47-52.

Hemingway, J. (1983) Biochemical studies on malathion resistance in A*nopheles arabiensis* from Sudan. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **77,** 477-480.

Hemingway, J. (2000) The molecular basis of two contrasting metabolic mechanisms of insecticide resistance. *Insect Biochemistry and Molecular Biology,* **30,** 1009-1015.

Hemingway, J. (2004) Taking aim at mosquitoes. *Nature,* **430,** 936.

Hemingway, J. and Karunaratne, S.H. (1998) Mosquito carboxylesterases: a review of the molecular biology and biochemistry of a major insecticide resistance mechanism. *Medical and Veterinary Entomology,* **12,** 1-12.

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

Hemingway, J., Field, L. and Vontas, J. (2002) An overview of insecticide resistance. *Science,* **298,** 96-97.

Hemingway, J., Hawkes, N.J., McCarroll, L. and Ranson, H. (2004) The molecular basis of insecticide resistance in mosquitoes. *Insect Biochemistry and Molecular Biology,* **34,** 653-665.

Hill, G.E. (1994) House finches are what they eat: a reply to Hudon. *Journal of Ornithology,* **111**, 221–225.

Hirsh, J. and Davidson, N. (1981) Isolation and characterization of the *dopa decarboxylase* gene of *Drosophila melanogaster*. *Molecular and Cellular Biology*, **1,** 475-485.

Hiruma, K. and Riddiford, L.M. (1985) Hormonal regulation of *dopa decarboxylase* during a larval molt. *Developmental Biology*, **110,** 509-513.

Hiruma, K., Carter, M.S. and Riddiford, L.M. (1995) Characterization of the *dopa decarboxylase* gene of *Manduca sexta* and its suppression by 20-hydroxyecdysone. *Developmental Biology*, **169,** 195-209.

Hiruma, K., Riddiford, L.M., Hopkins, T.L. and Morgan, T.D. (1985) Roles of *dopa decarboxylase* and *phenoloxidase* in the melanization of the tobacco hornworm and their control by 20-hydroxyecdysone. *Journal of Comparative Physiology*, **155,** 659-669.

Hirsh, J. and Davidson, N. (1981) Isolation and characterization of the *dopa decarboxylase* gene of *Drosophila melanogaster*. *Molecular and. Cellular Biology,* **1,**475–486

Hodgetts, R.B. and O'Keefe, S.L. (2006) *Dopa decarboxylase*: a model gene-enzyme system for studying development, behavior, and systematics. *Annual Review of Entomology*, **51,** 259-284.

Hodgetts, R.B., Clark,W.C., O'Keefe, S.L., Schouls, M. and Crossgrove, K. (1995) Hormonal induction of *dopa decarboxylase* in the epidermis of *Drosophila* is mediated by the *Broad-Complex*. *Development,* **121,**3913–3922.

Hoekstra, H.E. (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity,* **97,** 222-234.

Hoffman, S.L., Goh, L.M., Luke, T.C., Schneider, I., Le,T.P., Doolan, D.L., Sacci, J., de la Vega, P., Dowler, M., Paul, C., Gordon, D.M., Stoute, J.A., Church, L.W., Sedegah, M., Heppner, D.G., Ballou, W.R. and Richie, T.L. (2002) Protection of humans against malaria by immunization with radiation-attenuated *Plasmodium falciparum* sporozoites. *Journal of Infectious Diseases,* **185,** 1155–1164.

Hoffmann, A.A., Sgro, C.M. and Weeks, A.R. (2004) Chromosomal inversion polymorphisms and adaptation. *Trends in Ecology and Evolution,* **19,** 482-488.

Holl, S.M., Schaefer, J., Goldberg, W.M., Kramer, K.J., Morgan, T.D. and Hopkins,T.L. (1992) Comparison of black coral skeleton and insect cuticle by a combination of carbon-13 NMR and chemical analyses. *Archives of Biochemistry and Biophysics*, **292,** 107-111.

Hopkins, T. L. and Kramer, K. J. (1992) Insect Cuticle Sclerotization. *Annual Review of Entomology*, **37**: 273-302

Huang, C.-Y., Chou, S.-Y., Bartholomay, L.C., Christensen, B.M. and Chen, C.-C. (2005) The use of gene silencing to study the role of *dopa decarboxylase* in mosquito melanization reactions. *Insect Molecular Biology*, **14**, 237-244

Huang, H.S., Hu, N.T., Yao, Y.E., Wu, C.Y., Chiang, S.W. and Sun, C.N. (1998) Molecular cloning and heterologous expression of a *glutathione S-transferase* involved in insecticide resistance from the diamondback moth, *Plutella xylostella. Insect Biochemistry and Molecular Biology*, **28,** 651-658.

Hudson, R.R. (1987) Estimating the recombination parameter of a finite population model without selection. *Genetical Research,* **50,** 245-250.

Hudson, R.R., Kreitman, M. and Aguade, M. (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics,* **116,** 153-159.

Hudson, R.R., Boos, D.D. and Kaplan. N. L. (1992) A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution*, **9,** 138-151.

Hudson, R.R. and Kaplan, N.L. (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics,* **111**, 147-164.

Impoinvil, D.E., Cardenas, G.A., Gihture, J.I., Mbogo, C.M. and Beier, J.C. (2007) Constant temperature and time period effects on *Anopheles gambiae* egg hatching. *Journal of the American Mosquito Control Association,* **23**,124-130

Irvin, N., Hoddle, M.S., O'Brochta, D.A., Carey, B. and Atkinson, P.W. (2004) Assessing fitness costs for transgenic *Aedes aegypti* expressing the GFP marker and transposase genes. *Proceedings of the National Academy of Science,* USA, **101,** 891-896.

Jablonka, E., Oborny, B., Molnar, I., Kisdi, E., Hofbauer, J. and Czaran, T. (1995) The adaptive advantage of phenotypic memory in changing environments. *Philosophical Transactions of the Royal Society B: Biological Sciences,* **350,**133–141

Jong, P., Gussekloo, S. and Brakefield, P. (1996) Differences in thermal balance, body temperature and activity between non-melanic and melanic two-spot ladybird beetles (*Adalia bipunctata*) under controlled conditions. *Journal of Experimental Biology,* **199,** 2655-2666.

Jordan, K.W., Morgan, T.J. and Mackay, T. F. (2006). Quantitative Trait Loci for Locomotor Behavior in *Drosophila melanogaster. Genetics*, **174**, 271-284.

Joron, M., Papa, R., Beltran, M., Chamberlain, N., Mavarez, J., Baxter, S., Abanto, M., Bermingham, E., Humphray, S.J., Rogers, J., Beasley, H., Barlow, K., ffrench-Constant, R.H., Mallet, J., McMillan, W.O. and Jiggins, C.D. (2006) A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biology,* **4,** e303.

Kabayo, J.P. (2002) Aiming to eliminate tsetse from Africa. *Trends in Parasitology*, **18**, 473–475.

Kamau, L., Lehmann, T., Hawley, W.A., Orago, A.S. and Collins, F.H. (1998) Microgeographic genetic differentiation of *Anopheles gambiae* mosquitoes from Asembo Bay, western Kenya: a comparison with Kilifi in coastal Kenya. *American Journal of Tropical Medicine and Hygiene,* **58,** 64-69.

Kamau, L., Mukabana, W.R., Hawley,W.A., Lehmann, T., Irungu, L.W., Orago, A.A. and Collins, F.H. (1999) Analysis of genetic variability in *Anopheles arabiensis* and *Anopheles gambiae* using microsatellite loci. *Insect Molecular Biology*, **8,** 287-297.

Kaplan, N. and Hudson, R.R. (1985) The use of sample genealogies for studying a selectively neutral m-loci model with recombination. *Theoretical Population Biology*, **28,** 382-396.

Karan, D., Morin, J.P., Gibert, P., Moreteau, B., Scheiner, S.M. and David, J.R. (2000) The genetics of phenotypic plasticity. IX. Genetic architecture, temperature, and sex differences in *Drosophila melanogaster*. *Evolution International Journal of Organic Evolution,* **54,** 1035-1040.

Karban, R. and Baldwin, I.T. (1997) Induced responses to herbivory. University of Chicago Press, Chicago, Illinois, USA.

Kayondo, J.K., Mukwaya, L.G., Stump, A., Michel, A.P., Coulibaly, M.B., Besansky, N.J. and Collins, F.H. (2005) Genetic structure of *Anopheles gambiae* populations on islands in northwestern Lake Victoria, Uganda. *Malaria Journal*, **4,** 59.

Kazimirova, M. (1992) The role of physical contact in the induction of phase polymorphism of *Mamestra brassicae* (Lepidoptera, Noctuidae). *Acta Entomologica Bohemoslovaca*, **89**, 87–95.

Killeen, G.F., Fillinger, U. and Knols, B.G. (2002) Advantages of larval control for African malaria vectors: low mobility and behavioural responsiveness of immature mosquito stages allow high effective coverage. *Malaria Journal*, **1,** 8.

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

Kimura, M. and Takahata, N. (1983) Selective constraint in protein polymorphism: study of the effectively neutral mutation model by using an improved pseudosampling method. *Proceedings of the National Academy of Science*, USA. **80**, 1048-1052.

Kimura, M. (1968) Evolutionary rate at the molecular level. *Nature,* **217,** 624-626

Kirkpatrick, M. and Barton, N. (2006) Chromosome inversions, local adaptation and speciation. *Genetics,* **173,** 419–434.

Kitzmiller, J.B. (1976) Genetics, Cytogenetics, and Evolution of Mosquitoes. In: Advances in Genetics, Vol. 18 (Caspari E, ed). New York: Academic Press; 315–433

Koch, P. B., Keys, D.N., Rocheleau, T., Aronstein, K., Blackburn, M., Carroll, S.B. and ffrench-Constant, R.H. (1998) Regulation of *dopa decarboxylase* expression during colour pattern formation in wild-type and melanic tiger swallowtail butterflies. *Development,* **125,** 2303-2313.

Kolaczinski, K., Kolaczinski, J., Kilian, A. and Meek, S. (2007) Extension of indoor residual spraying for malaria control into high transmission settings in Africa. *Transaction of the Royal Society of Tropical Medicine and Hygiene*, **101,** 852-853.

Kooi, R.E. and Brakefield, P.M. (1999) The critical period for wing pattern induction in the polyphenic tropical butterfly *Bicyclus anynana* (Satyrinae). *Journal of Insect Physiology,* **45,** 201–212.

Konrad, K.D. and Marsh, J.L. (1987) Developmental expression and spatial distribution of *dopa decarboxylase* in *Drosophila. Developmental Biology,* **122,** 172–185

Kreitman, M. and Wayne, M.L. (1994) Organization of genetic variation at the molecular level: lessons from *Drosophila*. EXS, **69,** 157-183.

Krzywinski, J., Wilkerson, R.C. and Besansky, N.J. (2001) Toward understanding *Anophelinae* (Diptera, Culicidae) phylogeny: insights from nuclear single-copy genes and the weight of evidence. *Systems Biology*, **50,** 540-556.

Kulkarni, M.A., Rowland, M., Alifrangis, M., Mosha, F., Matowo,W.J., Malima, R., Peter, J., Kweka, E., Lyimo, I., Magesa, S., Salanti, A., Rau, M.E. and Drakeley, C. (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.

Kumar, S., Tamura, K. and Nei. M. (1994) MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Computer Applications in the Biosciences*, **10,** 189-191.

Kumar, S., Tamura, K. and Nei, M. (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics*, **5,** 150-163.

Lambrechts, L., Vulule, J.M. and Koella, J.C. (2004) Genetic correlation between melanization and antibacterial immune responses in a natural population of the malaria vector *Anopheles gambiae*. *Evolution,* **58,** 2377-2381.

Lehmann, T., Besansky, N.J., Hawley, W.A., Fahey, T.G., Kamau, L. and Collins, F.H. (1997) Microgeographic structure of *Anopheles gambiae* in western Kenya based on mtDNA and microsatellite loci. *Molecular Ecology*, **6,** 243-253.

Lehmann, T., Dalton, R., Kim, E.H., Dahl, E., Diabate, A., Dabire, R. and Dujardin, J.P. (2006) Genetic contribution to variation in larval development time, adult size, and longevity of starved adults of *Anopheles gambiae. Infection, genetics and Evolution*,**6**, 410-416

Lehmann, T., Hawley, W.A., Kamau, L., Fontenille, D., Simard, F. and Collins, F.H. (1996) Genetic differentiation of *Anopheles gambiae* populations from East and west Africa: comparison of microsatellite and allozyme loci. *Heredity,* **77,** 192-200.

Lehmann, T., Hawley, W.A., Grebert, H., Danga, M., Atieli, F. and Collins, F.H. (1999) The Rift Valley complex as a barrier to gene flow for *Anopheles gambiae* in Kenya. *Journal of Hereditary,* **90,** 613-621.

Lehmann, T., Licht, M.N., Elissa, B.T., Maega, J.M., Chimumbwa, F.T., Watsenga, Wondji, C.S. Simard, F. and Hawley, W.A. (2003) Population Structure of *Anopheles gambiae* in Africa. *Journal of Hereditary,* **94,** 133-147.

Lengeler, C. and Snow, R.W. (1996) From efficacy to effectiveness: insecticide-treated bednets in Africa. *Bulletin of the World Health Organization*, **74,** 325-332.

Li, J. and Christensen, B.M. (1993) Involvement of L-tyrosine and phenol oxidase in the tanning of *Aedes aegypti* eggs. *Insect Biochemistry and Molecular Biology* **23,**739–748

Li, J.S. and Li, J. (2006) Major chorion proteins and their crosslinking during chorion hardening in *Aedes aegypti* mosquitoes. *Insect Biochemistry and Molecular Biology,* **36**, 954-964

Lindblade, K.A., Eisele, T.P., Gimnig, J.E., Alaii, J.A., Odhiambo, F., ter Kuile, F.O., Hawley, W.A., Wannemuehler, K.A., Phillips-Howard, P.A., Rosen, D.H., Nahlen, B.L., Terlouw, D.J., Adazu, K., Vulule, J.M. and Slutsker, L. (2004) Sustainability of reductions in malaria transmission and infant mortality in western Kenya with use of insecticide-treated bednets: 4 to 6 years of follow-up. *The Journal of the American Medical Association,* **291,** 2571-2580.

Lindsay, S.W., Parson, L. and Thomas, C.J. (1998) Mapping the ranges and relative abundance of the two principal African malaria vectors, *Anopheles gambiae sensu stricto* and *Anopheles arabiensis*, using climate data. *Proceedings of the Royal Society of Biological Sciences*, **265,** 847-854.

Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* **25**, 402-408.

Lönn, M., Sandberg, A. and Redbo-Torstensson, P. (1998) Fitness-related traits of allozyme genotypes in *Bromus hordeaceus* L. (Poaceae) associated with field habitat and experimental flooding. *Biological Journal of the Linnean Society,* **64,** 207–222.

Louis, J.P. and Albert, J.P. (1988) Malaria in the Republic of Djibouti. Strategy for control using a biological antilarval campaign: indigenous larvivorous fishes (*Aphanius dispar*) and bacterial toxins. *Medicina Tropical,* **48,** 127-131.

Low, W.Y., Ng, H.L., Morton, C.J., Parker, M.W., Batterham, P. and Robin, C. (2007) Molecular evolution of glutathione s-transferases in the genus *Drosophila*. *Genetics,* **177,** 1363-1375.

Luikart, G., England, P.R., Tallmon, D., Jordan, S. and Taberlet, P. (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews of Genetics,* **4,** 981-994.

Lumjuan, N., McCarroll, L., Prapanthadara, L.A., Hemingway, J. and Ranson, H. (2005) Elevated activity of an Epsilon class glutathione transferase confers DDT resistance in the dengue vector, *Aedes aegypti*. *Insect Biochemistry Molecular Biology*, **35,** 861-871.

Lumjuan, N., Stevenson, B.J., Prapanthadara, L.A., Somboon, P., Brophy, P.M., Loftus, B.J., Severson, D.W. and Ranson, H. (2007) The *Aedes aegypti* glutathione transferase family. *Insect Biochemistry and Molecular Biology*, **37,** 1026-1035.

Lynch, W. (2007) Contributions of Anopheles larval control to malaria supression in tropical Africa: review of achievements and potentials. *Medical and Veterinary Entomology*, **21**, 2-21

Lyimo, E.O., Takken, W. and Koella, J.C. (1992) Effect of rearing temperature and larval density on larval survival, age at pupation and adult size of *Anopheles gambiae*. *Entomologia Experimentalis et Applicata,* **63,** 265–271.

Ma, W., Chen, L., Wang, M.O. and Li, X. (2007) Trade-offs between melanisation and life-history traits in *Helicoverpa armigera*. *Ecological Entomology*, **33**,37 - 44.

Mackay, T.F.C. (2001) The genetic architecture of quantitative traits. *Annual Reviews of Genetics*, **35,** 303-339

Maher, B. (2008) Malaria: The end of the beginning. *Nature*, **451**, 1042-1046

Majerus, M.E.N. (1998) Melanism: Evolution in Action. Oxford university press

Majerus, M.E.N. and Mundy, N. I. (2003) Mammalian melanism: natural selection in black and white. *Trends in Genetics*, **19**, 585-588

Mannervik, B. (1985) The isoenzymes of glutathione transferase. *Advances in Enzymology and Related Areas of Molecular Biology,* **57,** 357-417.

Marrelli, M.T., Li, C., Rasgon, J.L. and Jacobs-Lorena, M. (2007) Transgenic malaria-resistant mosquitoes have a fitness advantage when feeding on *Plasmodium*-infected blood. *Proceedings of the National Academy of Sciences*, **104,** 5580-5583.

Martinez-Torres, D., Chandre, F., Williamson, M.S., Darriet, F., Berge, J.B., Devonshire, A.L., Guillet, P., Pasteur, N. and Pauron, D. (1998) Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector *Anopheles gambiae s.s. Insect Molecular Biology,* **7,** 179-184.

Mason, G.F. (1967) Genetic studies on mutations in species A and B of the *Anopheles gambiae* complex.*Genetics Research*, **10,** 205-217

Matambo, T.S., Abdalla, H., Brooke, B.D., Koekemoer, L.L., Mnzava, A., Hunt, R.H. and Coetzee, M. (2007) Insecticide resistance in the malarial mosquito Anopheles arabiensis and association with the *kdr* mutation. *Medical and Veterinary Entomology,* **21,** 97-102.

Maxwell, C.A., Msuya, E., Sudi, M., Njunwa, K.J., Carneiro, I.A. and Curtis, C.F. (2002) Effect of community-wide use of insecticide-treated nets for 3-4 years on malarial morbidity in Tanzania. *Tropical Medicine and International Health,* **7,** 1003-1008.

McAbee, R.D., Kang, K.D., Stanich, M.A., Christiansen, J.A., Wheelock, C.E., Inman, A.D., Hammock, B.D. and. Cornel, A.J. (2004) Pyrethroid tolerance in *Culex pipiens pipiens var molestus* from Marin County, California. *Pest Management Science*, **60**, 359-368.

McDonald, J.H. and Kreitman, M. (1991) Adaptive protein evolution at the Adh locus in *Drosophila. Nature,* **351,** 652-654

McGrane, V., Carlson, J.O., Miller, B.R. and Beaty, B.J. (1988) Microinjection of DNA into *Aedes triseriatus* ova and detection of integration. A*merican Journal of Tropical Medicine and Hygiene*, **39,** 502-510.

McGraw, K.J. (2007) Dietary mineral content influences the expression of melanin-based ornamental coloration. *Behavioral Ecology*,**18**, 137-142

McMillan,W.O., Monteiro, A. and Kapan, D.D. (2002) Development and evolution on the wing. *Trends in Ecology and Evolution*, **17**,125-133

Merila, J., Laurila, A. and Lindgren, B. (2004) Variation in the degree and costs of adaptive phenotypic plasticity among *Rana temporaria* populations. *Journal of Evolutionary Biology*, **17,** 1132-1140.

Mitchell, S.E. and Seawright, J.A. (1984) A red stripe mutant and its relationship in an allelic series in *Anopheles quadrimaculatus*. *Journal of Hereditary,* **75,** 421-422

Moreira, L.A., Edwards, M.J., Adhami, F., Jasinskiene, N., James, A.A. and Jacobs-Lorena, M. (2000) Robust gut-specific gene expression in transgenic *Aedes aegypti* mosquitoes. *Proceedings of the National Academy of Sciences,* USA **97,** 10895-10898.

Moreira, L.A., Ghosh, A.K., Abraham, E.G. and Jacobs-Lorena, M. (2002) Genetic transformation of mosquitoes: a quest for malaria control. *International Journal of Parasitology*, **32,** 1599-1605.

Morgan, B.A., Johnson, W.A. and Hirsh, J. (1986) Regulated splicing produces different forms of *dopa decarboxylase* in the central nervous system and hypoderm of *Drosophila melanogaster*. *European Molecular Biology Organization Journal,* **5,**:3335–3342

Moore, R.C. and Purugganan, M.D. (2003) The early stages of duplicate gene evolution. *Proceedings of the National Academy of Sciences,* USA, **100,** 15682-15687.

Morris, A.C., Eggleston, P. and Crampton, J.M. (1989) Genetic transformation of the mosquito *Aedes aegypti* by micro-injection of DNA. *Medical and Veterinary Entomology*, **3,** 1-7.

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

Munday, S. (2007) Review of intermittent preventive treatment for malaria in infants and children. *Journal of Paediatrics and Child Health,* **43,** 424-428.

Najera, J.A. (2001) Malaria control: achievements, problems and strategies. *Parassitologia,* **43,** 1-89.

Nakashima, M.J., Georghiou, G.P. and Yermanos, D.M. (1975) A red stripe character and its inheritance in *Anopheles albimanus*. Wiedemann. *Mosquito News,* **35,** 334-337.

Nappi, A.J. and Christensen, B.M. (2005) Melanogenesis and associated cytotoxic reactions: applications to insect innate immunity. *Insect Biochemistry and Molecular Biology,* **35**, 443-459.

Narahashi, T. (1996) Neuronal ion channels as the target sites of insecticides. *Pharmacology and Toxicology,* **79,** 1-14.

Navarro, A., Betran, E., Barbadilla, A. and Ruiz, A. (1997) Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics,* **146,** 695-709.

Navarro, A. and Barton, N.H. (2003) Accumulating postzygotic isolation genes in parapatry: A new twist on chromosomal speciation. *Evolution International Journal of Organic Evolution,* **57,** 447–459.

Nei, M. (1982). Evolution of human races at the gene level, In: Bonne-Tamir,B., Cohen,T. and Goodman,R.M. (eds.), Human genetics, part A: The unfolding genome. Alan, R. Liss, New York, NY.

Nevill, C.G., Some, E.S., Mung'ala, V.O., Mutemi,W., New, L., Marsh, K., Lengeler, C. and Snow, R.W. (1996) Insecticide-treated bednets reduce mortality and severe morbidity from malaria among children on the Kenyan coast. *Tropical Medicine and International Health,* **1,** 139-146.

N'Guessan, R., Corbel, V., Akogbeto, M. and 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.

Nielsen, R. (2005) Molecular signatures of natural selection. *Annual Review of Genetics,* **39,** 197-218.

Nikou, D., Ranson, H. and 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.

Noguchi, H., Tsuzuki, S., Tanaka, K., Matsumoto, H., Hiruma, K. and Hayakawa, Y. (2003) Isolation and characterization of a *dopa decarboxylase* cDNA and the induction of its expression by an insect cytokine, growth-blocking peptide in *Pseudaletia separata*. *Insect Biochemistry and Molecular Biology,* **33,** 209-217.

Noor, M.A.F., Grams, K.L., Bertucci, L.A. and Reiland, J. (2001) Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences,* USA, **98,** 12084–12088

Nuwaha, F. (2001) The challenge of chloroquine-resistant malaria in sub-Saharan Africa. *Health Policy Plan,* **16**, 1-12.

Nyanjom, S.R., Chen, H., Gebre-Michael, T., Bekele, E., Shililu, J., Githure, J., Beier, J.C. and Yan, G. (2003) Population genetic structure of *Anopheles arabiensis* mosquitoes in Ethiopia and Eritrea. *Journal of Hereditary*, **94,** 457-463.

Nylin, S. and Gotthard, K. (1998) Plasticity in life-history traits. *Annual Review of Entomology,* **43,** 63-83.

Oakeshott, J.G., Claudianos, C., Russell, R.J. and Robin, G.C. (1999) Carboxyl/cholinesterases: a case study of the evolution of a successful multigene family. *Bioessays,* **21,** 1031-1042.

Oda, S., Hanazato, T. and Fujii, K. (2007) Changes in phenotypic plasticity of a morphological defence in *Daphnia galeata* (Crustacea: *Cladocera*) in selection experiment, *Journal of Limnology*, **66**, 142-152.

Ohta, T. (2003) Evolution by gene duplication revisited: differentiation of regulatory elements versus proteins. *Genetica,* **118,** 209-16.

Onyabe, D. Y. and Conn, J.E. (2001) Genetic differentiation of the malaria vector *Anopheles gambiae* across Nigeria suggests that selection limits gene flow. *Heredity,* **87,** 647-658.

Ortelli, F., Rossiter, L.C., Vontas, J., Ranson, H. and 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.

Ottenheim, M.M., Wertheim, B., Holloway, G.J. and Brakefield, P.M (1999) Survival of colour-polymorphic *Eristalis arbustorum* hoverflies in semi-field conditions. *Functional Ecology*, **13,**72-77.

Paskewitz, S. and Riehle, M.A. (1994) Response of Plasmodium refractory and susceptible strains of *Anopheles gambiae* to inoculated Sephadex beads. *Developmental and Comparative Immunology,* **18,** 369-375.

Paskewitz, S.M., Schwartz, A.M. and Gorman, M.J. (1998) The role of surface characteristics in eliciting humoral encapsulation of foreign bodies in *Plasmodium*-refractory and -susceptible strains of *Anopheles gambiae*. *Journal of Insect Physiology*, **44,** 947-954.

Paskewitz, S.M., Reese-Stardy, S. and Gorman, M.J. (1999) An easter-like serine protease from *Anopheles gambiae* exhibits changes in transcript abundance following immune challenge. *Insect Molecular Biology*, **8,** 329-337.

Pearse, D.E. and Pogson, G.H. (2000) Parallel evolution of the melanic form of the California legless lizard, *Anniella pulchra*, inferred from mitochondrial DNA sequence variation. *Evolution International Journal of Organic Evolution,* **54,** 1041-1046.

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

Petrarca, V., Carrara, G.C., Di Deco, M.A. and Petrangeli, G. (1983) The *Anopheles gambiae* complex in Guinea Bissau. *Parassitologia,* **25,** 29-39.

Petrarca, V., Nugud, A.D., Ahmed, M.A., Haridi, A.M., Di Deco, M.A. and Coluzzi, M. (2000) Cytogenetics of the *Anopheles gambiae* complex in Sudan, with special reference to *An. arabiensis*: relationships with East and West African populations. *Medical and Veterinary Entomology,* **14,** 149-164.

Pigliucci, M. (2001) Phenotypic Plasticity: Beyond Nature and Nurture. Baltimore, MD: The Johns Hopkins University Press

Pinto, J., Lynd, A., Elissa, N., Donnelly, M.J., Costa, C., Gentile, G., Caccone, A. and do Rosario, V.E. (2006) Co-occurrence of East and West African *kdr* mutations suggests high levels of resistance to pyrethroid insecticides in *Anopheles gambiae* from Libreville, Gabon. *Medical and Veterinary Entomology,* **20,** 27-32.

Pinto, J., Sousa, C.A., Gil, V., Ferreira, C., Goncalves, L., Lopes, D., Petrarca, V., Charlwood, J.D. and do Rosario, V.E. (2000) Malaria in Sao Tome and Principe: parasite prevalence and vector densities. *Acta Tropica,* **76,** 185-193.

Powell, J.R., Petrarca, V., della Torre, A., Caccone, A. and Coluzzi, M. (1999) Population structure, speciation, and introgression in the *Anopheles gambiae* complex. *Parassitologia,* **41,** 101-113.

Prapanthadara, L.A. and Ketterman, A.J. (1993) Qualitative and quantitative changes in *glutathione S-transferases* in the mosquito *Anopheles gambiae* confer DDT-resistance. *Biochemical Society Transactions,* **21,** 304S.

Prapanthadara, L.A., Kuttastep, S., Hemingway, J. and Ketterman, A.J. (1995) Characterization of the major form of *glutathione transferase* in the mosquito *Anopheles dirus A. Biochemical Society Transactions,* **23**, 81S.

Price, T.D. (2006) Phenotypic plasticity, sexual selection and the evolution of colour patterns. *Journal of Experimetal Biology,* **209,** 2368-2376.

Rai, K. (1996) Genetic Control of Vectors. In: *The Biology of Disease Vectors, Beaty, B. and Marquardt, W.* pp. 564–574. University Press of Colorado

Ranson, H., Claudianos, C., Ortelli, F., Abgrall, C., Hemingway, J., Sharakhova, M.V., Unger, M.F., Collins, F.H. and Feyereisen, R. (2002b) Evolution of supergene families associated with insecticide resistance. *Science,* **298,** 179-181.

Ranson, H., Jensen, B., Vulule, J.M., Wang, X., Hemingway, J. and Collins. F.H. (2000b) 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., Prapanthadara, L., Hemingway, J. and Collins, F.H. (2000a) Genetic mapping of two loci affecting DDT resistance in the malaria vector *Anopheles gambiae. Insect Molecular Biology,* **9,** 499-507.

Ranson, H., Nikou, D., Hutchinson, M., Wang, X., Roth, C.W., Hemingway, J. and Collins, F.H. (2002a) Molecular analysis of multiple cytochrome *P450* genes from the malaria vector, *Anopheles gambiae. Insect Molecular Biology*, **11,** 409-418.

Ranson, H., Prapanthadara, L. and Hemingway, J. (1997) Cloning and characterization of two *glutathione S-transferases* from a DDT-resistant strain of *Anopheles gambiae. Biochemical Journal,* **324,** 97-102.

Ranson, H., Rossiter, L., Ortelli, F., Jensen, B., Wang, X., Roth, C.W., Collins, F.H. and Hemingway, J. (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.

Relyea, R.A. and Hoverman, J.T. (2003) The impact of larval predators and competitors on the morphology and fitness of juvenile treefrogs. *Oecologia,* **134,** 596-604.

Reilly, S.J., Oum, R. and Heideman, P.D. (2006) Phenotypic plasticity of reproductive traits in response to food availability and photoperiod in white-footed mice (*Peromyscus leucopus*), Oecologia, **150**, 373-392.

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M.M. and Buckler IV E.S. (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences of the USA,* **98**, 11479–11484.

Reznick, D.N. and Ghalambor, C.K. (2001) The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica,* **112,** 183-98

Rieseberg, L.H. (2001) Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution,* **16,** 351–358.

Rishikesh, N., Di Deco, M.A., Petrarca, V. and Coluzzi, M. (1985) Seasonal variations in indoor resting *Anopheles gambiae* and *Anopheles arabiensis* in Kaduna, Nigeria. *Acta Tropica*, **42,** 165-170.

Roff, D.A. (1992) The Evolution of Life Histories: Theory and Analysis. New York: Chapman and Hall

Romi, R., Razaiarimanga, M.C., Raharimanga, R., Rakotondraibe, E.M., Ranaivo, L.H., Pietra, V., Raveloson, A. and Majori, G. (2002) Impact of the malaria control campaign (1993-1998) in the highlands of Madagascar: parasitological and entomological data. *American Journal of Tropical Medicine and Hygiene,* **66,** 2-6.

Roskam, J.C. and Brakefield, P. M. (1999) Seasonal polyphenism in Bicyclus (Lepidoptera: *Satyridae*) butterflies: different climates need different cues. *Biological Journal of the Linnean Society*, **66**, 345-356.

Rossjohn, J., Board, P.G., Parker, M.W. and Wilce, M.C. (1996) A structurally derived consensus pattern for theta class *glutathione transferases*. *Protein Engineering*, **9,** 327-332.

Rozen, S. and Skaletsky, H. (1999) "Primer3 on the WWW for general users and for biologist programmers." In: *Bioinformatics Methods and Protocols* in the series *Methods in Molecular Biology*. Krawetz, S. Misener, S. (eds). Humana Press, Totowa, NJ, 2000, pages 365-386

Rozas, J. and Rozas, R. (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics,* **15,** 174-175.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X. and Rozas, R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics,* **19,** 2496-2497.

Rubin, G.M. and Spradling, A.C. (1982) Genetic transformation of *Drosophila* with transposable element vectors. *Science,* **218,** 348-53.

Sabchareon, A., Burnouf, T., Ouattara, D., Attanath, P., Bouharoun-Tayoun, H., Chantavanich, P., Foucault, C., Chongsuphajaisiddhi, T. and Druilhe, P. (1991) Parasitologic and clinical human response to immunoglobulin administration in *falciparum* malaria. *American Journal of Tropical Medicine and Hygiene,* **45,** 297-308.

Sachs, J. D. (2001) A new global commitment to disease control in Africa. *Nature Medicine*, **7,** 521-3.

Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution,* **4,** 406-425.

Schapira, A. (2004) DDT still has a role in the fight against malaria. *Nature,* **432,** 439.

Schapira, A. (2006) DDT: a polluted debate in malaria control. *Lancet,* **368,** 2111-2113.

Schellenberg, D., Menendez, C., Aponte, J.J., Kahigwa, E., Tanner, M., Mshinda, H. and Alonso, P. (2005) Intermittent preventive antimalarial treatment for Tanzanian infants: follow-up to age 2 years of a randomised, placebo-controlled trial. *Lancet,* **365,** 1481-1483.

Schellenberg, D., Menendez, C., Kahigwa, E., Aponte, J., Vidal, J., Tanner, M., Mshinda, H. and Alonso, P. (2001) Intermittent treatment for malaria and anaemia control at time of routine vaccinations in Tanzanian infants: a randomised, placebo-controlled trial. *Lancet,* **357,** 1471-1477.

Scheiner, S. M. and Callahan, H. S. (1999) Measuring Natural Selection on Phenotypic Plasticity. *Evolution*, **53**, 1704-1713

Schneider, S.M. and Berrigan, D. (1998) The genetics of phenotypic plasticity. VIII. The cost of plasticity in *Daphnia pulex*. *Evolution*, **52,** 368–378.

Schlichting, C.D. (1986) The evolution of phenotypic plasticity in plants. *Annual Review of Ecology and Systematics,* **17,** 667–693.

Scholnick , S.B., Bray, S.J., Morgan, B.A., McCormick, C.A. and Hirsh, J. (1986) CNS and hypoderm regulatory elements of the *Drosophila melanogaster dopa decarboxylase* gene. *Science,* **234**, 998-1002.

Scott, J.A., Brogdon, W.G. and Collins, F.H. (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, J.G. and Georghiou, G.P. (1986) Malathion-specific resistance in *Anopheles stephensi* from Pakistan. *Journal of American Mosquito Control Associations*, **2,** 29-32.

Scott, J.G., Liu, N. and Wen, Z. (1998) Insect cytochromes P450: diversity, insecticide resistance and tolerance to plant toxins. *Comparative Biochemistry and Physiology Part C: Pharmacology Toxicology and Endocrinology,* **121,** 147-55.

Scott, J.G., Roush, R.T. and Liu, N. (1991) Selection of high-level abamectin resistance from field-collected house flies, *Musca domestica. Experientia,* **47,** 288-291.

Scott, T.W., Takken, W., Knols, B.G. and Boete, C. (2002) The ecology of genetically modified mosquitoes. *Science,* **298,** 117-119.

Seawright, J.A., Benedict, M.Q. and Narang, S. (1985) Color mutants in *Anopheles albimanus* (Diptera: Culicidae). *Annals of the Entomological Society of America,* **78,**177-181

Seawright, J.A. and Benedict, M.Q. (1985) Recessive lethal mutations in *Anopheles albimanus. Journal of the American Mosquito Control Association,* :**1,**227-232

Sharakhov, I.V., White, B.J., Sharakhova, M.V., Kayondo, J., Lobo, N.F., Santolamazza, F., Della Torre, A., Simard, F., Collins, F.H. and Besansky, N.J. (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,* USA, **103,** 6258-6262.

Sharma, S.N., Shukla, R.P., Raghavendra, K. and Subbarao, S.K. (2005) Impact of DDT spraying on malaria transmission in Bareilly District, Uttar Pradesh, India. *Journal of Vector Borne Diseases,* **42,** 54-60.

Sharp, B., van Wyk, P., Sikasote, J.B., Banda, P. and Kleinschmidt, I. (2002) Malaria control by residual insecticide spraying in Chingola and Chililabombwe, Copperbelt Province, Zambia. *Tropical Medicine and International Health*, **7,** 732-736

Sharp, B.L., Ridl, F.C., Govender, D., Kuklinski, J. and Kleinschmidt, I. (2007) Malaria vector control by indoor residual insecticide spraying on the tropical island of Bioko, Equatorial Guinea. *Malaria Journal*, **6,** 52.

Sheeban, D., Meade, G., Foley, M.V. and Dowd, C.A. (2001) Structure, function and evolution of *glutathione transferases*: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical Journal,* **360,** 1-16

Shiffman, J., Beer, T. and Wu, Y. (2002) The emergence of global disease control priorities. *Health Policy Planning,* **17,** 225-234.

Shousha, A.T. (1948) The eradication of *Anopheles gambiae* from Upper Egypt 1942-1945. *Bulletin of the World Health Organisation*, **1,** 309-342

Simonsen, K.L., Churchhill, G.A. and Aquadro, C.A. (1995) Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*, **141,**413-429

Siva-Jothy, M.T. (2000) A mechanistic link between parasite resistance and expression of a sexually selected trait in a damselfly. *Proceedings of the Royal Society of Biological Sciences,* **267,** 2523-2527.

Slatkin, M. and Hudson, R.R. (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics,* **129,** 555-562

Slatkin, M. and Wiehe, T. (1998) Genetic hitch-hiking in a subdivided population. *Genetics Research,* **71,** 155-160.

Slotman, M.A., Della Torre, A., Calzetta, M. and Powell, J.R. (2005) Differential introgression of chromosomal regions between *Anopheles gambiae* and *Anopheles arabiensis*. *American Journal of Tropical Medicine and Hygiene*, **73,** 326-335.

Solensky, M.J. and Larkin, E. (2003) Temperature-induced variation in larval coloration in *Danaus plexippus* (Lepidoptera: *Nymphalidae*), *Annals of the Entomological Society of America,* **96** , 211-216

Soper, F.L. and Wilson, D.B. (1943) *Anopheles gambiae* in Brazil. 1943: The Rockefeller Foundation

Stearne, S.C. (1992) The evolution of Life- histories, Oxford University Press, Oxford pp 249

Stillwell, R. C., Wallin, W.G., Hitchcock, L. J. and Fox, C.W. (2007) Phenotypic plasticity in a complex world: interactive effects of food and temperature on fitness components of a seed beetle. *Oecologia*, **153**, 309-321

Storz, J.F. (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**: 671-688.

Storz, J.F. and Nachman, M.W. (2003) Natural selection on protein polymorphism in the rodent genus *Peromyscus*: evidence from interlocus contrasts. *Evolution International Journal of Organic Evolution,* **57,** 2628-2635.

Storz, J.F. and Dubach, J.M. (2004) Natural selection drives altitudinal divergence at the albumin locus in deer mice, *Peromyscus maniculatus*. *Evolution International Journal of Organic Evolution,* **58,** 1342-1352.

Stump, A.D., Atieli, F.K., Vulule, J.M. and Besansky, N.J. (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.

Sultan, S.E. (2000) Phenotypic plasticity for plant development, function and life history. *Trends in Plant Sciences,* **5,** 537-542.

Sultan, S.E. (2001) Phenotypic plasticity for fitness components in *Polygonum* species of contrasting ecological breadth, *Ecology,* **82,** 328–343 |

Takahashi, A., Takahashi, K., Ueda, R. and Takano-Shimizu, T. (2007) Natural variation of ebony gene controlling thoracic pigmentation in *Drosophila melanogaster*. *Genetics,* **177,** 1233-1237.

Talloen, W., Van Dyck, H. and Lens, L. (2004) The cost of melanization: butterfly wing coloration under environmental stress. *Evolution International Journal of Organic Evolution*, **58,** 360-366.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics,* **123,** 585-595.

Tang, A.H. and Tu, C.P. (1994) Biochemical characterization of *Drosophila glutathione S-transferases* D1 and D21. *Journal of Biological Chemistry*, **269,** 27876-27884.

Tatarenkov, A. and Ayala, F.J. (2007) Nucleotide variation at the *dopa decarboxylase* (Ddc) gene in natural populations of *Drosophila melanogaster*. *Journal of Genetics*, **86,** 125-137.

Tatarenkov, A., Zurovcova, M. and Ayala, F.J. (2001) DDC and AMD sequences resolve phylogenetic relationships of *Drosophila*. *Molecular Phylogenetics and Evolution,* **20,** 321-325.

Taylor, C., Toure, Y.T., Carnahan, J., Norris, D.E., Dolo, G., Traore, S.F., Edillo, F.E. and Lanzaro, G.C. (2001) Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, West Africa. *Genetics,* **157,** 743-750.

Tempel, B.L., Livingstone, M.S. and Quinn, W.G. (1984) Mutations in the *Dopa decarboxylase* gene affect learning in *Drosophila*. *Proceedings of the National Academy of Sciences of the USA,* **81**, 3577–3581.

Temu, E. A. and Yan, G. (2005) Microsatellite and mitochondrial genetic differentiation of *Anopheles arabiensis* (Diptera: *Culicidae*) from western Kenya, the Great Rift Valley, and coastal Kenya. *American Journal of Tropical Medicine and Hygiene,* **73,** 726-733.

Thompson, J.D. (1991) Phenotypic plasticity as a component of evolutionary change. *Trends in Ecology and Evolution,* **6**, 246-249.

Tollrian, R. and Harvell, C.D. (1999) The Evolution of inducible defenses: current ideas, In: *The Ecology and Evolution of Inducible Defenses*. Tollrian, R. and Harvell, C.D. (eds). Princeton University Press, Princeton , New Jersey.

Toure, Y.T., Petrarca, V., Traore, S.F., Coulibaly, A., Maiga, H.M., Sankare, O., Sow, M., Di Deco, M.A. and Coluzzi, M. (1994) Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae* s.str. in Mali, West Africa. *Genetica,* **94,** 213-223.

Toure, Y.T., Petrarca, V., Traore, S.F., Coulibaly, A., Maiga, H.M., Sankare, O., Sow, M., Di Deco, M.A. and Coluzzi, M. (1998a). The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia,* **40,** 477-511.

Toure, Y.T., Doumbo, O., Toure, A., Bagayoko, M., Diallo, M., Dolo, A., Vernick, K.D., Keister, D.B., Muratova, O. and Kaslow, D.C. (1998b) Gametocyte infectivity by direct mosquito feeds in an area of seasonal malaria transmission: implications for Bancoumana, Mali as a transmission-blocking vaccine site. *American Journal of Tropical Medicine and Hygiene*, **59,** 481-486.

Trape, J.F., Rogier, C., Konate, L., Diagne, N., Bouganali, H., Canque, B., Legros, F., Badji, A., Ndiaye, G., Ndiaye, P., Brahimi, K., Faye, O., Druilhe, P. and Pereira Da Silva, L.P. (1994) The Dielmo project: a longitudinal study of natural malaria infection and the mechanisms of protective immunity in a community living in a holoendemic area of Senegal. *American Journal of Tropical Medicine and Hygiene,* **51,** 123-137.

Travis, J. (1994). Ecological genetics of life-history traits: variation and its evolutionary significance. In: Real, L. A. (ed.) *Ecological Genetics*, pp. 171–204. Princeton University Press, Princeton, NJ.

Tripet, F., Wright, J., Cornel, A., Fofana, A., McAbee, R., Meneses, C., Reimer, L., Slotman, M., Thiemann, T., Dolo, G., Traore, S. and Lanzaro, G. (2007). Longitudinal survey of knockdown resistance to pyrethroid (kdr) in Mali, West Africa, and evidence of its emergence in the Bamako form of *Anopheles gambiae s.s. American Journal of Tropical Medicine and Hygiene*, **76,** 81-87.

True, J.R., Yeh, S.D., Hovemann, B.T., Kemme, T., Meinertzhagen, I.A., Edwards, T.N., Liou, S.R., Han, Q. and Li, J. (2005) *Drosophila* tan encodes a novel hydrolase required in pigmentation and vision. *PLoS Genetics,* **1,** e63.

True, J.R. (2003) Insect melanism: the molecules matter. *Trends in Ecology and Evolution*, **8,**640-647

True, J.R., Edwards, K.A., Yamamoto, D. and Carroll, S.B. (1999) *Drosophila* wing melanin patterns form by vein-dependent elaboration of enzymatic prepatterns. *Current Biology,* **9,** 1382-1391.

Tufto, J. (2000) The Evolution of Plasticity and Nonplastic Spatial and Temporal Adaptations in the Presence of Imperfect Environmental Cues. *The American Naturalist,* **156,** 121-130.

Turner, T.L., Hahn, M.W. and Nuzhdin, S.V. (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology,* **3,** e285.

Verhaeghen, K., Van Bortel, W., Roelants, P., Backeljau, T. and 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.

Via, S., Gomulkiewicz, R., De Jong, G., Scheiner, S.M., Schlichting, C.D. and Van Tienderen, P.H. (1995). Adaptive phenotypic plasticity: consensus and controversy. *Trends in Ecology and Evolution*, **5**: 212–217.

Vontas, J.G., Small, G.J. and Hemingway, J. (2001) *Glutathione S-transferases* as antioxidant defence agents confer pyrethroid resistance in *Nilaparvata lugens*. *Biochemical Journal,* **357,** 65-72.

Vulule, J.M., Beach, R.F., Atieli, F.K., Roberts, J.M., Mount, D.L. and Mwangi, R.W. (1994) Reduced susceptibility of *Anopheles gambiae* to permethrin associated with the use of permethrin-impregnated bednets and curtains in Kenya. *Medical and Veterinary Entomoloy*, **8,** 71-75.

Vulule, J.M., Beach, R.F., Atieli, F.K., McAllister, J.C., Brogdon, W.G., Roberts, J.M., Mwangi, R.W. and Hawley, W.A. (1999) Elevated oxidase and esterase levels associated with permethrin tolerance in *Anopheles gambiae* from Kenyan villages using permethrin-impregnated nets. *Medical and Veterinary Entomology,* **13,** 239-244.

Wang, D., Marsh, J.L. and Ayala, F.J. (1996) Evolutionary changes in the expression pattern of a developmentally essential gene in three *Drosophila* species. *Proceedings of the National Academy of Sciences,* **93,** 7103-7107.

Weill, M., Chandre, F., Brengues, C., Manguin, S., Akogbeto, M., Pasteur, N., Guillet, P. and Raymond, M. (2000) The *kdr* mutation occurs in the Mopti form of *Anopheles gambiae s.s.* through introgression. *Insect Molecular Biology*, **9,** 451-455.

Weir, B.S. (1979) Inferences about linkage disequilibrium. *Biometrics,* **35,** 235-254.

West-Eberhard, M.J. (2003) Developmental plasticity and evolution. Oxford University Press, Oxford

Wheelock, C.E., Eder, K.J., Werner, I., Huang, H., Jones, P.D., Brammell, B.F., Elskus, A.A. and Hammock, B.D. (2005) Individual variability in esterase activity and *CYP1A* levels in Chinook salmon (*Oncorhynchus tshawytscha*) exposed to esfenvalerate and chlorpyrifos. *Aquatic Toxicology*, **74,** 172-192.

White, B.J., Santolamazza, F., Kamau, L., Pombi, M., Grushko, O., Mouline, K., Brengues, C., Guelbeogo, W., Coulibaly, M., Kayondo, J.K., Sharakhov, I., Simard, F., Petrarca, V., Della Torre, A. and Besansky, N.J. (2007) Molecular karyotyping of the 2La inversion in *Anopheles gambiae. American Journal of Tropical Medicine and Hygiene*, **76,** 334-339.

White, B.J, Matthew, W.H., Pombi, M., Cassone, B.J., Lobo, N.F., Simard, F. and Besansky N.J. (2007) Localization of Candidate Regions Maintaining a Common Polymorphic Inversion (2La) in *Anopheles gambiae. PLoS Genetics,* **3**, 217

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

Wiernasz, D.C (1989) Female Choice and Sexual Selection of Male Wing Melanin Pattern in *Pieris occidentalis* (Lepidoptera) *Evolution,* **43**:1672-1682

Wilder, J.A., Dyreson, E.G., O'Neill, R.J., Spangler, M.L., Gupta, R., Wilder, A.S. and Hollocher, H. (2004) Contrasting modes of natural selection acting on pigmentation genes in the *Drosophila dunni* subgroup. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution,* **302,** 469-482.

Wilson, K., Cotter, S.C., Reeson, A.F. and Pell, J.K. (2001) Melanism and disease resistance in insects. *Ecology letters*, **4**, 637-649

Windig, J.J. (1998) Evolutionary genetics of fluctuating asymmetry in the peacock butterfly *(Inachis io). Heredity,* **80,** 382-392.

Windig, J.J. (1999) Trade-offs between melanization, development time and adult size in *Inachis io* and *Araschnia levana* (Lepidoptera: *Nymphalidae*)? *Heredity,* **82,** 57-68

Wittkopp, P.J., Carroll, S.B. and Kopp, A. (2003a) Evolution in black and white: genetic control of pigment patterns in *Drosophila. Trends in Genetic,* **19,** 495-504.

Wittkopp, P.J., Williams, B.L., Selegue, J.E. and Carroll, S.B. (2003b) *Drosophila* pigmentation evolution: divergent genotypes underlying convergent phenotypes. *Proceedings of the National Academy of Sciences*, USA, **100,** 1808-1813.

Wright, S. (1946) Isolation by Distance under Diverse Systems of Mating. *Genetics,* **31,** 39-59.

Wright, T.R.F. (1996) Phenotypic analysis of the *Dopa decarboxylase* gene cluster mutants in *Drosophila. The journal of Heredity*, **87,** 175-190

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

World Health Organization, (1957). Sixth Report, Expert Committee on Malaria. *World Health Organization Technical Report Series*, **123**.

World Health Organization, (2003) The Africa malaria report, *World Health Organization,* **69,**1-115

Xu, Q., Wang, H., Zhang, L. and Liu, N. (2006) Sodium channel gene expression associated with pyrethroid resistant house flies and German cockroaches. *Gene,* **379,** 62-67.

Yawson, A.E., McCall, P.J., Wilson, M.D. and Donnelly, M.J. (2004) Species abundance and insecticide resistance of *Anopheles gambiae* in selected areas of Ghana and Burkina Faso. *Medical and Veterinary Entomology,* **18,** 372-377.

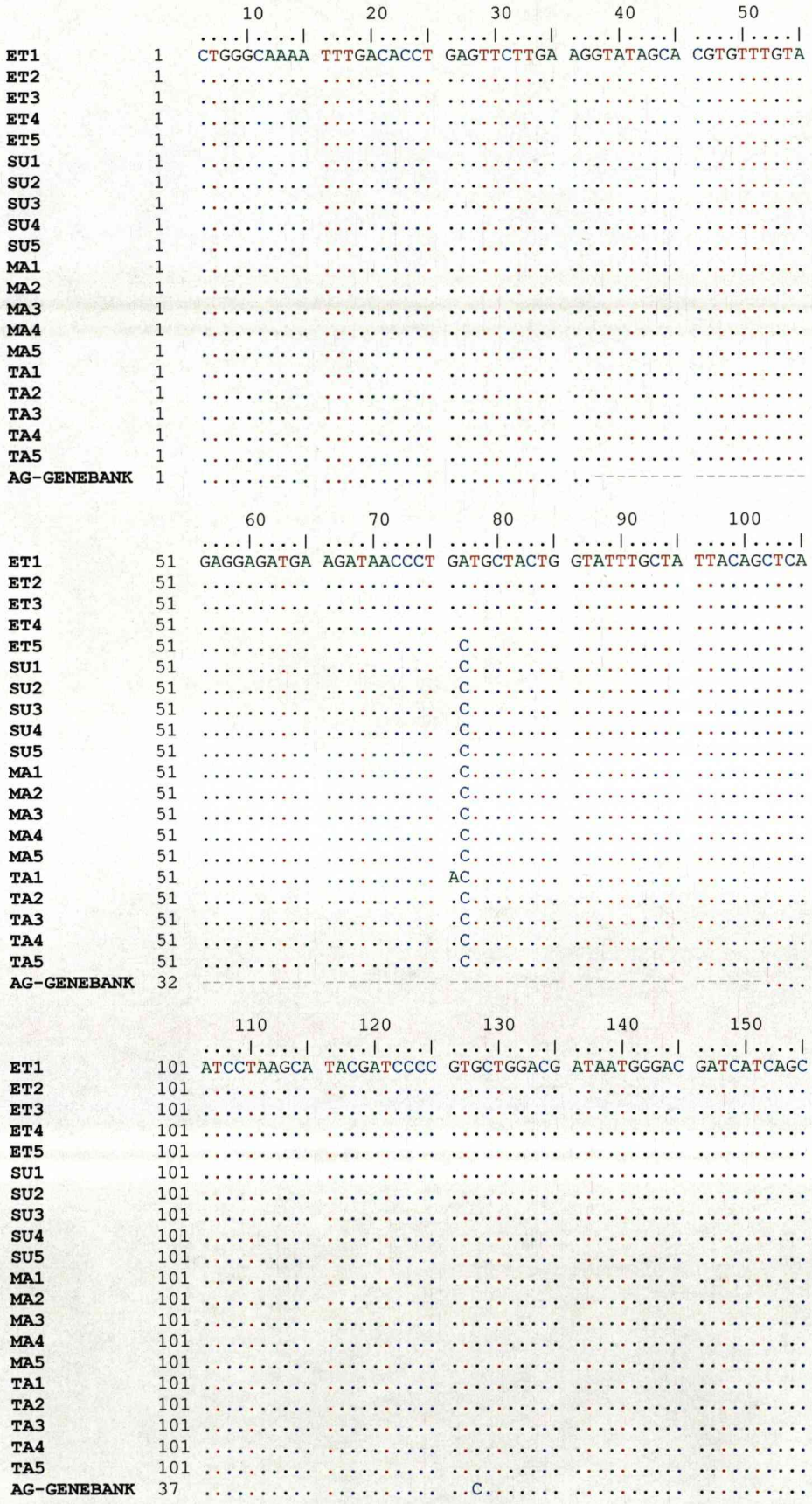
Zahar, A.R. (1984) Vector control operations in the African context. *Bulletin of the World Health Organization*, **62,** 89-100.

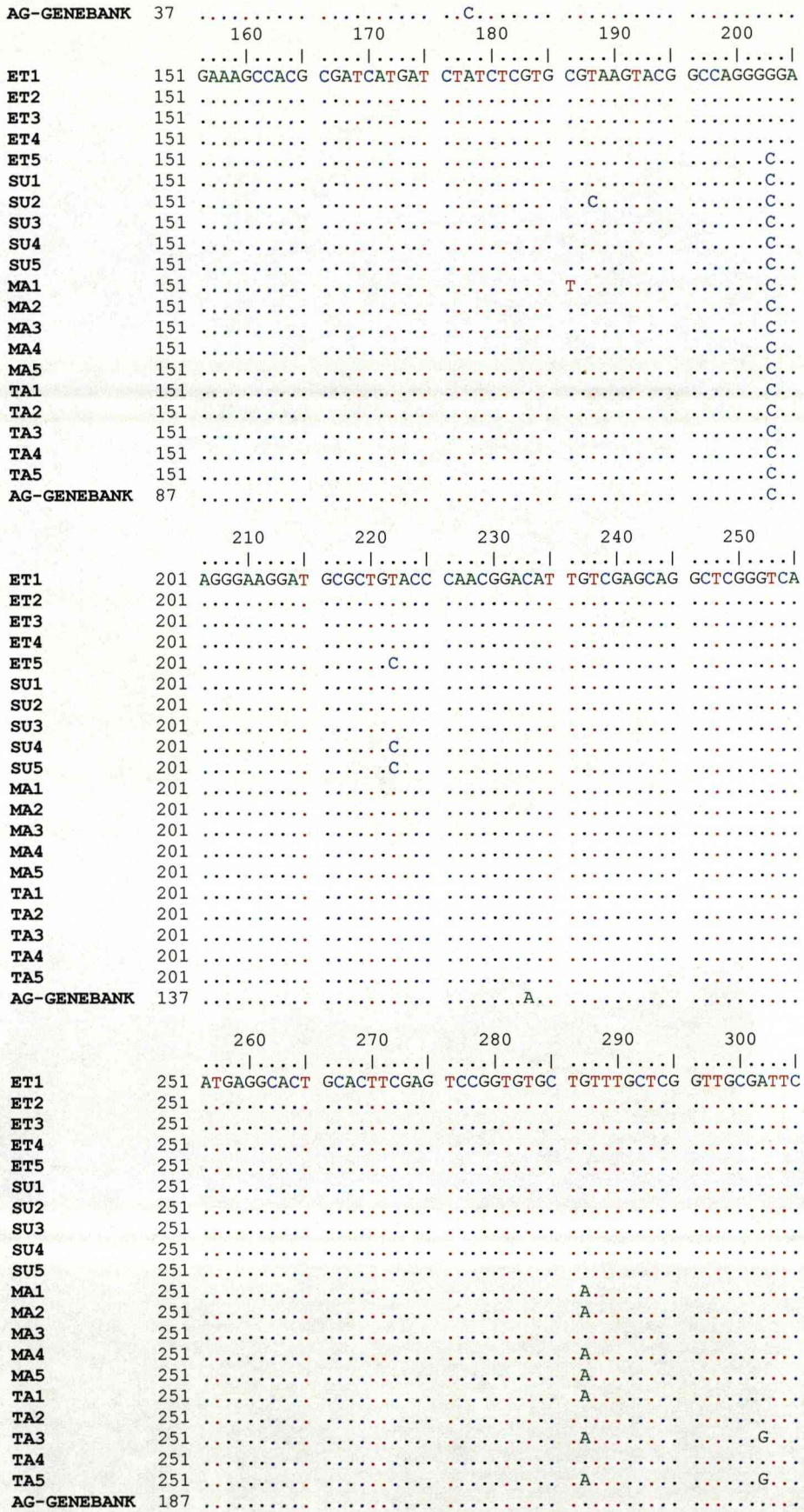
Zaim, M., Aitio, A. and Nakashima, N. (2000) Safety of pyrethroid-treated mosquito nets. *Medical and Veterinary Entomology*, **14,** 1-5.

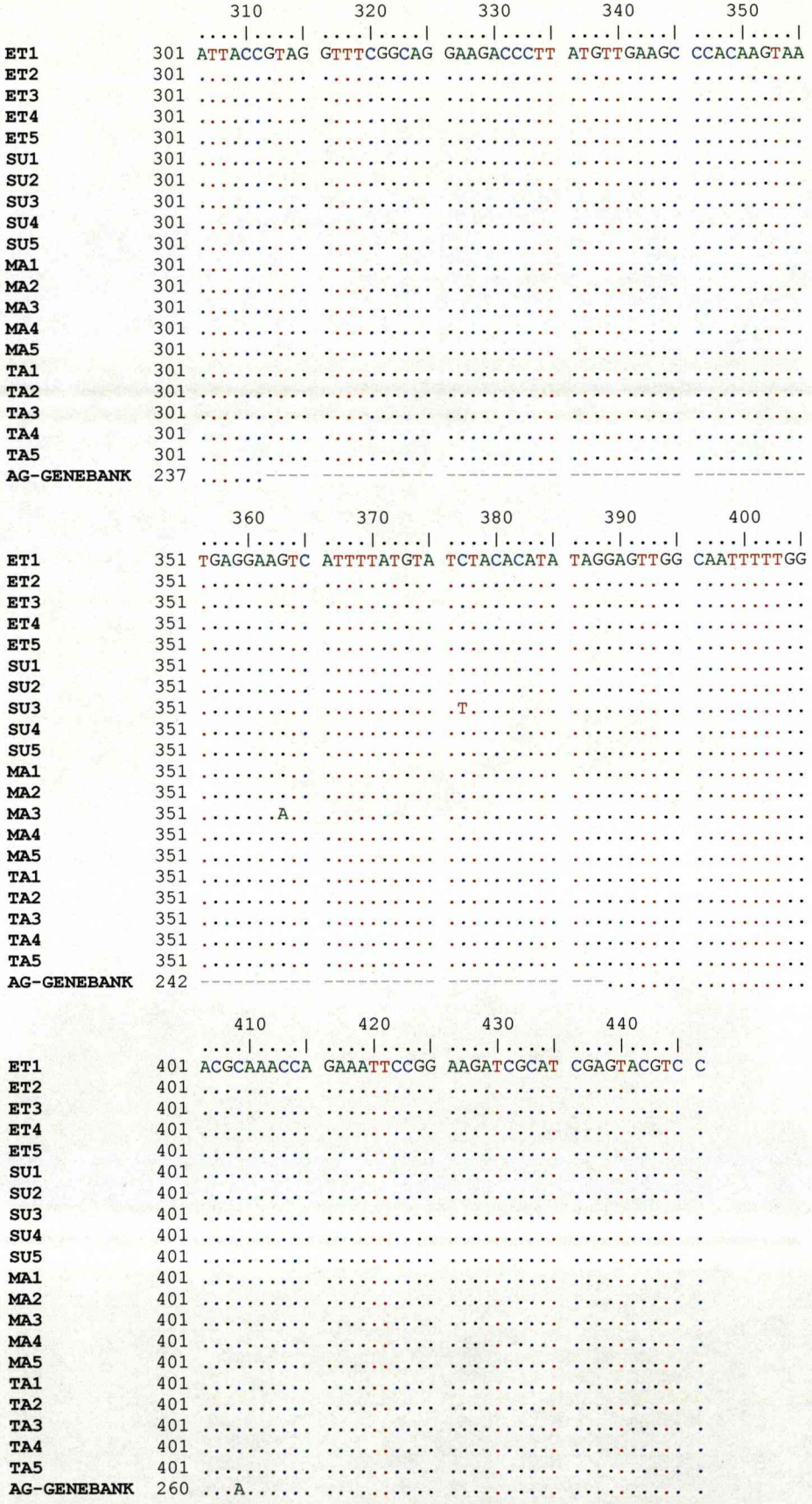
Zhang, J. 2003. Evolution by gene duplication: an update. *Trends in Ecology and Evolution,* **18,** 292-298.

APPENDICES

Appendix 1 DNA sequence alignment of GSTe1

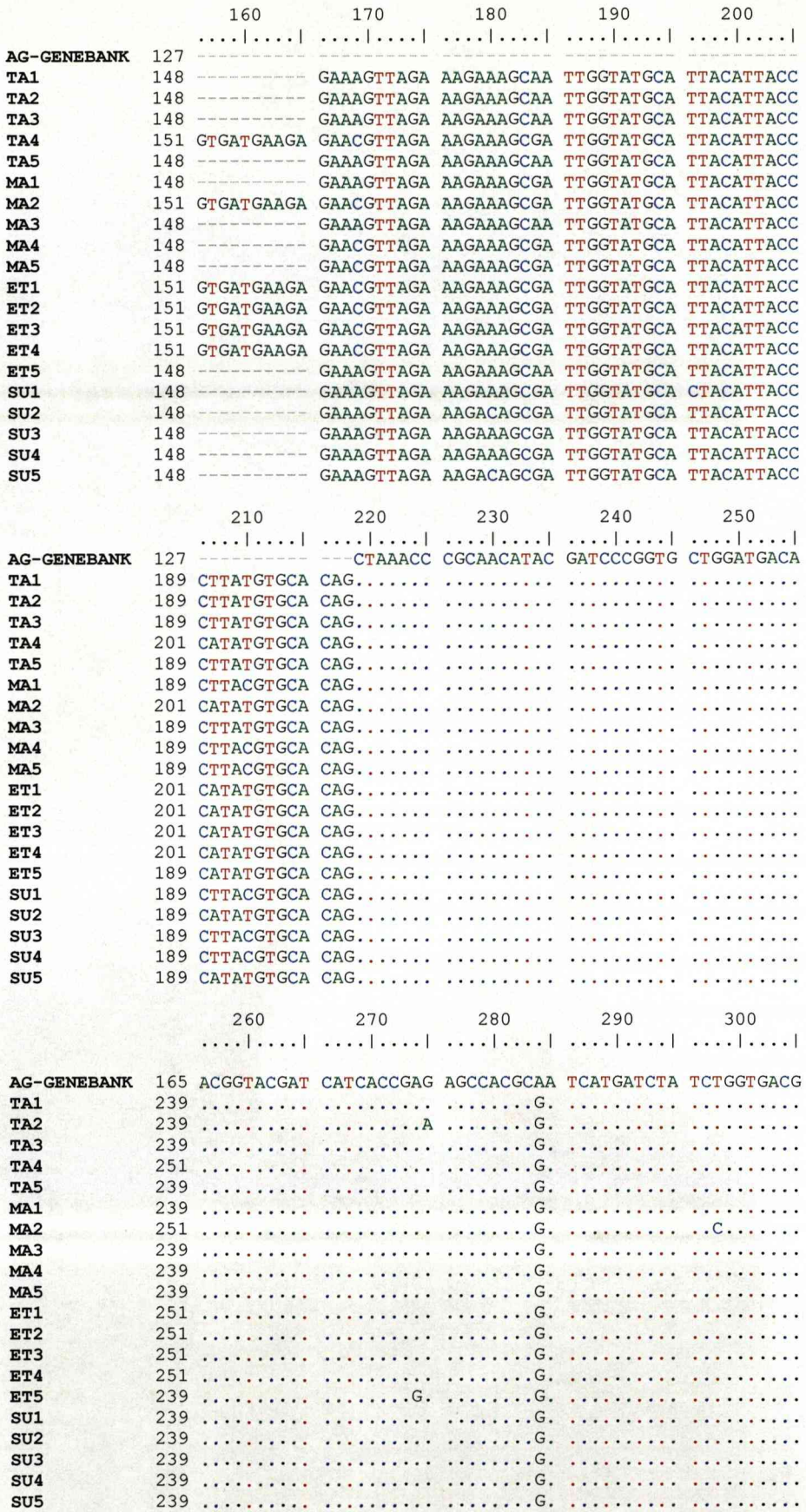


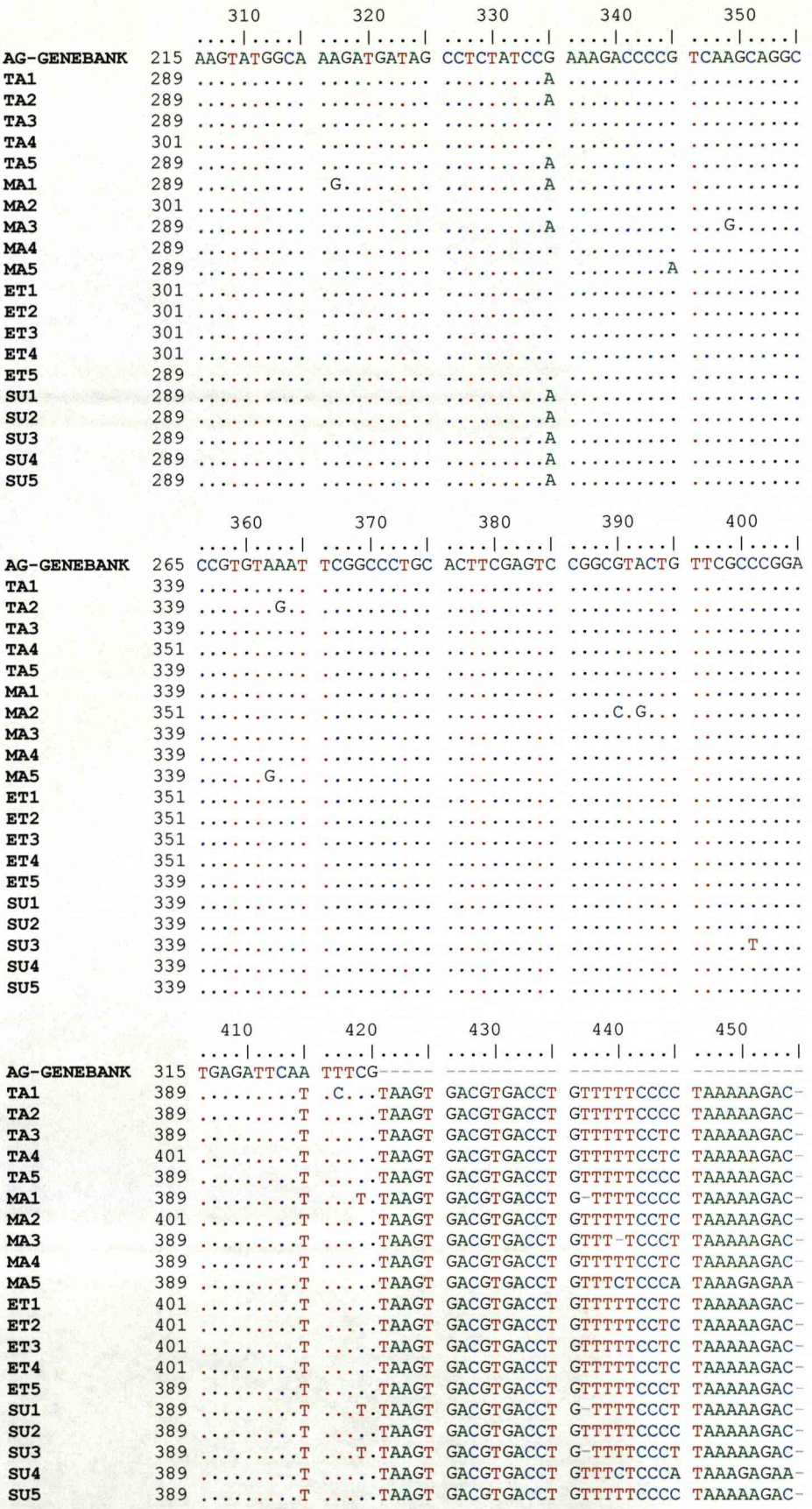

AG-GENEBANK 37
160 170 180 190 200
ET1 151 GAAAGCCACG CGATCATGAT CTATCTCGTG CGTAAGTACG GCCAGGGGGA
ET2 151
ET3 151
ET4 151
ET5 151
SU1 151
SU2 151
SU3 151
SU4 151
SU5 151
MA1 151
MA2 151
MA3 151
MA4 151
MA5 151
TA1 151
TA2 151
TA3 151
TA4 151
TA5 151
AG-GENEBANK 87
210 220 230 240 250
ET1 201 AGGGAAGGAT GCGCTGTACC CAACGGACAT TGTCGAGCAG GCTCGGGTCA
ET2 201
ET3 201
ET4 201
ET5 201
SU1 201
SU2 201
SU3 201
SU4 201
SU5 201
MA1 201
MA2 201
MA3 201
MA4 201
MA5 201
TA1 201
TA2 201
TA3 201
TA4 201
TA5 201
AG-GENEBANK 137
260 270 280 290 300
ET1 251 ATGAGGCACT GCACTTCGAG TCCGGTGTGC TGTTTGCTCG GTTGCGATTC
ET2 251
ET3 251
ET4 251
ET5 251
SU1 251
SU2 251
SU3 251
SU4 251
SU5 251
MA1 251
MA2 251
MA3 251
MA4 251
MA5 251
TA1 251
TA2 251
TA3 251
TA4 251
TA5 251
AG-GENEBANK 187

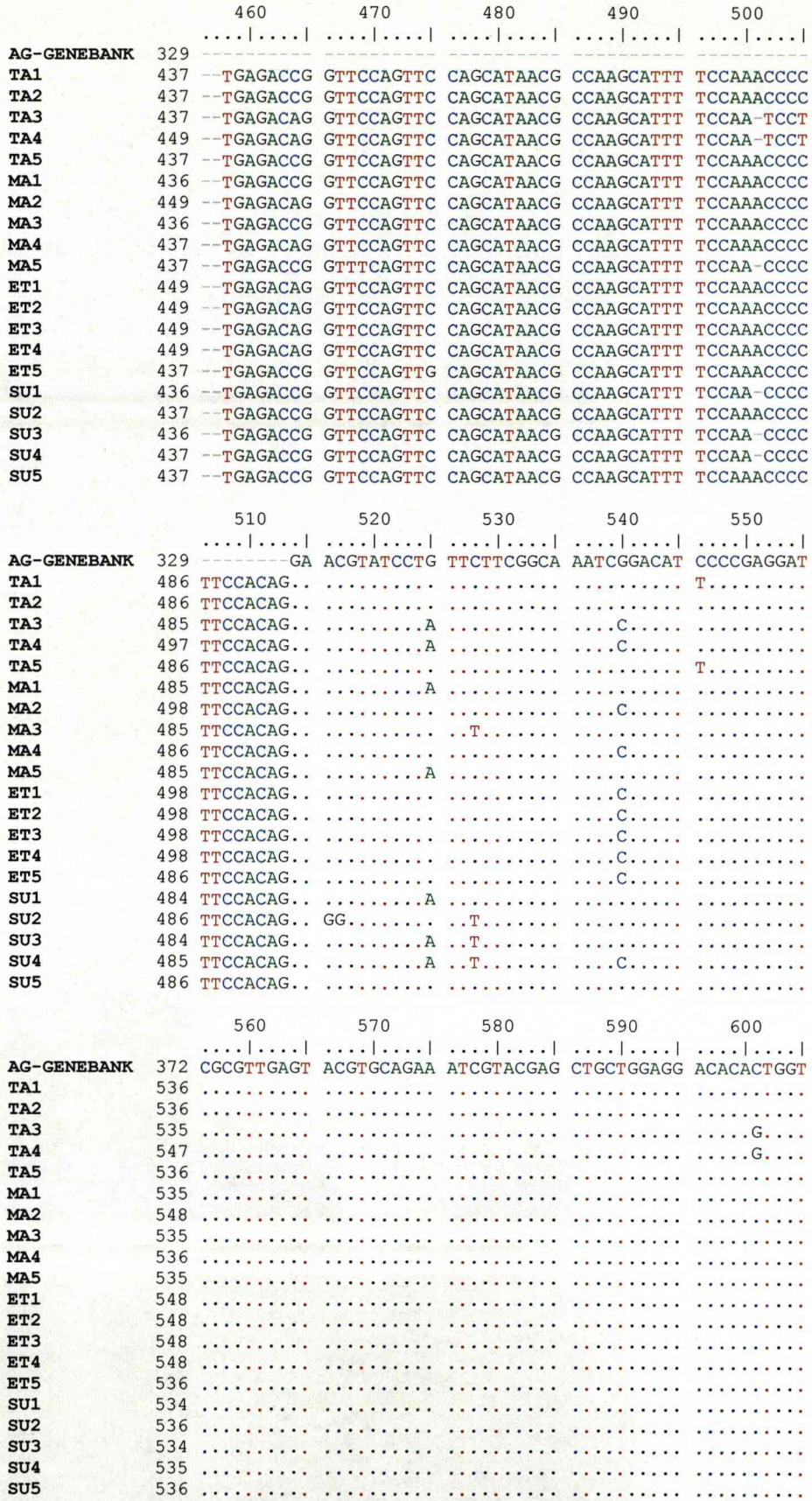

310 320 330 340 350
ET1 301 ATTACCGTAG GTTTCGGCAG GAAGACCCTT ATGTTGAAGC CCACAAGTAA
ET2 301 ....
ET3 301 ....
ET4 301 ....
ET5 301 ....
SU1 301 ....
SU2 301 ....
SU3 301 ....
SU4 301 ....
SU5 301 ....
MA1 301 ....
MA2 301 ....
MA3 301 ....
MA4 301 ....
MA5 301 ....
TA1 301 ....
TA2 301 ....
TA3 301 ....
TA4 301 ....
TA5 301 ....
AG-GENEBANK 237 ......---- ---------- ---------- ---------- ----------
360 370 380 390 400
ET1 351 TGAGGAAGTC ATTTTATGTA TCTACACATA TAGGAGTTGG CAATTTTTGG
ET2 351 ....
ET3 351 ....
ET4 351 ....
ET5 351 ....
SU1 351 ....
SU2 351 ....
SU3 351 .... .T........ ....
SU4 351 ....
SU5 351 ....
MA1 351 ....
MA2 351 ....
MA3 351 ....A.. ....
MA4 351 ....
MA5 351 ....
TA1 351 ....
TA2 351 ....
TA3 351 ....
TA4 351 ....
TA5 351 ....
AG-GENEBANK 242 ---------- ---------- ---------- ---..... ....
410 420 430 440
ET1 401 ACGCAAACCA GAAATTCCGG AAGATCGCAT CGAGTACGTC C
ET2 401 ....
ET3 401 ....
ET4 401 ....
ET5 401 ....
SU1 401 ....
SU2 401 ....
SU3 401 ....
SU4 401 ....
SU5 401 ....
MA1 401 ....
MA2 401 ....
MA3 401 ....
MA4 401 ....
MA5 401 ....
TA1 401 ....
TA2 401 ....
TA3 401 ....
TA4 401 ....
TA5 401 ....
AG-GENEBANK 260 ...A...... ....

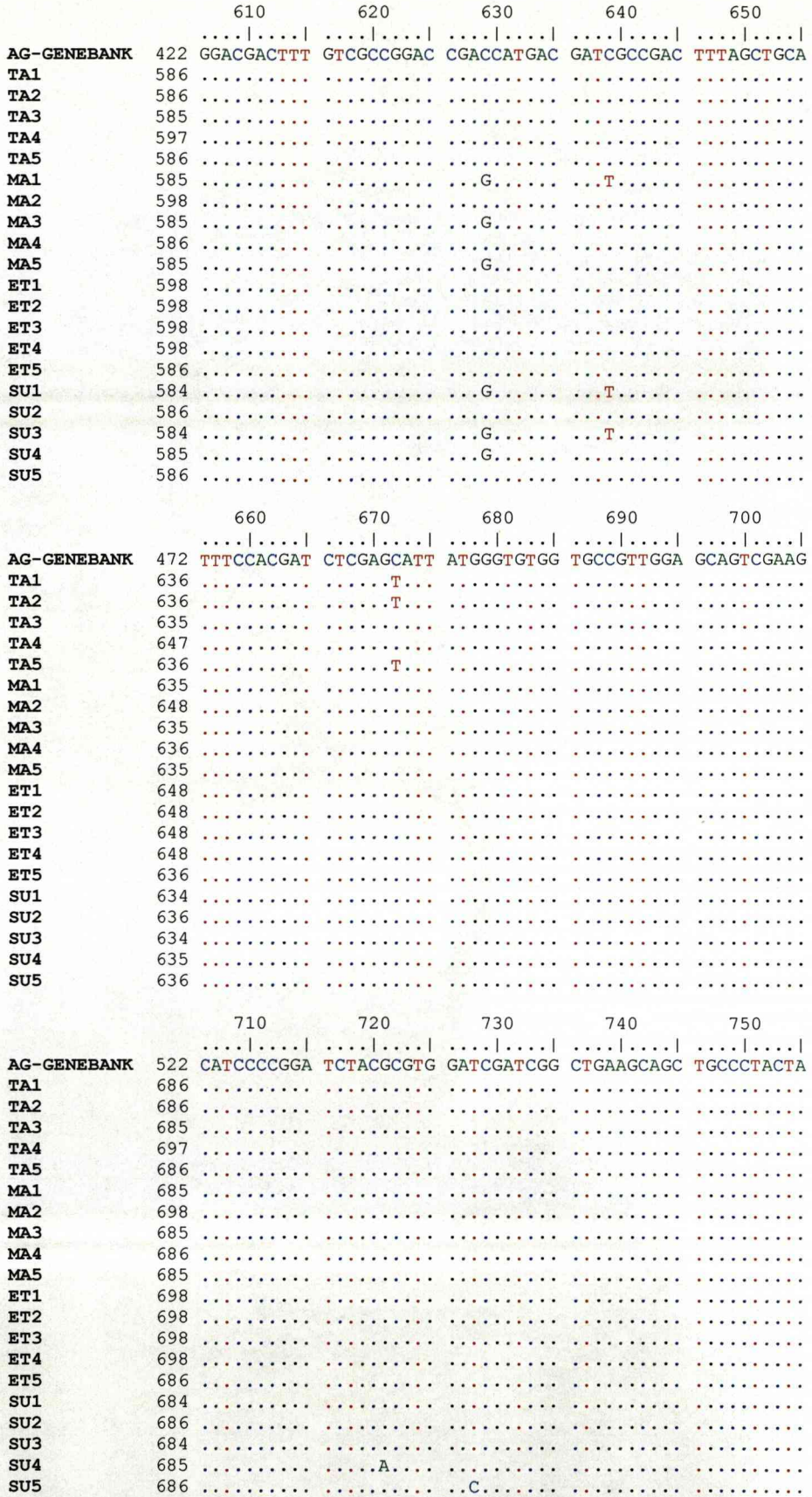
APPENDIX 2 DNA sequence alignment of GSTe2

```
                      10         20         30         40         50
                      ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK       1   GTACACCCTG CACCTTAGCC CACCGTGCCG TGCCGTGGAG CTGACGGCCA
TA1               1   .......... .......... .......... .......... ...G......
TA2               1   .......... .......... .......... .......... ...G......
TA3               1   .......... .......... .......... .......... ...G......
TA4               1   .......... .......... .......... .......... ..........
TA5               1   .......... .......... .......... .......... ...G......
MA1               1   .......... .......... ....A..... .......... ...G......
MA2               1   .......... .......... .......... .......... ..........
MA3               1   .......... .......... .......... .......... ..........
MA4               1   .......... .......... .......... .......... ..........
MA5               1   .......... .......... .......... .......... ..........
ET1               1   .......... .......... .......... .......... ..........
ET2               1   .......... .......... .......... .......... ..........
ET3               1   .......... .......... .......... .......... .........G
ET4               1   .......... .......... .......... .......... ..........
ET5               1   .......... .......... .......... .......... ...G......
SU1               1   .......... .......... .......... .......... ...G......
SU2               1   .......... .......... .......... .......... ...G......
SU3               1   .......... .......... .......... .......... ...G......
SU4               1   .......... .......... .......... .......... ...G......
SU5               1   .......... .......... .......... .......... ...G......

                      60         70         80         90         100
                      ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK      51   AAGCATTGGG CTTGGAGCTG GAGCAGAAGA CCATTAATCT GCTAACGGGT
TA1              51   ....G..... .......... .......... .......... ...G......
TA2              51   ....G..... .......... .......... .......... ...G......
TA3              51   ....G..... .......... .......... .......... ...G......
TA4              51   .......... .......... .......... .......... ...G......
TA5              51   ....G..... .......... .......... .......... ...G......
MA1              51   ....G..... .......... .......... .......... ...G......
MA2              51   .......... .......... .......... .......... ...G......
MA3              51   ....G...A. .......... .......... .......... ...G......
MA4              51   ........A. .......... .......... .......... ...G......
MA5              51   ........A. .......... .......... .......... ...G......
ET1              51   .......... .......... .......... .......... ...G......
ET2              51   .......... .......... .......... .......... ...G......
ET3              51   .......... .......... .......... .......... ...G......
ET4              51   .......... .......... .......... .......... ...G......
ET5              51   ....G..... .......... .......... .......... ...G......
SU1              51   ....G..... .......... .......... .......... ...G......
SU2              51   ....G..... .......... .......... .......... ...G......
SU3              51   ....G..... .......... .......... .......... ...G......
SU4              51   ....G..... .......... .......... .......... ...G......
SU5              51   ....G..... .......... .......... .......... ...G......

                      110        120        130        140        150
                      ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK      101  GACCATTTGA AGCCGGAATT TGTGAAG--- ---------- ----------
TA1              101  .......... .......... .......GTA CGTAATGGGA TTGAGAGA--
TA2              101  .......... .......... .......GTA CGTAATGGGA TTGAGAGA--
TA3              101  .......... .......... .......GTA CGTAATGGGA TTGAGAGA--
TA4              101  .......... .......... .T.....GTA CGTAATGGGA TTGAGAGAAA
TA5              101  .......... .......... .......GTA CGTAATGGGA TTGAGAGA--
MA1              101  .......... .......... .......GTA CGTAATGGGA TTGAGAGA--
MA2              101  .......... .......... .T.....GTA CGTAATGGGA TTGAGAGAAA
MA3              101  .......... .......... .......GTA CGTAAAGGGA TTGAGAGA--
MA4              101  .......... .......... .......GTA CGTAATGGGA TTGAGAGA--
MA5              101  .......... .......... .......GTA CGTAATGGGA TTGAGAGA--
ET1              101  .......... .......... .T.....GTA CGTAATGGGA TTGAGAGAAA
ET2              101  .......... .......... .T.....GTA CGTAATGGGA TTGAGAGAAA
ET3              101  .......... .......G.. .T.....GTA CGTAATGGGA TTGAGAGAAA
ET4              101  .......... .......... .T.....GTA CGTAATGGGA TTGAGAGAAA
ET5              101  .......... .......... .......GTA CGTAATGGGA TTGAGAGA--
SU1              101  .......... .......... .......GTA CGTAATGGGA TTGAGAGA--
SU2              101  .......... .......... .......GTA CGTAATGGGA TTGAGAGA--
SU3              101  .......... .......... .......GTA CGTAATGGGA TTGAGAGA--
SU4              101  .......... .......... .......GTA CGTAATGGGA TTGAGAGA--
SU5              101  .......... .......... .......GTA CGTAATGGGA TTGAGAGA--
```


160 170 180 190 200
AG-GENEBANK 127 ---------- ---------- ---------- ---------- ----------
TA1 148 ---------- GAAAGTTAGA AAGAAAGCAA TTGGTATGCA TTACATTACC
TA2 148 ---------- GAAAGTTAGA AAGAAAGCAA TTGGTATGCA TTACATTACC
TA3 148 ---------- GAAAGTTAGA AAGAAAGCAA TTGGTATGCA TTACATTACC
TA4 151 GTGATGAAGA GAACGTTAGA AAGAAAGCGA TTGGTATGCA TTACATTACC
TA5 148 ---------- GAAAGTTAGA AAGAAAGCAA TTGGTATGCA TTACATTACC
MA1 148 ---------- GAAAGTTAGA AAGAAAGCGA TTGGTATGCA TTACATTACC
MA2 151 GTGATGAAGA GAACGTTAGA AAGAAAGCGA TTGGTATGCA TTACATTACC
MA3 148 ---------- GAAAGTTAGA AAGAAAGCAA TTGGTATGCA TTACATTACC
MA4 148 ---------- GAACGTTAGA AAGAAAGCGA TTGGTATGCA TTACATTACC
MA5 148 ---------- GAACGTTAGA AAGAAAGCGA TTGGTATGCA TTACATTACC
ET1 151 GTGATGAAGA GAACGTTAGA AAGAAAGCGA TTGGTATGCA TTACATTACC
ET2 151 GTGATGAAGA GAACGTTAGA AAGAAAGCGA TTGGTATGCA TTACATTACC
ET3 151 GTGATGAAGA GAACGTTAGA AAGAAAGCGA TTGGTATGCA TTACATTACC
ET4 151 GTGATGAAGA GAACGTTAGA AAGAAAGCGA TTGGTATGCA TTACATTACC
ET5 148 ---------- GAAAGTTAGA AAGAAAGCAA TTGGTATGCA TTACATTACC
SU1 148 ---------- GAAAGTTAGA AAGAAAGCGA TTGGTATGCA CTACATTACC
SU2 148 ---------- GAAAGTTAGA AAGACAGCGA TTGGTATGCA TTACATTACC
SU3 148 ---------- GAAAGTTAGA AAGAAAGCGA TTGGTATGCA TTACATTACC
SU4 148 ---------- GAAAGTTAGA AAGAAAGCGA TTGGTATGCA TTACATTACC
SU5 148 ---------- GAAAGTTAGA AAGACAGCGA TTGGTATGCA TTACATTACC
210 220 230 240 250
AG-GENEBANK 127 ---------- ---CTAAACC CGCAACATAC GATCCCGGTG CTGGATGACA
TA1 189 CTTATGTGCA CAG..... ....
TA2 189 CTTATGTGCA CAG..... ....
TA3 189 CTTATGTGCA CAG..... ....
TA4 201 CATATGTGCA CAG..... ....
TA5 189 CTTATGTGCA CAG..... ....
MA1 189 CTTACGTGCA CAG..... ....
MA2 201 CATATGTGCA CAG..... ....
MA3 189 CTTATGTGCA CAG..... ....
MA4 189 CTTACGTGCA CAG..... ....
MA5 189 CTTACGTGCA CAG..... ....
ET1 201 CATATGTGCA CAG..... ....
ET2 201 CATATGTGCA CAG..... ....
ET3 201 CATATGTGCA CAG..... ....
ET4 201 CATATGTGCA CAG..... ....
ET5 189 CATATGTGCA CAG..... ....
SU1 189 CTTACGTGCA CAG..... ....
SU2 189 CATATGTGCA CAG..... ....
SU3 189 CTTACGTGCA CAG..... ....
SU4 189 CTTACGTGCA CAG..... ....
SU5 189 CATATGTGCA CAG..... ....
260 270 280 290 300
AG-GENEBANK 165 ACGGTACGAT CATCACCGAG AGCCACGCAA TCATGATCTA TCTGGTGACG
TA1 239 ....G. ....
TA2 239 ....A .....G. ....
TA3 239 ....G. ....
TA4 251 ....G. ....
TA5 239 ....G. ....
MA1 239 ....G. ....
MA2 251 ....G. ....C.....
MA3 239 ....G. ....
MA4 239 ....G. ....
MA5 239 ....G. ....
ET1 251 ....G. ....
ET2 251 ....G. ....
ET3 251 ....G. ....
ET4 251 ....G. ....
ET5 239 ....G. .....G. ....
SU1 239 ....G. ....
SU2 239 ....G. ....
SU3 239 ....G. ....
SU4 239 ....G. ....
SU5 239 ....G. ....


310 320 330 340 350
AG-GENEBANK 215 AAGTATGGCA AAGATGATAG CCTCTATCCG AAAGACCCCG TCAAGCAGGC
TA1 289 A
TA2 289 A
TA3 289
TA4 301
TA5 289 A
MA1 289 .G..... A
MA2 301
MA3 289 A ...G.....
MA4 289
MA5 289 A
ET1 301
ET2 301
ET3 301
ET4 301
ET5 289
SU1 289 A
SU2 289 A
SU3 289 A
SU4 289 A
SU5 289 A
360 370 380 390 400
AG-GENEBANK 265 CCGTGTAAAT TCGGCCCTGC ACTTCGAGTC CGGCGTACTG TTCGCCCGGA
TA1 339
TA2 339 .....G..
TA3 339
TA4 351
TA5 339
MA1 339
MA2 351 ....C.G...
MA3 339
MA4 339
MA5 339 .....G...
ET1 351
ET2 351
ET3 351
ET4 351
ET5 339
SU1 339
SU2 339
SU3 339 .....T....
SU4 339
SU5 339
410 420 430 440 450
AG-GENEBANK 315 TGAGATTCAA TTTCG----- ---------- ---------- ----------
TA1 389 .....T .C...TAAGT GACGTGACCT GTTTTTCCCC TAAAAAGAC-
TA2 389 .....T .....TAAGT GACGTGACCT GTTTTTCCCC TAAAAAGAC-
TA3 389 .....T .....TAAGT GACGTGACCT GTTTTTCCTC TAAAAAGAC-
TA4 401 .....T .....TAAGT GACGTGACCT GTTTTTCCTC TAAAAAGAC-
TA5 389 .....T .....TAAGT GACGTGACCT GTTTTTCCCC TAAAAAGAC-
MA1 389 .....T ...T.TAAGT GACGTGACCT G-TTTTCCCC TAAAAAGAC-
MA2 401 .....T .....TAAGT GACGTGACCT GTTTTTCCTC TAAAAAGAC-
MA3 389 .....T .....TAAGT GACGTGACCT GTTT-TCCCT TAAAAAGAC-
MA4 389 .....T .....TAAGT GACGTGACCT GTTTTTCCTC TAAAAAGAC-
MA5 389 .....T .....TAAGT GACGTGACCT GTTTCTCCCA TAAAGAGAA-
ET1 401 .....T .....TAAGT GACGTGACCT GTTTTTCCTC TAAAAAGAC-
ET2 401 .....T .....TAAGT GACGTGACCT GTTTTTCCTC TAAAAAGAC-
ET3 401 .....T .....TAAGT GACGTGACCT GTTTTTCCTC TAAAAAGAC-
ET4 401 .....T .....TAAGT GACGTGACCT GTTTTTCCTC TAAAAAGAC-
ET5 389 .....T .....TAAGT GACGTGACCT GTTTTTCCCT TAAAAAGAC-
SU1 389 .....T ...T.TAAGT GACGTGACCT G-TTTTCCCT TAAAAAGAC-
SU2 389 .....T .....TAAGT GACGTGACCT GTTTTTCCCC TAAAAAGAC-
SU3 389 .....T ...T.TAAGT GACGTGACCT G-TTTTCCCT TAAAAAGAC-
SU4 389 .....T .....TAAGT GACGTGACCT GTTTCTCCCA TAAAGAGAA-
SU5 389 .....T .....TAAGT GACGTGACCT GTTTTTCCCC TAAAAAGAC-


460 470 480 490 500
AG-GENEBANK 329 ---------- ---------- ---------- ---------- ----------
TA1 437 --TGAGACCG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAAACCCC
TA2 437 --TGAGACCG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAAACCCC
TA3 437 --TGAGACAG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAA-TCCT
TA4 449 --TGAGACAG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAA-TCCT
TA5 437 --TGAGACCG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAAACCCC
MA1 436 --TGAGACCG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAAACCCC
MA2 449 --TGAGACAG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAAACCCC
MA3 436 --TGAGACCG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAAACCCC
MA4 437 --TGAGACAG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAAACCCC
MA5 437 --TGAGACCG GTTTCAGTTC CAGCATAACG CCAAGCATTT TCCAA-CCCC
ET1 449 --TGAGACAG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAAACCCC
ET2 449 --TGAGACAG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAAACCCC
ET3 449 --TGAGACAG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAAACCCC
ET4 449 --TGAGACAG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAAACCCC
ET5 437 --TGAGACCG GTTCCAGTTG CAGCATAACG CCAAGCATTT TCCAAACCCC
SU1 436 --TGAGACCG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAA-CCCC
SU2 437 --TGAGACCG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAAACCCC
SU3 436 --TGAGACCG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAA-CCCC
SU4 437 --TGAGACCG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAA-CCCC
SU5 437 --TGAGACCG GTTCCAGTTC CAGCATAACG CCAAGCATTT TCCAAACCCC
510 520 530 540 550
AG-GENEBANK 329 --------GA ACGTATCCTG TTCTTCGGCA AATCGGACAT CCCCGAGGAT
TA1 486 TTCCACAG.. .... T.....
TA2 486 TTCCACAG.. ....
TA3 485 TTCCACAG.. .....A ....C..... .....
TA4 497 TTCCACAG.. .....A ....C..... .....
TA5 486 TTCCACAG.. .... T.....
MA1 485 TTCCACAG.. .....A ....
MA2 498 TTCCACAG.. ....C..... .....
MA3 485 TTCCACAG.. ....T..... ....
MA4 486 TTCCACAG.. ....C..... .....
MA5 485 TTCCACAG.. .....A ....
ET1 498 TTCCACAG.. ....C..... .....
ET2 498 TTCCACAG.. ....C..... .....
ET3 498 TTCCACAG.. ....C..... .....
ET4 498 TTCCACAG.. ....C..... .....
ET5 486 TTCCACAG.. ....C..... .....
SU1 484 TTCCACAG.. .....A ....
SU2 486 TTCCACAG.. GG..... ..T..... ....
SU3 484 TTCCACAG.. .....A ..T..... ....
SU4 485 TTCCACAG.. .....A ..T..... ....C..... .....
SU5 486 TTCCACAG.. ....
560 570 580 590 600
AG-GENEBANK 372 CGCGTTGAGT ACGTGCAGAA ATCGTACGAG CTGCTGGAGG ACACACTGGT
TA1 536 ....
TA2 536 ....
TA3 535 ....G....
TA4 547 ....G....
TA5 536 ....
MA1 535 ....
MA2 548 ....
MA3 535 ....
MA4 536 ....
MA5 535 ....
ET1 548 ....
ET2 548 ....
ET3 548 ....
ET4 548 ....
ET5 536 ....
SU1 534 ....
SU2 536 ....
SU3 534 ....
SU4 535 ....
SU5 536 ....

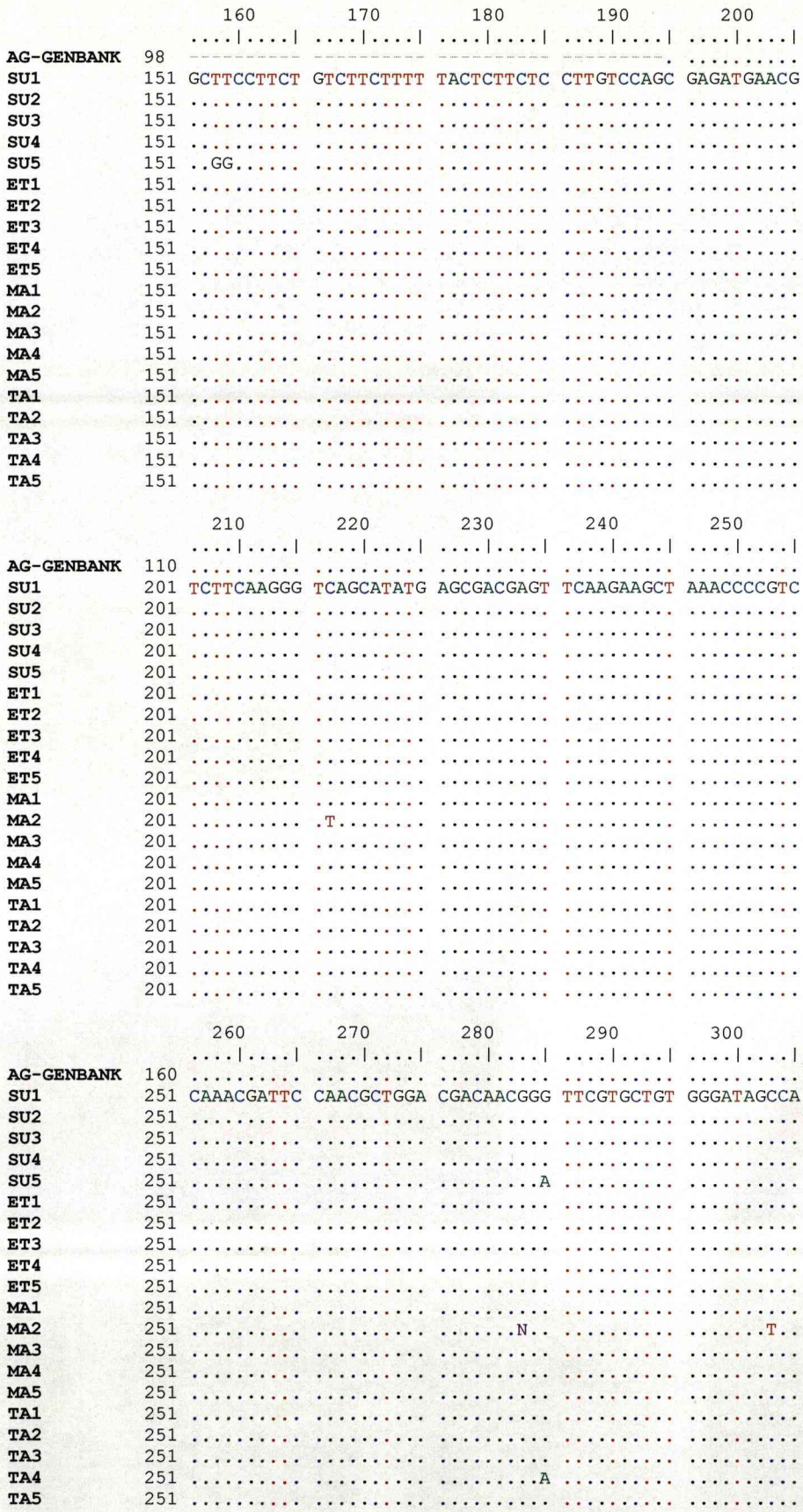

610 620 630 640 650
AG-GENEBANK 422 GGACGACTTT GTCGCCGGAC CGACCATGAC GATCGCCGAC TTTAGCTGCA
TA1 586
TA2 586
TA3 585
TA4 597
TA5 586
MA1 585 ...G... ...T...
MA2 598
MA3 585 ...G...
MA4 586
MA5 585 ...G...
ET1 598
ET2 598
ET3 598
ET4 598
ET5 586
SU1 584 ...G... ...T...
SU2 586
SU3 584 ...G... ...T...
SU4 585 ...G...
SU5 586
660 670 680 690 700
AG-GENEBANK 472 TTTCCACGAT CTCGAGCATT ATGGGTGTGG TGCCGTTGGA GCAGTCGAAG
TA1 636 .....T...
TA2 636 .....T...
TA3 635
TA4 647
TA5 636 .....T...
MA1 635
MA2 648
MA3 635
MA4 636
MA5 635
ET1 648
ET2 648
ET3 648
ET4 648
ET5 636
SU1 634
SU2 636
SU3 634
SU4 635
SU5 636
710 720 730 740 750
AG-GENEBANK 522 CATCCCCGGA TCTACGCGTG GATCGATCGG CTGAAGCAGC TGCCCTACTA
TA1 686
TA2 686
TA3 685
TA4 697
TA5 686
MA1 685
MA2 698
MA3 685
MA4 686
MA5 685
ET1 698
ET2 698
ET3 698
ET4 698
ET5 686
SU1 684
SU2 686
SU3 684
SU4 685 .....A....
SU5 686 ..C.....

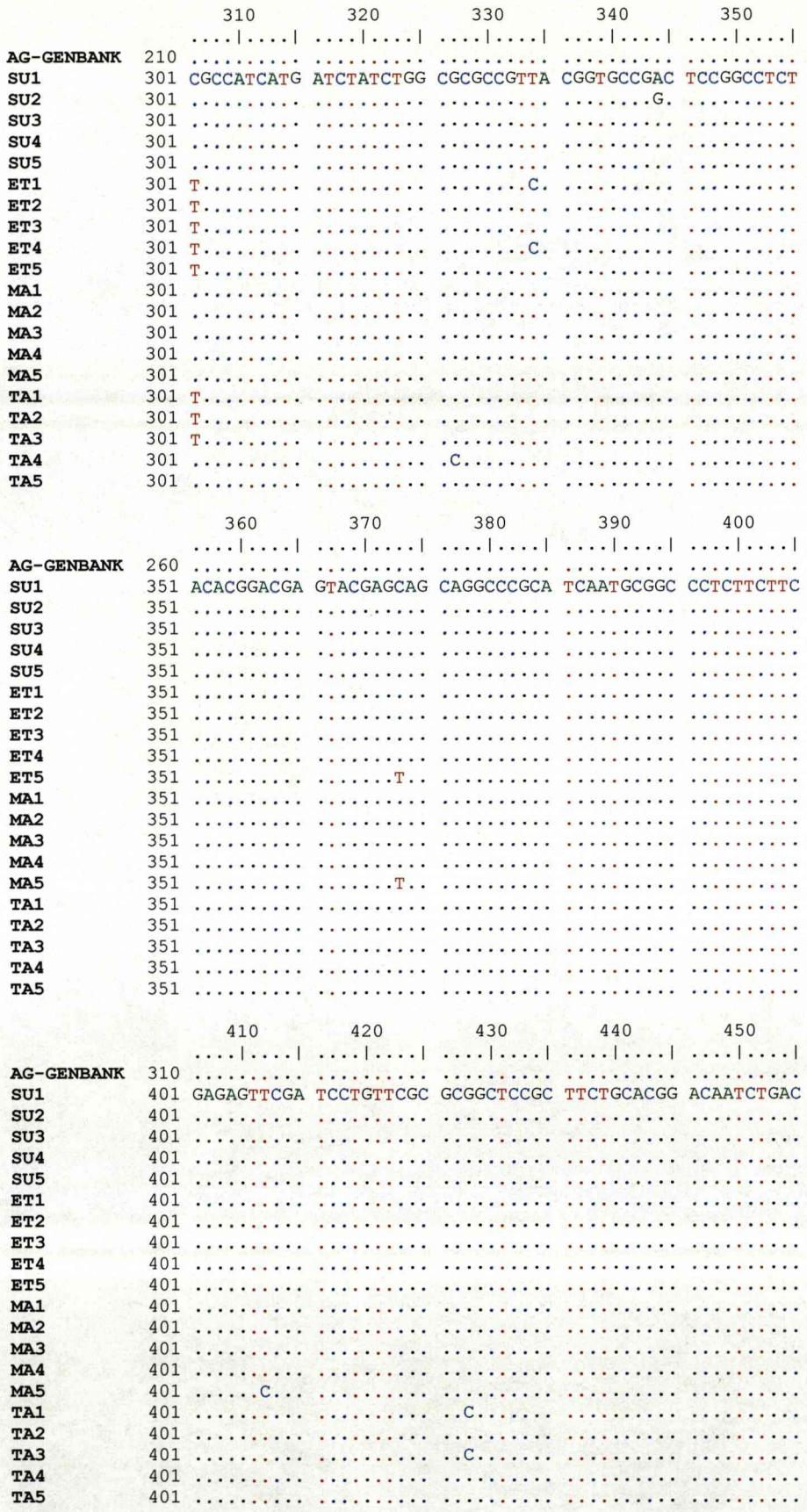
```
                        760         770         780         790         800
                   ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK    572 CGAGGAAGCG AACGGTGGCG GTGGCACCGA TCTGGGCAAG TTTGTGCTAG
TA1            736 .......... .......... .......... .......... ..........
TA2            736 .......... .......... .A........ .......... ..........
TA3            735 ......G... .........T. .A........ .......... ..........
TA4            747 ......G... .........T. ....T..... .......... ..........
TA5            736 .......... .......... .......... .......... ..........
MA1            735 .......... .......... .A........ .......... ..........
MA2            748 ......G... .........T. ....T..... .......... ..........
MA3            735 .......... .......... .......... .......... ..........
MA4            736 ......G... .........T. ....T..... .......... ..........
MA5            735 .......... .......... .......... .......... ..........
ET1            748 ......G... .........T. ....T..... .......... ..........
ET2            748 ......G... .........T. ....T..... .......... ..........
ET3            748 ......G... .........T. ....T..... .......... ..........
ET4            748 ......G... .........T. ....T..... .......... ..........
ET5            736 .......... .......... .......... .......... ..........
SU1            734 .......... .......... .A........ .......... ..........
SU2            736 .......... .......... .......... .......... ..........
SU3            734 .......... .......... .A........ .......... ..........
SU4            735 .......... .......... .......... .......... ..........
SU5            736 .......... .......... .......... .......... ..........
```

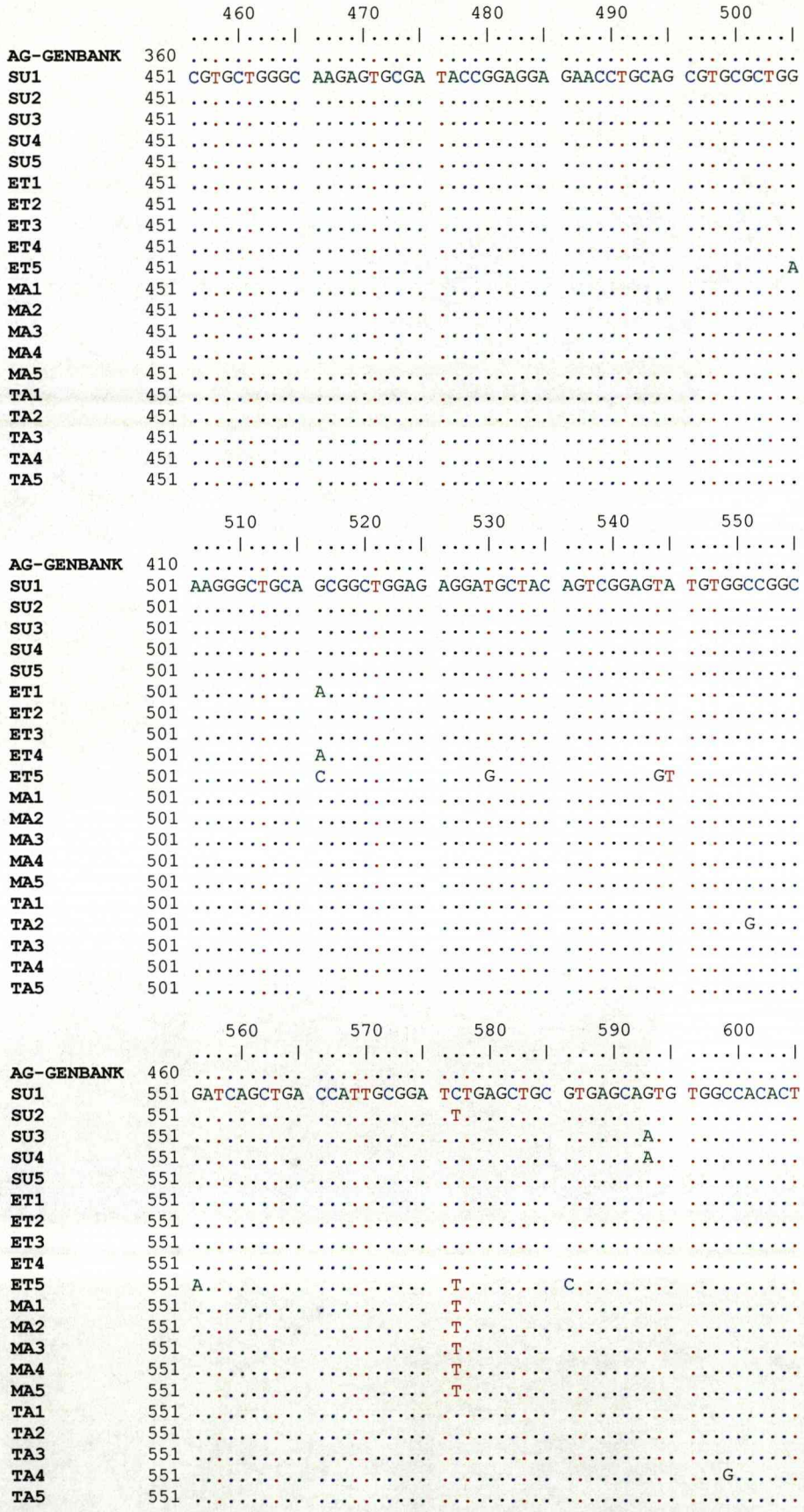
```
AG-GENEBANK    621
TA1            785
TA2            785
TA3            784
TA4            796
TA5            785
MA1            784
MA2            797
MA3            784
MA4            785
MA5            784
ET1            797
ET2            797
ET3            797
ET4            797
ET5            785
SU1            783
SU2            785
SU3            783
SU4            784
SU5            785
```

Appendix 3 DNA sequence alignment of GSTe6

```
                    10         20         30         40         50
                    ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENBANK      1   .......... .......... .......... .......... ..........
SU1             1   ATGTCGAGCA AGCCGGTCCT GTACACGCAC ACGATTAGTC CCGCCGGCCG
SU2             1   .......... .......... .......... .......... ..........
SU3             1   .......... .......... .......... .......... ..........
SU4             1   .......... .......... .......... .......... ..........
SU5             1   .......... .......... .......... .......... ..........
ET1             1   .......... .......... .......... .T........ ..........
ET2             1   ..A....... .......... .......... .......... ..........
ET3             1   .......... .......... .......... .......... ..........
ET4             1   .......... .......... .......... .T........ ..........
ET5             1   .......... .......... .......... .T........ ..........
MA1             1   .......... .......... .......... .......... ..........
MA2             1   .......... .......... .......... .......... ..........
MA3             1   .......... .......... .......... .......... ..........
MA4             1   .......... .......... .......... .......... ..........
MA5             1   .......... .......... .......... .......... ..........
TA1             1   .......... .......... .......... .......... ..........
TA2             1   .......... .......... .......... .......... ..........
TA3             1   .......... .......... .......... .......... ..........
TA4             1   .......... .......... .......... .......... ..........
TA5             1   .......... .......... .......... .......... ..........

                    60         70         80         90         100
                    ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENBANK     51   .......... .......... .......... .......... ........--
SU1            51   TGCGGTCGAG CTGACCGTGA AGGCGTTGAA CCTTGACGTC GATGTTCGGT
SU2            51   .......... .......... .......... .......... ..........
SU3            51   .......... .......... .......... .......... ..........
SU4            51   .......... .......... .......... .......... ..........
SU5            51   .......... .......... .......... .......... ..........
ET1            51   A......... ........A. .......... .......... ..........
ET2            51   A......... ........A. .......... .......... ..........
ET3            51   A......... ........A. .......... .......... ..........
ET4            51   A......... ........A. .......... .......... ..........
ET5            51   A......... ........A. .......... .......... ..........
MA1            51   A......... ........A. .......... .......... ..........
MA2            51   .......... .......... .......... .......... ..........
MA3            51   A......... ........A. .......... .......... ..........
MA4            51   A......... ........A. .......... .......... ..........
MA5            51   A......... ........A. .......... .......... ..........
TA1            51   A......... ........A. .......... .......... ..........
TA2            51   A......... ........A. .......... .......... ..........
TA3            51   A......... ........A. .......... .......... ..........
TA4            51   .......... .......... .......... .......... ..........
TA5            51   A......... ........A. .......... .......... ..........

                    110        120        130        140        150
                    ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENBANK     98   ---------- ---------- ---------- ---------- ----------
SU1            101  AAGCCACTGA AGAAGAGGAG TCGCCTTTTC ACGTGTGTCG TAAAACGAAT
SU2            101  .......... .......... .......... .......... ......T...
SU3            101  .......... .......... .......... .......... ..........
SU4            101  .......... .......... .......... .......... ..........
SU5            101  .......... .......... .......... ..A....... ......T...
ET1            101  .......... .......... .....C.... .......... ......T...
ET2            101  .......... .......... ..A..C.... .......... ......T...
ET3            101  .......... .......... ..A..C.... .......... ......T...
ET4            101  .......... .......... .....C.... .......... ......T...
ET5            101  .......... .......... .....C.... .......... ......T...
MA1            101  .......... ........G. ..A..C.... .......... ......T...
MA2            101  .......... .......... .......... .......... ......T...
MA3            101  .......... .......... ..A..C.... .......... ......T...
MA4            101  .......... ........G. ..A..C.... .......... ......T...
MA5            101  .......... .......... ..A..C.... .......... ......T...
TA1            101  .......... .......... ..A..C.... .......... ......T...
TA2            101  .......... .......... ..A..C.... .......... ......T...
TA3            101  .......... .......... ..A..C.... .......... ......T...
TA4            101  .......... .......... .......... ..A....... ......T...
TA5            101  .......... .......... ..A..C..A. .......... ......T...
```


160 170 180 190 200
AG-GENBANK 98
SU1 151 GCTTCCTTCT GTCTTCTTTT TACTCTTCTC CTTGTCCAGC GAGATGAACG
SU2 151
SU3 151
SU4 151
SU5 151 ..GG.....
ET1 151
ET2 151
ET3 151
ET4 151
ET5 151
MA1 151
MA2 151
MA3 151
MA4 151
MA5 151
TA1 151
TA2 151
TA3 151
TA4 151
TA5 151
210 220 230 240 250
AG-GENBANK 110
SU1 201 TCTTCAAGGG TCAGCATATG AGCGACGAGT TCAAGAAGCT AAACCCCGTC
SU2 201
SU3 201
SU4 201
SU5 201
ET1 201
ET2 201
ET3 201
ET4 201
ET5 201
MA1 201
MA2 201 .T.....
MA3 201
MA4 201
MA5 201
TA1 201
TA2 201
TA3 201
TA4 201
TA5 201
260 270 280 290 300
AG-GENBANK 160
SU1 251 CAAACGATTC CAACGCTGGA CGACAACGGG TTCGTGCTGT GGGATAGCCA
SU2 251
SU3 251
SU4 251
SU5 251 .....A.....
ET1 251
ET2 251
ET3 251
ET4 251
ET5 251
MA1 251
MA2 251 .....N.. .....T..
MA3 251
MA4 251
MA5 251
TA1 251
TA2 251
TA3 251
TA4 251 .....A.....
TA5 251

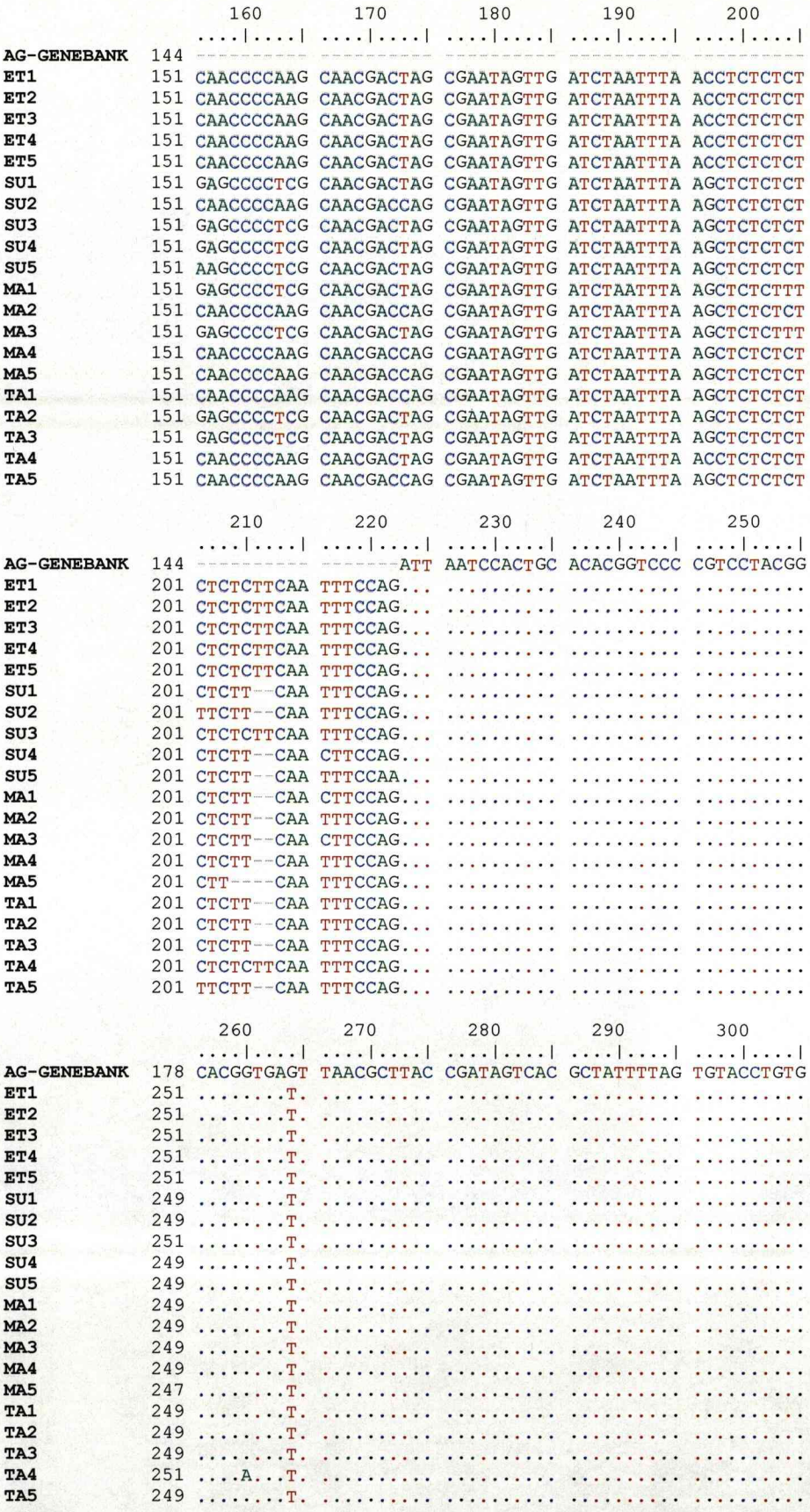

310 320 330 340 350
AG-GENBANK 210
SU1 301 CGCCATCATG ATCTATCTGG CGCGCCGTTA CGGTGCCGAC TCCGGCCTCT
SU2 301
SU3 301
SU4 301
SU5 301
ET1 301
ET2 301
ET3 301
ET4 301
ET5 301
MA1 301
MA2 301
MA3 301
MA4 301
MA5 301
TA1 301
TA2 301
TA3 301
TA4 301
TA5 301
360 370 380 390 400
AG-GENBANK 260
SU1 351 ACACGGACGA GTACGAGCAG CAGGCCCGCA TCAATGCGGC CCTCTTCTTC
SU2 351
SU3 351
SU4 351
SU5 351
ET1 351
ET2 351
ET3 351
ET4 351
ET5 351
MA1 351
MA2 351
MA3 351
MA4 351
MA5 351
TA1 351
TA2 351
TA3 351
TA4 351
TA5 351
410 420 430 440 450
AG-GENBANK 310
SU1 401 GAGAGTTCGA TCCTGTTCGC GCGGCTCCGC TTCTGCACGG ACAATCTGAC
SU2 401
SU3 401
SU4 401
SU5 401
ET1 401
ET2 401
ET3 401
ET4 401
ET5 401
MA1 401
MA2 401
MA3 401
MA4 401
MA5 401
TA1 401
TA2 401
TA3 401
TA4 401
TA5 401

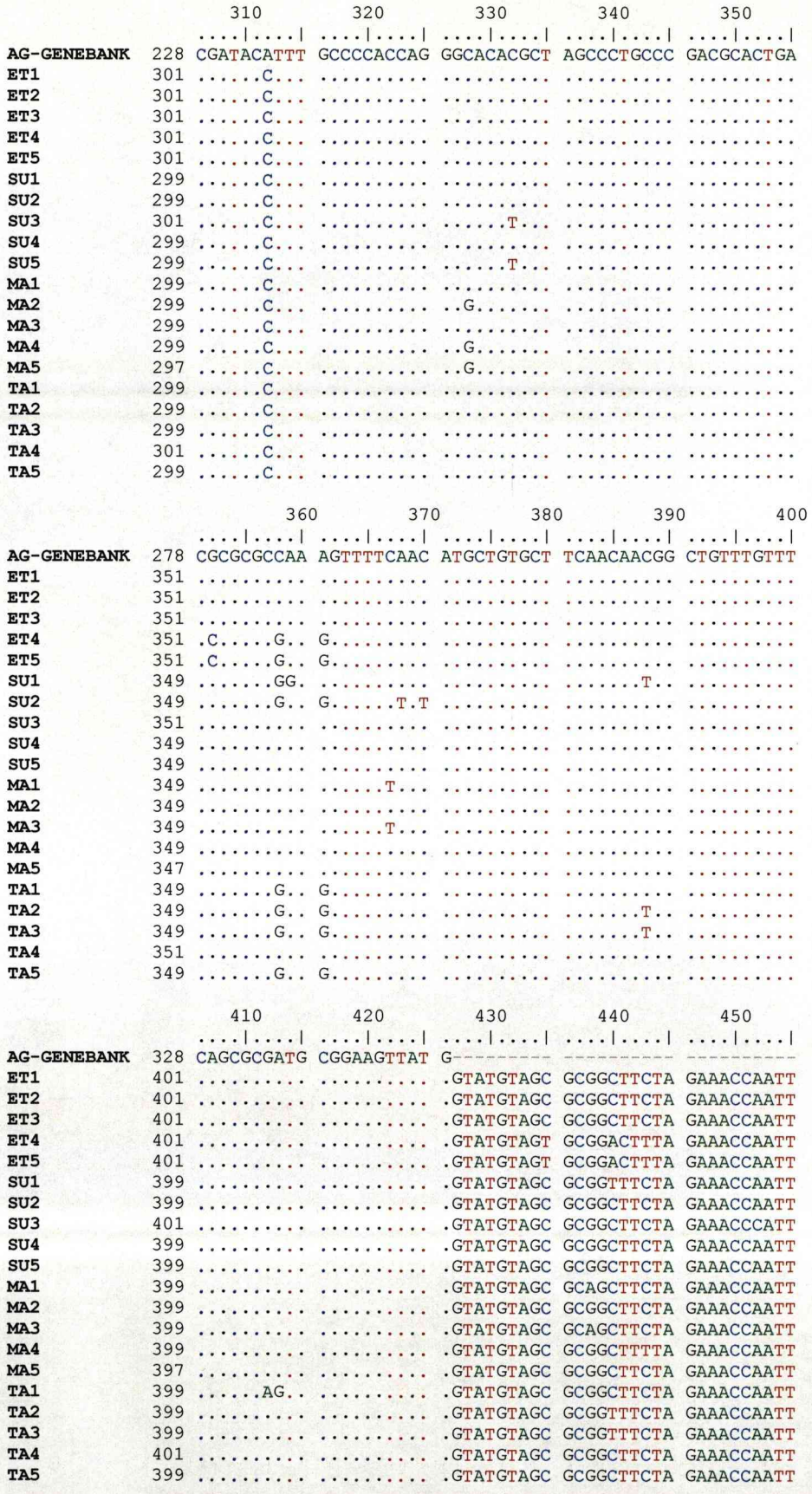

460 470 480 490 500
AG-GENBANK 360
SU1 451 CGTGCTGGGC AAGAGTGCGA TACCGGAGGA GAACCTGCAG CGTGCGCTGG
SU2 451
SU3 451
SU4 451
SU5 451
ET1 451
ET2 451
ET3 451
ET4 451
ET5 451 .A
MA1 451
MA2 451
MA3 451
MA4 451
MA5 451
TA1 451
TA2 451
TA3 451
TA4 451
TA5 451
510 520 530 540 550
AG-GENBANK 410
SU1 501 AAGGGCTGCA GCGGCTGGAG AGGATGCTAC AGTCGGAGTA TGTGGCCGGC
SU2 501
SU3 501
SU4 501
SU5 501
ET1 501 A
ET2 501
ET3 501
ET4 501 A
ET5 501 C ....G..... .....GT
MA1 501
MA2 501
MA3 501
MA4 501
MA5 501
TA1 501
TA2 501 .....G....
TA3 501
TA4 501
TA5 501
560 570 580 590 600
AG-GENBANK 460
SU1 551 GATCAGCTGA CCATTGCGGA TCTGAGCTGC GTGAGCAGTG TGGCCACACT
SU2 551 .T
SU3 551 .....A..
SU4 551 .....A..
SU5 551
ET1 551
ET2 551
ET3 551
ET4 551
ET5 551 A .T..... C.....
MA1 551 .T
MA2 551 .T
MA3 551 .T
MA4 551 .T
MA5 551 .T
TA1 551
TA2 551
TA3 551
TA4 551 ...G.....
TA5 551

```
                      610        620        630        640        650
                 ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENBANK   510 .......... .......... .......... .......... ..........
SU1          601 GCACCTGATG CTGAAACCGT CGGCGGAAGA GTTCCCCAAA ACGTTCGCCT
SU2          601 .......... .......... .......... .......C.. ..........
SU3          601 .......... .......... .......... .......... ..........
SU4          601 .......... .......... .......... .......... ..........
SU5          601 .......... .......... .......... .......... ..........
ET1          601 .......... .......... .......... .......... ..........
ET2          601 .......... .......... .......... .......... ..........
ET3          601 .......... .......... .......... .......... ..........
ET4          601 .......... .......... .......... .......... ..........
ET5          601 .......... .......... .......... .......... ..........
MA1          601 .......... .......... .......... .......... ..........
MA2          601 .......... .......... .......... .......... ..........
MA3          601 .......... .......... .......... .......C.. ..........
MA4          601 .......... .......... .......... .......... ..........
MA5          601 .......... .......... .......... .......C.. ..........
TA1          601 .......... .......... .......... .......... ..........
TA2          601 .......... .......... .......... .......... ..........
TA3          601 .......... .......... .......... .......... ..........
TA4          601 .......... .......... .......... .......... ..........
TA5          601 .......... .......... .......... .......... ..........

                      660        670        680        690        700
                 ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENBANK   560 .G....C.. .T........ .......... .......... ..........
SU1          651 GAATGGAACG GGTGTCGAAG TTGCCGTACT ACGGGGAGGT GATGGGACGG
SU2          651 .G........ .T........ .......... .......... ..........
SU3          651 .G........ .T........ .......... .......... ..........
SU4          651 .G........ .T........ .......... .......... ..........
SU5          651 .G........ .T........ .......... .......... ..........
ET1          651 .G........ .T........ .......... .......... ..........
ET2          651 .G........ .......... .......... .......... T.........
ET3          651 .G........ .......... .......... .......... T.........
ET4          651 .G........ .T........ .......... .......... ..........
ET5          651 .G........ .T........ .......... .......... ..........
MA1          651 .......... .T........ .......... .......... ..........
MA2          651 .......... .T........ .......... .......... ..........
MA3          651 .G........ .T........ .......... .......... ..........
MA4          651 .......... .T........ .......... .......... ..........
MA5          651 .G........ .T........ .......... .......... ..........
TA1          651 .G........ .T........ .......... .......... ..........
TA2          651 .G........ .T........ .......... .......... ..........
TA3          651 .G........ .T........ .......... .......... ..........
TA4          651 .G........ .T........ .......... .......... ..........
TA5          651 .G........ .T........ .......... .......... ..........

                      710        720        730
                 ....|....| ....|....| ....|....| ....|
AG-GENBANK   610 ....T....G ...C....G. ........AG .....
SU1          701 GGGCCTAAAA CGGGCGGAAA GCTGATGCCA AACGC
SU2          701 ....T....G ...C....G. ........AG .....
SU3          701 ....T....G ...C....G. ........AG .....
SU4          701 ....T....G ...C....G. ........AG .....
SU5          701 ....T....G ...C....G. ........AG .....
ET1          701 ....T....G ...C....G. ........AG ....A
ET2          701 ....T....G ...C...... ........AG .....
ET3          701 ....T....G ...C...... ........AG .....
ET4          701 ....T....G ...C....G. ........AG ....A
ET5          701 ....T....G ...C....G. ........AG .....
MA1          701 .........G ...C....G. ........AG .....
MA2          701 .........G ...C....G. ........AG .....
MA3          701 ....T....G ...C....G. ........AG .....
MA4          701 .........G ...C....G. ........AG .....
MA5          701 ....T....G ...C....G. ........AG .....
TA1          701 ....T....G ...C....G. ........AG .....
TA2          701 ....T....G ...C....G. ........AG .....
TA3          701 ....T....G ...C....G. ........AG .....
TA4          701 ....T....G ...C....G. ........AG .CGCT
TA5          701 ....T....G ...C....G. ........AG .....
```

Appendix 4 DNA sequence alignment of GSTe8

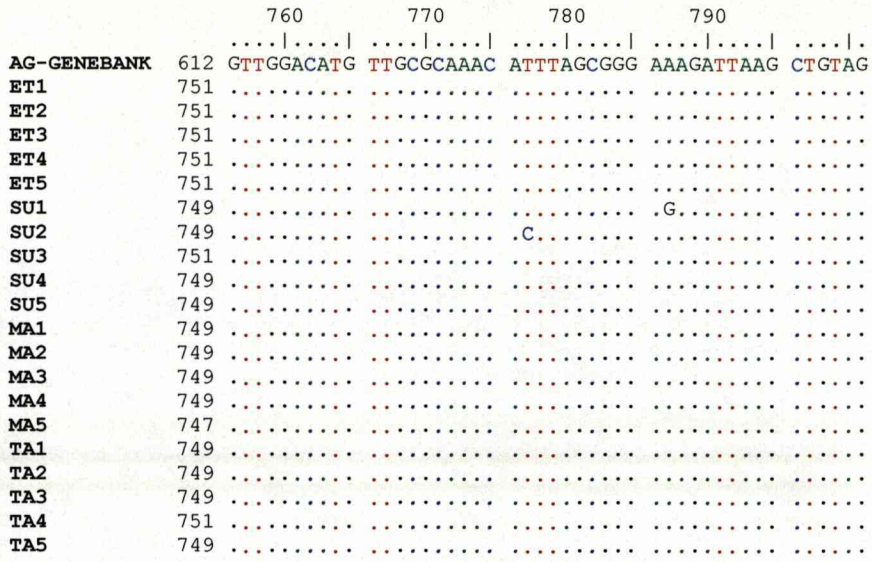
```
                        10        20        30        40        50
                   ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK      1 GCCATGATTC TGTACTACGA CGAGGTCAGC CCACCGGTTC GGGGCGTCCT
ET1              1 .......... .......... .........T .......... ..........
ET2              1 .......... .......... .........T .......... ..........
ET3              1 .......... .......... .........T .......... ..........
ET4              1 .......... .......... .........T .......... ..........
ET5              1 .......... .......... .........T .......... ..........
SU1              1 .......... ......T... .........T .......... ..........
SU2              1 .......... .......... .....C...T .......... ..........
SU3              1 .......... .......... .........T .......... ..........
SU4              1 .......... .......... .........T .......... ..........
SU5              1 .......... .......... .........T .......... ..........
MA1              1 .......... .......... .........T .......... ..........
MA2              1 .......... .......... .........T .......... ..........
MA3              1 .......... .......... .........T .......... ..........
MA4              1 .......... .......... .........T .......... ..........
MA5              1 .......... .......... .........T .......... ..........
TA1              1 .......... .......... .........T .......... ..........
TA2              1 .......... .......... .........T .......... ..........
TA3              1 .......... .......... .........T .......... ..........
TA4              1 .......... .......... .........T .......... ..........
TA5              1 .......... .......... .........T .......... ..........

                        60        70        80        90       100
                   ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK     51 GCTAGCGATT GCAGCACTCG GTGTGAAGGA CCGCATCAAG CTCGAGTACA
ET1             51 .......... ......A... .......... .......... ..........
ET2             51 .......... ......A... .......... .......... ..........
ET3             51 .......... ......A... .......... .......... ..........
ET4             51 .......... ......A... .......... .......... ..........
ET5             51 .......... ......A... .......... .......... ..........
SU1             51 .......... .......... .......... .......... ..........
SU2             51 .......... .......... .......... ........G. ..........
SU3             51 .....T.... .......... .......... .......... ..........
SU4             51 .......... .......... .......... .......... ..........
SU5             51 ....A..... .......... .......... .......... ..........
MA1             51 .......... .......... .......... .......... ..........
MA2             51 .......... .......... .......... .......... ..........
MA3             51 .......... .......... .......... .......... ..........
MA4             51 .......... .......... .......... .......... ..........
MA5             51 .......... .......... .......... .......... ..........
TA1             51 .......... .......... .......... ..A....... ..........
TA2             51 .......... .......... .......... .......... ..........
TA3             51 .......... .......... .......... .......... ..........
TA4             51 .......... ......A... .......... .......... ..........
TA5             51 .......... .....G.... .......... .......... ..........

                       110       120       130       140       150
                   ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK     101 TCGATCTCTT TAAGGGTGGC CATTTAAGTA GCGATTATCT TAAG------
ET1             101 .......... .......... .......... .......... ....GTAGTA
ET2             101 .......... .......... .......... .......... ....GTAGTA
ET3             101 .......... .......... .......... .......... ....GTAGTA
ET4             101 .......... .......... .......... .......... ....GTAGTA
ET5             101 .......... .......... .......... .......... ....GTAGTA
SU1             101 .......... .......... .......... .......... ....GTAGTA
SU2             101 .......... .......... .......... .......... ....GTAGTA
SU3             101 .......... .......... .......... .......... ....GTAGTA
SU4             101 .......... .......... .......... .......... ....GTAGTA
SU5             101 .......... .......... .......... .......... ....GTAGTA
MA1             101 .......... ........T. .......... .......... ....GTAGTA
MA2             101 .......... .......... .......... .......... ....GTAGTA
MA3             101 .......... ........T. .......... .......... ....GTAGTA
MA4             101 .......... .......... .......... .......... ....GTAGTA
MA5             101 .......... .......... .......... .......... ....GTAGTA
TA1             101 .......... .......... .......... .......... ....GTAGTA
TA2             101 .......... .......... .......... .......... ....GTAGTA
TA3             101 .......... .......... .......... .......... ....GTAGTA
TA4             101 .......... .......... .......... .......... ....GTAGTA
TA5             101 .......... .......... .......... .......... ....GTAGTA
```


160 170 180 190 200
AG-GENEBANK 144
ET1 151 CAACCCCAAG CAACGACTAG CGAATAGTTG ATCTAATTTA ACCTCTCTCT
ET2 151 CAACCCCAAG CAACGACTAG CGAATAGTTG ATCTAATTTA ACCTCTCTCT
ET3 151 CAACCCCAAG CAACGACTAG CGAATAGTTG ATCTAATTTA ACCTCTCTCT
ET4 151 CAACCCCAAG CAACGACTAG CGAATAGTTG ATCTAATTTA ACCTCTCTCT
ET5 151 CAACCCCAAG CAACGACTAG CGAATAGTTG ATCTAATTTA ACCTCTCTCT
SU1 151 GAGCCCCTCG CAACGACTAG CGAATAGTTG ATCTAATTTA AGCTCTCTCT
SU2 151 CAACCCCAAG CAACGACCAG CGAATAGTTG ATCTAATTTA AGCTCTCTCT
SU3 151 GAGCCCCTCG CAACGACTAG CGAATAGTTG ATCTAATTTA AGCTCTCTCT
SU4 151 GAGCCCCTCG CAACGACTAG CGAATAGTTG ATCTAATTTA AGCTCTCTCT
SU5 151 AAGCCCCTCG CAACGACTAG CGAATAGTTG ATCTAATTTA AGCTCTCTCT
MA1 151 GAGCCCCTCG CAACGACTAG CGAATAGTTG ATCTAATTTA AGCTCTCTTT
MA2 151 CAACCCCAAG CAACGACCAG CGAATAGTTG ATCTAATTTA AGCTCTCTCT
MA3 151 GAGCCCCTCG CAACGACTAG CGAATAGTTG ATCTAATTTA AGCTCTCTTT
MA4 151 CAACCCCAAG CAACGACCAG CGAATAGTTG ATCTAATTTA AGCTCTCTCT
MA5 151 CAACCCCAAG CAACGACCAG CGAATAGTTG ATCTAATTTA AGCTCTCTCT
TA1 151 CAACCCCAAG CAACGACCAG CGAATAGTTG ATCTAATTTA AGCTCTCTCT
TA2 151 GAGCCCCTCG CAACGACTAG CGAATAGTTG ATCTAATTTA AGCTCTCTCT
TA3 151 GAGCCCCTCG CAACGACTAG CGAATAGTTG ATCTAATTTA AGCTCTCTCT
TA4 151 CAACCCCAAG CAACGACTAG CGAATAGTTG ATCTAATTTA ACCTCTCTCT
TA5 151 CAACCCCAAG CAACGACCAG CGAATAGTTG ATCTAATTTA AGCTCTCTCT
210 220 230 240 250
AG-GENEBANK 144 ATT AATCCACTGC ACACGGTCCC CGTCCTACGG
ET1 201 CTCTCTTCAA TTTCCAG...
ET2 201 CTCTCTTCAA TTTCCAG...
ET3 201 CTCTCTTCAA TTTCCAG...
ET4 201 CTCTCTTCAA TTTCCAG...
ET5 201 CTCTCTTCAA TTTCCAG...
SU1 201 CTCTT--CAA TTTCCAG...
SU2 201 TTCTT--CAA TTTCCAG...
SU3 201 CTCTCTTCAA TTTCCAG...
SU4 201 CTCTT--CAA CTTCCAG...
SU5 201 CTCTT--CAA TTTCCAA...
MA1 201 CTCTT--CAA CTTCCAG...
MA2 201 CTCTT--CAA TTTCCAG...
MA3 201 CTCTT--CAA CTTCCAG...
MA4 201 CTCTT--CAA TTTCCAG...
MA5 201 CTT----CAA TTTCCAG...
TA1 201 CTCTT--CAA TTTCCAG...
TA2 201 CTCTT--CAA TTTCCAG...
TA3 201 CTCTT--CAA TTTCCAG...
TA4 201 CTCTCTTCAA TTTCCAG...
TA5 201 TTCTT--CAA TTTCCAG...
260 270 280 290 300
AG-GENEBANK 178 CACGGTGAGT TAACGCTTAC CGATAGTCAC GCTATTTTAG TGTACCTGTG
ET1 251 ....T.
ET2 251 ....T.
ET3 251 ....T.
ET4 251 ....T.
ET5 251 ....T.
SU1 249 ....T.
SU2 249 ....T.
SU3 251 ....T.
SU4 249 ....T.
SU5 249 ....T.
MA1 249 ....T.
MA2 249 ....T.
MA3 249 ....T.
MA4 249 ....T.
MA5 247 ....T.
TA1 249 ....T.
TA2 249 ....T.
TA3 249 ....T.
TA4 251 ....A...T.
TA5 249 ....T.


310 320 330 340 350
AG-GENEBANK 228 CGATACATTT GCCCCACCAG GGCACACGCT AGCCCTGCCC GACGCACTGA
ET1 301 ......C... ....
ET2 301 ......C... ....
ET3 301 ......C... ....
ET4 301 ......C... ....
ET5 301 ......C... ....
SU1 299 ......C... ....
SU2 299 ......C... ....
SU3 301 ......C... ....T... ....
SU4 299 ......C... ....
SU5 299 ......C... ....T... ....
MA1 299 ......C... ....
MA2 299 ......C... ....G....... ....
MA3 299 ......C... ....
MA4 299 ......C... ....G....... ....
MA5 297 ......C... ....G....... ....
TA1 299 ......C... ....
TA2 299 ......C... ....
TA3 299 ......C... ....
TA4 301 ......C... ....
TA5 299 ......C... ....
360 370 380 390 400
AG-GENEBANK 278 CGCGCGCCAA AGTTTTCAAC ATGCTGTGCT TCAACAACGG CTGTTTGTTT
ET1 351 ....
ET2 351 ....
ET3 351 ....
ET4 351 .C.....G.. G......... ....
ET5 351 .C.....G.. G......... ....
SU1 349 ....GG. ....T.. ....
SU2 349 ....G.. G......T.T ....
SU3 351 ....
SU4 349 ....
SU5 349 ....
MA1 349 ....T... ....
MA2 349 ....
MA3 349 ....T... ....
MA4 349 ....
MA5 347 ....
TA1 349 ....G.. G......... ....
TA2 349 ....G.. G......... ....T.. ....
TA3 349 ....G.. G......... ....T.. ....
TA4 351 ....
TA5 349 ....G.. G......... ....
410 420 430 440 450
AG-GENEBANK 328 CAGCGCGATG CGGAAGTTAT G---------- ---------- ----------
ET1 401 .... .GTATGTAGC GCGGCTTCTA GAAACCAATT
ET2 401 .... .GTATGTAGC GCGGCTTCTA GAAACCAATT
ET3 401 .... .GTATGTAGC GCGGCTTCTA GAAACCAATT
ET4 401 .... .GTATGTAGT GCGGACTTTA GAAACCAATT
ET5 401 .... .GTATGTAGT GCGGACTTTA GAAACCAATT
SU1 399 .... .GTATGTAGC GCGGTTTCTA GAAACCAATT
SU2 399 .... .GTATGTAGC GCGGCTTCTA GAAACCAATT
SU3 401 .... .GTATGTAGC GCGGCTTCTA GAAACCCATT
SU4 399 .... .GTATGTAGC GCGGCTTCTA GAAACCAATT
SU5 399 .... .GTATGTAGC GCGGCTTCTA GAAACCAATT
MA1 399 .... .GTATGTAGC GCAGCTTCTA GAAACCAATT
MA2 399 .... .GTATGTAGC GCGGCTTCTA GAAACCAATT
MA3 399 .... .GTATGTAGC GCAGCTTCTA GAAACCAATT
MA4 399 .... .GTATGTAGC GCGGCTTTTA GAAACCAATT
MA5 397 .... .GTATGTAGC GCGGCTTCTA GAAACCAATT
TA1 399 ......AG.. .... .GTATGTAGC GCGGCTTCTA GAAACCAATT
TA2 399 .... .GTATGTAGC GCGGTTTCTA GAAACCAATT
TA3 399 .... .GTATGTAGC GCGGTTTCTA GAAACCAATT
TA4 401 .... .GTATGTAGC GCGGCTTCTA GAAACCAATT
TA5 399 .... .GTATGTAGC GCGGCTTCTA GAAACCAATT

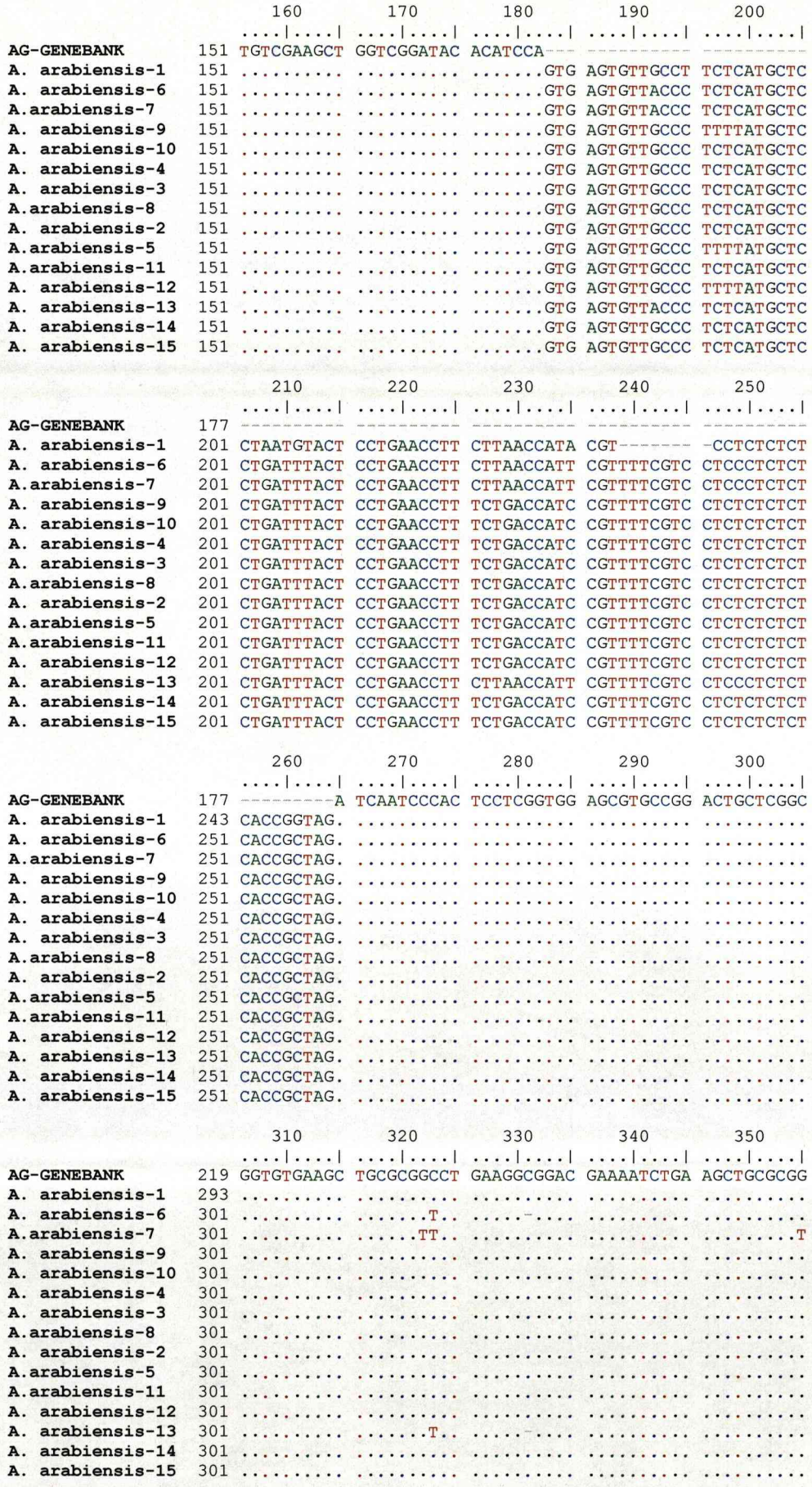
```
                          460        470        480        490        500
                     ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK      348 ---------- ---------- ---------- ------CGT AAAATCTTCA
ET1              451 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
ET2              451 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
ET3              451 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
ET4              451 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
ET5              451 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
SU1              449 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
SU2              449 CCCTTTCCTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
SU3              451 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
SU4              449 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... .....T....
SU5              449 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
MA1              449 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
MA2              449 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
MA3              449 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
MA4              449 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
MA5              447 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
TA1              449 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
TA2              449 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
TA3              449 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
TA4              451 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........
TA5              449 CCCTTTCTTT GCCATTTAAA ACCGTTCTCT TGTGTAG... ..........

                          510        520        530        540        550
                     ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK      362 GCGGTGCCAT TACCGACCCA ACGCAGCATC TGAAACCGAT CGAGGCAGCG
ET1              501 .......... .......... .......... .......... ..........
ET2              501 .......... .........C .......... .......... ..........
ET3              501 .......... .......... .......... .......... ..........
ET4              501 .......... .......... .......... .......... ..........
ET5              501 .......... .......... .......... .......... ..........
SU1              499 .......... .......... T......... .C........ ........A.
SU2              499 .......... .......... .......... .......... ..........
SU3              501 .......... .......... T......... .C........ ..........
SU4              499 .......... .......... T......... .C........ ..........
SU5              499 .......... .......... T......... .C........ ..........
MA1              499 .......... ....T..... T......... .C........ ..........
MA2              499 .......... .......... T......... .C..C..... ..........
MA3              499 .......... ....T..... T......... .C........ ..........
MA4              499 .......... .......... T......... .C........ ..........
MA5              497 .......... .........C .......... .......... ..........
TA1              499 .....A.... .......... .......... .......... ..........
TA2              499 .......... .......... T......... .C........ ..........
TA3              499 .......... .......... T......... .C........ ..........
TA4              501 .......... .......... .......... .......... ..........
TA5              499 .......... .......... .......... .......... ..........

                          560        570        580        590        600
                     ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK      412 ATCGATGCGC TGGAGCAGTT TCTGCAGCGA TCGCGCTACA CCGCACACGA
ET1              551 .......... .......... .......... .......... ..........
ET2              551 .......... .......... .......... .......... ..........
ET3              551 .......... .......... .......... .......... ..........
ET4              551 .......... .......... .......... .......... ..........
ET5              551 .......... .......... .......... .......... ..........
SU1              549 .......... .......... .......... .......... ..........
SU2              549 .......... .......... .......... .......... ..........
SU3              551 .......... .......... .......... ...T...... ..........
SU4              549 .......... .......... .......... A......... ..........
SU5              549 .......... .......... .......... .......... ..........
MA1              549 .......... .......... .......... .......... ..........
MA2              549 .......... .......... ....A..... .......... ..........
MA3              549 .......... .......... .......... .......... ..........
MA4              549 .......... .......... ....A..... .......... ..........
MA5              547 .......... .......... .......... .......... ..........
TA1              549 .......A.. .......... .......... .......... ..........
TA2              549 .......... .......... .......... .......... ......G...
TA3              549 .......... .......... .......... .......... ..........
TA4              551 .......... .......... .......... .......... ..........
TA5              549 .......... .......... .......... .......... ..........
```

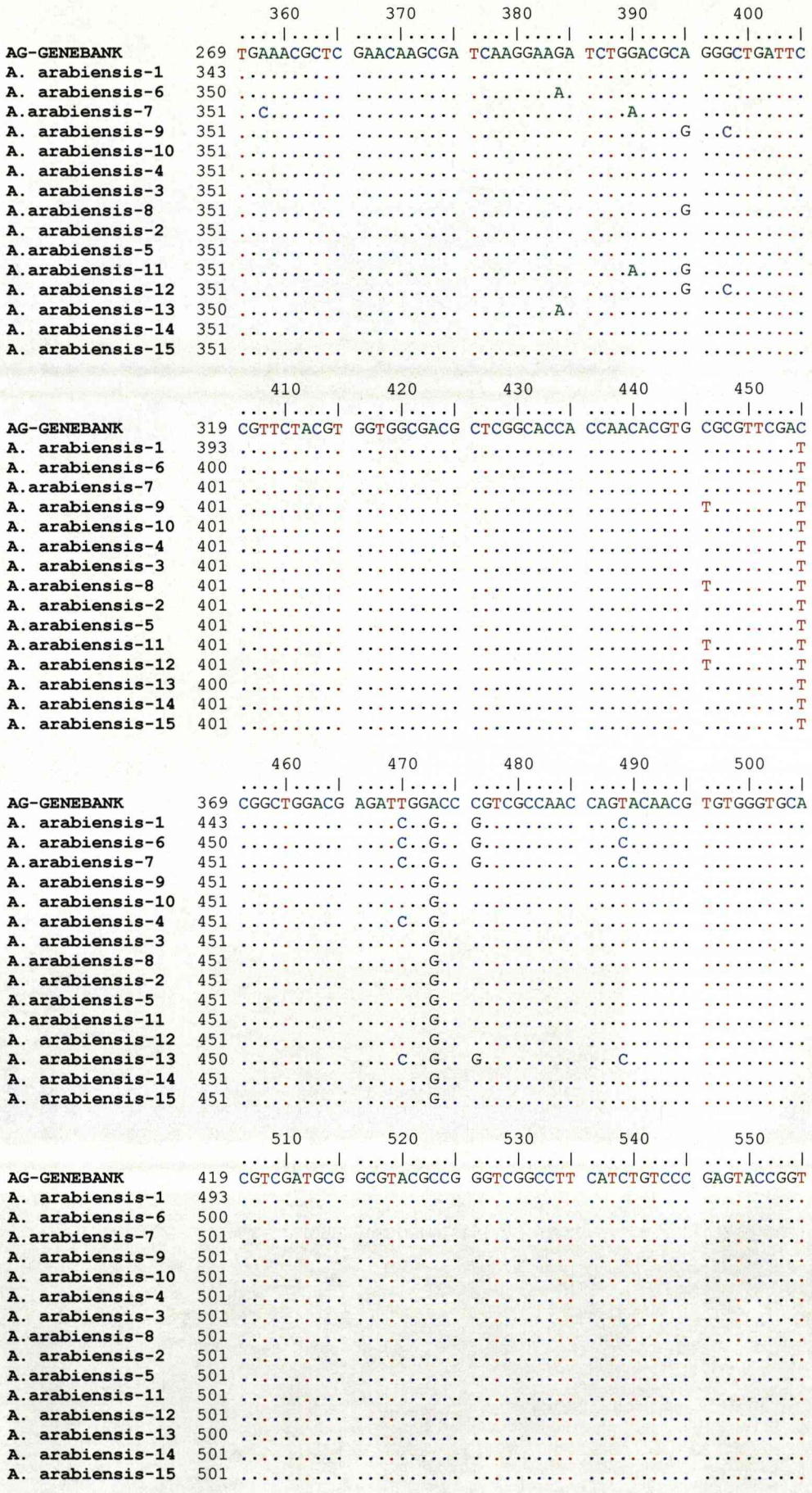
```
                          610        620        630        640        650
                     ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK      462 TCAGCTTTCG GTGGCAGATT TCGCAATCGT CGCGACACTC AGCACGGTGG
ET1              601 .......... .......... .......... .......... ..........
ET2              601 .......... .......... .......... .......... ..........
ET3              601 .......... .......... .......... .......... ..........
ET4              601 .......... .......... .......... .......... ..........
ET5              601 .......... .......... .......... .......... ..........
SU1              599 .......... .......... .......... ....G..... ..........
SU2              599 .......... .......... .......... .......... ..........
SU3              601 .......... .......... ....G..... .......... ..........
SU4              599 .......... .......... .......... .......... ..........
SU5              599 .......... .......... ....G..... .......... ..........
MA1              599 .......... .......... ....G..... .......... ..........
MA2              599 .......... .......... .......... .......... ........T.
MA3              599 .......... .......... ....G..... .......... ..........
MA4              599 .......... .......... .......... .......... ..........
MA5              597 .....C.... ......A... C......... ......C... .C.C......
TA1              599 .......... .......... .......... .......... ..........
TA2              599 .......... .......... .......... .......... ..........
TA3              599 .......... .......... .......... .......... ..........
TA4              601 .......... .......... .......... .......... ..........
TA5              599 .......... .......... .......... .......... ..........

                          660        670        680        690        700
                     ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK      512 CCATTTTTGT GCCGCTCCCG GCGGATCGTT GGCCGCGGGT ATGCGAGTGG
ET1              651 .......... .......GTT .......... .......... ..........
ET2              651 .......... .......GTT .......... .......... ..........
ET3              651 .......... .......GTT .......... .......... ..........
ET4              651 .......... .......GTT .......... .......... ..........
ET5              651 .......... .......GTT .......... .......... ..........
SU1              649 .......... .......GTT .......... .......... ..........
SU2              649 .......... .......GTT .......... .......... ..........
SU3              651 .......... ......T.GT .......... .......... ..........
SU4              649 .......... .......GTT .......... .......... ..........
SU5              649 .......... .......GT. .......... .......... ..........
MA1              649 ........G.. .......GTT .......... .......... ..........
MA2              649 G......G.. .......GTT .......T.. ..G.C..... T...A.A...
MA3              649 .......G.. .......GTT .......... .......... ..........
MA4              649 .......... .......GTT .......... .......... ..........
MA5              647 .......... .......GTT .......C.. ..G.C..... ....A.A...
TA1              649 .......... .......GTT .......... .......... ..........
TA2              649 .......... ......TGTT .......... .......... ..........
TA3              649 .......... .......GTT .......... .......... ..........
TA4              651 .......... .......GTT .......... .......... ..........
TA5              649 .......... .......GTT .......... .......... ..........

                          710        720        730        740        750
                     ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK      562 TTCGCGGTGA TGGAAGCGCT GCCATACTAC AACGACCAGA ACCGTGTTGG
ET1              701 .......... ....G..... .......... .......... ..........
ET2              701 .......... ....G..... .......... .......... ..........
ET3              701 .......... ....G..... .......... .......... ..........
ET4              701 .......... ....G..... .......... .......... ..........
ET5              701 .......... ....G..... .......... .......... ..........
SU1              699 .......... .......... A..C...... .......... ..........
SU2              699 .......... ....G..... .......... .......... ..........
SU3              701 .......... .......... .......... .....A.... ..........
SU4              699 .......... .......... A..C...... .......... ..........
SU5              699 .......... .......... .T........ .....A.... ..........
MA1              699 .......... .......... ...C...... .......... ..........
MA2              699 .......... .......C.. A......... .......... ..........
MA3              699 .......... .......... ...C...... .......... ..........
MA4              699 .......... .......... A..C...... .......... ..........
MA5              697 ....G..... .......C.. A......... .......... ..........
TA1              699 .......... .......... A..T...... .......... ..........
TA2              699 .......... .......... A..C...... .......... ..........
TA3              699 .......... .......... A..C...... .......... ..........
TA4              701 .......... ....G..... .......... .......... ..........
TA5              699 .......... ....G..... .......... .......... ..........
```

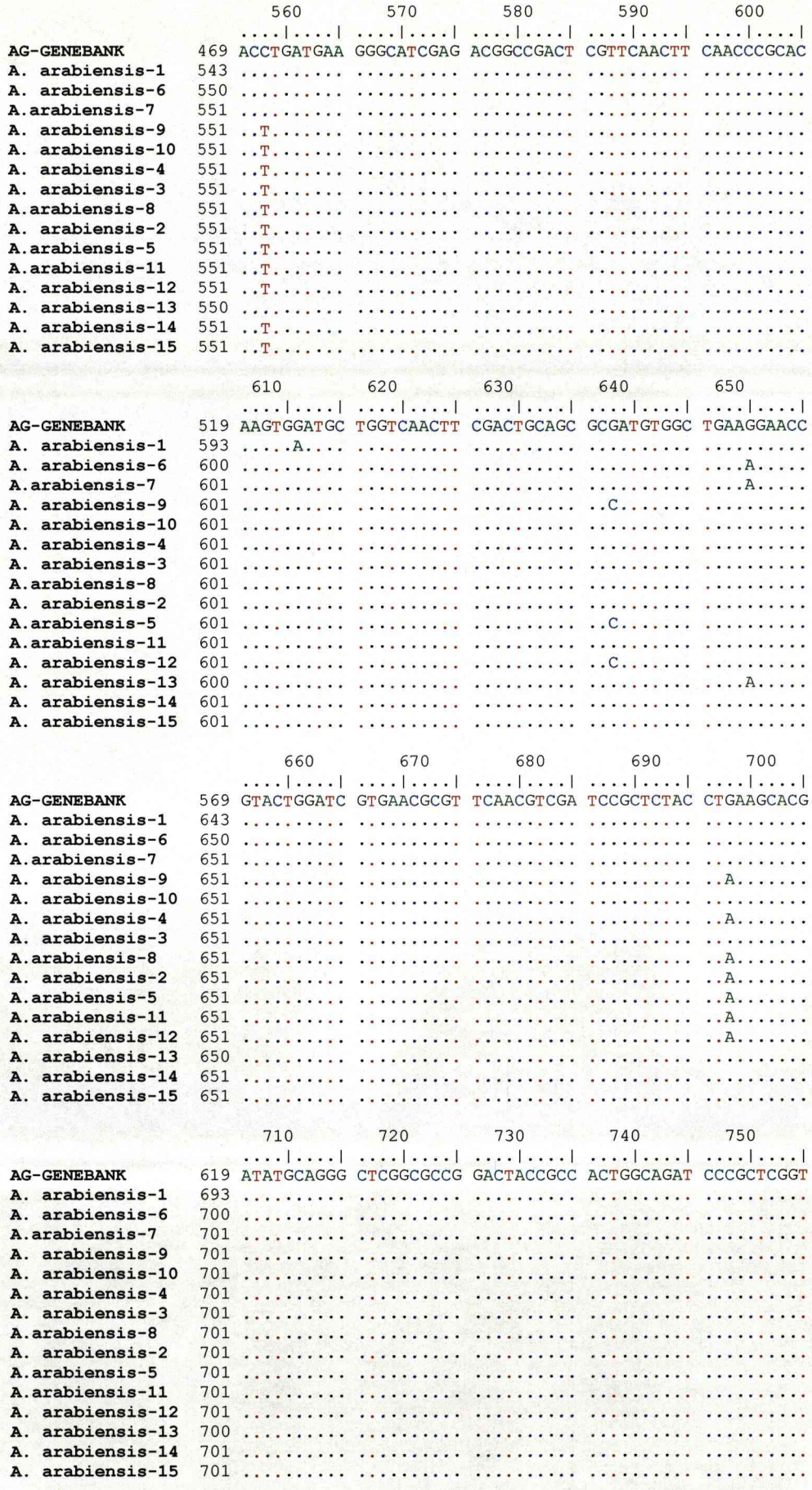

760 770 780 790
AG-GENEBANK 612 GTTGGACATG TTGCGCAAAC ATTTAGCGGG AAAGATTAAG CTGTAG
ET1 751 .....
ET2 751 .....
ET3 751 .....
ET4 751 .....
ET5 751 .....
SU1 749 ....
SU2 749 ..... .C..... ....
SU3 751 .....
SU4 749 .....
SU5 749 .....
MA1 749 .....
MA2 749 .....
MA3 749 .....
MA4 749 .....
MA5 747 .....
TA1 749 .....
TA2 749 .....
TA3 749 .....
TA4 751 .....
TA5 749 .....

Appendix 5 DNA sequence alignment of *Dopa decaboxylase* gene

```
                      10        20        30        40        50
                      ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK           1   AGGAGTTCCT GGCCTGCTCC GGTGGGCAGG GTGGCGGTGT CATTCAGGGC
A. arabiensis-1       1   .......... .......... .......... .......... ..........
A. arabiensis-6       1   .......... .......... .......... .......... ..........
A.arabiensis-7        1   .......... .......... .......... .......... ..........
A. arabiensis-9       1   .......... .......... .......... .......... ..........
A. arabiensis-10      1   .......... .......... .......... .......... ..........
A. arabiensis-4       1   .......... .......... .......... .......... ..........
A. arabiensis-3       1   .......... .......... .......... .......... ..........
A.arabiensis-8        1   .......... .......... .......... .......... ..........
A. arabiensis-2       1   .......... .......... .......... .......... ..........
A.arabiensis-5        1   .......... .......... .......... .......... ..........
A.arabiensis-11       1   .......... .......... .......... .......... ..........
A. arabiensis-12      1   .......... .......... .......... .......... ..........
A. arabiensis-13      1   .......... .......... .......... .......... ..........
A. arabiensis-14      1   .......... .......... .......... .......... ..........
A. arabiensis-15      1   .......... .......... .......... .......... ..........

                          60        70        80        90        100
                          ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK           51  ACGGCCAGCG AGGCAACGCT GGTCGCGTTG CTCGGCGCTA AGGCGAAGGC
A. arabiensis-1       51  .......... .......... .......... .......... ..........
A. arabiensis-6       51  ....G..... .......... .......... ..A....... ..........
A.arabiensis-7        51  ....G..... .......... .......... ..A....... ..........
A. arabiensis-9       51  .......... .......... .......... .......... ..........
A. arabiensis-10      51  .......... .......... .......... .......... ..........
A. arabiensis-4       51  .......... .......... .......... .......... ..........
A. arabiensis-3       51  ....G..... .......... .......... .......... ..........
A.arabiensis-8        51  .......... .......... .......... .......... ..........
A. arabiensis-2       51  .......... .......... .......... ..A....... ..........
A.arabiensis-5        51  .......... .......... .......... .......... ..........
A.arabiensis-11       51  .......... .......... .......... .......... ..........
A. arabiensis-12      51  .......... .......... .......... .......... ..........
A. arabiensis-13      51  ....G..... .......... .......... .......... ..........
A. arabiensis-14      51  ....G..... .......... .......... .......... ..........
A. arabiensis-15      51  .......... .......... .......... ...A...... ..........

                         110       120       130       140       150
                          ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK           101 GATGAAGCGC GTCAAGGAGG AGCATCCCGA CTGGGACGAT AATACGATCG
A. arabiensis-1       101 .......... .......... .......... .......... ..........
A. arabiensis-6       101 .......... .......... .......... .......... ..........
A.arabiensis-7        101 .......... .......... .......... .......... ..........
A. arabiensis-9       101 .......... .......... .......... .......... ..........
A. arabiensis-10      101 .......... .......... .......... .......... ..........
A. arabiensis-4       101 .......... .......... .......... .......... ..........
A. arabiensis-3       101 .......... .......... .......... .......... ..........
A.arabiensis-8        101 .......... .......... .......... .......... ..........
A. arabiensis-2       101 .......... .......... .......... .......... ..........
A.arabiensis-5        101 .......... .......... .......... .......... ..........
A.arabiensis-11       101 .......... .......... .......... .......... ..........
A. arabiensis-12      101 .......... .......... .......... .......... ..........
A. arabiensis-13      101 .......... .......... .......... .......... ..........
A. arabiensis-14      101 .......... .......... .......... .......... ..........
A. arabiensis-15      101 .......... .......... .......... .......... ..........
```


160 170 180 190 200
AG-GENEBANK 151 TGTCGAAGCT GGTCGGATAC ACATCCA
A. arabiensis-1 151 .....GTG AGTGTTGCCT TCTCATGCTC
A. arabiensis-6 151 .....GTG AGTGTTACCC TCTCATGCTC
A.arabiensis-7 151 .....GTG AGTGTTACCC TCTCATGCTC
A. arabiensis-9 151 .....GTG AGTGTTGCCC TTTTATGCTC
A. arabiensis-10 151 .....GTG AGTGTTGCCC TCTCATGCTC
A. arabiensis-4 151 .....GTG AGTGTTGCCC TCTCATGCTC
A. arabiensis-3 151 .....GTG AGTGTTGCCC TCTCATGCTC
A.arabiensis-8 151 .....GTG AGTGTTGCCC TCTCATGCTC
A. arabiensis-2 151 .....GTG AGTGTTGCCC TCTCATGCTC
A.arabiensis-5 151 .....GTG AGTGTTGCCC TTTTATGCTC
A.arabiensis-11 151 .....GTG AGTGTTGCCC TCTCATGCTC
A. arabiensis-12 151 .....GTG AGTGTTGCCC TTTTATGCTC
A. arabiensis-13 151 .....GTG AGTGTTACCC TCTCATGCTC
A. arabiensis-14 151 .....GTG AGTGTTGCCC TCTCATGCTC
A. arabiensis-15 151 .....GTG AGTGTTGCCC TCTCATGCTC
210 220 230 240 250
AG-GENEBANK 177
A. arabiensis-1 201 CTAATGTACT CCTGAACCTT CTTAACCATA CGT -CCTCTCTCT
A. arabiensis-6 201 CTGATTTACT CCTGAACCTT CTTAACCATT CGTTTTCGTC CTCCCTCTCT
A.arabiensis-7 201 CTGATTTACT CCTGAACCTT CTTAACCATT CGTTTTCGTC CTCCCTCTCT
A. arabiensis-9 201 CTGATTTACT CCTGAACCTT TCTGACCATC CGTTTTCGTC CTCTCTCTCT
A. arabiensis-10 201 CTGATTTACT CCTGAACCTT TCTGACCATC CGTTTTCGTC CTCTCTCTCT
A. arabiensis-4 201 CTGATTTACT CCTGAACCTT TCTGACCATC CGTTTTCGTC CTCTCTCTCT
A. arabiensis-3 201 CTGATTTACT CCTGAACCTT TCTGACCATC CGTTTTCGTC CTCTCTCTCT
A.arabiensis-8 201 CTGATTTACT CCTGAACCTT TCTGACCATC CGTTTTCGTC CTCTCTCTCT
A. arabiensis-2 201 CTGATTTACT CCTGAACCTT TCTGACCATC CGTTTTCGTC CTCTCTCTCT
A.arabiensis-5 201 CTGATTTACT CCTGAACCTT TCTGACCATC CGTTTTCGTC CTCTCTCTCT
A.arabiensis-11 201 CTGATTTACT CCTGAACCTT TCTGACCATC CGTTTTCGTC CTCTCTCTCT
A. arabiensis-12 201 CTGATTTACT CCTGAACCTT TCTGACCATC CGTTTTCGTC CTCTCTCTCT
A. arabiensis-13 201 CTGATTTACT CCTGAACCTT CTTAACCATT CGTTTTCGTC CTCCCTCTCT
A. arabiensis-14 201 CTGATTTACT CCTGAACCTT TCTGACCATC CGTTTTCGTC CTCTCTCTCT
A. arabiensis-15 201 CTGATTTACT CCTGAACCTT TCTGACCATC CGTTTTCGTC CTCTCTCTCT
260 270 280 290 300
AG-GENEBANK 177 A TCAATCCCAC TCCTCGGTGG AGCGTGCCGG ACTGCTCGGC
A. arabiensis-1 243 CACCGGTAG. .....
A. arabiensis-6 251 CACCGCTAG. .....
A.arabiensis-7 251 CACCGCTAG. .....
A. arabiensis-9 251 CACCGCTAG. .....
A. arabiensis-10 251 CACCGCTAG. .....
A. arabiensis-4 251 CACCGCTAG. .....
A. arabiensis-3 251 CACCGCTAG. .....
A.arabiensis-8 251 CACCGCTAG. .....
A. arabiensis-2 251 CACCGCTAG. .....
A.arabiensis-5 251 CACCGCTAG. .....
A.arabiensis-11 251 CACCGCTAG. .....
A. arabiensis-12 251 CACCGCTAG. .....
A. arabiensis-13 251 CACCGCTAG. .....
A. arabiensis-14 251 CACCGCTAG. .....
A. arabiensis-15 251 CACCGCTAG. .....
310 320 330 340 350
AG-GENEBANK 219 GGTGTGAAGC TGCGCGGCCT GAAGGCGGAC GAAAATCTGA AGCTGCGCGG
A. arabiensis-1 293 .....
A. arabiensis-6 301 .....T.. .....
A.arabiensis-7 301 .....TT.. .....T
A. arabiensis-9 301 .....
A. arabiensis-10 301 .....
A. arabiensis-4 301 .....
A. arabiensis-3 301 .....
A.arabiensis-8 301 .....
A. arabiensis-2 301 .....
A.arabiensis-5 301 .....
A.arabiensis-11 301 .....
A. arabiensis-12 301 .....
A. arabiensis-13 301 .....T.. .....
A. arabiensis-14 301 .....
A. arabiensis-15 301 .....

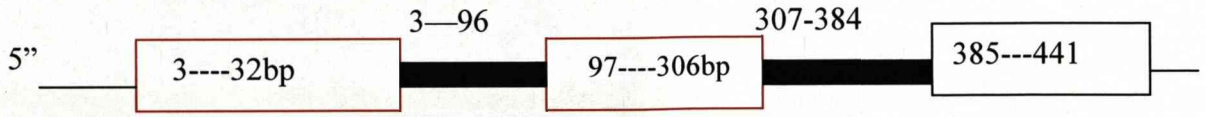

360 370 380 390 400
AG-GENEBANK 269 TGAAACGCTC GAACAAGCGA TCAAGGAAGA TCTGGACGCA GGGCTGATTC
A. arabiensis-1 343
A. arabiensis-6 350
A.arabiensis-7 351
A. arabiensis-9 351
A. arabiensis-10 351
A. arabiensis-4 351
A. arabiensis-3 351
A.arabiensis-8 351
A. arabiensis-2 351
A.arabiensis-5 351
A.arabiensis-11 351
A. arabiensis-12 351
A. arabiensis-13 350
A. arabiensis-14 351
A. arabiensis-15 351
410 420 430 440 450
AG-GENEBANK 319 CGTTCTACGT GGTGGCGACG CTCGGCACCA CCAACACGTG CGCGTTCGAC
A. arabiensis-1 393
A. arabiensis-6 400
A.arabiensis-7 401
A. arabiensis-9 401
A. arabiensis-10 401
A. arabiensis-4 401
A. arabiensis-3 401
A.arabiensis-8 401
A. arabiensis-2 401
A.arabiensis-5 401
A.arabiensis-11 401
A. arabiensis-12 401
A. arabiensis-13 400
A. arabiensis-14 401
A. arabiensis-15 401
460 470 480 490 500
AG-GENEBANK 369 CGGCTGGACG AGATTGGACC CGTCGCCAAC CAGTACAACG TGTGGGTGCA
A. arabiensis-1 443
A. arabiensis-6 450
A.arabiensis-7 451
A. arabiensis-9 451
A. arabiensis-10 451
A. arabiensis-4 451
A. arabiensis-3 451
A.arabiensis-8 451
A. arabiensis-2 451
A.arabiensis-5 451
A.arabiensis-11 451
A. arabiensis-12 451
A. arabiensis-13 450
A. arabiensis-14 451
A. arabiensis-15 451
510 520 530 540 550
AG-GENEBANK 419 CGTCGATGCG GCGTACGCCG GGTCGGCCTT CATCTGTCCC GAGTACCGGT
A. arabiensis-1 493
A. arabiensis-6 500
A.arabiensis-7 501
A. arabiensis-9 501
A. arabiensis-10 501
A. arabiensis-4 501
A. arabiensis-3 501
A.arabiensis-8 501
A. arabiensis-2 501
A.arabiensis-5 501
A.arabiensis-11 501
A. arabiensis-12 501
A. arabiensis-13 500
A. arabiensis-14 501
A. arabiensis-15 501


560 570 580 590 600
AG-GENEBANK 469 ACCTGATGAA GGGCATCGAG ACGGCCGACT CGTTCAACTT CAACCCGCAC
A. arabiensis-1 543
A. arabiensis-6 550
A.arabiensis-7 551
A. arabiensis-9 551 ..T.....
A. arabiensis-10 551 ..T.....
A. arabiensis-4 551 ..T.....
A. arabiensis-3 551 ..T.....
A.arabiensis-8 551 ..T.....
A. arabiensis-2 551 ..T.....
A.arabiensis-5 551 ..T.....
A.arabiensis-11 551 ..T.....
A. arabiensis-12 551 ..T.....
A. arabiensis-13 550
A. arabiensis-14 551 ..T.....
A. arabiensis-15 551 ..T.....
610 620 630 640 650
AG-GENEBANK 519 AAGTGGATGC TGGTCAACTT CGACTGCAGC GCGATGTGGC TGAAGGAACC
A. arabiensis-1 593 .....A....
A. arabiensis-6 600 ....A.....
A.arabiensis-7 601 ....A.....
A. arabiensis-9 601 ..C.....
A. arabiensis-10 601
A. arabiensis-4 601
A. arabiensis-3 601
A.arabiensis-8 601
A. arabiensis-2 601
A.arabiensis-5 601 ..C.....
A.arabiensis-11 601
A. arabiensis-12 601 ..C.....
A. arabiensis-13 600 ....A.....
A. arabiensis-14 601
A. arabiensis-15 601
660 670 680 690 700
AG-GENEBANK 569 GTACTGGATC GTGAACGCGT TCAACGTCGA TCCGCTCTAC CTGAAGCACG
A. arabiensis-1 643
A. arabiensis-6 650
A.arabiensis-7 651
A. arabiensis-9 651 ..A.....
A. arabiensis-10 651
A. arabiensis-4 651 ..A.....
A. arabiensis-3 651
A.arabiensis-8 651 ..A.....
A. arabiensis-2 651 ..A.....
A.arabiensis-5 651 ..A.....
A.arabiensis-11 651 ..A.....
A. arabiensis-12 651 ..A.....
A. arabiensis-13 650
A. arabiensis-14 651
A. arabiensis-15 651
710 720 730 740 750
AG-GENEBANK 619 ATATGCAGGG CTCGGCGCCG GACTACCGCC ACTGGCAGAT CCCGCTCGGT
A. arabiensis-1 693
A. arabiensis-6 700
A.arabiensis-7 701
A. arabiensis-9 701
A. arabiensis-10 701
A. arabiensis-4 701
A. arabiensis-3 701
A.arabiensis-8 701
A. arabiensis-2 701
A.arabiensis-5 701
A.arabiensis-11 701
A. arabiensis-12 701
A. arabiensis-13 700
A. arabiensis-14 701
A. arabiensis-15 701

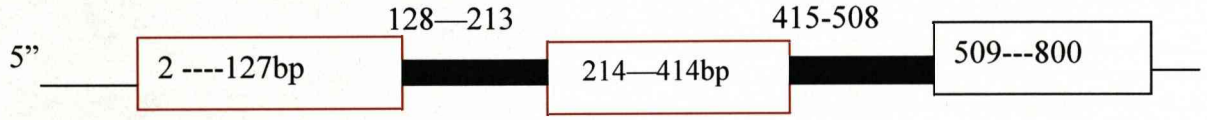
```
                          760        770        780        790        800
                     ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK          669 CGCCGGTTCC GTGCGCTGAA GCTGTGGTTC GTGCTGCGCC TGTACGGGGT
A. arabiensis-1      743 .......... .......... .......... .......... ..........
A. arabiensis-6      750 ..T....... .......... .......... .......... ..........
A.arabiensis-7       751 ..T....... .......... .......... .......... ..........
A. arabiensis-9      751 .......... .......... .......... .......... ..........
A. arabiensis-10     751 .......... .......... .......... .......... ..........
A. arabiensis-4      751 .......... .......... .......... .......... ..........
A. arabiensis-3      751 .......... .......... .......... .......... ..........
A.arabiensis-8       751 .......... .......... .......... .......... ..........
A. arabiensis-2      751 .......... .......... .......... .......... ..........
A.arabiensis-5       751 .......... .......... .......... .......... ..........
A.arabiensis-11      751 .......... .......... .......... .......... ..........
A. arabiensis-12     751 .......... .......... .......... .......... ..........
A. arabiensis-13     750 ..T....... .......... .......... .......... ..........
A. arabiensis-14     751 .......... .......... .......... .......... ..........
A. arabiensis-15     751 .......... .......... .......... .......... ..........

                          810        820        830        840        850
                     ....|....| ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK          719 GGACAATCTG CAGGCCCACA TTCGGCGCCA CTGCGGCTTC GCCAAGCAGT
A. arabiensis-1      793 .......... .......... .......... ......G... ..........
A. arabiensis-6      800 .......... .......... .......... ......G... ..........
A.arabiensis-7       801 .......... .......... .......... ......G... ..........
A. arabiensis-9      801 .......... .......... .......... ......G... ..........
A. arabiensis-10     801 .......... .......... .......... ......G... ..........
A. arabiensis-4      801 .......... .......... .......... ......G... ..........
A. arabiensis-3      801 .......... .......... .......... ......G... ..........
A.arabiensis-8       801 .......... .......... .......... ......G... ..........
A. arabiensis-2      801 .......... .......... .......... ......G... ..........
A.arabiensis-5       801 .......... .......... .......... ......G... ..........
A.arabiensis-11      801 .......... .......... .......... ......G... ..........
A. arabiensis-12     801 .......... .......... .......... ......G... ..........
A. arabiensis-13     800 .......... .......... .......... ......G... ..........
A. arabiensis-14     801 .......... .......... .......... ......G... ..........
A. arabiensis-15     801 .......... .......... .......... ......G... ..........

                          860        870        880        890
                     ....|....| ....|....| ....|....| ....|....|
AG-GENEBANK          769 TCGAGGCGCT GTGCCGGGCG GACGACCGGT TCGAGATATT
A. arabiensis-1      843 .......... .........A .......... ..........
A. arabiensis-6      850 .......... .......... .......... ..........
A.arabiensis-7       851 .......... .......... .......... ..........
A. arabiensis-9      851 .......... .......... .......... ..........
A. arabiensis-10     851 .......... .......... .......... ..........
A. arabiensis-4      851 .......... .......... .......... ..........
A. arabiensis-3      851 .......... .......... .......... ..........
A.arabiensis-8       851 .......... .......... .......... ..........
A. arabiensis-2      851 .......... .......... .......... ..........
A.arabiensis-5       851 .......... .......... .......... ..........
A.arabiensis-11      851 .......... .......... .......... ..........
A. arabiensis-12     851 .......... .......... .......... ..........
A. arabiensis-13     850 .......... .......... .......... ..........
A. arabiensis-14     851 .......... .......... .......... ..........
A. arabiensis-15     851 .......... .......... .......... ..........
```

APPENDIX 6: GENE ANNOTATION FOR GSTE1,GSTE2, GSTE6 AND GSTE8

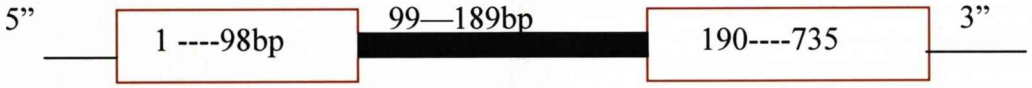
A) GSTe1 (441bp)



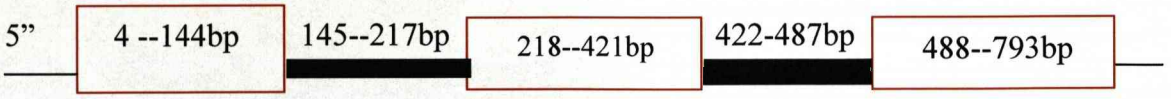
B) GSTe2 (800bp)



C) GSTe6 (783bp)



D) GSTe8 (796bp)



Legend

Exon

Intron

APPENDIX 7: Primer information for GST genes amplified

Locus	forward sequence	reverse sequence	Annealing temperature °C	Expected band size	GeneBank Accession No
GSTe1	5'CTGGGCAAAATTTGACACCT 3'	3'GGACGTACTCGATGCGATCT 5'	55 °C	441	AY063776
GSTe2	5'GTACACCCTGCACCTTAGCC 3'	3'CTAGCACAAACTTGCCCAGA 5'	55 °C	788	AF316636
GSTe6	5'GCGAAGATCAACGGTAGTGGTG 3'	3' GTAAGCTCTACAGCTCGTTC 5'	52 °C	788	AY070256
GSTe8	5'GCCATGATTCTGTACTACG 3'	3'CTACAGCTTAATCTTTCCCG 5'	52 °C	796	AY070257
AS-STe8	5-ACTACGACCAGGTCAGTCCA-3	5-TCTTTCCCGCTAAATGTTTG-3	58 °C	744	