GENE FLOW ANALYSIS OF *ANOPHELES ARABIENSIS* (DIPTERA: CULICIDAE) POPULATIONS IN SOUTHERN AFRICA USING MICROSATELLITE DNA MARKERS

Joel Claude Mouatcho

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science.

Johannesburg, 2005.

DECLARATION

I, Joel Claude Mouatcho declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

……..

31st day of March 2005

ABSTRACT

Anopheles arabiensis is considered an important vector of human malaria in the southern African region where the disease is the major cause of morbidity and mortality. Gene flow plays an important role in malaria control with the spread of insecticide resistance. The main objectives of this study were to (i) measure the genetic variability within and between five populations (Botswana, Namibia, South Africa, Swaziland and Zimbabwe) of wild *An. arabiensis* and (ii) estimate the level of gene flow between natural populations across the Southern-limits of *An. arabiensis.* A total of 1225 *An. arabiensis* specimens were identified out of 1300 mosquitoes collected from 2000-2003 with the sample sizes ranging from 180-292 per country. Variation at four microsatellite markers was investigated on non-denaturing polyacrylamide gels. The results showed fewer variations between populations (2.96%) than within populations (82.60%) suggesting considerable homogeneity. However, estimates of gene flow (Nm) calculated from mean F_{ST} and R_{ST} statistics were relatively low, 1.14 and 1.19 respectively, suggesting somewhat restricted gene flow between populations. The occurrence of gene flow within subpopulations of *An. arabiensis* in Zimbabwe but not in South Africa is interesting with regard to the spread of insecticide resistance in Zimbabwe.

The results presented here are obviously subject to the limitations inherent in manual, silver staining method of analysing microsatellite DNA markers. It is possible that a different set of results would be obtained if an Automated Sequencing Analyzer were used.

To Papa Remy Mouatcho and the almighty God

Publications and Presentations Arising from this Study

Presentations

Mouatcho, J. C., Koekemoer, L. L. and Coetzee, M. 2003. Genetic analysis of *Anopheles arabiensis* (Diptera: Culicidae) populations from Southern Africa using microsatellite DNA markers. *Proceedings of the 14th Entomological Society of Southern Africa Conference,* 6-9 July, Pretoria, South Africa.

Mouatcho, J. C., Koekemoer, L. L. and Coetzee, M. 2003. Population structure of *Anopheles arabiensis,* the primary vector of malaria in Southern Africa, using microsatellite DNA markers. Invited seminar at the *Department of Biochemistry, Rand Afrikaans University, South Africa,* 7 November.

ACKNOWLEDGMENTS

My profound gratitude to Prof Maureen Coetzee, Head of the Vector Control Reference Unit (VCRU) at the National Institute for Communicable Disease (NICD) of the National Health Laboratory Service, (NHLS) for introducing me the interesting field of Medical Entomology, for initiating this study, teaching me about morphology and rearing mosquito colony. I warmly thank her for the financial support, supervision and encouragement throughout my studies and her patience with me. I will always remember the two "green books" on mosquitoes she gave me just after one day in the unit.

I wish to thank Dr. Lizette. L. Koekemoer for training in DNA technology, microsatellite work and analysis, in sequencing, and in various domains, and for providing excellent working facilities in the laboratory. I am most grateful for her urge skilful technical assistance in the molecular laboratory and her deeply supervision. She always said "Joe don't hesitate to contact me especially on week-ends or after hours if you need help".

I also wish to thank the following persons: Benson Ntowma (Namibia), Mrs Goodness Nkosi and John Govere (Mpumalanga, South Africa), Keith Hargreaves (KwaZulu Natal, South Africa), Ron Masendu (Zimbabwe), Dr Mulenda Basimike and team mates (Swaziland) for their assistance with fields caught material, Dr. Basil Brooke for revising the language. Mrs Vanessa Wiseman and Guy Hall of the illustration department are acknowledged for computer scanning of photographs.

Special thank goes to Prof. Himla Soodyall, Department of Human Genetic, NHLS, for inspiring and supervising the sequencing part of this study, for use of equipments and chemicals and for introducing me to Dr Debra Barkhan at the HIV unit, NICD. I extended my thank Maharaj Akashnie from the same department, for teaching me the Arlequin software program.

Thanks to Dr Debra Barkham for her valuable advice and suggestions on the analysis of my data.

I am much obliged to Prof Hendrik J. Koornhof, head of Sexual Transmitted Diseases (STD) at the NICD for introducing me to Profs Maureen Coetzee and Himla Soodyall. He is also thanking for his guidance and support during this study.

Last but not least, I want to express my gratitude to my family, to Mathews Ngounou, to Nestor Djomathui, to staffs and students at the Vector Control Reference Unit (Entomology), NICD, and my friends for their endless support and encouragement throughout my studies.

This study was supported by the South Africa National Research Found and the WHO/AFRO Multicentre Research Project.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

CHAPTER TWO

MATERIALS AND METHODS

CHAPTER THREE

RESULTS

CHAPTER FOUR

GENERAL DISCUSSION

CHAPTER FIVE

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1 GENERAL INTRODUCTION

Despite advances in preventive and public health measures as well as in therapy, malaria, one of the oldest diseases known to mankind, remains the principal killer mainly in tropical and sub-tropical areas. Africa has the largest number of people exposed to stable malaria transmission and the greatest burden of malaria morbidity and mortality in the world (WHO, 1999).

The malaria burden in tropical and sub-tropical zones can be attributed to many causes. These include the development and rapid spread of resistance in the human *Plasmodium* to antimalarial drugs, resistance of vectors to insecticide, social causes (e.g. high birth rates, poverty, marginalized communities), transmigration, environmental changes (global warming may have contributed to the spread of malaria into some previously malaria free areas) and political conflicts affecting efficient application of control measures (Coetzee and Hunt, 1998; Mouchet *et al.*, 1998; Gallup and Sachs, 2001). In addition to these, the effect of global climatic

change may be significant in the long-term future (Lindsay and Martens, 1998; Mouchet *et al.,* 1998; Coetzee *et al.*, 2000).

Over the last decade of the twentieth century, malaria killed about ten times as many children as did all the wars in that period (WHO, 1999). The World Health Organization (WHO) estimates approximately 300-500 million clinical cases every year (this is five times higher than AIDS, TB, measles and leprosy combined), with over a million deaths. Over 90% of these occur in tropical Africa, with 1.5-2.7 millions fatal cases reported, most of which are children and pregnant women (Guyatt and Snow, 2001). Tropical and sub-tropical areas have the highest level of poverty with an average per capita gross domestic product of 0.4% per year compared to 2.3% in other countries (Gallup and Sachs, 2001).

The female *Anopheles* mosquito is the natural vector of this protozoan vector borne disease. *Anopheles* is the only genus that contains important human malaria vectors and belongs to the subfamily Anophelinae, Family Culicidae (Gillies and De Meillon, 1968). The principal vectors in sub-Saharan Africa are *An. gambiae* Giles and *An. arabiensis* Patton, both members of the *An. gambiae* complex, and *An. funestus* Giles (Gillies and De Meillon, 1968; Coluzzi, 1970; Gillies and Coetzee, 1987). These species are efficient malaria vectors due to their marked preference for human blood meals and rapid adaptation to environmental changes induced by human shelter and agriculture (Gillies and De Meillon, 1968; Coluzzi *et al*., 1979; Lindsay and Martens, 1998).

Protozoan parasites of the genus *Plasmodium* are the causative organisms of malaria. Human malaria is caused by four species of *Plasmodium: P. falciparum* is responsible for most malaria mortality and severe morbidity, predominant in Africa, New Guinea and Haiti; *P. vivax* causes tertian malaria and is rare in Africa; *P. ovale,* also rare, is found occasionally in South Africa, causes ovale tertian fever; *P. malariae* is found in most endemic areas worldwide, distributed especially in sub-Saharan Africa, and causes quartan fever. It is not as prevalent in Africa as *P. falciparum.* (Gillies and De Meillon, 1968; Davis and Icke, 2002).

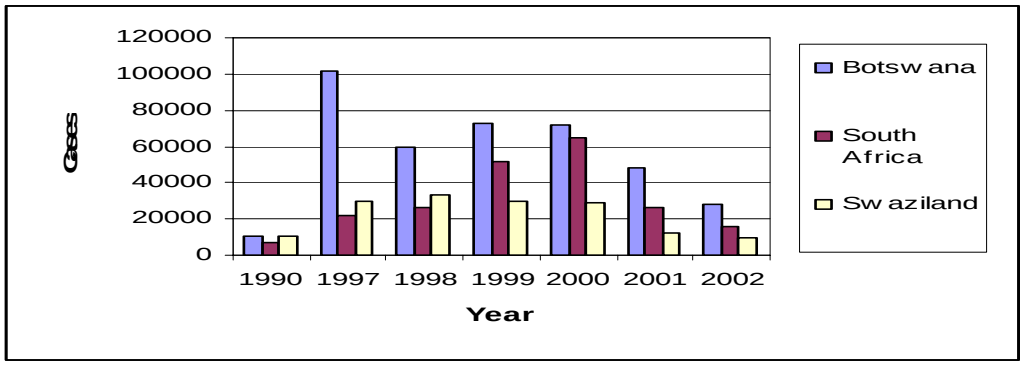
1.2 MALARIA IN SOUTHERN AFRICA

Malaria is the major cause of morbidity in this region with an average of 250,000 deaths annually (report from the Southern Africa Malaria Control, 2000, www. malaria.org.zw). The intensity of malaria transmission in southern Africa varies considerably and includes malaria-free zones as well as stable and unstable areas. Malaria is said to be stable where transmission is marked throughout the year with little seasonal variation and unlikely epidemics, while in unstable malaria zones, transmission level is low and varies from year to year with a high potential for epidemics. (Pampana, 1963).

Unstable malaria is encountered below latitude 20^{o}S in areas encompassing South Africa, Botswana, Swaziland, Namibia and Zimbabwe (Smith *et al.,* 1977). These areas are prone to epidemics that can result in high levels of morbidity and mortality

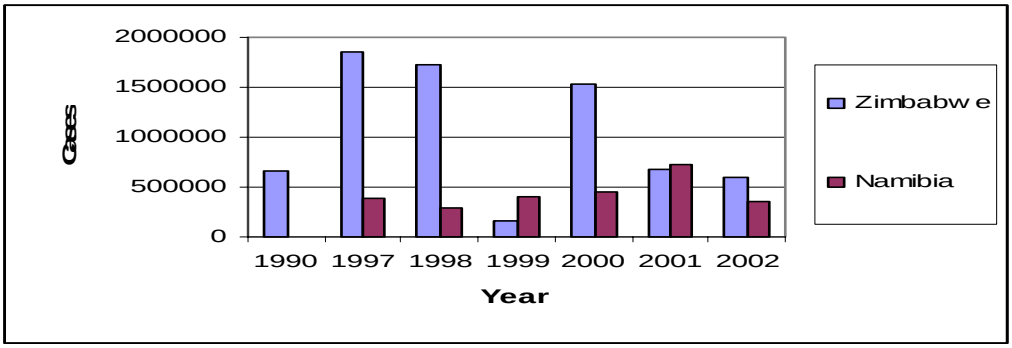
(Figs 1.1 and 1.2). Transmission peaks are generally observed between February and April during the rainy season, but this period is extended as one moves north closer to the equator (Coetzee *et al*., 2000).

Four members of the *Anopheles gambiae* complex and seven species of the *An. funestus* group are present in southern Africa, with *P. falciparum* the predominant species of parasite. The main vectors of malaria in the southern regions are *An. gambiae, An. funestus* and *An. arabiensis.* (Gillies and De Meillon, 1968; Gillies and Coetzee, 1987)



Data are from the WHO-Regional Office for Africa (www.malaria.org.zw) (December 2003)

Figure 1.1 Malaria profile in three southern African countries 1990-2002



Data are from the WHO-Regional Office for Africa (www.malaria.org.zw) (December 2003)

Figure 1.2 Malaria profile for Zimbabwe and Namibia 1990-2002

1.3 HISTORY OF THE *ANOPHELES GAMBIAE* GILES COMPLEX

Giles first described *An.gambiae* in 1902 (cited in Gillies and De Meillon, 1968). This mosquito was originally known as *An. costalis* Loew (De Meillon, 1947; Gillies and De Meillon, 1968). However, it was only in 1932 that the name *An. gambiae* came into general use after years of discussions among malariologists (Gillies and De Meillon, 1968). During the following years, publications on *An. gambiae* were focused mainly on the biology and morphological differences noticed in different regions of Africa (e.g. De Meillon, 1947). Based on environmental analysis, the larval habitats of *An. gambiae* varied from small rainwater puddles completely exposed to sun (De Meillon, 1947; Haddow *et al*., 1947) to pools with high salinity (Evans, 1938; De Meillon, 1947; Muirhead-Thomson, 1951). The adult biology of *An. gambiae* was also highly variable, from endophilic (mainly indoors) and anthropophilic (mainly feeding on man) to exophilic (mainly outdoors) and zoophilic (mainly cattle feeding) (Gillies and De Meillon, 1968). In 1947, De Meillon noted that, "the distribution of *gambiae* is not continuous, and hence it is most likely that patterns of behaviour will not be the same everywhere quite apart from environmental influences". Behavioural differences of *An. gambiae* were observed in Swaziland by Mastbaum (1954, 1957a, 1957b) with the re-emergence of this species after six years of indoor-residual insecticide spraying.

Anopheles gambiae was once regarded as a single freshwater-breeding species with salt-water variants, *An. melas* Theobald in West Africa and *An. merus* Dönitz in East

Africa (Evans, 1938; Muirhead-Thomson, 1947). *Anopheles melas* could be differentiated from *An. gambiae* by its extreme dark pigmentation (Evans, 1938) and both species by larval salinity test (Ribbands, 1944; Muirhead-Thomson 1951). These morphological differences were expanded by Muirhead-Thomson (1945, 1947) using eggs of both species and confirmed by Paterson (1962). In addition Muirhead-Thomson (1945, 1947, 1951) presented clear evidence from crossing experiments that the two *An. gambiae* salt-water types, *An. melas* and *An. merus* respectively were in fact distinct species. Crossing between each of those salt-water species and the fresh-water breeding *An.gambiae* always yielded sterile males in the F1 generation (Muirhead-Thomson, 1947; Paterson, 1962).

Paterson (1964) found that three fresh-water forms did not mate in nature. He therefore assigned the names A, B and C to them using a non-Linnean nomenclature. Crossing between these forms yielded sterile F1 male and fertile female hybrids (Davidson *et al*., 1967). Davidson and Hunt (1973) presented evidence of a new sixth species, species D, from hot mineral water springs in Bwamba County, Uganda. In 1977, Mattingly assigned the following scientific names to the six species:

Species	**Scientific names**
West Africa salt-water breeder	*An. melas* Theobald
East Africa salt-water breeder	*An. merus* Dönitz
A	*An. gambiae* Giles
B	*An. arabiensis* Patton
C	*An. quadriannulatus* Theobald
D	*An. bwambae* White

A new member of the group *An. quadriannulatus* species B, was described from Ethiopia by Hunt *et al.* (1998). Furthermore, polytene chromosome studies carried out on populations of *An. gambiae s.s.* in West Africa revealed the existence of distinct populations based on paracentric inversion polymorphisms on the second chromosome (Coluzzi *et al*., 1979; Toure *et al.,* 1998) and subsequently the presence of two molecular forms (M and S) by genotyping the X-linked ribosomal DNA (rDNA) of *An. gambiae s.s.* (Favia *et al.,* 2001; della Tore *et al*., 2001; Gentile *et al.,* 2001).

1.4 SPECIES DESCRIPTION, BIOLOGY, DISTRIBUTION, AND VECTOR STATUS OF THE *ANOPHELES GAMBIAE* COMPLEX

1.4.1 *Anopheles gambiae* Giles

This fresh-water breeder is widespread in most African countries south of the Sahara including neighbouring islands. It is the most anthropophilic member of the complex and one of the main malaria vectors in Africa (Gillies and De Meillon, 1968). The species predominates in vegetation zones of forest and humid Savannah areas and breeding occurs throughout the year (Coluzzi *et al*., 1979). The larvae are found in sunlit pools, puddles burrow pits and rice-fields (Gillies and De Meillon, 1968). This species is closely related to man and his environment (Coluzzi *et al*., 1979).

Among the seven members of the *An. gambiae* complex, only *An. gambiae* and *An. arabiensis* have shown a complex pattern of inversion polymorphism, principally on chromosome 2R (Coluzzi *et al.,* 1979). Furthermore, polytene chromosomes (Coluzzi *et al.,* 1979; Toure *et al.,* 1998) and molecular (Favia *et al.,* 2001; della Tore *et al*., 2001; Gentile *et al.,* 2001; Fannello *et al.,* 2003) studies carried out on populations of *An. gambiae* in West Africa showed the existence of subpopulations within the species. This species is now composed of sympatric taxa. These subpopulations are morphologically indistinguishable, with no reproductive barriers

under laboratory conditions but genetically distinct (Coluzzi *et al.,* 1979; Toure *et al.,* 1998).

1.4.1.1Chromosomal forms of *Anopheles gambiae.*

Five chromosomal forms have been identified within this species and named by non-Linnean nomenclature as Mopti, Bamako, Forest, Bissau, and Savannah (Coluzzi et al., 1985; Toure et al., 1998). These forms have been characterized on the basis of different ecological and geographical habitats corresponding to frequencies of paracentric inversion polymorphisms. These inversions mostly occur on the right arm of the second chromosome 2R. Deviation from Hardy-Weinberg equilibrium was recorded (Coluzzi et al., 1979; Bryan et al., 1982) suggesting non-random mating. The following summary has been adapted from the cytogenetic studies of Coluzzi et al. (1985) and Toure et al. (1998).

-The **Forest** form has the standard arrangement of chromosome 2R or occasionally with inversions 2Rb and 2Rd. This form occurs in all the rain forest areas and humid Savannah. It integrates with the Savannah form.

-The **Savannah** form involves many alternative chromosome 2 inverted arrangements (2Rb/cu/bcu) plus the standard arrangement. It is the most widespread form occurring in the Savannah habitat throughout Africa. The Savannah and the Forest forms share the 2La inversion.

-The **Bissau** form (2Rd inversion) is restricted to the coastal rice cultivated areas in The Gambia, South Senegal, Guinea-Bissau and Guinea Conakry. This form shows partial integration with Savannah.

-The **Bamako** form (inversion 2Rj that maybe in association with 2Rcu and 2Rbcu) is found along the upper Niger River and its tributaries in southern Mali and northern Guinea Conakry. It is frequently in sympatry with Savannah and Mopti but hybridize with only Savannah or Forest-Savannah integrating forms.

-The **Mopti** form (2Rbc/u) is found in the riverine or irrigated habitats in Mali, Guinea Conakry, Côte d'Ivoire and Burkina Faso. Sympatrically with Savannah and /or Bamako, Mopti's larvae are well adapted in man-made breeding places. This is the only form that totally exploits the breeding opportunity during the dry season. Toure *et al.* (1994) state "The 2R chromosomal polymorphism of *An. gambiae* Mopti is so precise in its response to the ecological conditions that it could be used as a biological indicator to predict the degree of aridity at the sampling site".

1.4.1.2 Molecular forms of *An.gambiae*

A PCR-RFLP assay was developed by Favia *et al.* (1994, 1997, 2001) and Gentile *et al.* (2001, 2002) using the rDNA located on the X chromosome of *An. gambiae.* (Collins *et al*., 1987). Two molecular forms called M and S were identified using the intergenic spacer of rDNA (IGS) (della Torre *et al.,* 2001), and types I and II using

the internal transcribed spacer (ITS) region (Gentile *et al.,* 2001, 2002). The S form corresponds to the type I and the M form to type II.

In Mali, the M and S forms are equivalent to the Mopti and Bamako/Savannah chromosomal forms respectively (Favia *et al.,* 1997) but this association breaks down in other regions (Chandre *et al*., 1999; Weill *et al.,* 2000; della Torre *et al.,* 2001; Wondji *et al*., 2002).

1.4.2 *Anopheles arabiensis* Patton

Anopheles arabiensis, the second anthropophilic member of the complex, tends to be more frequent than *An gambiae* in hot and dry conditions. This fresh-water breeder occurs sympatrically with *An. gambiae* in many Afrotropical areas (Gillies and De Meillon, 1968; Gillies and Coetzee, 1987; Coetzee *et al.,* 2000) and shares the same larval habitat (Service, 1970). It is the only member of the group present in the Aden hinterland of Arabia (Mattingly, 1977).

The numbers of *An. arabiensis* tend to increase towards the end of the rainy season and during the dry season as have been observed in northern Nigeria and Tanzania (White *et al*., 1972). *Anopheles arabiensis* is found resting both indoors and outdoors and feeds on cattle and humans (Gillies and De Meillon, 1968; Gillies and Coetzee, 1987). Due to the variations found in its resting behaviour, it is less susceptible than *An. gambiae* to the effect of house spraying (Service *et al.,* 1978). Both *Anopheles*

arabiensis and *An. gambiae* have been shown to be vectors of *Wuchereria brancrofti* in tropical Africa as well as malaria (Gillies and Coetzee, 1987).

1.4.3 *Anopheles quadriannulatus* Theobald

This fresh water species was discovered in the Transvaal region, South Africa (Gillies and De Meillon, 1968; Gillies and Coetzee, 1987). It has a limited distribution and is of no importance as a malaria vector due to being markedly zoophilic (Gillies and De Meillon, 1968).

1.4.4 *Anopheles quadriannulatus* B Hunt, Fettene and Coetzee

This new fresh water species was described from Jimma, Ethiopia (Hunt *et al.,* 1998) using cross- mating experiments. It cannot be distinguished from *An. quadriannulatus* Theobald on the basis of chromosome arrangement nor on the standard PCR assay (Hunt *et al*., 1998). It is unknown whether the southern African *An. quadriannulatus* also occurs in Ethiopia (Coetzee *et al*., 2000). It is of no medical importance, feeding mainly on cattle (Gillies and De Meillon, 1968).

1.4.5 *Anopheles bwambae* White

This species was described from samples found in mineral springs in the Semiliki forest of Uganda (Davidson and Hunt, 1973; Davidson and White, 1972). It has not

been recorded elsewhere. *Anopheles bwambae* is anthropophilic and exophilic but is not considered an important malaria vector, although it can transmit malaria within its very restricted range (White, 1985).

1.4.6 *Anopheles melas* Theobald and *An. merus* Dönitz

Anopheles melas and *An. merus* are the saltwater-breeding forms of the *An. gambiae* complex. They are allopatric with *Anopheles melas* restricted to coastal West Africa while *An. merus* is found in the eastern and southern parts of the continent occurring far inland in Zimbabwe and South Africa (Coetzee *et al.,* 2000).

Adults of both species bite humans and animals and rest indoors and outdoors. They appear less important medically than local freshwater vectors, but are not to be considered insignificant (Gillies and De Meillon, 1968; Temu *et al.,* 1998). *Anopheles merus* has been implicated in malaria transmission at localised sites in Zimbabwe (H. Masendu, unpublished data).

1.5 IDENTIFICATION OF THE SPECIES

One of the most critical points during epidemiological investigations and effective control is the proper identification of the vector, as species or cryptic species express different behavioural and physiological characteristics and consequently different

vector capacities. Many techniques have been used and these are briefly described below.

1.5.1 Morphological analysis

Classical morphological analysis (Gillies and Coetzee, 1987) is the baseline of any identification but was found to be of limited value at the cryptic species level (Coluzzi, 1964; Coetzee, 1989). No morphological distinction has been found among *An. arabiensis, An. gambiae* and *An. quadriannulatus* A (Green, 1971). Ribbands (1944), Muirhead-Thomson (1945), Paterson (1962) and Coluzzi (1964) distinguished the two salt water breeding forms from the fresh water species of the complex on the basis of their eggs. Coetzee *et al*. (1982) and Coetzee (1989) discriminated among *An. arabiensis, An. gambiae s.s., An. merus* and *An. quadriannulatus* using all life stages. However, identification of individual mosquitoes is not possible and morphology is not used beyond the group level for identification of members of the *An. gambiae* complex.

1.5.2 Salinity tolerance tests

This technique was used by Ribbands (1944) and Muirhead-Thomson (1951) to distinguish between the salt water and fresh water members of the *An. gambiae* complex. This method involves exposing first instars larvae to 75% seawater (23.8g

sodium chloride per litre of water). The fresh water species die within one hour while the salt-water breeders survive.

1.5.3 Cross- mating experiments

This method was first used by Muirhead-Thomson (1947) to differentiate fresh water *An. gambiae* from salt-water *An. melas* in Lagos, Nigeria, but the results were ignored. It was only when Paterson (1962) showed that *An. merus* was a good biological species that the technique was widely used to elucidate cryptic species of the *An. gambiae* complex (Davidson, 1962; Paterson, 1964; Davidson and Hunt, 1973; Hunt, 1973). The concept is based on the sterility of the F1 male progeny while females are fertile. Hunt *et al.* (1998) used cross-mating techniques to demonstrate a new member of the complex from Ethiopia, *An. quadriannulatus* species B.

This technique is time consuming and laborious, requires the laboratory stocks of all species, and is not suitable for large-scale use.

1.5.4 Chromosomal banding patterns

Identification of most members of the *An. gambiae* complex can be obtained using the banding patterns on the giant polytene chromosomes found in fourth instar larvae (Coluzzi and Sabatini, 1967, 1968, 1969) and the ovarian nurse cells of half gravid

females. This method uses fixed inversion differences on the X chromosomes and the autosomes to identify species.

Disadvantages include the use of half- gravid females or fourth instar larvae only and also cannot distinguish between *An. quadriannulatus* A and B. Analysis of results requires a high level of expertise. However, the technique has some advantages such as the stability of the characteristics used for identification, low-cost of reagents and only requires very simple laboratory facilities.

1.5.5 Isoenzyme electrophoresis

Isoenzyme electrophoresis describes the migration of charged enzyme molecules through a gel matrix under the influence of an electric field. Depending on their charges, molecules will migrate either to the cathode or the anode end of the gel (Wilson and Walker, 2000).

Mahon *et al*. (1976) were first to describe isoenzyme electrophoresis for the identification of members of the *An. gambiae* complex. This method is applicable to adults of both sexes (Mahon *et al.,* 1976; Miles, 1978) but is no longer used due to overlap of diagnostic characters (Hunt and Coetzee, 1986), and the development of new molecular methods such as the species-specific PCR assay (Scott *et al.,* 1993).

Other drawbacks of this technique are the requirement of live material or dead material preserved and transported at low temperature (e.g. in liquid nitrogen) (Miles, 1979).

1.5.6 DNA- based methods

The identification of insect vectors at any life stage regardless of sexes through DNA analysis is a major advance in the study of malaria vector mosquitoes. In addition to fresh and frozen materials, alcohol preserved and dried specimens can be identified.

1.5.6.1 DNA hybridization methods

DNA probes were the first molecular method used to identify members of the *An. gambiae* complex at the species level (Collins *et al.,* 1987; Gale and Crampton, 1987; Hill *et al.,* 1991). DNA probes are mainly single- stranded DNA sequences. Three members of the *An. gambiae* complex, *An.gambiae s.s. An. arabiensis* and *An. melas* were distinguished using a single probe. The probe was constructed using the EcoR1 / SaII fragment from the 3' end of the 28Sβ coding region of the rDNA cistron of *An. gambiae* (Collins *et al.,* 1987). Ribosomal DNA (rDNA) are a family of tandem-repeated sequences with at least 100 copies per chromosome in all organisms and are separated by intergenic spacer regions (IGS) (Collins *et al.,* 1987). The IGS differ significantly between species. This technique was 100% accurate on wild specimens that were also chromosomally identified (Collins *et al*., 1988).

However, this technique was shown to be limited (Gale and Crampton, 1987; Hill *et al.,* 1991) and required sophisticated equipment. Also, species-specific probes are often difficult and costly to isolate. Hill *et al*. (1992) improved the method for rapid identification of members of the *An. gambiae* complex but this was never used also due to the cost of equipment and reagents.

1.5.6.2 Species-specific polymerase chain reaction (PCR) identification methods

Polymerase chain reaction is an enzymatic DNA amplification procedure. The identification is based on either ribosomal genes (rDNA) (Collins *et al.,* 1987, 1988; Paskewitz and Collins, 1990; Scott *et al*., 1993) or by using random amplified polymorphic DNA (RAPD) markers (William *et al.,* 1990; Kambhampati *et al*., 1992; Favia *et al.,* 1994).

Paskewitz and Collins (1990) developed three non-radiometric/radioactive polymerase chain reaction primers that could distinguish *An. arabiensis* from *An. gambiae s.s.* These three PCR primers were derived from the IGS of the rDNA genes. The technique used a universal plus-strand primer derived from a conserved region at the 3' end of the 28S rDNA-coding region and two species-specific minus-strand primers derived from the IGS. The universal primer works with either of the species-specific primers to produce a 0.5kb fragment if *An. arabiensis* DNA is used as template and a 1.3kb fragment in the case of *An. gambiae* DNA. Both 1.3kb and

0.5kb fragments will be produced from *An. gambiae*/*An. arabiensis* hybrids and nothing at all if the DNA from other members of the complex is used as the template.

This identification method was taken further by Scott *et al*. (1993) by developing four species-specific primers to identify *An. quadriannulatus* (153bp), *An. arabiensis* (315bp), *An. gambiae* (390bp), *An. melas* (466 bp), and *merus* (467 bp). The rDNA of the later two species are amplified by the same minus strand primer and as a result they produce closely the same size fragment of DNA, but due to their allopatric distribution, there is no requirement to separate them. Townson and Onapa (1994) identified the sixth member of the complex, *An. bwambae*. The species is characterized by two fragments of 390bp and 690bp. Paskewitz *et al*. (1993) proved the reliability of the rDNA-PCR method using chromosomally and electrophoretically identified field and colony material from southern Africa.

Fettene *et al.* (2002) also Fettene and Temu (2003) have developed a PCR based method to distinguish between the sibling species A and B of *An. quadriannulatus.* Also, Favia *et al*. (1994) developed a PCR assay that can identify the two molecular forms M and S of *An. gambiae.*

1.6 RATIONALE ON GENE FLOW

The fight against malaria and other pathogen vectors are commonly based on insecticides. But the development of resistance to insecticides has prompted changes

in control, particularly for malaria vectors (Chandre *et al.,* 1999; Weill *et al.,* 2000; Hargreaves *et al.,* 2000, 2003; Fanello *et al.,* 2003). Knowledge of geographic dispersion of a particular mosquito population and its susceptibility to an insecticide is important (Fannello *et al.,* 2003*)*. Once resistance has been detected, analysis of gene flow will give some idea as to how the resistance will or will not spread to neighbouring populations. Gene flow can be defined as the movement of genes within and between populations (Ferris *et al.,* 1983), and thus include the movement of gametes, individuals or groups of individuals from one place to another (Slatkin, 1985). It is influenced by several factors such as natural selection, random genetic drift, and mutations, which determine genetic differences between and within species populations (Ferris *et al.,* 1983; Slatkin, 1995).

Gene flow provides information on the dispersal patterns of natural populations or subpopulations that is critical for vector control, indicating the distance of dispersal of vector (s) (Taylor *et al*., 1993; Donnelly and Townson, 2000; Kamau *et al.,* 1999; Pinto *et al.,* 2003). Geographic distribution of knock- down resistance (kdr) has been described using gene flow analysis (Weill *et al.*, 2000; Fanello *et al.,* 2003)

Gene flow cannot be measured (Slatkin, 1995). It can be estimated using either a direct approach or an indirect approach (Slatkin, 1995). The indirect model is based on the estimation of allele frequencies, usually obtained by electrophoretic survey of proteins or DNA sequence using molecular markers; this effect is cumulative acting over all temporal and spatial scales (Lanzaro *et al.,* 1998; Walton *et al.,* 1998;

Donnelly and Townson, 2000). In contrast, direct estimates of gene flow result from the observation of dispersal within a defined range of time and space (Taylor *et al.,* 2001; Edillo *et al*., 2002).

1.7 MOLECULAR MARKERS USED IN POPULATION GENETICS

1.7.1 Protein markers

Isozymes / allozymes are the most widely used protein markers. Allozymes can be assayed and surveyed in nearly any species. They are separated on starch, polyacrylamide or cellulose media and stained using specific enzyme reaction mixtures. Allozymes markers detect both homo-and heterozygous banding patterns (Loxdale and Lushai, 1998). They segregate under electrophoretic conditions, according to Mendelian genetics with codominance in heterozygotes. The rates of mutations that produce detectable differences in allozyme alleles are in the order of 10-6 (Voelker *et al.,* 1980). Allozymes have been used for various purposes, e.g. species identification (Mahon *et al.,* 1976; Miles, 1978; 1979), and population genetics (Aagaard *et al.,* 1998).

Allozyme analysis is cheap, often much quicker to isolate and develop and is easy to use. The drawbacks of this method include low detectable variability and a need for fresh or frozen material (-20 degree Celsius). Because allozymes do not represent a random sample of the genome they may bias some population genetic interferences (Loxdale and Lushai, 1998).

1.7.2 DNA markers.

1.7.2.1 Restriction fragment length polymorphism (RFLP)

The restriction fragment length polymorphism marker reflects the differences in the sequences of DNA samples. They are regions of nuclear DNA isolated through PCR or other means and digested with restriction enzymes, which recognize specific four or six base pair sequences (Wilson and Walker, 2000). They are codominant and able to detect analogous DNA on both pairs of homologous chromosomes (Loxdale and Lushai, 1998). They have been used in many studies due to their polymorphic character (Favia *et al.,* 1997; della Torre *et al.,* 2001). Despite the importance of RFLPs as molecular markers (Botstein *et al.,* 1980), they suffer from several drawbacks. The level of variation associated with most RFLPs is low and is usually limited by the number of alleles (Goodfellow, 1992). In addition, they are time consuming and difficult to handle.

1.7.2.2 Random amplified DNA polymorphism (RADP-PCR)

This method was first described by Williams *et al.* (1990) and uses a single nucleotide primer from an 8-10 base pairs long sequence to amplify random sections of nuclear DNA. The method exhibits dominant banding patterns (compared to allozymes and microsatellites that are codominant) whereby both homozygotes and

heterozygotes produce a band that can limit their use in mapping studies (Black, 1993).

The method requires stringent standardization before the markers can be used reliably and consistently (Black, 1993). Also, results are not easily reproduced making them difficult to interpret. However, this method presents some advantages. It does not employ radioactive markers; presents high levels of polymorphism, is inexpensive, and results can be visualized within 24h from initial extraction of genomic DNA (Black, 1993).

This method has been used widely for different purposes that include identification of sub-species and cryptic species (Black *et al.*, 1992; Kambhampati *et al.,* 1992), genome mapping (Favia *et al.,* 1994) and genetic fingerprinting (Black *et al.,* 1992).

1.7.2.3 Microsatellite markers (short tandem repeats (STR) or simple sequence repeats (SSRs)).

Microsatellite markers are also called simple sequence repeats (SSRs) (Tautz, 1989) or short tandem repeats (STRs) (Edward *et al.,* 1991). They are short stretches of DNA that occur throughout the genomes of many eukaryotic organisms, including insects (Weber and May, 1989). They are typically tandemly arranged with a motif of 1 to 6bp repeated up to 100 times (specific DNA base sequence or nucleotides can contain mono, di, tri, or tetra tandem repeats) (Beckman and Weber, 1992).

Mutation rates and mode of inheritance of microsatellites determine the sensitivity of a DNA locus. If microsatellite loci are long enough and uninterrupted, they are excellent genetic markers (Powell *et al.,* 1996). Microsatellite mutation is very high and varies from 10^{-4}-10^{-5} in Yeast (Henderson and Petes, 1992), $6x10^{-6}$ in Drosophila (Shug *et al.,* 1997) up to 10^{-3} events per locus per generation in humans (Weber and Wong, 1983). Microsatellites, with their high levels of length polymorphism generated by a high mutation rate should give them great power to resolve differentiation within and between populations (Walton *et al.,* 1998; Kamau *et al.,* 1999; Simard *et al.,* 1999). They are inherited in a Mendelian fashion and expressed co-dominantly (Litt and Luty, 1989).

Slipped-Strand misparing or polymerase slippage (Levinson and Gutman, 1987; Schlötterer and Tautz, 1992) and unequal crossing-over during meiosis (Wolff *et al*., 1989) can explain why microsatellites are highly polymorphic (Fig 1.3). The polymerase slippage occurs during the replication of DNA. During the process, the polymerase detaches from the template, resulting in displacement of the replicated strand from the complementary strand. The result is a new strand with a different number of repeats to the parent strand. This is obtained via "looping" during reannealing of the two strands (Levinson and Gutman, 1987; Schlötterer and Tautz, 1992). The repeat sequences vary continuously between and among populations in a species or closely related populations (Lanzaro *et al.,* 1995).

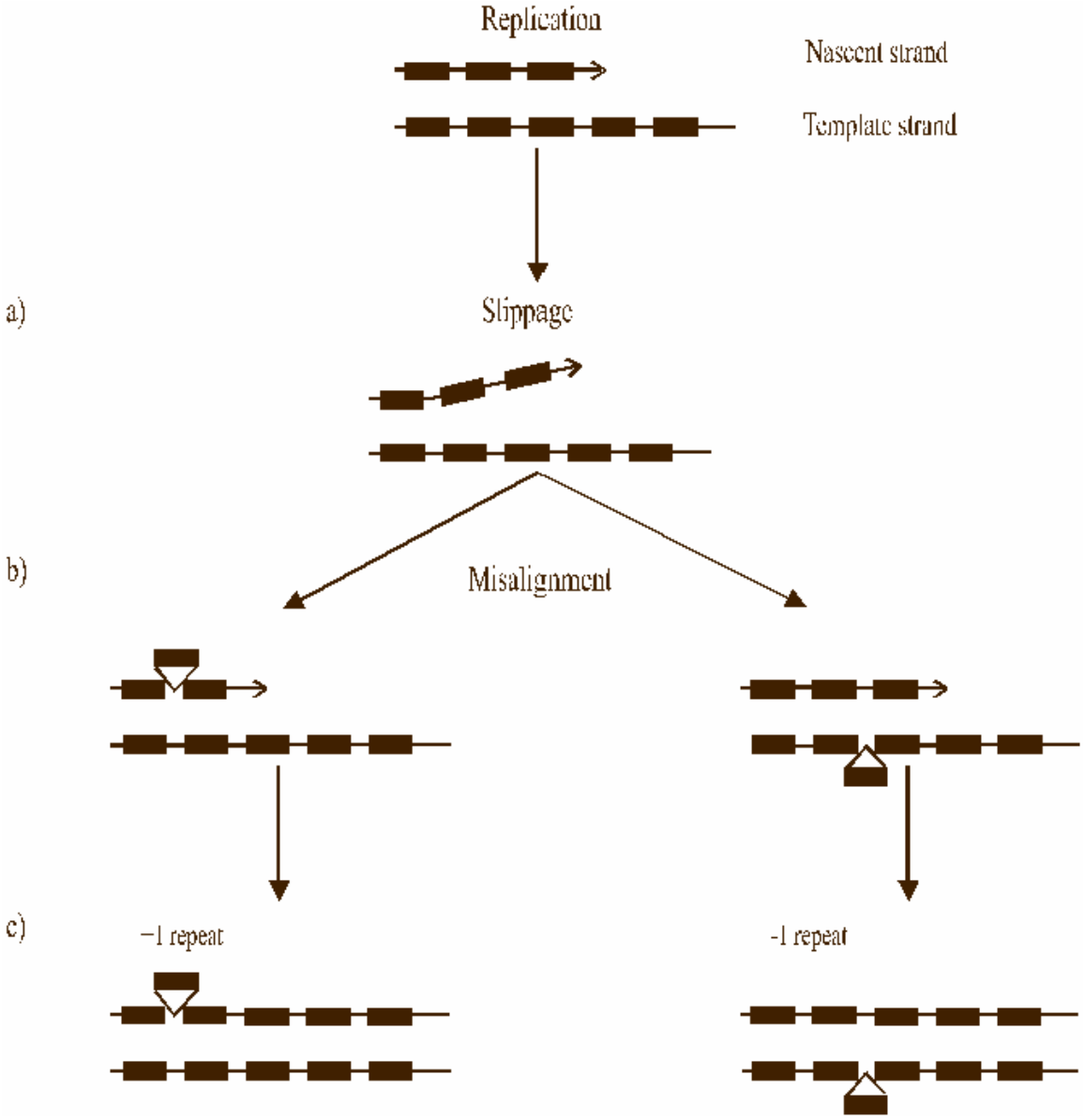


Figure 1.3 Model of mutation process at microsatellite loci. (a) Slippage of the DNA polymerase during replication. (b) Misalignment of the template or the newly replicated strand. (c) Continuation of replication. DNA strands are represented by lines, repeat units by small boxes, and direction of replication by small arrows (http//hekules.oula.fit/)

Microsatellite markers use PCR (Saiki *et al.,* 1988) to amplify regions of interest. The size of amplified microsatellite alleles are determined by electrophoresis on denaturing or non-denaturing high-resolution gels that are easy to score. Because microsatellites are scored in a Mendelian codominant fashion (Litt and Luty, 1989), each amplicon from a diploid (2n chromosomes) genome is expected to produce two bands, to be scored as genotype. If the individual tested is homozygous at that particular locus, then both alleles will have the same mobility and the genotype would appear as a single band while two bands should appear for a heterozygous individual.

Microsatellite markers are considered neutral (non-coding) fragments of DNA. They are unique in that many are closely associated with conserved loci that contain coding regions (flanking regions) (Loxdale and Lushai, 1998). These flanking regions (5 prime and 3 prime) can be used as PCR primers when amplifying microsatellite loci and can be conserved across genera. (Loxdale and Lushai, 1988).

Microsatellites as genetic markers have been extensively used for population studies in Anopheline mosquitoes (e.g. Rongnoparut *et al.,* 1999) and principally *An. gambiae* in Africa (e.g. Lanzaro *et al.,* 1995; 1998). However, recent works on *An. arabiensis* demonstrated the power of microsatellite markers to detect barriers to gene flow in these mosquitoes (Walton *et al.,* 1998). Donnelly and Townson (2000) showed a significant level of differentiation between populations from localities in Tanzania and Mozambique using short tandem repeats, while Simard *et al.* (2000)

demonstrated extensive gene flow between two populations of *An. arabiensis* 250 km apart in Senegal during dry season conditions.

1.8 AIMS AND OBJECTIVES OF THE STUDY

Recent studies from West and East Africa have revealed considerable population structuring, even within taxa that have been long recognized as single panmictic species (Toure *et al.,* 1998; Kamau *et al.,* 1999; Donnelly and Townson, 1999; 2000). A multi-centre project, entitled the "characterization of vector populations and capacity building for vector control in seasonal, epidemic-prone, malaria transmission areas of southern Africa" was initiated in 2002 with the support of WHO/AFRO. This study was prompted by malaria epidemics experienced in Namibia and Zimbabwe during April/May 2001 after heavy rains, with an extension in Botswana. A similar epidemic was encountered in South Africa in 2000. Collections were carried out in five countries in southern Africa (Fig 1.4) that formed part of the multi-centre collaborative project.

The study aimed to estimate the level of gene flow in *An. arabiensis* across the southern-most limits of its distribution using microsatellite DNA markers. A better understanding of population genetic structure of *An. arabiensis* will provide information regarding the spread of genes such as those involved in insecticide resistance (e.g. Hargreaves *et al.,* 2003). In addition, such knowledge is relevant to the development of genetic strategies for vector control using genetically engineered

mosquitoes with reduced vectorial competence (e.g. Crampton *et al.,* 1994). Generally, this information will indirectly have an impact on the existing malaria control programme.

Objectives of the study

The overall objective of this study was to use four highly polymorphic microsatellite loci (Donnelly and Townson, 2000) to determine the genetic structure of *An. arabiensis* populations within and between malaria affected areas in five southern African countries.

Specific objectives include:

1- Identification of wild populations of *An. arabiensis* from five southern African countries (Botswana, Namibia, South Africa, Swaziland and Zimbabwe) using the PCR method developed by Scott *et al*. (1993).

2-Optimization of the microsatellite loci in *An. arabiensis* colonized in the Botha De Meillon insectary (Vector Control Reference Unit, NICD).

3- Measure of the genetic variability within and between subpopulations of wild *An arabiensis*

4- Estimation of the level of gene flow between natural populations across the Southern- limits of *An. arabiensis* distribution.

•Collection sites

<u>Figure 1.4</u> Map of study areas in southern Africa

CHAPTER TWO

MATERIALS AND METHODS

2.1 SPECIMEN COLLECTION

2.1.1 Study sites

Villages within five southern African countries were selected for this study (Table 2.1). The selection was done in concordance with the severe malaria epidemics encountered in these countries between 2000 and 2001.

2.1.2 Mosquito collection

Wild material was collected either at the larval or adult stage or both (Table 2. 1) using various techniques described by the World Health Organization (1975). Collections were carried out both indoors and outdoors (Table 2. 1).

Table 2.1 Wild mosquito collections showing country, locality, collection period and methods, and life stage of specimens/mosquitoes collected.

Country	Areas	Collection period	Collection methods	Life stage
SOUTH AFRICA	Mamfene (30^{o}47'E, 27^{o}23'S) Malahlapanga, KNP (31^{o}03'E, 22^{o}53'S) Limpopo (24°E,23°55'S) Mpumalanga	July 2002 February 2002 May2000 &May2002	Window trap Outdoors landing catches	**Adults** **Adults**
ZIMBABWE	Save-valley (Gokwe) (18^{o}-19^{o}E, 28^{o}-29^{o}S) Zambezi-valley (Kanyemba) (30^{o}-31^{o}E, 15^{o}-16^{o}S)	April 2002 November 2001	Outdoors landing catches Larval	**Adults** **Larval**
SWAZILAND	Big-Bend (31^{o}65'E, 26^{o}60'S)	Feb-March 2003	Resting outdoors	**Adults**
NAMIBIA	Okaku (16^{o}12'E, 17^{o} 80'S) Caprivi (24°15'E,17°30'S) Rundu (18^{o}16'E, 19^{o}90'S) Ishanaputa	Feb 2002 March 2002 April 2002 May2002	Larval Resting indoors Resting Indoors Resting Indoors	**Larvae** **Adults** **Adults** **Adults**

BOTSWANA	Chobe (24°-25°E, 18-19°S) Okavango (22°-23°E, 19°&20°S) Tutume (27°10'E, 20°25'S)	Feb 2002 March-2002 March 2002	Larval	**Larvae**

Table 2.2 Laboratory colonies of *Anopheles arabiensis* used in the optimization of the microsatellite markers.

Colony name	Origin	Year established
KGB	Kanyemba, Zimbabwe	1975
NAM	Oshakati, Namibia	2000
MAN	Mananga, Swaziland	1984

2.2 SPECIES IDENTIFICATION

2.2.1 Morphological identification

Members of the *An. gambiae* complex were separated from the *An. funestus* group at the collection sites using the morphological keys detailed in Gillies and Coetzee (1987) prior to preservation on desiccant (silica gel). These specimens were then sent

to the Vector Control Reference Unit (National Institute for Communicable Diseases, Johannesburg, South Africa) for further analysis**.**

2.2.2 Species-specific identification: Polymerase Chain Reaction (PCR)

2.2.2.1 Genomic DNA extraction

Extractions were done on specimens identified as *An. gambiae s.l* in order to perform species-specific identification and genotyping. DNA was extracted from either a single mosquito leg or both wings using the protocol of Collins *et al*. (1987). Sample tissue was homogenized in 50µl Grinding Buffer (0.08M NaCl, 0.16M sucrose, 0.06M EDTA, 0.1M Tris-HCl, 0.5%SDS and distilled water). The negative control excluding DNA was prepared concurrently and was called "NO DNA".

After homogenization, specimens were centrifuged at 13000 revolutions per minute (rpm) or 16060Xg for a few seconds (using a Biofuge fixed angle rotor, Heraeus) and incubated at 65^{o}C for 30 minutes. This ensured total inactivation of nucleases so as to avoid the degradation of DNA. After incubation 8M KAc (Potassium Acetate, Sigma) was added to a final concentration of 1.12 M, mixed and incubated on ice for 30 minutes. This salt solution precipitates insoluble materials as well as proteins denatured by Sodium Dodecyl Sulphate (SDS). After incubation the homogenate was centrifuged at 16060Xg for 20 minutes and the supernatants were transferred to new sterile 1.5ml microcentrifuge tubes without disturbing the pellets. Double the volume

of ice-cold (–20^{o}C) 100% ethanol was added and mixed by inverting the tube. DNA was precipitated for 5 minutes on the bench prior to centrifugation or kept at–20^{o}C overnight. Tubes were then centrifuged at 16060Xg for 30 minutes to pellet the DNA. Ethanol was then pipetted off and the DNA pellets were washed again with one volume of ice-cold 70% ethanol and re-centrifuged for 20 minutes at 16060Xg. The ethanol was removed and the pellets air-dried on the bench. The pellets were resuspended in 50µl 1X TE buffer (or 50µl of sterile water) and kept at -20°C (or –70°C for long term storage) until needed.

2.2.2.2. PCR identification

Field-collected mosquitoes were identified as *An. arabiensis* using the polymerase chain reaction (PCR) - based identification method developed by Scott *et al.* (1993) with few modifications (Van Rensburg *et al*., 1996). The PCR assay is based on species-specific sequences in an rDNA intergenic spacer (IGS) region, (Collins *et al.*, 1988).

A set of 5 oligonucleotide primers (Table 2.3) was used to identify *An. gambiae* s.s., *An. arabiensis*, *An. melas/merus* and *An. quadriannulatus.* The primers consist of one universal primer that has sequence identity to all 5 species and 4 species-specific primers for above-mentioned species of the *An. gambiae* complex. Both the universal primer and one of the species-specific primers anneal to produce a DNA fragment

with a distinctive length for each species (Scott *et al*., 1993). The amplified diagnostic fragment sizes are 315 base pairs (bp) for *An. arabiensis*, 390 bp for *An. gambiae*, 153 bp for *An. quadriannulatus* and 464/466 bp for *An. melas/ merus*. In addition, Van Rensburg *et al*. (1996) reported the presence of two bands (153 and 466bp) on *An. merus*.

Table 2.3 Primer sequences of the four *An. gambiae* complex species-specific primers as well as the universal primer. Melting temperatures are given (Scott *et al*., 1993).

Primer name	DNA sequence (5'to 3')	Tm (^{o}C)
UN (universal)	GTG TGC CCC TTC CTC AT GT	58.3
GA (gambiae)	CTG GTT TGG TCG GCA CGT TT	59.3
ME (merus)	TGA CCA ACC CAC TCC CTT GA	57.2
AR (arabiensis)	AAG TGT CCT TCT CCA TCC TA	47.4
QD (quadriannulatus)	CAG ACC AAG ATG GTT AGT AT	42.7

Tm=4Σ (A+T) + 2Σ (G+C) or [69.3ºC + (0.41 x GC %)] - [650÷ primer length] for primer with > 14 bases (Sambrook *et al.,* 1989).

One twentieth of the extracted DNA was used in a 12.5µl PCR reaction. The master mix contained the following chemicals: 10X reaction buffer (100mM Tris-HCl pH8.3, 500mM KCl), 125mM of each of the four deoxynucleotide triphosphates (dATP, dCTP, dTTP, dGTP), 1mM MgCl2 0.5 units (5units/µl) Taq polymerase (all

these chemicals were from Takara Shuzo CO., LTD Japan), 0.3μM of each primer (see Table 2.3) and 0.15μM of quadriannulatus primer (synthesized by Integrated DNA Technologies, INC. USA). Finally, double distilled water was added to make total volume of 12.5μl.

Amplification was carried out using an Eppendorf thermocycler (Mastercycler® gradient, by Eppendorf AG, Hamburg) for a total number of 30 cycles. The 30 cycles comprise an initial denaturation at 94ºC for 30 sec following by 29 cycles of (1) Denaturation at 94ºC for 30 sec; (2) Annealing at 50ºC for 30 sec; (3) Extension at 72ºC for 30 sec and a final auto-extension at 72°C for 5min to ensure the completion of double stranded amplicons.

After amplification, 10μl of each PCR product was mixed with 4μl of loading or ficoll dye (50% sucrose, 0.05M EDTA pH7, 0.1% bromophenol blue, 10% ficoll powder).

Ten microlitres of the mixture was loaded on a 2.5% agarose gel stained with 0.03mg/100ml ethidium bromide (Gibco BRL, UK) (Sambrook *et al*., 1989) and electrophoresed horizontally in 1X Tris-Acetic acid-EDTA (TAE) buffer. A DNA size marker of 1Kb- plus ladder (Gibco-BRL) was electrophoresed along side the PCR samples to estimate the average size of the amplified fragment.

The gel was run for 1.5h-2h at 103 V or until the Bromophenol blue had migrated 3cm from the origin. Amplicons were visualized using Ultra-Violet (UV) light on a transilluminator (Herolab Gmbh Laborgeräte, Germany).

The photograph was taken on a Polaroid Film (Fuji Instant, ISO 3000/30^{o} second processing at 75F).

In all amplifications, a negative control, which consisted of the master mix excluding the template DNA, was used to test for DNA contamination. The "NO DNA" was also included and reacted in the same fashion as the rest of the specimens amplified. A positive control consisting of species standards (DNA from a mosquito's leg) for each species (*An. arabiensis, An. merus, An. gambiae,* and *An. quadriannulatus)* colonized in the Botha De Meillon Insectary at National Institute for Communicable Diseases (NICD), Johannesburg was included during amplification. The colonized materials are kept at 25° C with relative humidity between 70-80%. The colony materials are routinely checked for purity at both chromosomal and molecular levels.

2.3 MICROSATELLITE GENOTYPE SCORING

Microsatellite analysis was performed using PCR. The products were amplified in an Eppendorf thermocycler (Mastercycler® gradient, by Eppendorf AG, Hamburg). Analysis of PCR product was performed using the high resolution vertical

polyacrylamide gel electrophoresis (PAGE) (Budowle and Allen, 1991). Bands were visualised after silver staining (Cairns and Murray, 1994).

Microsatellite loci were selected from published *An. gambiae* sequence data (Zheng *et al.,* 1993, 1996). Four loci were chosen (Table 2.4) for this study based on the level of polymorphism exhibited in *An. arabiensis* (Van Rensburg, 1997; Donnelly and Townson, 2000). Loci located on the X chromosome and the autosomal chromosome are considered neutral as opposed to the inversion polymorphism present on chromosome 2 (Coluzzi *et al.,* 1979).

2.3.1 Optimization of the microsatellite-PCR using colony materials

Prior to PCR amplification of wild materials, optimal PCR conditions were determined for each pair of primers (Table 2.4) in a few individuals per colonised mosquitoes (Table 2.2). Various parameters were tested including the PCR reagents (units of enzyme, concentration of each pair of primers, concentration of Mg2+) and the numbering of cycles and annealing temperature. The basic challenge was to devise an optimization method that was efficient in terms of time, cost and quality of results.

Optimization was initially performed at a high annealing temperature of 60°C and subsequently decreased by 5°C intervals in association with different concentrations of magnesium, variation of units of enzyme (Taq), concentration of each primer pairs

and finally the number of cycles. The amplified products were resolved on a polyacrylamide gel (Budowle *et al.,* 1991) and amplicons visualized by silver staining (Cairns and Murray, 1991).

Table 2.4 Primer sequences of microsatellite loci (From Whitehead Scientific SA, No 10401)**. Repeat and primer sequences as well as melting temperatures are given. F= Forward primer sequence and R= reverse primer sequence.**

Locus (Cytological position) [a]	**Repeat no. & allele size (bp)***	**Tm (° C)**	**Primer sequence** F= 5′ primer (5′→3′) R=3′ primer (5′→3′)
AGXH25 (X: 3A)	(GT)9-134	57.3° C	(F) GCCGAAAACATTCCAACAGG (R) CAGTTATGTCGGCATGCTAC
33C1 (III: 33C)	(AGC)9-148	55° C	(F) TTGCGCAACAAAAGCCCACG (R) ATGAAACACCACGCTCTCGG
AG2H26 (II: 12)	(GT)40-154	58.4° C	(F) GGTTCCCTGTTACTTCCTGCC (R) CCGGCAACAAACAATCGG
AG3H88 (III: 35-38)	(GT)9-176	60° C	(F) TGCGGCGGTAAAGCATCAAC (R) CCGGTAACACTGCGCCGAC

*bp= base-pair. [a] all loci are located outside polymorphic chromosomal inversions (Zheng *et al.,* 1993, 1996).

2.3.2 Polyacrylamide gel system

Microsatellite analyses were performed using high-resolution vertical polyacrylamide gel electrophoresis (PAGE) (Budowle *et al.,* 1991) on a Protean® II Xi apparatus (Biorad) with 16 cm x 1.0mm spacers (Biorad).

A 10% non-denaturing gel was prepared from 30% acrylamide (30%w/v Thiamine) as described by Sambrook *et al*. (1989) (see Appendix). The PCR products (4µl) were mixed with an equal volume of 6X loading buffer containing 0.25% xylene, 0.25% bromophenol blue, 15% Ficoll (type 400) in water (Sambrook *et al*., 1989). A 6µl aliquot of this mixture was loaded onto a 10% non-denaturing gel and electrophoresed.

The electrophoresis was performed at constant voltage of 200 V for approximately 4h (or until the bromophenol front had disappeared off the plate) in 1X TBE buffer (89mM Tris-borate, 2mM EDTA (pH8.0) with cooling system (temperature 5-10°C). The allelic bands were visualized after rapid silver staining (refer to 2.3.3) and scored. The scoring of alleles was done on the basis of their electrophoretic mobility in reference to a specific band on the allelic ladder (containing all the loci), as Ratio Factor (RF) technique mainly employed in paper chromatography.

2.3.3 Gel staining

DNA fragments were recorded following the silver staining method developed by Cairns and Murray (1994). This photochemically derived silver stain method of nucleic acids in polyacrylamide gels was originally described by Merril *et al.* (1981). The staining procedure consisted of an initial pre-stain fixation (10% ethanol, 0.5% acetic acid) of the gel for 3min followed by incubation in fixing solution containing 0.2% silver nitrate for 5min. The gel was rinsed in distilled water for a maximum of 2min to remove excess silver, and the reduction step was performed with a solution of 0.75M sodium hydroxide containing 0.5% formaldehyde for approximately 5min or until bands appeared. Formaldehyde selectively reduces silver ions to metallic silver under alkaline conditions. After sufficient appearance of bands, the development was stopped in the fixing solution for a maximum of 1hour. The gel was washed gently with distilled water under a shaker for 30min, soaked in glycerol for 1h and finally sealed into a plastic bag for further use or scanned on a computer and saved on a diskette.

2.3.4 Construction of allelic ladders

The protocol was based on that of Cairns and Murray (1994). Alleles identified as polymorphic at each locus were excised from the gel and placed into sterile labelled 1.5 ml microfuge tubes containing 100µl of elution buffer (50mM KCL, 10mM Tris-HCL, pH9.0 and 0.1% Triton X-100). The mixtures were vortexed, incubated at 95°C

for 20min and amplified separately. The ladder was constructed by mixing equal volumes of PCR products of genotypes.

2.4 DNA SEQUENCING: MICROSATELLITE ALLELES

Sequencing was completed to unambiguously establish the number of repeats between and among templates and to ascertain the reliability of the ratio factor (RF) scoring method uses during the course of this study.

Four microsatellite markers (AgXH25, Ag2H26, Ag3H88, and 33C1) were screened using a few individuals that exhibited different alleles (homozygotes) at particular loci on the non-denaturing polyacrylamide gel to determine their number of base-pair repeats. Only amplicons with clear bands on the gel with an estimated 10ng /μDNA or more after purification were carried on for sequencing. The origins of those specimens are known (see Table 4.1).

Amplification: Protocol described in section 2.4.1was used with the number of PCR-cycles increased to 40. The PCR products were resolved on a 2% low melting temperature agarose gel (NuSieve®GTG® agarose from Whitehead Scientific) containing 0.03mg/100ml ethidium bromide and electrophoresed in into 1X TAE buffer.

Purification of template DNA: Amplified fragments were purified using the NucleoSpin®Extract II kit (from Macherey-Nagel). The DNA fragment was excised carefully from the gel using a clean, sharp scalpel to minimise the volume of the gel. The gel was weighed and transferred into a clean 1.5ml centrifuge tube. The gel was lysed using three volumes of NT1 buffer per 1 volume of gel (300µl to each 100mg of gel) and incubated at 50ºC until the gel slices were completely dissolved (5-10min). The sample was briefly centrifuged (2min) to ensure a complete dissolving of the gel. The sample was then transferred in a NucleoSpin®Extract column placed into a 2ml collecting tube and centrifuged for 1min at 8000Xg, this step was to allow the binding of the DNA to the silica membrane. The flow-through was discarded and the NucleoSpin®Extract column was placed back into the collecting tube. The silica membrane was washed twice in NT3 buffer followed by 2min incubation at 70ºC prior to elution to ensure a completely silica membrane free ethanol. Since the NT3 buffer contains ethanol that has an inhibition power toward subsequent reactions for the first wash, 600µl NT3 buffer was added to the silica membrane and centrifuged for 1min at 11000Xg. The flow-through was discarded and the NucleoSpin®Extract column placed back into the collecting tube. For the second washing, 200µl NT3 buffer was added to the membrane and centrifuged for 2min at 11000Xg to remove buffer NT3 and allowed to dry. To elute the DNA, the NucleoSpin®Extract column was placed into a clean 1.5ml microcentrifuge tube and eluted using elution NE buffer (25ml)**.** The microcentrifuge tube was centrifuged for 1min at 11000Xg. Thereafter, the column was then discarded and the DNA directly used or preserved at -20°C until needed.

Preparation of sequencing products: The DyeEx Spin Protocol for Dye-Terminator Removal from QIAGEN (DyeEx™) was used to prepare the sequencing product.

In order to resuspend the resin, the spin column was gently vortexed. To ensure a free vacuum inside the spin column, the cap of the column was a quarter turned off. The bottom closure of the spin column was snapped off and the spin column was placed in a 2ml collection tube provided by the manufacturer. The spin column was centrifuged for 3min at 750Xg to bring down the resin, and then transferred (spin column) into a clean 1.5ml microcentrifuge tube. Twenty microlitre of the sequencing reaction was then pipetted directly onto the centre of the slanted gel-bed surface and centrifuged for 3min at 750Xg. Thereafter, the spin column was removed from the 1.5ml microcentrifuge tube containing the eluate (purified DNA). The sample was dried in a vacuum centrifuge for a maximum period of 10min, resuspended in 25µl of Template Suppression Reagent (TSR), vortexed and spun for few second. The template DNA was then heated at 95°C for 2min and chilled on ice to denature the DNA.

Sequencing of template DNA: Sequencing of selected alleles was performed by the dideoxynucleotide chain terminator (Sanger dideoxy method) using an ABI PRISM™ 310 version 2.0 sequencer (Applied Biosystems). Both strands were sequenced for each allele.

2.5 DATA ANALYSIS

2.5.1 Colony material

Preliminary studies of *An. arabiensis* colony material (Table 2.2) before the study of field material provide an estimate of microsatellite marker polymorphism. The analysis of each population included a computation of allele frequencies, observed and expected heterozygosity per locus based on Hardy Weinberg Equilibrium (HWE) using GENEPOP version 3.4 (Raymond and Rousset, 1995). Genetic Bottleneck (Cornuet and Luikart, 1996) and the frequency of null alleles (Chakraborty *et al* (1992) were also assessed.

2.5.1.1 Genetic Bottleneck and null alleles

Population bottleneck was assessed through the heterozygosity tests. The heterozygosity tests compare two estimates of expected heterozygosity, one based on allele frequencies (He) and the other based on the number of alleles and sample size (Heq) (Cornuet and Luikart, 1996). If **He>Heq**, (heterozygosity excess) this may indicate a recent bottleneck and conversely this could be an expansion event. In other words if a positive difference (heterozygote excess) is observed more often the number of loci for which there is a heterozygote excess will be significantly larger than half the number of polymorphic loci (mathematically express as **L>L/2**, with L=number of polymorphic loci) resulting in a so called "shifted mode" if not the "L-

shape" will be observed (Cornuet and Luikart, 1996). Estimates of He and Heq were calculated under the Stepwise Mutational Model (SMM) that is more reliable to microsatellite than the Infinite Allele Model (Valdes *et al.,* 1993). The SMM assumes mutations change the state of an allele by one step forward or backward with equal probability (Valdes *et al.,* 1993) Tests for population bottlenecks were performed using the computational program called Bottleneck version 1.2.02 (Cornet and Luikart, 1996).

The frequency of null alleles was calculated manually in respect to Chakraborty *et al* (1992) as follows: $\mathbf{r= (H_E-H_O)/ (H_E+H_O)}$ where H_E and H_O are respectively the expected and observed heterozygosity relative to HWE. This method assumes that all heterozygote deficiencies relative to HWE are due to null alleles and that none is due to deme subdivisions.

2.5.2 Wild caught material

Each locus was scored against allelic ladders and an arbitrary number was assigned. Data were analysed using GENEPOP version.3.4 (Raymond and Rousset, 1995) and ARLEQUIN software package version 2.000 (http//anthro.unige.ch/arlequin/) developed by Schneider *et al* (2000). Sequence alignments of contigs were analysed using SeqMan™ II software (DNASTAR Inc). The analysis of each population included a computation of allele frequencies, observed and expected Heterozygosity per locus at HWE and additional measures of genetic variability.

2.5.2.2 Gene diversity and Theta (Hom)

These indices were performed using ARLEQUIN version 2.000. The gene diversity (the expected heterozygosity under HWE) at a locus is defined as the probability that two randomly chosen haplotypes alleles are different in the population sample. It was estimated as $h_e = 1-\sum p_i^2$ where p_i is the sample frequency of the *i*-th haplotype or allele and the mean over *n* loci is $He = \sum he/n$ with variance $[h_e - H_e]^2 / [n(n-1)]$; the effective number of alleles n_e is $1/\sum p_i^2$ according to Nei (1987).

Theta (Hom) defined as the expected homozygosity in a population at equilibrium between drift and mutation was calculated under the pure stepwise mutation model (SMM) after Zouros's (1979) as $E(\theta H)=\theta(1+2(1+\theta) / (2+\theta)(3+\theta))$ and the standard error $s.d.(\theta H)\approx(2+\theta)2(3+\theta)2s.d.(H) / H2 (1+\theta)[(2+\theta)(3+\theta)(4+\theta)+10(2+\theta]$ where s.d.(H) is the standard error of the gene diversity and an unknown parameter θ

2.5.2.3 Genetic diversity

Hardy Weinberg Equilibrium (HWE) Each locus was tested separately for deviation from HWE using Fisher's exact test by applying the Markov chain method to determine the level of differences in allele frequencies existing among individuals in a population sample. The test was performed using GENEPOP version 3.4 (Raymond and Rousset, 1995). The sequential Bonferroni procedure (Holms, 1979) was applied

to evaluate the overall significance of the values obtained when multiple tests were performed.

2.5.2.4 Estimate of population differentiation and gene flow

Genetic differentiation between and within populations or subpopulations, one of the first steps in the analysis of gene flow, was estimated using Wright's F-statistic (F_{ST}), developed by Sewall Wright (Wright, 1965) calculated according to Weir and Cockerman (1984). The microsatellites' equivalent *R* statistics (R_{ST}) Slatkin (1995) were calculated according to Slatkin (1995). Estimates of F_{ST} and R_{ST} were computed using analysis of molecular variance AMOVA (Excoffier *et al.*, 1992) with 1000 permutations implemented in ARLEQUIN (Schneider *et al*., 2000). The AMOVA procedure is an analysis of variance framework with the significance of the variance components tested by permutation. The F_{ST} and R_{ST} were estimated using the absolute frequencies of different alleles and the sum of squared number of repeat differences respectively.

The F_{ST} is defined as the variance in allele frequencies standardized by the mean allele frequency among demes (F_{ST}=1 shows total differentiation that there are no heterozygote in the subpopulations while F_{ST}=0 shows no genetic differentiation between the constituent subpopulations). This genetic distance statistic (F_{ST}) uses the Infinite Allele Model (IAM) and implies that each mutation can create any new allele randomly.

The R_{ST} measures variance of mean allelic lengths and assumes a generalized Stepwise Mutation Model (SMM) with the possibility of mutation events involving more than one repeat unit and being independent of allelic size (Slatkin, 1995). The R_{ST} is more reliable than F_{ST} when analyzing microsatellite data (Slatkin, 1995). There are two sources of variation in the computed values of F_{ST} for a population under study. One is the variation due to the sample size and the other is due to differences amongst alleles and amongst loci (Slatkin, 1995).

Other *F*-statistics include F_{IS}, and $F_{SC.}$ F_{IS} was computed using GENEPOP version 3.4 (Raymond and Rousset, 1995) and the rest using AMOVA. The F_{IS} statistics (Raymond and Rousset, 1995) was used to test the distortion from HWE. F_{IS} provides a measure of the degree of inbreeding in individuals relative to the population to which they belong. F_{IS} values that do not differ significantly from zero imply that random mating occurs within populations. F_{IS} values were also computed from each site separately. Large positive value of F_{IS} is synonymous of the presence of null alleles. $\boldsymbol{F_{IS}}$ ($= \boldsymbol{H_S - H_I / H_S . H_S}$) is the average expected Heterozygosity estimate from each population by $\boldsymbol{H_S = 1 - \sum p^2_i}$ and H_I is the average observed Heterozygosity, $\boldsymbol{H_I = \sum H_i / k}$ for k subpopulations). $\boldsymbol{F_{SC}}$ ($= \sigma^2_b / \sigma^2_b + \sigma^2_c$) is the frequency among populations among group. Also with AMOVA, $\boldsymbol{F_{ST} = \sigma^2_a + \sigma^2_b / \sigma^2_T = S_B - S_W / S_B}$

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

$$\bar{S} = \frac{2n-1}{2nd_S - 1} S_W + \frac{2n(d_S - 1)}{2nd_S - 1} S_B.$$

$$S_W = \frac{1}{d_s} \sum_{j=1}^{d_s} \frac{2}{2n(2n-1)} \sum_{i<i'} (a_{ij} - a_{i'j})^2,$$

$$S_B = \frac{2}{(2n)^2 d_s (d_s - 1)} \sum_{j>j'} \sum_{i>i'} (a_{ij} - a_{i'j'})^2,$$

Where S_W and $\bar{S}$ are the average sum of squares of the difference in allele size between pair of gene within population and between pairs of genes taken from a collection of P populations, respectively (Saltkin, 1995) while S_B is the average squared difference in allele size between pairs of genes from different populations.

σ^2_a is the covariance component due to differences among the group populations; σ^2_b the covariance component due to differences among haplotypes in different populations within a group, σ^2_c is the sum of the covariance component due to differences among haplotypes within a population and σ^2_T the total molecular variance.

Migration rates (Nm) were estimated from the F_{ST} values using the relationship **Nm= ¼ (1/ F_{ST} – 1)**. N is the effective size of the population, and m is the proportion of population or deme that are migrants (Wright, 1978). The Nm values were also derived from R_{ST} by substituting R_{ST} for F_{ST} (Slatkin, 1995). R_{ST} provides less biased estimates of demographic parameters for deme than does the analogue F_{ST} (Slatkin, 1995). A reason for estimating Nm is that this combination of parameters indicates the level of gene flow and genetic drift. Nm >1shows no genetic differentiation and

thus high gene flow. For example if a population receives three migrants per generation (Nm =3) F_{ST} will equilibrate at 0.077.

Is there any relationship between the level of differentiation and geographic distance? Rousset (1997) proposed the following regression, $\boldsymbol{F_{ST}/ (1-F_{ST})}$ on geographic distance computed using ARLEQUIN software package version 2.000 (Schneider *et al* (2000). This relationship was also tested with the *R* statistic by replacing F_{ST} by R_{ST}. There is a linear relationship between the logarithm of geographic distance and $\boldsymbol{F_{ST}/ (1-F_{ST})}$**.** This formula assumes independence (genetic equilibrium and neutrality) between loci and an island model of migration (Slatkin, 1995) arises from the consideration of the probabilities of the so-called homoplasy (or identity by decent of two alleles) and the correlation of alleles within populations.

CHAPTER THREE

RESULTS

3.1 OPTIMIZATION OF MICROSATELLITE MARKERS USING PCR

The following PCR conditions were found to be suitable for microsatellite amplification (see Figs 3.1-3.4). A total of 25 µl reaction volumes containing 100X dilution of template DNA were carried out in 0.5ml eppendorf tubes. The reaction mixture contained 1X reaction buffer, 200mM of each dNTP, 1.5mM $MgCl_2$, 0.5 Units of DNA polymerase (Taq) and 12.5pmol of each primer. For amplification: an initial denaturation at 94º C for 2 min followed by 25 cycles of 15 s at 94º C, 15 s at 55º C, 15 s at 72º C and a final elongation step of 2 min at 72º C.

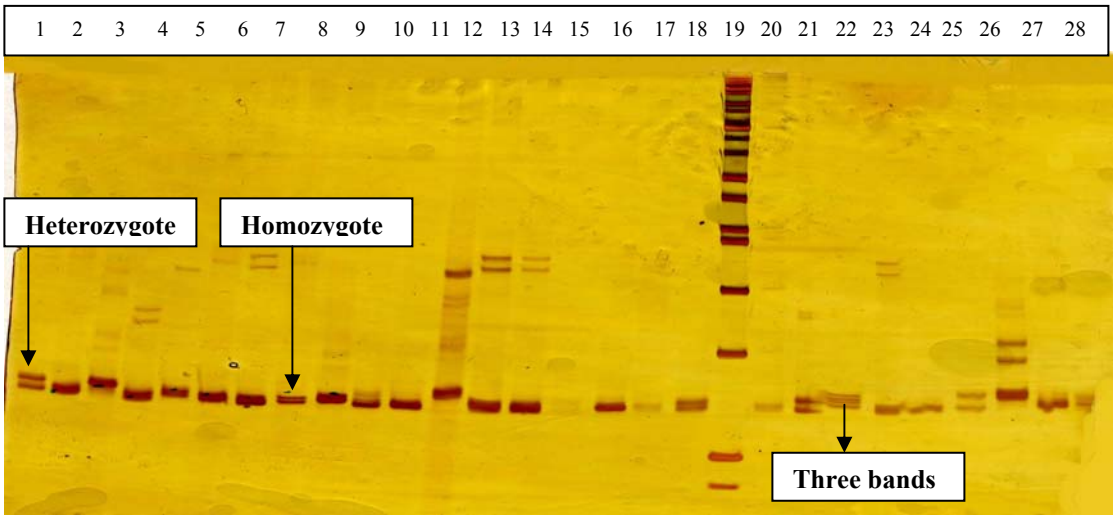


Figure 3.1 **Locus 33C1 tested on KGB and Nam. Lane 19= 1kb-plus size marker. The remaining lanes are extracted DNA with the presence of double and single bands. Lanes 1-16= NAM. 17-28= KGB. Lane 22= heterozygote with three banded pattern. No amplification on lane 15.**

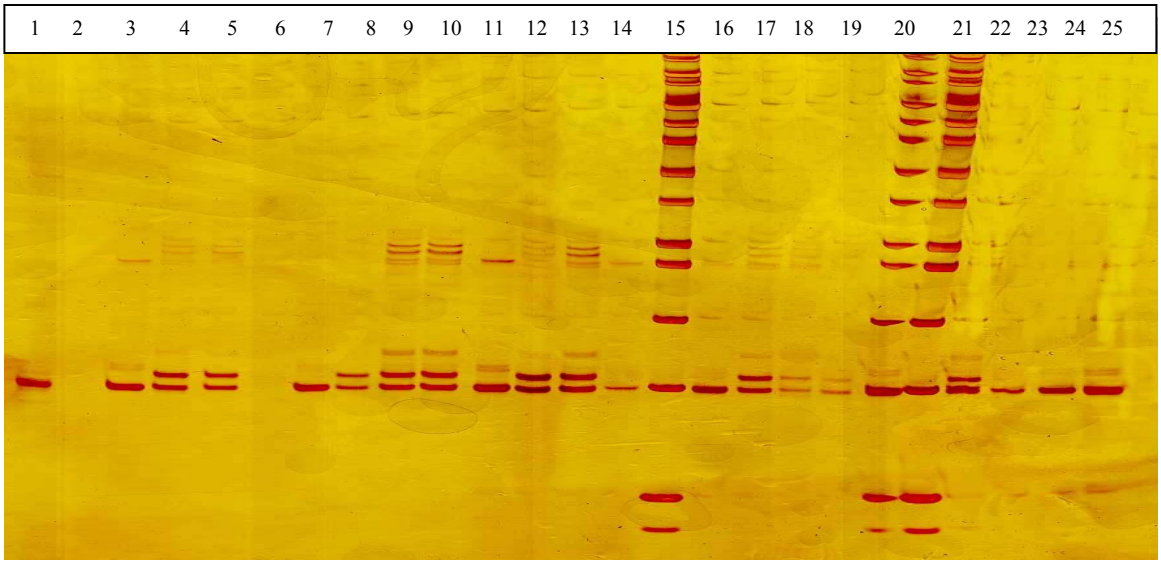


Figure 3.2 **Locus AG3H88. Lane 2=negative control. Lane 6= no amplification. Lanes 15, 20, 21=1 kb-plus. The remaining lanes are extracted DNA from MAN.**

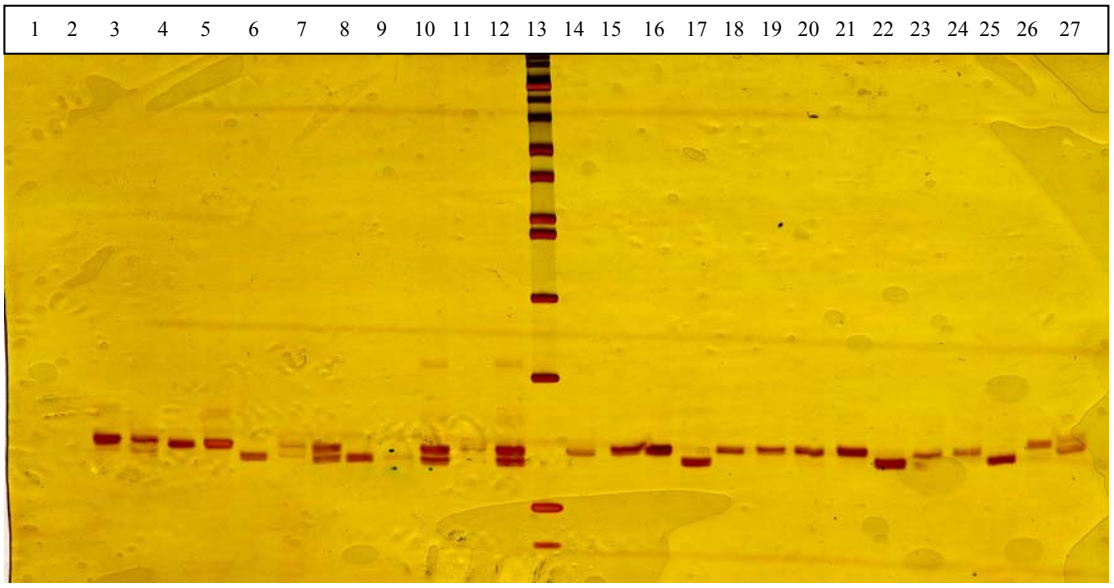


Figure 3.3 Locus AGXH25.Lane 1=negative control. Lane 13= 1kb-plus. Lane 2= no amplification. The remaining lanes are extracted DNA from NAM.

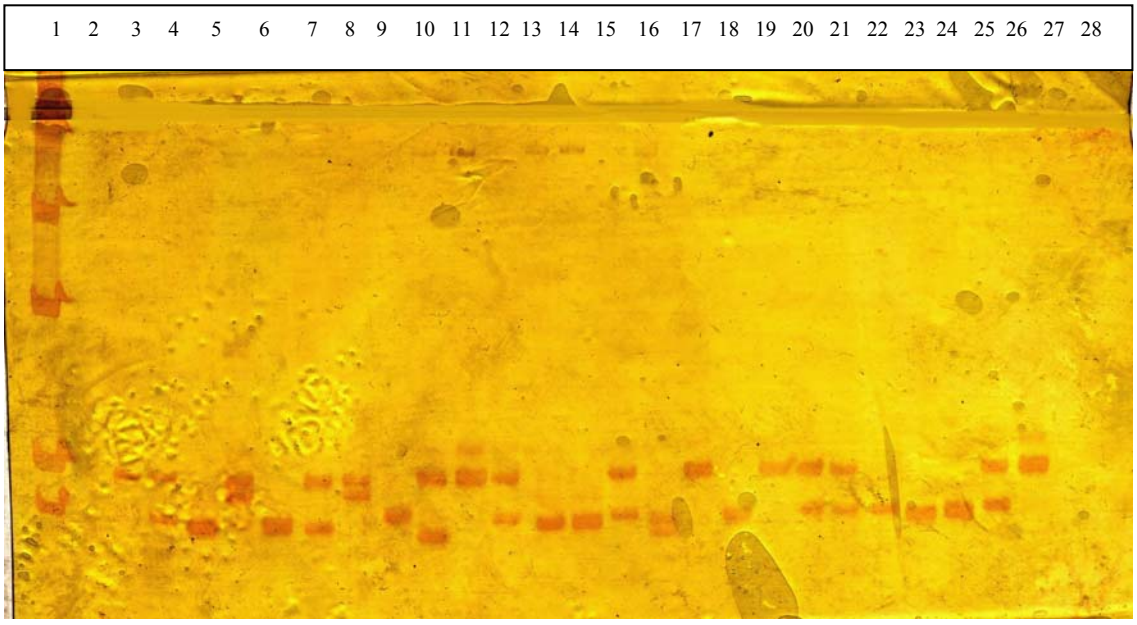


Figure 3.4 Locus AG2H26.Lane1= 1kb-plus marker. Lane2= negative control. Lane28= no amplification. The remaining lanes are extracted DNA from KGB.

3.1.1 Sequencing microsatellite alleles

Sequencing was done to confirm the scoring method used (ratio factor, RF).

A total of 32 sequencing reactions were carried out. Results of nine samples are shown in table 3.1.

Considerable variations were found in the number of repeats and therefore the length of each DNA fragment tested for each locus. The number of repeats was determined through sequencing. Unlike other markers, locus AG2H26 failed to produce "clean" sequences even after three independent sequencing attempts resulting in weak peaks with high background noise in all specimens analysed (see Walton *et al*., 1998).

Sequencing results showed phenomena of transitions (purine to pyrimedine) and inversions (e.g. G to C) within repeats compared with the original di or tri-nucleotide repeats on the *An. gambiae* markers (Zheng *et al.,* 1993, 1996). The allelic variation within species and between closely related species have been largely elucidated (Gillespies 2000, 2001). *Anopheles arabiensis* and *An. gambiae* are closely related, they are sympatric with a common evolutionary history, and the priming sites are conserved making comparisons of repeat sizes possible (Moore *et al*.1991).

Table 3.1 Sequence comparison between *An. arabiensis* and *An. gambiae*

	Locus			
	AG3H88	AGXH25	33C1	AG2H26#
*An.gambiae**	**(GT)** $_{7}$-	**(GT)** $_{7}$-	**(AGC)** $_{9}$-	**(GT)** $_{40}$
An.arabiensis				
Sample names				
Swa192	**(GT)** $_{12}$**,** 131	-	-	
Swa67	**(GT)** $_{11}$**,** 130	-	-	
Swa100	**(GT)** $_{14}$**,** 126			
Swa256	-	**(GT)** $_{7}$147	**(AGC)** $_{11}$**,** 149	
Nam	-	**(GT)** $_{11}$**,** 146	-	
KGB	-	**(GT)** $_{13}$**,** 144	-	
Ch1366	-	-	**(AGC)** $_{9}$**,** 150	
KZN100	-	-	**(AGC)** $_{7}$**,** 147	

Bold= nucleotide repeats. Underline= number of base-pair calculated from RF. # did not work. Data were analysed to estimate the effectiveness of the method used to score alleles/bands on the PAGE. *Adapted from Zheng *et al*., 1993, 1996 and Donnelly *et al.,* 2000).

3.1.3 Preliminary analysis of microsatellite polymorphism using colony material

A total of 150 mosquitoes were analysed (Table 2.2). All the four loci were polymorphic. The microsatellite PCR products appeared either as single bands (homozygote) or double bands (heterozygote) (Fig 3.1). Heterozygotes with three banded patterns were observed in the KGB colony (Loci 33C1, lane 22) (Fig3.1). This phenomenon was also reported by Lanzaro *et al.* (1995) in *An. gambiae s.s* in West Africa with homozygotes showing closely spaced doublet and heterozygotes with four-banded patterns who resolved it as a poly"A" tail artefact due to an addition of an extra "A" onto some, but not all of the products. The number of alleles ranged from 2-12 per locus per population with an average of 5.75 (Table 3.2) and a total average heterozygosity (H_T) of 0.24.

Table 3.2: **Genetic variability at all four microsatellite loci studied for the three *An. arabiensis* colonies at 95%CI (confidence interval).**

Populations	Locus AGXH25	AG2H26	AG3H88	33C1	Mean over populations
KGB (Kanyemba) (n=50)					
Nall	5	12	8	5	7.5
H_{obs}	0	10	6	3	4.75
H_{exp}	15.7	22.8	19.9	20.1	19.62
r	1.0	0.39	0.54	0.7	0.65
F_{IS}	+1	+0.57	+0.70	+0.85	+0.78
NAM (Namibia) (n=50)					
Nall	5	6	7	6	6
H_{obs}	5	9	8	4	6.5
H_{exp}	17.8	16.7	19.7	21.3	18.75
r	0.56	0.29	0.42	0.68	0.48
F_{IS}	+0.72	+0.47	+0.60	+0.81	+0.65
MAN (Swaziland) (n=50)					
Nall	3	7	7	2	4.75
H_{obs}	1	9	18	19	11.75
H_{exp}	22.9	16.7	22.4	14.0	19.0
r	0.91	0.29	0.12	-	nc
F_{IS}	+0.96	+0.47	+0.20	-0.36	+1.27
Mean over loci					
Nall	4.33	8.33	6	4.33	
r	**0.26**	**0.33**	**0.55**	**nc**	
F_{IS}	**+0.89**	**+0.41**	**+0.45**	**+0.43**	

- r =estimate of null alleles.
- N= sample size
- Nall= Number of alleles, H_{obs}=Heterozygosity observed, H_{exp}=Heterozygosity expected,
- (-) Heterozygosity excess, nc= not computed
- F_{IS} also known as **inbreeding coefficient** is the average over all subpopulations of the correlation between uniting gametes relative to those of their own subpopulation as formally defined by Wright (1968). **F_{IS}>0, alleles are said to be *identical in state* (autozygous). F_{IS}<0, alleles are said to be allozygous.**
-

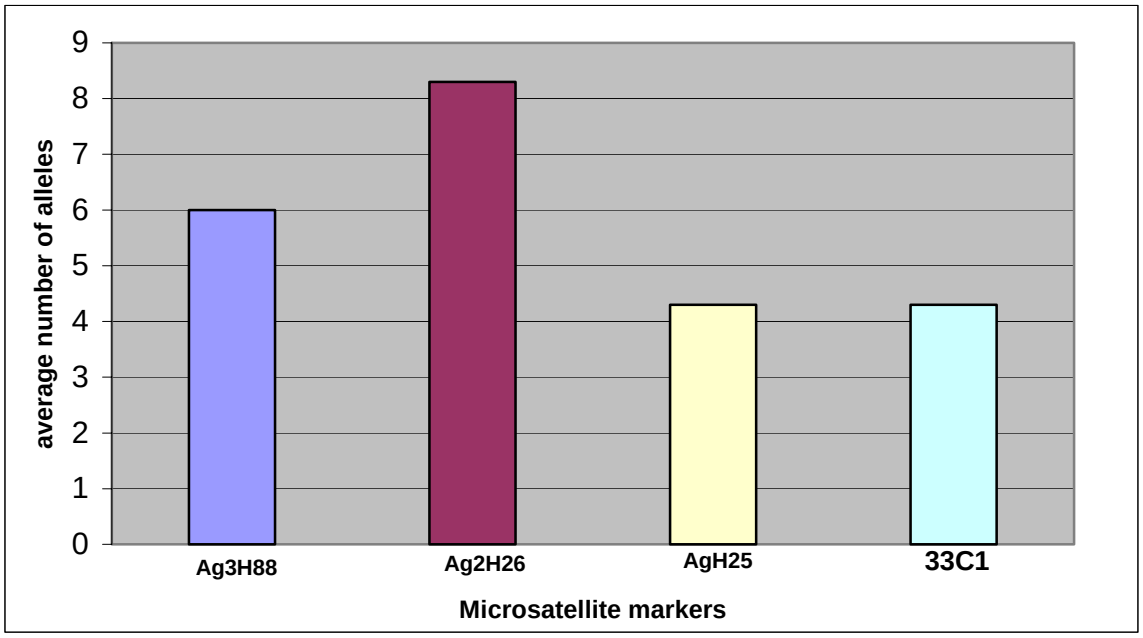


Figure 3.5 Distribution of alleles at four microsatellite loci in *An. arabiensis* colonies.

The number of alleles at each locus was significantly different. Moderate to high levels of polymorphism were found in all loci across all colonies. Locus AG2H26 with an average of 8.33 (from 6 to 12) alleles per colony was almost double that of AGXH25 and 33C1 with an average of 4.33 alleles. At locus AG3H88, the number of alleles ranged from 3-8 (average of 6 alleles per population). These alleles were all different in size and thus recognised as distinctive alleles (Excoffier *et al*., 1992). Alleles were distinguished from artefacts by the intensity and size of bands. And specimens showing faint bands were repeated before being scored as an allele or not. This was because alleles have been reported where there is inhibition of PCR reaction (Weber *et al.,* 1991).

The frequency of inbreeding was also determined to assess the heterogeneity between loci. Values of F_{IS} were consistently positive with a range of +0.06 -1 found in all four loci, although a value of -0.36 was found at locus 33C1, but in a single population, MAN (P>0.05). The frequency of null alleles (r) was estimated for each locus in each population (Table 3.2). The average frequency of null alleles was estimated to be 5.75%.

When data for all populations were examined there was a significant deviation from Hardy -Weinberg Equilibrium (HWE) at most loci (P<0.05) except in MAN which was in HWE. Many causes can be attributed to the departure from the HWE: effect of null alleles with a frequency above 5% (Chakraborty *et al.,* 1992, Callen *et al.,* 1993), the Wahlunds' effect, inbreeding or bottlenecks (Munstermann 1994, Lehmann *et al.,* 1996).

Null alleles refer to the reduction or complete loss of amplification of some alleles; they are locus specific, while the Wahlunds' effect and inbreeding have an affect on the entire genome. Bottlenecking is the consequence of drastic reduction in the numbers of individuals in a population or colony and results in loss of genetic variability (Nei, 1975).

The large and positive value of F_{IS} can be interpreted as the presence of null alleles (Raymond and Rousset, 1995). Whereas the other loci have a mixed elevated F_{IS} value in all populations, locus AGXH25 was the only locus to show very high values

for F_{IS} (values between 0.70 -1) in any population suggesting this locus has null alleles. Null alleles may fail to amplify due to the loss of a primer site or an inversion has caused the distance between primer sites to exceed the capacity of the PCR (Lynch and Milligan, 1994). They produce homozygote excess because an individual heterozygous for a null allele will be scored as a homozygote for the other amplified allele. In addition, a few individuals failed to amplify a PCR product principally at locus AG2H26 despite successful amplification at other loci, suggesting that they represent single band homozygotes. Also, the detection of no alleles could arise from technical problems that may include the staining sensitivity or the concentration of DNA.

Inbreeding by itself does not change the allele frequency if no selection process is applied (Hartl and Clark, 1997). As the colonies are maintained in the same insectary, this selection seems unlikely to happen. Therefore, heterozygote deficit could not be genome wide. It is possible that the deficit is due to the effect of a bottleneck (Hartl and Clark, 1997).

One of the common problems that affect laboratory colonies is the genetic bottleneck. Homozygote excess can be an indication that a recent population size reduction has occurred. Population bottleneck was assessed through the heterozygosity tests (available in Bottleneck Version 1.2.02). In my study only one population (NAM) showed a shifted mode or deviation from mutation-drift equilibrium (Table 3.3) but the probability was not significant ($P>0.05$).

Table 3.3 Estimate of heterozygosity tests at each laboratory population used

Mutation Model	NAM	MAN	KGB
SMM	**3**^{ns}	**3**^{ns}	**4**^{ns}
Mode-shift	Shifted	Normal L- shaped	Normal L-shaped

ns= non significant. Bold: number of locus with heterozygosity excess (He>Heq). SMM=Stepwise Mutation Model

3.1.4 Allelic ladder

An allelic ladder was constructed using the range of alleles of each locus from laboratory materials. There was an overlap between locus 33C1 and locus AGXH25 (Fig 3.6).

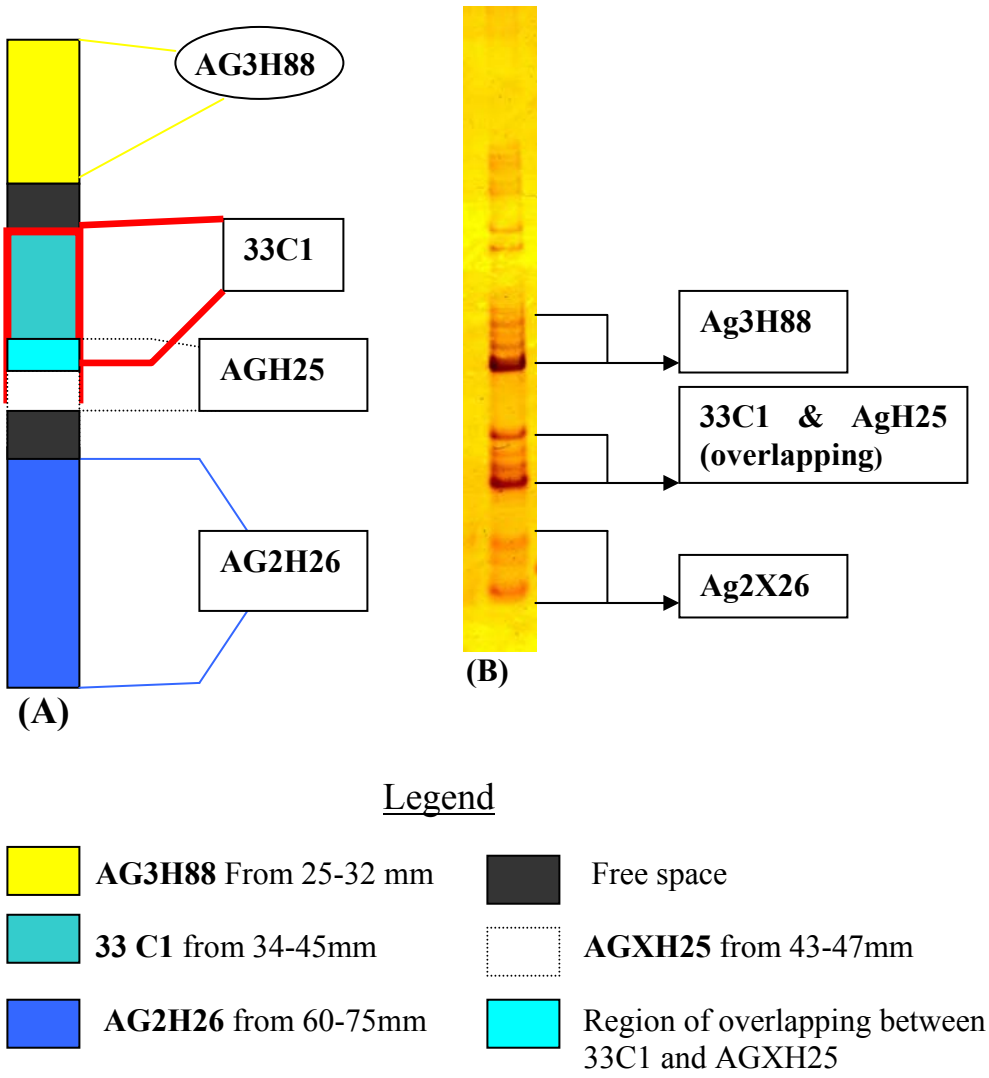


Figure 3.6 Allelic ladders. The alignment is as follows from the bottom to the top: Ag2H26-AgXH25-33C1-Ag3H88. (B)=Schematic approach. (A)= View on the PAGE.

3.2 TESTING OF FIELD CAUGHT SPECIMENS.

3.2.1 Samples Used

A total of 1225 specimens identified (Fig 3.7) as *An. arabiensis* were analysed. Localities, sample sizes and statistics are given in Table 3.4. Inclusion of the size marker in each running of field material (Figs 3.8-3.10) resulted in an estimation of the fragment size of each sample despite the inevitable variability of the gel caused by slight variation in the matrix pore size.

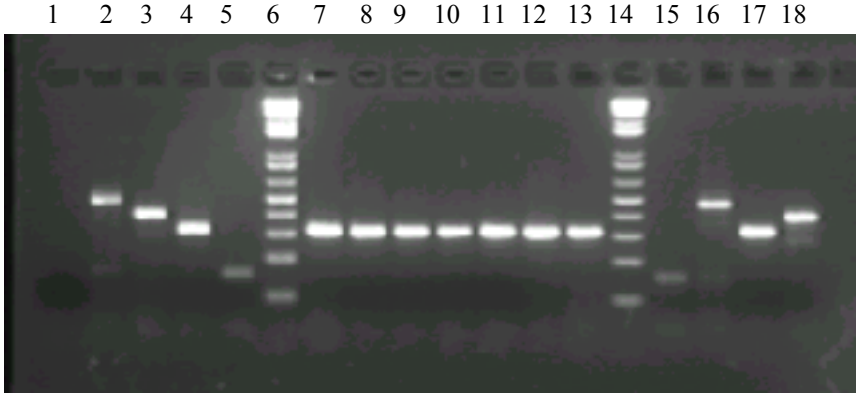


Figure 3.7 2% agarose gel of wild mosquitoes. Lane1=negative control, lanes 2-5 and 15-18= positive controls: *An.merus* (lanes 2 &16), *An. gambiae* (lanes 3 & 18), *An.arabiensis* (lanes 4 & 17) *and An. quadrianulatus* (lanes 5 & 15). Lanes 6 & 14= 1kb-plus ladder, lanes 7-13= field caught ***An. arabiensis.***

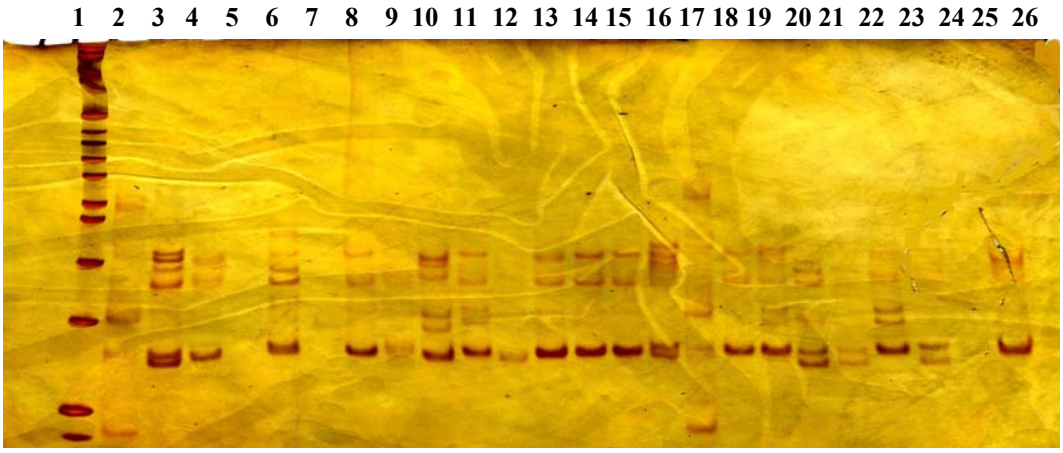


Figure 3.8 locus 33C1. Lane 1=negative control, lane 2=1kb-plus ladder, lanes 3 & 18= ladder (AG (XH25+2H26+3H88) +33C1) from colonised material (see Table2.4), the remaining lanes are samples from Okavango- Namibia with no amplification on lanes 6, 8 & 25.

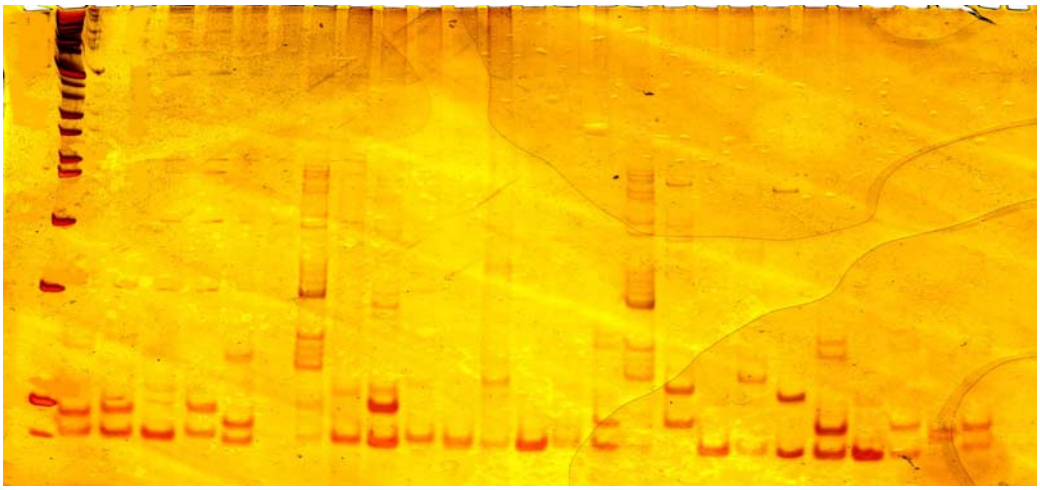


Figure 3.9 locus AG2H26. Lane1=1kb-plus ladder, lane 7=negative control, lanes 8&17= ladder (AG (XH25+2H26+3H88) +33C1) from colonised material (see Table2.4), the remaining lanes are samples from Gokwe-Zimbabwe.

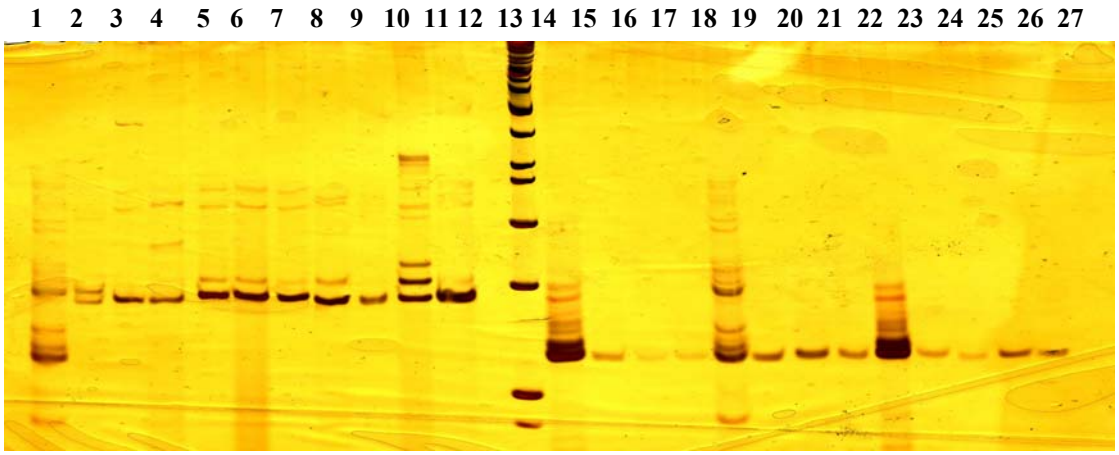


Figure 3.10 lanes 1 &18=ladder (AG (XH25+2H26+3H88) +33C1) from colonised mosquitoes (see Table2.4), lane 27=Negative control, lane 13=1kb-plus, lanes2-12=locus AG3H88, lane14-17&19-27=locus AGXH25, lane10=ladder from AG3H88, lanes14&22=ladder from AGXH25. Specimens were from KZN.

3.2.2 Allele Frequencies

Loci were considered polymorphic if they had more than one allele and all microsatellite loci tested here were polymorphic. Moderate to high levels of polymorphism were found in all loci across most populations (average number of 4.38 alleles (ranging from 1- –11) per locus per population (Table 3.4).

Allele frequency distributions were calculated from genotypes of individual mosquitoes. Figure 3.11 shows the allelic distribution in all the thirteen localities. In comparing all the populations, the number of alleles varied tremendously among

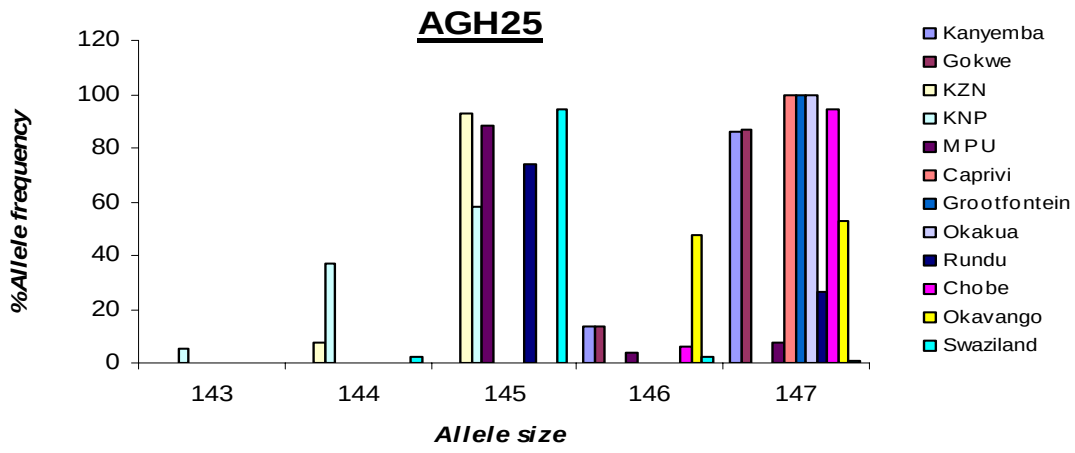
populations and also from locus to locus with two loci having identical or similar distributions (AG3H88 and AGXH25) while others had diverse patterns (AG2H26 and 33C1).

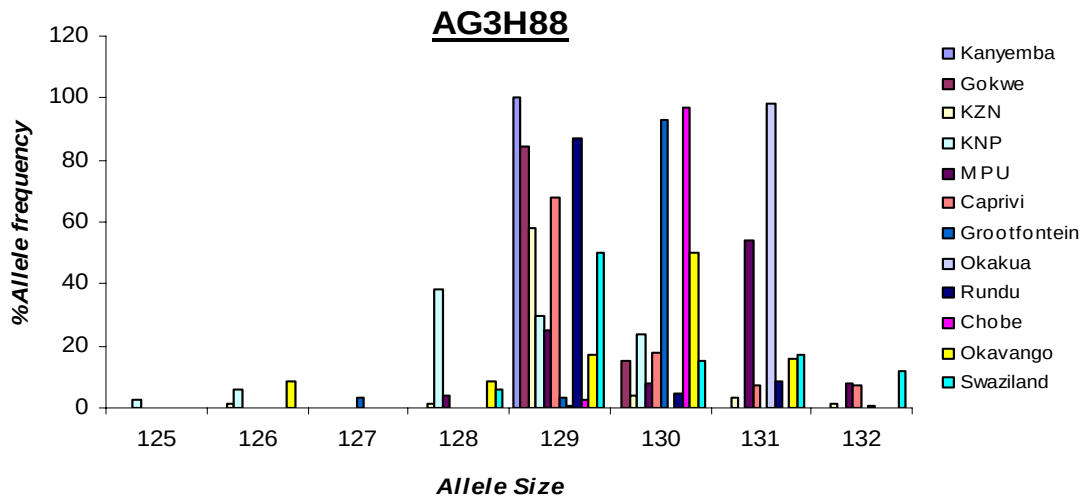
Table 3.4 Genetic diversity at four microsatellite loci in 13 subpopulations

		Locus				
		AG3H88	**AG2H26**	**AGXH25**	**33C1**	**Mean over subp**
POP	**Subpopulations**					
	Kanyemba (n=113)					
	Nall	1	5	2	4	3
	H_{Obs}	nc	15	0.0	2	5.7
	H_{Exp}	nc	35.8	16.8	3.9	14.3
	F_{IS}	nc	+0.58	+1.00	+0.49	+0.60
Zimbabwe						
	Gokwe (n=173)					
	Nall	3	11	2	5	7
	H_{Obs}	4	30	4	1	9.5
	H_{Exp}	3.6	40.6	5.6	19.8	17.4
	F_{IS}	-0.11*	+0.26	**+0.29**	+0.95	+0.46
	Chobe (n=176)					
	Nall	5	10	4	7	6.5
	H_{Obs}	18	48	11	33	27.5
	H_{Exp}	64.4	93.7	64.0	87.3	77.3
	F_{IS}	+0.72	+0.49	+0.83	+0.62	+0.65
	Okavango (n=62)					
Botswana	Nall	5	7	2	4	4.5
	H_{Obs}	6	12	3	3	6
	H_{Exp}	15.7	21.4	1.8	24.7	15.9
	F_{IS}	+0.62	+0.44	-1.00*	+0.88	+0.23
	Tutume (n=54)					
	Nall	4	3	4	7	4.5
	H_{Obs}	6	8	3	16	8.25
	H_{Exp}	15.7	15.3	13.5	17.7	15.5
	F_{IS}	+0.65	+0.48	+0.78	+0.09	+0.50
	Caprivi (n=31)					
	Nall	4	2	1	2	2.25
	H_{Obs}	6	17	nc	17	10
	H_{Exp}	7	9	nc	9	6.43

	F_{IS}	**+0.17**	-1.00	nc	-1.00	-0.89
	Okakua (n=118)					
	Nall	3	3	1	2	2.75
	H_{Obs}	2	44	nc	42	22
Namibia	H_{Exp}	3	24.6	nc	25	13.3
	F_{IS}	+0.49	-0.80	nc	-0.67	-4.35
	Grootfontein (n=57)					
	Nall	3	5	1	5	3.5
	H_{Obs}	2	62	nc	77	35.25
	H_{Exp}	44	43	nc	59	36.6
	F_{IS}	+0.95	-0.43	nc	-0.32	+0.04
	Rundu (n=26)					
	Nall	3	2	2	5	2.5
	H_{Obs}	0.0	17	10	16	10.75
	H_{Exp}	8.6	10.4	7.8	15.4	10.47
	F_{IS}	+1.00	-0.67	**-0.39**	-0.03	-0.15
	Big-Bend (n=180)					
	Nall	5	11	4	5	6.25
	H_{Obs}	7	12	1	17	9.25
Swaziland	H_{Exp}	35.4	43.3	4.8	16.1	24.92
	F_{IS}	+0.80	+0.72	+0.79	-0.06	+0.63
	Mpumalanga (n=30)					
	Nall	5	2	3	5	3.75
	H_{Obs}	5	0	3	3	5.5
	H_{Exp}	8.3	3.3	2.8	4.4	4.7
	F_{IS}	+0.41	+1.00	-0.05*	**+0.33**	+0.42
	KNP (n=33)					
South	Nall	5	6	3	2	4.75
Africa	H_{Obs}	1	5	0	2	2
	H_{Exp}	13.5	17	10	11.8	13.1
	F_{IS}	+0.92	+0.71	+1.0 0	+0.83	+0.85
	KZN (n=172)					
	Nall	6	11	2	8	6
	H_{Obs}	5	24	4	22	13.75
	H_{Exp}	31	66	3	37	34.3
	F_{IS}	+0.84	+0.64	-0.43	+0.41	+0.36
	Mean over loci					
	Nall	4.15	5.8	2.8	4.76	
	H_{Obs}	54	294	39	251	
	H_{Exp}	101.7	421	146.5	202.1	
	F_{IS}	+0.49	+0.31	+0.73	-0.24	

POP=populations, Nall=number of alleles per locus, HObs= observed heterozygosity and HExp= expected heterozygosity (Nei, 1978), *FIS* is calculated according to Weir and Cockerman (1984). Bold underlined values: P<0.005 after Bonferroni's correction for multiple tests * values: (P>0.05). nc= not computed, because of single allele at those loci.


AGH25
%Allele frequency
120
100
80
60
40
20
0
143
144
145
146
147
Allele size
Kanyemba
Gokwe
KZN
KNP
MPU
Caprivi
Grootfontein
Okakua
Rundu
Chobe
Okavango
Swaziland


AG3H88
%Allele frequency
120
100
80
60
40
20
0
125
126
127
128
129
130
131
132
Allele Size
Kanyemba
Gokwe
KZN
KNP
MPU
Caprivi
Grootfontein
Okakua
Rundu
Chobe
Okavango
Swaziland

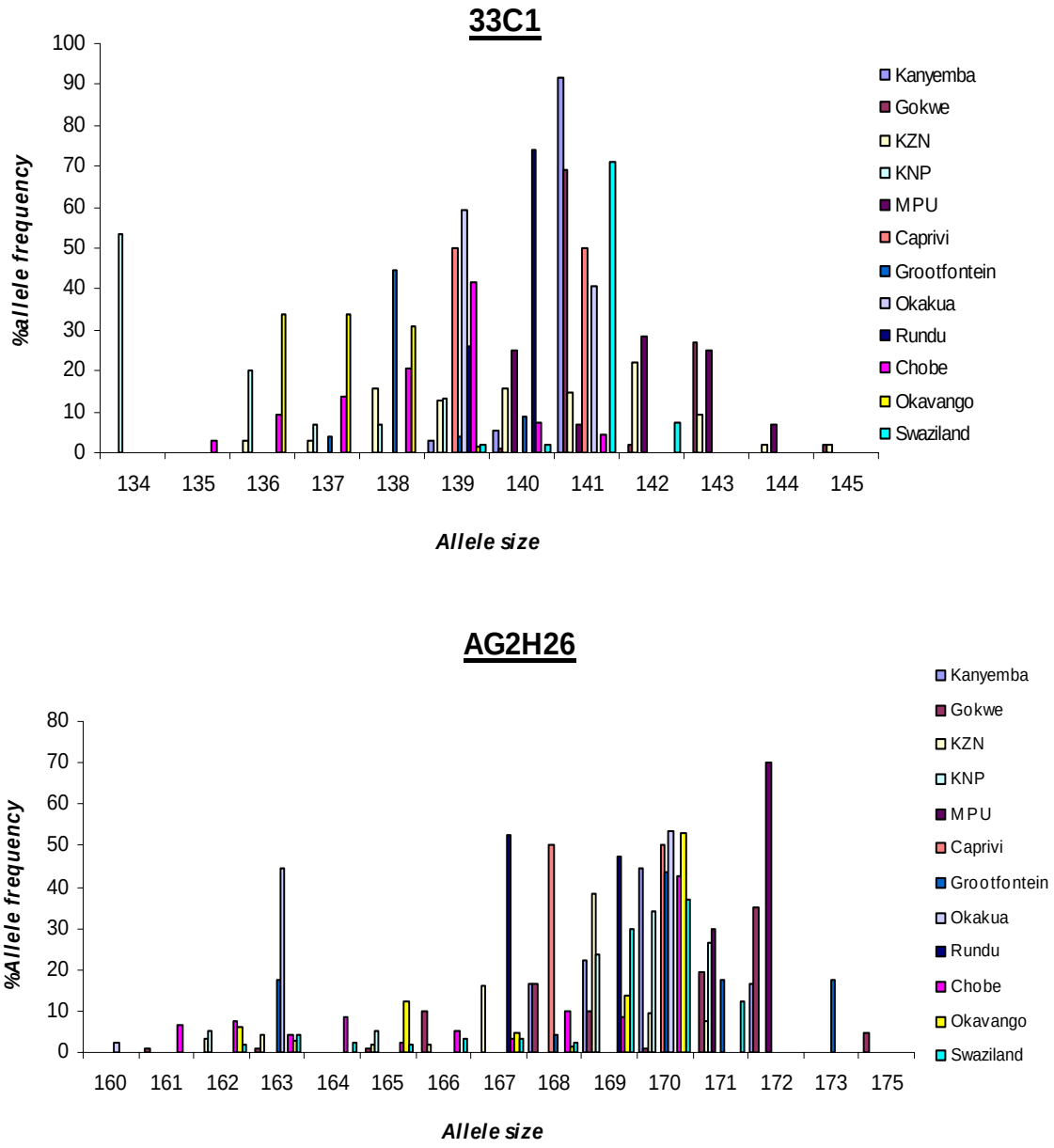


Figure 3.11 Allele frequency distribution in 13 subpopulations of *An. arabiensis* at four microsatellite locus. Alleles are denoted by their estimated fragment (allele) size in base pair.

3.2.3 Hardy-Weinberg Equilibrium

Four subpopulations, Caprivi, Okakua and Grootfontein (locus AGXH25) and Kanyemba (locus AG3H88) were monomorphic. All samples showed a significant departure from Hardy -Weinberg Equilibrium (HWE) for at least two loci ($P<0.05$) even after the sequential Bonferroni procedure was applied (to take into account the number of tests within each subpopulation). Three subpopulations conformed to HWE (Okavango and Mpumalanga at locus AGXH25 and Gokwe at locus AG3H88, $P>0.05$) (Table 3.4).

All significant tests were associated with either negative or positive F_{IS} for HWE with a total average of 0.35 (range -0.11-1). In most cases positive F_{IS} values indicated the reduction of heterozygotes. In cases where departure from HWE was observed, it was due to deficiency of heterozygotes as seen by positive F_{IS} values (e.g. Kanyemba with F_{IS}=+1 at locus AGXH25, P=0.00). From this positive F_{IS} value we can suggest the presence of null alleles or other bias in sample collections or non random mating. In other cases, departure from HWE was due to excess of heterozygotes as seen by negative F_{IS} values (e.g. Okakua with F_{IS}=-0.80 at locus AG2H26, $P<0.05$) (Table 3.4). Absence of PCR product was assumed to be caused by the presence of two null alleles at a given locus.

3.2.4 Molecular diversity within subpopulations

Gene diversity which is equivalent to average population gene heterozygosity (Nei, 1987) and the Theta which is equivalent to the expected homozygosity in a population at equilibrium between drift and mutation (Zouros, 1979) were high and comparable for all subpopulations (Table 3.5).

In comparing all the subpopulations, Kanyemba showed the least heterozygosity (0.69±0.02SE) with Theta (4.946±0.42SE), the highest heterozygosity was recorded in Swaziland (0.88 SE±0.00) and Okavango (0.88 SE±0.01) with Theta (34.59 ±0.81SE, 34.30±0.98SE respectively).

Table 3.5 Molecular diversity at the subpopulation level

Subpopulations	Gene diversity	Theta (Homo)	Countries
Big-Bend	0.881	34.591	**Swaziland**
Gokwe	0.843	19.773	**Zimbabwe**
Kanyemba	0.697	4.946	
Chobe	0.789	10.829	
Okavango	0.880	34.307	
Tutume	0.838	18.635	**Botswana**
Rundu	0.849	21.531	
Okakua	0.723	6.037	
Caprivi	0.776	9.440	**Namibia**
Grootfontein	0.855	23.177	
KNP	0.823	15.511	
KZN	0.869	28.534	**South Africa**
MP	0.832	17.139	

3.2.5 Intra-population estimates of differentiation

In the following analysis, subpopulations were pooled into single populations, i.e. Namibia was a group of four subpopulations (Rundu, Caprivi, Okakua and Grootfontein) (Refer map Fig 1.3). AMOVA analysis showed higher percentage of variation within than among subpopulations (Table 3.6) suggesting that subpopulations can be grouped together into populations, example of Zimbabwe with 95.44% of genetic variation found within subpopulations and 4.66% among subpopulations (F_{ST}= 0.065) An average of 85.74% of genetic variation was found within subpopulations, whereas 14.29% was attributed to differences among subpopulations. Swaziland was excluded from the analysis because it was constituted by a single subpopulation (degree of freedom =0). More variation was found among or between subpopulations in Namibia (7.65%) and South Africa (16.6%) than in Zimbabwe (4.66%) but subpopulations still can be grouped together.

Table 3.6 Weir and Cockerham (1984) estimates of sum of square size difference values for variation among and within subpopulations using AMOVA.

Source of Var	Populations				
	South Africa	Zimbabwe	Namibia	Botswana	All
%Var. among subp	16.60	4.66	7.65	28.25	14.29
%Var. within subp	83.40	95.44	92.35	71.75	85.74
R_{ST}	0.17	0.05	0.08	0.28	0.15

Var=Variation, subp=subpopulation, Number of permutation=1023, P=0.000

3.2.6 Genetic relationships between subpopulations

3.2.6.1 F_{ST} and R_{ST} statistic values

Estimates of pairwise F_{ST} and R_{ST} values across all loci between populations are compared (Table 3.7). The R_{ST} values obtained were low (0.00-0.39) compared to F_{ST} (the lowest value being 0.09) and non- significant (P>0.05) in almost all comparisons. The F statistic (F_{ST}), a measure of the amount of differentiation among populations showed a mean value of 0.180 (R_{ST}=0.174) (P<0.05). And the mean index of fixation of individuals relative to all subpopulations (F_{IS}) values of 0.026. The genetic differentiation among subpopulations within populations showed F_{SC} values of 0.158 (R_{SC}= 0.148).

The genetic structure was assessed using the Neighbour Joining (NJ) tree (Kumar *et al.*, 2002) (Fig 3.12) after the R_{ST} values of pairwise genetic distance matrix (Table 3.7). The thirteen subpopulations were grouped into two clusters. KNP, Chobe and Rundu grouped together and the remaining subpopulations forming the second cluster.

Table 3.7 Population pairwise genetic distances between subpopulations used in this study.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
1		0.350	0.273	0.208	0.253	0.289	0.030	0.388	0.251	0.121	0.388	0.453	0.449
2	0.182		0.108	0098	0.099	0.088	0.257	0.109	0.127	0.141	0.110	0.047	0.111
3	0.158	0.093		-0.002	0.013	-0.000	0.219	0.048	0.008	0.042	0.011	0.134	0.076
4	0.122	0.156	0.095		0.006	0.009	0.159	0.066	0.006	0.010	0.036	0.157	0.114
5	0.131	0.129	0.111	0.087		0.006	0.177	0.076	0.045	0.062	0.020	0.168	0.107
6	0.181	0.161	0.124	0.074	0.124		0.226	0.038	0.027	0.071	-0.001	0.110	0.053
7	0.250	0.333	0.296	0.282	0.328	0.347		0.336	0.221	0.112	0.311	0.410	0.399
8	0.293	0.171	0.133	0.119	0.162	0.129	0.430		0.045	0.144	0.026	0.073	0.016
9	0.203	0.169	0.117	0.033	0.119	0.099	0.336	0.066		0.034	0.046	0.135	0.093
10	0.163	0.124	0.096	0.127	0.152	0.174	0.238	0.198	0.149		0.121	0.220	0.206
11	0.253	0.145	0.095	0.147	0.167	0.114	0.315	0.117	0.141	0.102		0.117	0.026
12	0.291	0.091	0.127	0.185	0.213	0.184	0.384	0.114	0.161	0.176	0.142		0.053
13	0.298	0.115	0.109	0.168	0.191	0.155	0.383	0.077	0.127	0.141	0.064	0.072	

***F*ST values in the lower left-hand matrix and *R*ST *values* in the upper right-hand matrix.** Underline are the pairs of population with P>0.05 based on 110 permutation. Bold are populations: 1=Chobe, 2= Okavango, 3=Tutume, 4= Caprivi, 5=Grootfontein, 6=Okakua, 7=Rundu, 8=Gokwe, 9=Kanyemba, 10=Kruger National Park (KNP), 11=Mpumalanga, 12=KwaZuluNatal, 13=Swaziland.

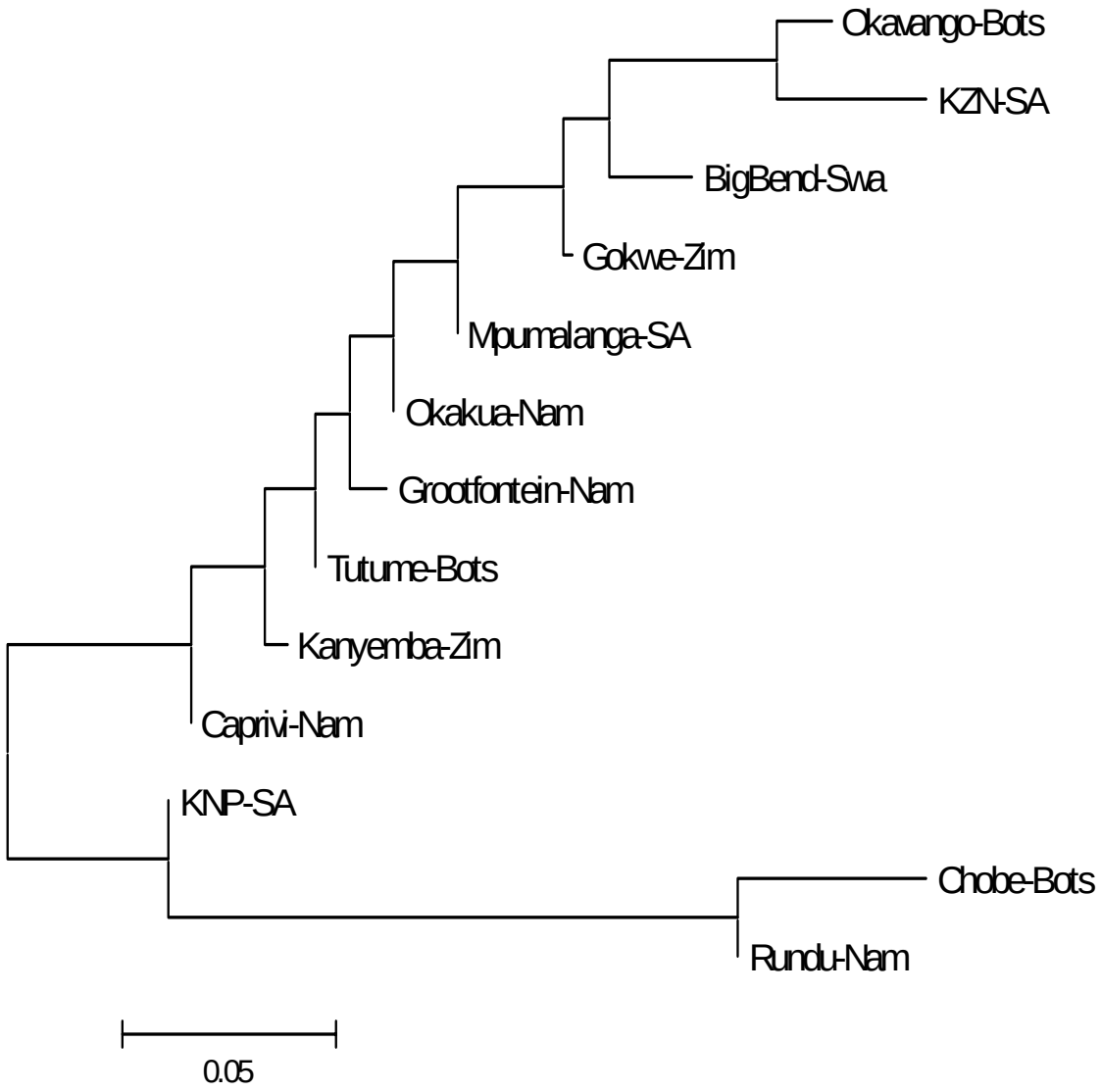


Figure 3.12 **NJ tree for *Anopheles arabiensis* subpopulations (generated by *MEGA* version 2.2) using the pairwise genetic distances (R_{ST}) of each subpopulation (Table 3.7).**

3.2.6.2 AMOVA analysis

Samples from South Africa, Zimbabwe, Botswana and Namibia were subjected to AMOVA analysis in which each pair of populations was considered (Tables 3.8). In each pairwise combination of populations (or for each analysis) less variation was found among populations than variation among subpopulations within populations. Therefore all populations are quite similar to each other and can be grouped; this however is seen by the total values (2.96 versus 14.44).

Table 3.8 Percentage variation and genetic distance within and among *An. arabiensis* populations using AMOVA (1023 permutations tested)

Source of variation	Zim-SA	Nam-Zim	Bots-SA	Nam-SA	Bots-Nam	Bots-Zim	All
Among pop (%)	-0.71	1.61	6.92	2.52	-0.29	7.96	2.96
Among subpop within pop (%)	9.58	6.19	22.55	10.83	18.33	16.43	14.44
Within pop (%)	86.29	92.20	78.52	86.29	81.96	75.93	82.60
R_{ST}	0.088	0.078	0.295	0.133	0.180	0.241	0.174
R_{SC}	0.095	0.063	0.242	0.111	0.095	0.177	0.148
F_{ST}	0.137	0.155	0.215	0.197	0.169	0.219	0.180

Sample size: Zim=286, SA=235, Nam=232, Bots=292; All means 13 populations including Swaziland=1225.

3.2.7 Isolation by distance

Isolation by distance was tested. No positive correlation (linear regression) was found between genetic variation (pairwise F_{ST}/ (1-F_{ST}) and R_{ST}/ (1-R_{ST}) and geographic distance ($P<0.05$) when comparing one subpopulation versus another (refer to table 3.7).

3.2.8 Estimates of gene flow (Nm)

The number of migrants per population per generation (Nm) calculated from F_{ST} and R_{ST} were compared (Table 3.9). The average migration rate Nm with a value of 1.17 (range from 0.64-5.31(R_{ST}) and from 1.05-3.59 (F_{ST})) across all populations indicated low gene flow (Table 3.9). Negligible gene flow was also observed between populations (Table 3.9). In their study in West Africa, Taylor *et al.* (2001) found a relatively low gene flow between species across the border

At the population level, estimates of gene flow (Nm) in Zimbabwe calculated from F_{ST} and R_{ST} statistics were relatively high, 3.59 and 5.31 respectively, suggests that gene flow is only minimally restricted between Kanyemba and Gokwe subpopulations. In Namibia estimates of Nm gave converse results using both F_{ST} (=1.05) and R_{ST} (=3.09). However this value of Nm (R_{ST}) implies high level of gene flow. F_{ST}, which measures variation across the allele array, is more sensitive to

variation in allele frequencies across the array than is R_{ST} which may give undue weight to alleles (Wright, 1965; Slatkin, 1995).

Table 3.9 Nm values (in bracket) for microsatellite loci among *An. arabiensis* populations in different regions.

Populations	R_{ST} (Nm)	F_{ST} (Nm)	Populations	R_{ST} (Nm)	F_{ST} (Nm)
Namibia	0.076 (3.09)	0.192 (1.05)	Bots-SA	0.295 (0.59)	0.215 (0.91)
Botswana	0.282 (0.64)	0.163 (1.28)	Nam-Zim	0.078 (2.95)	0.155 (1.36)
South Africa	0.166 (1.26)	0.152 (1.39)	Nam-SA	0.133 (1.63)	0.155 (1.36)
Zimbabwe	0.045 (5.31)	0.065 (3.59)	Zim-SA	0.088 (2.59)	0.137 (1.57)
All	**0.174 (1.19)**	**0.180 (1.14)**	Bots-Nam	0.180 (1.14)	0.169 (1.23)

CHAPTER FOUR

GENERAL DISCUSSION

There were three objectives of this study: (1) optimization of the microsatellite loci using *An. arabiensis* colonized in the Botha De Meillon insectary; (2) measure genetic variability within and between populations or subpopulations ; and (3) estimation of gene flow in wild caught *An. arabiensis* from five southern African countries.

4.1 OPTIMIZATION OF THE MICROSATELLITE-PCR USING COLONISED *AN. ARABIENSIS*

The work presented here aimed to optimize reaction conditions for each of the four microsatellite DNA markers as well as to evaluate the level of polymorphism within each *An. arabiensis* colony used. Laboratory material was used as a standard in comparing other populations.

All microsatellite loci tested were polymorphic. The number of alleles at locus AG2H26 was high with an average of almost twice that of AGXH25 and 33C1. Locus AG2H26 is inside the 2Rb polymorphic chromosomal inversion (Zheng *et al.*, 1996, Petrarca *et al.*, 2000) and therefore selection pressure may be influencing

variation at this locus. All colonies showed massive heterozygote deficiency for marker AGXH25 with a high F_{IS} average of +0.89 (from -0.36 to1). Since it is on the X chromosome, it is possible that it is trapped in an inversion or linked to some sex-specific gene that is disadvantageous at the heterozygote state. The deficiency of heterozygotes in colony material could also be attributable to the artificial conditions in which these mosquitoes are kept and little is known of the selection pressures exerted on them.

The frequency of null alleles is proportional to the heterozygote deficiency because an individual heterozygous for a null allele will be scored as homozygote for the other amplified allele. Since the same pattern of null alleles was not expressed at all the four loci in a single colony, the null allele effect cannot be totally responsible for the low heterozygosity observed in the sense of HWE (Chakraborty *et al.*, 1992).

The average heterozygosity increases slower than the average number of alleles per locus after a bottleneck (Nei *et al.,* 1975) and since laboratory colonies regularly undergo bottlenecking; this may be the cause of the heterozygote deficiency (Van Rensburg, 1997). There is no apparent reason to postulate non-random mating of mosquitoes in a standard cage.

However, Pinto *et al.* (2004) did not find any consistency between the heterozygosity tests and a recent bottleneck in colonies kept in the same environment. They concluded that the bottleneck was so recent that not enough generations had passed to

eliminate the effect. Precision tests for detecting a bottleneck (founder effect) as being responsible for heterozygosity excess/deficiency, require a minimum of 20 microsatellite loci and is generation dependent (Cornuet and Luikart, 1996).

We can speculate that the manual method of the vertical PAGE used in this study is not sensitive enough to pick up heterozygotes of alleles that have very similar sizes, maybe one or two base-pair differences. This problem could be resolved using an Automated Sequencer for running the PCR product.

4.2 FIELD CAUGHT MATERIAL

In this study, 13 subpopulations of *An. arabiensis* from five countries were investigated for genetic structure and gene flow using four microsatellite DNA markers described for *An. gambiae.* Locus AGXH25 displayed a low mean number of alleles and possibly low sensitivity which may be the result of selection pressure at the X-chromosome and also the power of the analysis used (Donnelly and Townson, 2000). Genotypic frequencies in all four populations did not conform to Hardy-Weinberg expectation. The number of alleles or subpopulations that did not conform was greater than that by chance alone, indicating violation of some Hardy-Weinberg equilibrium assumptions. These deviations may have been caused by the Wahlund effect (pooling of separate genes), resulting in excess of homozygotes (F_{IS}>0) at all loci. Mosquitoes were collected randomly at different time periods and this may also be responsible for deviations from HWE (Simard *et al.,* 2000). The presence of null

alleles is another possible cause of excess homozygotes (Chakraborty *et al.,* 1992). In 1975, Nagylaki pointed out that natural selection can favour alleles that confer adaptation to local conditions in one population but not in others.

The level of gene diversity was high in subpopulations as indicated by high estimates of differentiation between subpopulations and a low estimate of average migration rate (Nm<3) with both F_{ST} and R_{ST}. There was no detectable correlation between the average F_{ST} and R_{ST} values across all loci or populations. Gene flow estimated by Nm was low when all data were pooled (Nm (F_{ST}) =1.14, Nm (R_{ST}) = 1.19). This can be explained by the existence of geographic barriers such as rivers and mountains. There was no significant linear regression between the genetic distance measures and the geographic distance (assumed to approximate to road distance). This was also concluded by Donnelly and Townson (2000) in a study carried out in East Africa.

Low estimates of genetic differentiation between the Zimbabwean subpopulations (Gokwe and Kanyemba) separated by approximately 310km (F_{ST} =0.065 and R_{ST}=0.045) suggested extensive gene flow (Nm=3.59 (F_{ST}) and 5.31 (R_{ST})). However, we would argue that these results are more likely to reflect large effective population sizes. Similar results were recorded in Senegal by Simard *et al.* (2000) although their study was done during the severe dry conditions. The low gene flow detected in Botswana (Nm=0.64 (R_{ST})) may reflect the distances between northern and central localities sampled in this study (>600km) thought four loci coupled with low allelic diversity may limit the ability to detect distance effects as well the possibility that the

gene pools were limited because of the larval sampling (e.g. many larvae in a single pool of water would have come from the same parents and therefore represent only two wild genomes). This could be tested using mitochondrial DNA. However this was outside of the scope of this study.

CHAPTER FIVE

GENERAL CONCLUSION

Despite the fact that all microsatellite loci were isolated from *An. gambiae,* all markers were proven to be adequate for use on field specimens of *An. arabiensis* from southern Africa. Although there are obvious constraints on their use, colonies remain the principal source of material for basic research and for reference standards.

Based on the differentiation indices, low gene flow was observed in the 13 subpopulations as a whole. This may be the result of random collecting methods, or recent or historical events rather than ongoing gene flow. However, gene flow (Nm>4) was observed in Zimbabwe between Kanyemba and Gokwe. This requires special attention to prevent the spread of insecticide resistance detected in Gokwe (H. Masendu, unpublished data)

This is the first study to be carried out in southern Africa and at this point there is insufficient information to explain the level of genetic differentiation of *An. arabiensis* among five populations of southern Africa. Extensive studies on population structure of this vector in the region using additional microsatellite loci and more knowledge on factors that influence gene flow are needed.

APPENDIX

Preparation of non-denaturing gel

Chemicals and solutions

- **30% Acrylamide**

  Into 100ml ddH_2O, 29g acrylamide powder and 1g N, N'-methylenebisacrylamide were added and heated to 37^oC to dissolve the chemicals.

- **1 X TBE**

  It was obtained from 5X TBE by dilution.

- **5 X TBE for electrophoresis**

  Prepared by mixing **54g** Tris base, 27.5g boric acid and 4.65g of Disodium Ethylene Diamine Tetra-Acetate (Na_2EDTA) dissolved in ddH_2O and brought up to a litre.

- **10% Ammonium persulfate (APS)**

  1g of APS in 10ml ddH_2O

- **10% Non denaturing polyacrylamide gel**

  Prepared by mixing 33.3ml of 30% Acrylamide and 20ml of 5X TBE. This was made up to 100ml with ddH_2O. 0.7ml of 10% Ammonium persulfute and 35µl of TEMED (N,N,N', N'-tetramethylethylenediamine) were added just before pouring. The gel was poured using a syringe to inject the solution between two glass plates. The plates were separated

with two 16cm x 1.0mm spacers. The comb was inserted and the acrylamide was allowed to polymerase for two hours at room temperature. After polymerisation the comb was removed and the wells rinsed with double distilled water (ddH_2O) before loading the microsatellite PCR products.

REFERENCES

Aagaard, J. E., Krutovskii, K. V. and Strauss, S. H. 1998. RAPDs and Allozymes exhibit similar levels of diversity and differentiation among populations and races of Douglas-fir. *Heredity* **81**: 69-78.

Anyanwu, G. I., Davies, D. H., Molyneux, D. H. and Priestman, A. 2001. Cuticular-hydrocarbon discrimination between *Anopheles gambiae* s.s. and *An. arabiensis* larval karyotypes. *Annals of Tropical Medicine and Parasitology* **95**: 843-852.

Beckman, J. S. and Weber, J. L. 1992. Survey of human and rat microsatellites. *Genomics* **12**: 627-631.

Black, W. C IV. 1993. PCR with arbitrary primers: approach with care. *Insect Molecular Biology* **2**: 1-6.

Black, W. C IV., Du Teau, N. M., Puterka, G. J., Nechols, J. R. and Pettorini, J. M. 1992. Use of the Random Amplified Polymorphic DNA Polymerase Chain Reaction (RAPD-PCR) to detect DNA polymorphisms in aphids (Homopetra: Amphidae). *Bulletin of Entomological Research* **82**: 151-159.

Botstein, D., White, R. L., Skolnick, M. and Davis, R. W. 1980. Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *American Journal of Human Genetics* **32**: 314-331.

Bryan, J. H., Di Deco, M. A., Petrarca, V. and Coluzzi, M. 1982. Inversion

Polymorphism and incipient speciation in *Anopheles gambiae s.str.* in the Gambia, West Africa. *Genetica* **59**: 167-176.

Budowle, B., Chakraborty, R., Giusti, A.M., Eisenberg, A.J. and Allen, R.C. 1991. Analysis of the VNTR locus DIS8O by the PCR followed by high resolution PAGE. *American Journal of Genetics* **48**: 137-144.

Cairns, M. J., and Murray, V. 1994. Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques* **17**: 915-919.

Callen, D. F. Thompson, A. D. Shen, Y., Phillips, H. A., Richards, R. I., Mulley, J. C., and Sutherland, G. R. 1993. Incidence and origin of "null" alleles in the AC)$_n$ microsatellite marker. *American Journal of Human Genetics* **52**: 922-927.

Carnahan, J., Zheng, L., Taylor, C. E., Touré, Y. T., Norris, D. E., Dolo, G., Duik-Wasser, M., and Lanzaro, G. C. 2002. Genetic differentiation of *Anopheles gambiae s.s.* populations in Mali, West Africa using microsatellite loci. *The Journal of Heredity* **93**: 249-253.

Chakraborty, R., de Andrade, Diager, M., and Bulowe, B. 1992. Apparent Heterozygote deficiencies observed in DNA typing data and their implication in Forensic applications. *Annals of Human Genetic* **56**: 45-57.

Chambers, G. K. and MacAvoy, E. S. 2000. Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B* **126**: 455-476.

Chandre, F., Manguin, S., Brengues, C., Dossou-Yovo, J., Darriet, F., Diabaté, A., Carnevale, P. and Guillet, P. 1999. Current distribution of pyrethroid resistance gene (kdr) in *Anopheles gambiae* complex from West Africa and further evidence for reproductive isolation of the Mopti form. *Parassitologia* **41**: 319-322.

Coetzee, M. 1989. Comparative morphology and multivariate analysis for the discrimination of four members of the *Anopheles gambiae* group in Southern Africa. *Mosquito Systematics* **21**: 100-116.

Coetzee, M. and Hunt, R. H. 1998. Malaria at its southern-most fringe in Africa. *Research and Reviews in Parasitology* **58**: 175-179.

Coetzee, M., Craig, M. and le Sueur, D. 2000. Distribution of African mosquitoes belonging to the *Anopheles gambiae* complex. *Parasitology Today*: **no 2**.

Coetzee, M., Newberry, K. and Durand, D. 1982. A preliminary report on a morphological character distinguishing important malaria vectors in the *Anopheles gambiae* Giles complex in Southern Africa. *Mosquito Systematics* **14**: 88-93.

Collins, F. H., Finnerty, V. and Petrarca, V. 1988. Ribosomal DNA probes differentiate five cryptic species in the *Anopheles gambiae* complex. *Parassitologia* **30**: 231-240.

Collins, F.H., Mendez, M.A., Rasmussen, M.O., Mehaffeh, P.C., Besansky, N.J. and Finnerty, V. 1987. A ribosomal RNA genes probe differentiates member species of the *Anopheles gambiae* complex. *American Journal of Tropical Medicine and Hygiene* **37**: 37-41.

Coluzzi, M. 1964. Morphological divergences in the *Anopheles gambiae* complex. *Rivista Di Malariologia.*

Coluzzi, M.1970. Sibling species in *Anopheles* and their importance in malariology. *Miscellaneous Publications of the Entomological Society of America* **7**: 63-77.

Coluzzi, M. 1982. Spatial distribution of chromosomal inversions and speciation in anopheline mosquitoes. In: *Mechanism of speciation.* (Edition Barigozzi, C.) pp143-153. Alan R. Liss, New York.

Coluzzi, M. 1993. Advances in the study of Afrotropical malaria vectors. *Parassitologia* **35**: 23-29.

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. and Sabatini, A. 1967. Cytogenic observations on species A and B of the *Anopheles gambiae* complex. *Parassitologia* **9**: 73-88.

Coluzzi, M. and Sabatini, A. 1968. Cytogenetic observations on species C of the *Anopheles gambiae* complex. *Parassitologia* **10**: 155-165.

Coluzzi, M. and Sabatini, A. 1969. Cytogenetic observations on the salt water species, *An. merus* and *melas* of the *Anopheles gambiae* complex. *Parassitologia* **11**: 177-187.

Coluzzi, M., Sabatini, A., Petrarca, V. and Di Deco, M. A. 1979. Chromosomal differentiation and adaptation to human environments in the *Anopheles*

gambiae complex. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **73**: 483-497.

Cornuet, J. M., and Luikart, G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001-2014.

Crampton, J. M., Warren, A., Lycett, G. l., Hughes, M. A., Comley, I. P. and Eggleston, P. 1994. Genetic manipulation of insect vectors as a strategy for the control of vector-borne disease. *Annals of Tropical and Medical Parasitology* **88**: 3-12.

Davidson, G. 1962. *Anopheles gambiae* complex. *Nature* **196**: 907.

Davidson, G. and Hunt, R. H. 1973. The crossing and chromosome characteristics of a new 6th species in the *Anopheles gambiae* complex. *Parassitologia* **15**: 121-128.

Davidson, G., Paterson, H. E., Coluzzi, M., Mason, G. F. and Micks, D. W. 1967. The *Anopheles gambiae* complex. In: *Genetics of Insect Vectors of disease.* (editions J W Wright and R. Pal) pp 211-250. Elsevier Publication and Company, Amsterdam.

Davidson, G. and White, G. B. 1972. The crossing characteristics of a new, sixth species in the *Anopheles gambiae* complex. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **66**: 531-532.

Davis, R. and Icke G. 2002. *Malaria : An on-line Resource*. Division of Laboratory Medicine at Royal Perth Hospital.

della Torre, A. Fanello, C., Akogbeto, M., Dossou-yoyo, 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.

De Meillon, B. 1947. *The Anophelini of the Ethiopian Geographical Region*. Publications of the South African Institute for Medical Research, Johannesburg **no. 49**.

Donnelly, M.J., Cuamba, N., Charlwood, J.D., Collins, F. H. and Townsom, H. 1999. Population structure in the malaria vector, *Anopheles arabiensis* (Patton), in East Africa. *Heredity* **83***: 521-528.*

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

Edillo, F. E., Touré, Y. T., Lanzaro, G. C., Dolo, G. and Taylor, C. E. 2002. Spatial and Habitat Distribution of *Anopheles gambiae* and *Anopheles arabiensis* (Diptera: Culicidae) in Banambani Village, Mali. *Journal of Medical Entomology* **39**: 70-77.

Edwards, A., Civitello, A., Hammond, H. A. and CAskey, C. T. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics* **49**: 746-756.

Evans, A. M. 1938. *Mosquitoes of the Ethiopian Region.* II. *Anophelini adults and early stages.* British Museum (National History), London.

Excoffier, L., Smousse, P., and Quattro, J. 1992. Analysis of molecular variance

inferred from metric distances among DNA haplotypes: Application to human mitochondria DNA restriction data. *Genetics* **136**: 343-359.

Fannelo, C., Petrarca, V., della Torre, A., Santolamazza, F., Dollo, G., Coulibaly, M., Alloueche, A., Curtis, C. F., Touré, Y. T. and Coluzzi, M. 2003. The pyrethroid knock-down resistance gene in the *Anopheles gambiae* complex in Mali and further indication of incipient speciation within *An. gambiae s.s. Insect Molecular Biology* **12**: 241-245.

Favia, G., della Torre, A., Bagayoko, M., Lanfrancotti, A., Sagnon, N'F., 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.

Favia, G., Dimopoulos, G., della Torre, A., Touré, Y. T., Coluzzi, M. and Louis, C. 1994. Polymorphisms detected by random PCR distinguish between different chromosomal forms of *Anopheles gambiae*. *Proceeding of the National Academy of Sciences, USA* **91**: 10315-10319.

Favia, G., Lanfrancotti, A., Spanos, L., Sidén-Kiamos, I.and Louis C. 2001. Molecular characterisation of ribosomal DNA polymorphisms discriminating among chromosomal forms of *Anopheles gambiae* s. s. *Insect Molecular Biology* **10:** 19-23.

Ferris, S. D., Sage, R. D., Huang, C. M., Nielsen, J. T., Ritte, U. and Wilson, A. C. 1983. Flow of mitochondrial DNA across a species boundary. *Proceeding of the National Academy of Sciences, USA* **91**: 2290-2294.

Fettene, M., Koekemoer, L. L., Hunt, R. H. and Coetzee, M. 2002. PCR assay for identification of *Anopheles quadriannulatus* species B from Ethiopia and

other sibling species of the *Anopheles gambiae* complex. *Medical and Veterinary Entomology* **16**: 214-217.

Fettene, M. and Temu, E. 2003. Species-specific primer for identification of *Anopheles quadriannulatus* sp. B (Diptera: Culicidae) from Ethiopia using a multiplex polymerase chain reaction assay. *Medical and Veterinary Entomology* **40**: 112-115.

Gale, K. R. and Crampton, J. M. 1987. DNA probes for species identification of mosquitoes in the *Anopheles gambiae* complex. *Medical and Veterinary Entomology* **1**: 127-136.

Gallup, J. L. and Sachs, J. D. 2001. The economic burden of malaria. *American Journal of Tropical Medicine and Hygiene* **64**: 85-96.

Gentile, G., della Torre, A., Maegga, B., Powell, J. 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.

Gillespie, J. H. 2000. The neutral theory in an infinite population. *Gene* **261**: 11-18.

Gillespie, J. H. 2001. Genetic drift in an infinite population: the pseudohitchhiking model. *Genetic* **155**: 909-919.

Gillies, M.T. and Coetzee, M. 1987. *A supplement to the Anophelinea of Africa south of the Sahara*. Publications of the South African Institute for Medical Research, Johannesburg **no.55**.

Gillies, M.T. and De Meillon, B. 1968. *The Anophelinae of Africa south of the Sahara. Second Edition*. Publications of the South African Institute for Medical Research**,** Johannesburg **no.54**.

Goodfellow, P. N. 1992. Variation is now the theme. *Nature* **359**: 777-778.

Green, C. A. 1971. The practical problems of identifying members of the *Anopheles gambiae* complex in autoecological studies. *Parassitologia* **13**: 421-427.

Green, C. A. and Miles, S. J. 1980. Chromosomal evidence for sibling species of malaria vector *Anopheles* (Cellia) culicifacies Giles. *Journal of Tropical Medicine and Hygiene* **83**: 75-78.

Guyatt, H. L. and Snow, R. W. 2001. Malaria in pregnancy as an indirect cause of infant mortality in sub-Saharan Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **95**: 569-576.

Haddow, A. J., Gillett, J. D. and Highton, R. B. 1947. The mosquitoes of Bwambae County, Uganda. The vertical distribution and biting cycle of mosquitoes in rain-forest, with further observations on microclimates. *Bulletin of Entomological Research* **37**: 301-330.

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

Hargreaves, K., Hunt, R. H., Brooke, B. D., Mthembu, J., Weeto, M. M., Awolola, T. S. and Coetzee, M. 2003. *Anopheles arabiensis* and *An. quadriannulatus* resistance to DDT in South Africa. *Medical and Veterinary Entomology* **17**: 417-422.

Hartl, D. L. and Clark, A. G.1997. *Principles of population genetics*, Third edition. Sinauer Associates, Inc, Sunderland, Massachusetts. Pp 135-137.

Henderson, S. T. and Petes, T. D. 1992. Instability of simple sequence DNA in Saccharomyces cerevisiae. *Molecular and Cell Biology* **12**: 2749-2757.

Hill, S. M., Urwin, R. and Crampton, J. M. 1992. A simplified non-radioactive DNA probe protocol for the field identification of insect vector specimens. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **86**: 213-215.

Hill, S. M. Urwin, R., Knapp, T. F. and Crampton, J. M. 1991. Synthetic DNA probes for the identification of sibling within the *Anopheles gambiae* complex. *Medical and Veterinary Entomology* **5**: 455-463.

Holm, H. 1979. A simple sequential rejective multiple test procedure. *Scandinavian Journal of Statistic* **6**: 65-70.

Hunt, R. H. 1973. A cytological technique for the study of the *Anopheles gambiae* complex. *Parassitologia* **15**: 137-139.

Hunt, R. H. and Coetzee, M. 1986. Field sampling of *Anopheles* mosquitoes for correlated Cytogenetics, electrophoretic and morphological studies. *Bulletin of the World Health Organization,* **64**: 897-900.

Hunt, R.H., Coetzee, M. and Fettene, M. 1998. The *Anopheles gambiae* complex: new species from Ethiopia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **92**: 231-235.

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

Kambhampati, S., Black, W. C. IV. and Rai, K. S. 1992. Random amplified Polymorphic DNA of Mosquito Species and Populations (Diptera: Culicidae): Techniques, Statistical Analysis, and Applications. *Journal of Medical Entomology* **29**: 939-945.

Lanzaro, G.C., Touré, Y.T., Carnahans, J., Zheng, L., Dolo, G., Traoré, S., Petrarca, V., Vernick, K.D. and Taylor, C.E. 1998. Complexities in the genetic structure of *Anopheles gambiae* populations in West Africa as revealed by microsatellite DNA analysis. *Proceeding of the National Academy of Sciences* **95**: 14260-14265.

Lanzaro, G. C., Zheng, L., Touré, Y. T., Kafatos, F. C., and Vernick, k. D. 1995. Microsatellite DNA and isozyme variability in a West African population *Anopheles gambiae*. *Insect Molecular Biology* **4**: 105-112.

Lehmann, T., Hawley, W. A., and Collins, F. H. 1996. An evolution of evolutionary constraints on microsatellite loci using null alleles. *Genetics* **144**: 1155-1163.

Levinson, G., and Gutman, G. A. 1987. Slipped-strand mispairing: a major

mechanism for DNA sequence evolution. *Molecular Biology and Evolution* **4**: 203-221.

Lindsay, S. W. and Martens, W. J. 1998. Malaria in the African highlands: past, present and future. *Bulletin of the World Health Organization.* **76**: 33-45.

Litt, M. and Luty, J. A. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac actin gene. *American Journal of Human Genetics* **44**: 397-401.

Loxdale, H. D. and Lushai, G. 1998. Molecular markers in Entomology. *Bulletin of Entomological Research* **88**: 577-600.

Lynch, M. and Milligan, B. G. 1994. Analysis of population genetic structure with RAPD markers. *Molecular Ecology* **3**: 91-99.

Mahon, R. J., Green, C. A. and Hunt, R. H. 1976. Diagnostic allozymes for routine identification of adults of the *Anopheles gambiae* complex (Diptera: Culicidae). *Bulletin of Entomological Research* **66**: 25-31.

Mastbaum, O. 1954. Observations of two epidemic malaria seasons (1946 and 1953) before and after malaria control in Swaziland. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **48**: 325-331.

Mastbaum, O. 1957a. Past and present position of malaria in Swaziland. *Journal of Tropical Medicine and Hygiene* **May 1957**: 119-127.

Mastbaum, O. 1957b. Malaria control in Swaziland: Some observations during the first year of partial discontinuation of insecticides. *Journal of Tropical Medicine and Hygiene* **August 1957**: 190-192.

Mattingly, P. F. 1977. Names for *Anopheles gambiae* complex. *Mosquitoes Systematic* **9**: 323-328.

Merril, C. R. 1981. Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science* **211**: 1437-1438.

Miles, S. J. 1978. Enzyme variation in the *Anopheles gambiae* Giles group of species (Diptera: Culicidae). *Bulletin of Entomological Research* **68**: 85-96.

Miles, S. J. 1979. A biochemical key to adult members of the *Anopheles gambiae* group of species (Diptera: Culicidae). *Journal of Medical Entomology* **15**: 297-299.

Miller, L. H. 1992. The challenge of malaria. *Science* **257**: 36-38.

Milligan, P. J. M., Phillips, A., Broomfield, G. and Molyneux, D. H. 1993. A study of the use of gas chromatography for identifying members of *Anopheles gambiae* (Diptera: Culicidae) complex. *Bulletin of Entomological Research* **83**: 613-624.

Moore, S. S., Sargeant, L. L., King, T. L., Mattick, J. S., Georges, M. and Hetzel D. J. S. 1991. The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics* **10**: 654-660.

Mouchet, J., Manguin, S., Sircoulon, J., Laventure, S., Faye, O., Onapa, A. W., Carnevale, P., Julvez, J. and Fontenille, D. 1998. Evolution of malaria in Africa for the past forty years: Impact of climatic and human factors. *Journal of American Mosquitoes Control Association* **14**: 121-130.

Muirhead-Thomson, R. C. 1945. Studies on the breeding places and control of *Anopheles gambiae* and *A. gambiae* var *melas* in coastal districts of Sierra Leone. *Bulletin of Entomological Research* **36**: 185-252.

Muirhead-Thomson, R. C. 1947. Studies on *Anopheles gambiae* and *A. melas* in and around Lagos. *Bulletin of Entomological Research* **38**: 527-558.

Muirhead-Thomson, R. C. 1951. Studies on salt-water and fresh- water *Anopheles gambiae* on the East African coast. *Bulletin of Entomological Research* **41**: 487-502.

Nagylaki, T. 1975. Conditions for the existence of clines. *Genetics* **80**: 595-615.

Nei, M., Maruyama, T. and Chakraborty, R. 1975. The bottleneck effect and gene variability in populations. *Evolution* **29**: 1-10.

Norris, D. E., Shurtleff, A. C., Toure, Y. T., and Lanzaro, G. C. 2001. Microsatellite DNA polymorphism and heterozygosity among field and laboratory populations of *Anopheles gambiae* s.s. *Journal of Medical Entomology* **38**: 336-346.

Pampana, E. 1936. *A Textbook o f Malaria eradication.* Oxford University Press, Amen house, London. Pp 76-89.

Paskewitz, S. M. and Collins, F. H. 1990. Use of polymerase chain reaction to identify mosquito species of the *Anopheles gambiae* complex. *Medical and Veterinary Entomology* **4**: 367-373.

Paskewitz, S. M., Karin, N. G., Coetzee, M. and Hunt, R. H. 1993. Evaluation of the polymerase chain reaction method for identifying members of the

Anopheles gambiae (Diptera: Culicidae) complex in Southern Africa. *Journal of Medical Entomology* **30**: 953-957.

Paterson, H. E. 1962. On the status of the East African Saltwater- breeding variant of *Anopheles gambiae* Giles. *Nature* **195**: 469-470.

Paterson, H. E. 1964. Direct evidence for the specific distinctness of forms A, B and C of the *Anopheles gambiae* complex. *Rivista Di Malariologia* **43**: 191-196.

Paterson, H. E. 1968. *Evolutionary and population genetical studies of certain Diptera.* PhD thesis, University of the Witwatersrand, Johannesburg.

Petrarca, V., Nugud, A. D., Ahmed, M. A. E. 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 Africa populations. *Medical and Veterinary Entomology* **14**: 149-164.

Pinto, J. Donnelly, M. J., Sousa, C. A., Malta-Vacas, J., Gil, V., Ferreira, C., Petrarca, V., do Rosário, V. E. and Charlwood, J. D. 2004. An island within and island: genetic differentiation of *Anopheles gambiae* in São Tomé, West Africa, and its relevance to malaria vector control. *Heredity* **91**: 407-414.

Powell, W., Machray, G. C. and Provan, J. 1996. Polymorphism revealed by simple sequence repeats. *Trends in Plant Science* **7**: 215-222.

Raymond, M. and Rousset, F. 1995. An exact test for population differentiation. *Evolution* **49**: 1280-1283.

Ribbands, C. R. 1944. Differences between *Anopheles melas* (*An. gambiae* var. *melas*) and *Anopheles gambiae. Annals of Tropical Medicine and Parasitology* **38**: 85-99.

Rongnoparut, R., Sirichotpakorn, N., Rattannarithikul, S. Y. and Linthicum, K.J. 1999. Estimate of gene flow among *Anopheles maculates* populations in Thailand using microsatellite analysis. *American Journal of Tropical Medicine and Hygiene* **60**: 508-515.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219-1228.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1988. Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.

Sambrook, J., Fritsch, E. F. and Maniatis, T. 1989. DNA sequencing In*: Molecular Cloning, a laboratory manual.* Nolan, C. (editor) New York: Cold Spring Harbour Laboratory Press, pp 13.3-13.10.

Schlötterer, C. and Tautz, D. 1992. Slippage synthesis of simple sequences DNA. *Nucleic Acids Research* **20**: 211-215.

Schlötterer, C., Vogl, C., and Tautz, D. 1997. Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. *Genetics* **146**: 309-320.

Schneider, S., Roesseli, D. and Excoffier, L. 2000. *Arlequin ver 2.000: A software for population genetics data analysis.* Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Schug, M. D., Mackay, T. F. C. and Aquadro, C. F. 1997. Low mutation rates of microsatellite loci in Drosophila melanogaster. *Nature Genetics* **15**: 99-102.

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.

Service, M. W. 1970. Ecological notes on species A and B of the *Anopheles gambiae* complex in Kisumu area of Kenya. *Bulletin of Entomological Research* **60**: 105-108.

Service, M. M. W., Joshi, G. P. and Pradhan, G. D. 1978. A survey of *Anopheles gambiae* (sp A) and *An. arabiensis (sp* B) of the *An. gambiae* Giles complex in the Kisumu area of Kenya following insecticide spraying with OMS-43 (fenitrothion). *Annals of Tropical Medicine and Parasitology* **72**: 377-386.

Simard, F., Fontenille, D., Girod, R., Brutus, L., Gopaul, R., Dournon, C. and Collins, F. H. 1999. High amounts of genetic differentiation between populations of the malaria vector *Anopheles arabiensis* from the West Africa and Eastern outer island. *American Journal of Tropical Medicine and Hygiene* **60**: 1000-1009.

Simard, F., Lehman, T., Lemasson, J-J., Diatta, M. and Fontenille, D. 2000. Persistence of *Anopheles arabiensis* during the severe dry season conditions in Senegal: an indirect approach using microsatellite loci. *Insect Molecular Biology* **9**: 467-469.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457-462.

Smith, A., Hansford, C. F. and Thomson, J. F. 1977. Malaria along the southernmost fringe of its distribution in Africa: epidemiology and control. *Bulletin of the World Health Organization.* **55**: 95-103.

Sudhir K., Tamura, K., Jakobsen, I. B., and Nei, M. 2001. *MEGA* 2: Molecular Evolutionary Genetics Analysis software, Arizona State University, Tempe, Arizona, USA.

Tabachnick, W. J. 1991. Evolutionary genetics and arthropod-borne disease: the yellow fever mosquito. *American Entomology* **37**: 14-24.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids and Research* **17**: 6463-6471.

Taylor, C.E., Touré, 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.

Taylor, C.E., Touré, Y.T., Coluzzi, M. and Petrarca, V. 1993. Effective population size and persistence of *Anopheles arabiensis* during the dry season in West Africa. *Medical and Veterinary Entomology* **7**: 351-357.

Temu, E. A., Minjas, J. N., Coetzee, M., Hunt, R. H. and Shiff, C. J. 1998. The role of four anophelina species (Diptera: Culicidae) in malaria transmission in

coastal Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **92**: 152-158.

Touré, Y. T., Petrarca, V., Traoré, S. F., Coulibaly, A., Maiga, H. M., Sankaré, O., Sow, M., DI Deco, M. A., and Coluzzi, M. 1998. The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia* **40**: 477-511.

Townson, H. and Onapa, A. W. 1994. Identification by rDNA-PCR of *Anopheles bwambae,* a geothermal spring species of the *Anopheles gambiae* complex. *Insect Molecular Biology* **3**: 279-282.

Valdes, A.M., Slatkin, M and Freiner, N. B. 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* **133:** 737-749.

Van Rensburg, A. J. 1997. *Molecular biological studies on mosquitoes.* Msc dissertation. University of the Witwatersrand, Johannesburg.

Van Rensburg, A. J., Hunt, R. H., Koekemoer, L. L., Coetzee, M., Shiff, C. J. and Minjas, J. 1996. The polymerase chain reaction as a tool for identifying members of the *Anopheles gambiae* complex (Diptera: Culicidae) in Northeastern Tanzania. *Journal of American Mosquito Control Association* **12**: 271-274.

Voelker, R. A., Schaffer, H. E. and Mukai, T. 1980. Spontaneous allozymes mutations in *Drosophila melanogaster*: rate of occurrence and nature of the mutants. *Genetics* **94**: 961-968.

Walton, C., Thelwell, N. J., Priestman, A. and Butlin, R. K. 1998. The use of microsatellites to study gene flow in natural populations of *Anopheles* malaria vectors in Africa: Potential and Pitfalls. *Journal of the American Mosquito Control Association* **14**: 266-272.

Weber, J. L and May, P. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**: 388-396.

Weber, J. L., Polymeropoulos, M. H., May, P. E., Kwitek, A. E., Xiao, H., McPherson, J. D., and Wasmuth, J. J. 1991. Mapping of human chromosome 5 microsatellite DNA polymorphism. *Genomics* **11**: 695-700.

Weber, J. L. and Wong, C. 1993. Mutation of human short tandem repeats. *Human Molecular Genetics* **2**: 1123-1128.

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. and Cockerham, C.C. 1984. Estimating F-statistics for analysis of population structure. *Evolution* **38**: 1358-1370.

White, G. B. 1985. *Anopheles bwambae* n. sp., a malaria vector in the Semliki valley, Uganda, and its relationships with other sibling species of the *An. gambiae* complex (Diptera: Culicidae). *Systematic Entomology* **10**: 501-522.

White, G. B., Magayuka, S. A. and Boreham, P. F. L.1972. Comparative studies on sibling species of the *Anopheles gambiae* Giles complex (Diptera: Culicidae). II. Ecology of species A and B in savannah around Kaduna, Nigeria, during transition from wet to dry season. *Bulletin of Entomological Research* **62**: 295-317.

Williams, J. G. K., Kubeklik, A. R., Livak, K. J., Rafalski, J. A. and Tingey, S. V. 1990. DNA polymorphisms amplified by arbitrary primers are useful genetic markers. *Nucleic Acids Research* **18**: 6531-6535.

Wilson, K. and Walker, J. 2000. *Practical biochemistry: Principles and techniques.* Fifth edition, Cambridge University Press, UK.

Wolff, R. K. Plaetke, R., Jeffreys, A. J. and White, R. 1989. Unequal crossing over between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. *Genomics* **5**: 382-384.

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

World Health Organization. 1975. Manual on practical entomology in malaria. *Part I: Bionomics and organization of anti-malaria activities*. Geneva.

World Health Organization. 1999. The World Health Report 1999: Making a difference. *Geneva: World Health Organisation* 49-63

Wright, S. 1978. *Evolution and the genetics of population. Volume 4, Variability within and among natural populations*. Chicago: University of Chicago Press.

Zheng, L., Benedict, M. Q., Cornel, A. J., Collins, F. H., and Kafatos, F. C. 1996. An integrated genetic map of the African human vector mosquito, *Anopheles gambiae. Genetics* **143**: 941-952.

Zheng, L., Collins, F. H., Kumar, V., and Kafatos, F. C.1993. A detailed genetic map for the X chromosome of the malaria vector *Anopheles gambiae. Science* **261**: 605-608.