POPULATION STRUCTURE OF *ANOPHELES ARABIENSIS* BREEDING IN TRANSIENT AND PERMANENT INLAND HABITATS AT TWO SITES IN WESTERN KENYA.

By

Godwil Otsyula Munyekenye (B.Sc. Hons)

PG/MSC/044/99

A Thesis Submitted In Partial Fulfilment For The Requirements Of M.Sc. Degree In Cell And Molecular Biology Of Maseno University, Kenya

MAY, 2003

ABSTRACT

Malaria is one of the main diseases that cause great morbidity and mortality in sub-Saharan Africa region. The prevalence and incidence of malaria in this region is determined mostly by the mosquito vector species. Several measures and programmes have been undertaken to control malaria vectors and reduce malaria incidence in Africa. However, most of these measures have not been very successful. This has led to proposals to use mosquitoes that are refractory to malaria parasite infection as one of the options in malaria control programmes. Such a control strategy requires that population genetic structure of the vector species is understood. The understanding of population genetic structure of vector species will enable malaria entomologists to predict how genes associated with refractoriness will spread through the vector populations. Knowledge of vector population genetic structure is also useful for predicting the spread of insecticide-resistance genes in a given vector population. Studies on the population genetic structure of vector species have been carried out using various techniques and markers. These include restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), chromosomal inversions, and microsatellite DNA.

In this study, *Anopheles arabiensis* populations from two different types of breeding habitats in Western Kenya were analysed using a total of eight microsatellite loci. The aim of this study was to determine whether *Anopheles arabiensis* species partitioned larval breeding habitats based on the habitat type. Two habitat types namely, transient small size larval breeding habitats and permanent/semi-permanent large size larval breeding habitats were considered and sampling done in three sites, two in Ahero and

one in Nyakach. Specimen collection was done by pipetting and dipping using standard World Health Organisation (WHO) procedures. A total of 500 mosquito larvae were collected. The specimens were processed by DNA extraction and then analysed using polymerase chain reaction (PCR) and polyacrylamide gel electrophoresis (PAGE). The data were analysed using GENEPOP version 3.2a software programme, which evaluated Hardy-Weinberg equilibrium, allele frequencies, and genotypic frequencies. Levels of genetic differentiation was analysed using G-like test available in GENEPOP. Significance of multiple tests was evaluated using sequential rejective Bonferroni test and binomial test.

There was polymorphism at six out of eight studied loci. Overall allelic and genotypic frequency distribution at all loci deviated significantly from Hardy-Weinberg equilibrium and pooled data from similar habitats and those from all breeding habitats at Ahero also deviated from Hardy-Weinberg equilibrium ($P < 0.05$). The deviation from Hardy-Weinberg equilibrium suggested lack of random mating and presence of population sub-divisions. This meant that there was restricted gene flow between *An. arabiensis* populations from similar habitat types in the study sites.

The G-like test showed genotypic differentiation at all loci across all populations. The observed genotype distribution frequency was not in agreement with the expected frequency suggesting the presence of non-random mating and population sub-division. The presence of sub-populations and non-random mating suggested that *Anopheles arabiensis* did not partition their larval breeding habitats based on habitat type. The non-random mating and presence of sub-population may suggest presence of genetic variability within *Anopheles arabiensis* populations in the study area. This may

contribute to the diversity and stability of malaria transmission in study area. The variation may also cause the proposed control measure based on vector replacement to be less effective. Therefore studies such as this, which, apply markers that can show association with breeding habitats, can be useful in understanding whether the type of breeding habitat may have effect on genetic variability of *Anopheles arabiensis* populations at a local scale.

CHAPTER ONE: INTRODUCTION AND LITERATURE REVIEW

1.1 INTRODUCTION

Malaria is one of the most important parasitic diseases in the world affecting over one million people each year (Rojas *et al.*, 1992; WHO, 2000). Malaria impedes industrial and agricultural development in sub-Saharan Africa (WHO, 2000). It is estimated that nine out of ten cases of malaria worldwide occur in sub-Saharan Africa and the economic cost of malaria in Africa is estimated to exceed US$ 12 billion per year (WHO, 2000). Malaria contributes substantially to impoverishment of the sub-Saharan Africa, even though some rich countries also experience malaria for instance, Iran where 45% of the population is at risk of malaria infection (Zakeri S *et al.*, 2002) the effects of malaria are much felt in sub-Saharan Africa.

In the last decade, the prevalence of malaria has been escalating at an alarming rate, especially in Africa. Estimated 300 to 500 million clinical cases occur each year causing over one million deaths annually, 90% of these deaths occur in Africa. Malaria is ranked third among major infectious disease threats in Africa after pneumococcal acute respiratory infections, which kills over 2.6 million people per year and tuberculosis (TB), which kills two million people annually (WHO, 2000). In Kenya, number of malaria cases reported between the year 1999 and 2000 were over 4 million (WHO, 2000). The risk of malaria infection in Kenya

is high and it is throughout the year in almost all parts of the country (WHO, 2000).

Only anopheline mosquitoes can transmit human malaria although not all of them do transmit the disease to humans. The anopheline mosquitoes belong to: order *Diptera*, sub-order *Nematocera,* family *Culicidae* and genus *Anopheles.* The *Anopheles* mosquitoes have four developmental stages during their growth, the stages include: egg, larva, pupa and adult. The anopheline larva hatches from the egg and has distinct head, thorax and abdomen. The thorax is broader than the head or abdomen and has several hairs that are used in species identification. The larval abdomen is composed of nine segments; the eighth segment bears respiratory apparatus that consist of a pair of spiracular openings, while the ninth segment bears two to four tapering membranous appendages, known as anal gills. The larvae have mouth brushes, which sweep food particles into the mouth. Feeding takes place at water surface and larval head is able to turn 180 degrees. The body of anopheline larvae lie parallel to water surface (Clements, 1992) and this is one characteristic that distinguishes anopheline larvae from other mosquitoes. The larvae undergo three successive moulting during their growth and the different larval stages are called instars.

Morphologically, anopheline larva can be identified based on following characteristics: the head has frontal plumose hairs, abdomen has plumose lateral hairs on first six segments, there are clypeal hairs on the thorax which may be branched or simple, basal tubercle of inner anterior clypeal hairs may be close together or separated from each other by a distance much less than between inner

and out clypeal. The thorax and abdomen laterally and ventrally have numerous spicules and sides of thorax and abdomen are without spicules. They have palmate hairs, which may be functional or rudimentary (Gilles M.T and Coetzee M, 1987; King M.V and Bradley G.H, 1941). Some anopheline like *An. gambiae* have basal spine of pleural hairs, large, curved and sharp (Gilles M.T and Coetzee M, 1987).

The anopheline pupa differs a lot from the larva, the pupal head and thorax are enlarged and enclosed in a sheath, and usually pupa does not feed. The anopheline adult has head, thorax and abdomen. *Anopheles* adult male differs from female by having hairy antennae. The anophelines rest at an angle to resting surface, which is one of characteristic that distinguishes anopheline adults from other mosquitoes.

Malaria is caused by protozoan parasites of the genus *Plasmodium.* There are four *Plasmodium* species that cause human malaria. These are *Plasmodium falciparum*, *Plasmodium malariae*, *Plasmodium ovale* and *Plasmodium vivax* (WHO, 2002) Clinically, malaria is characterized by fever (which is often periodic), varying degrees of anaemia, splenic enlargement, and various syndromes resulting from the physiological and pathological involvement of certain organs including the brain, the liver, and the kidneys.

Efforts to control malaria, mainly in Tropical Africa, have been impeded by emergence of drug-resistant parasites, the emergence of insecticide-resistant mosquitoes (Vulule *et al.,* 1994), and environmental concerns over the application of insecticides (WHO, 2002). The exophilic behaviour among some vector populations (Mnzava, *et al.,* 1995) has, in part, also limited the control of malaria.

Main malaria vectors in sub-Saharan Africa are mosquitoes of the *Anopheles gambiae* complex and the *Anopheles funestus* group (Collins *et al.*, 1988; Petrarca *et al.*, 1991). *Anopheles gambiae sensu stricto*, (s.s), *An. arabiensis* and *Anopheles funestus* are the primary malaria vectors in Western Kenya region (Taylor *et al.*, 1990). These three African mosquitoes are the most efficient malaria vectors in the world (Besansky, 1999). This is because of their marked preference for human environments and for humans as hosts, and due to their rapid adaptation to changes in the environment induced by human habitation and agriculture (Collins and Besansky, 1994; Powell *et al.*, 1999).

Dynamics of malaria transmission involve interactions of various factors that affect the parasite, the vector and the human host all of which have to be taken into account in order to control malaria. Therefore, thorough analyses of malaria vectors at intra-species level can contribute to understanding of how a single species plays role in transmission dynamics of malaria. The analysis of vectors at the intra-species level requires the use of markers that are powerful in elucidating the genetic structure. Examples of such markers are microsatellite loci, which are relatively short tracts of tandemly repeated DNA sequences of 2-6 base pairs (Budowle *et al.*, 1991). The microsatellite loci are useful markers for population genetics because of their abundance throughout eukaryotic genome, their high polymorphism and co-dominance, and relative ease of scoring (Budowle *et al.*, 1991). The microsatellite loci generally have higher mutation rates than other regions of the genome and thus they are highly polymorphic in populations and therefore favourable for population genetic studies (Collins *et al.*, 2000).

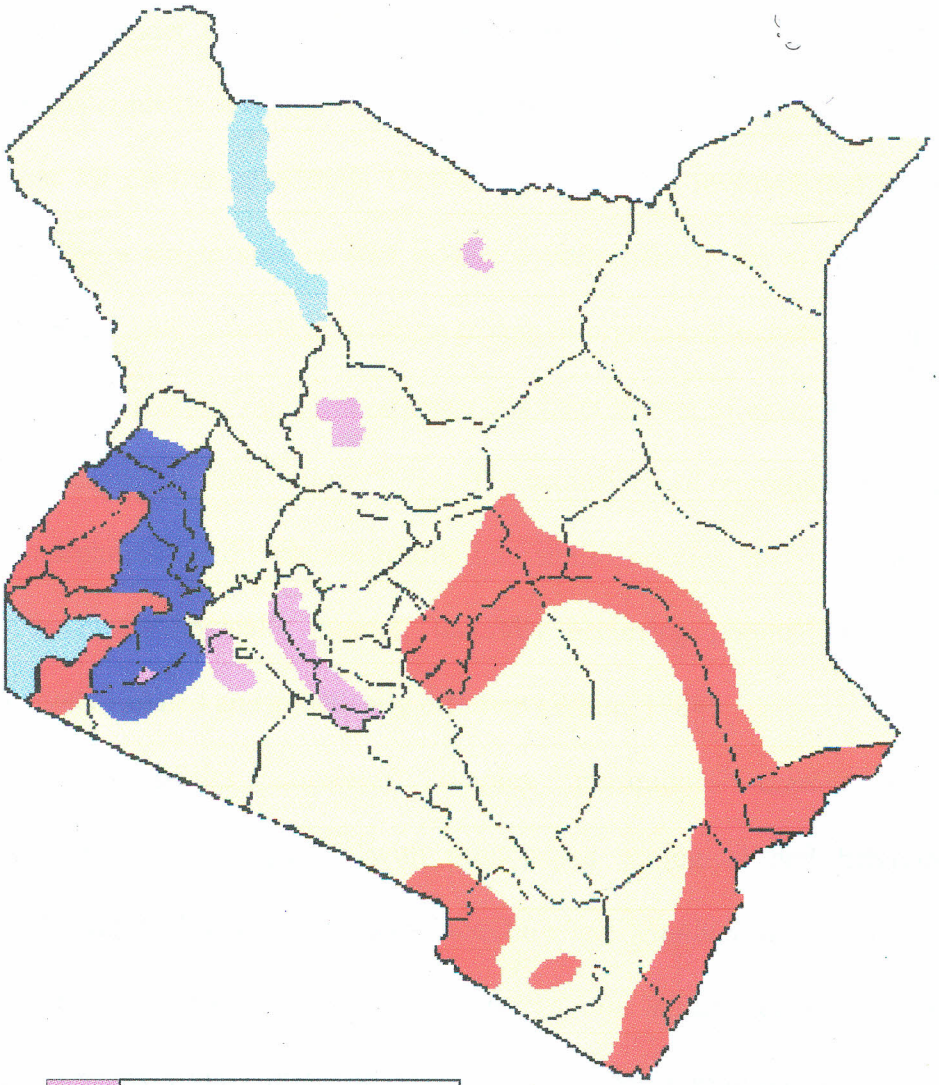
This study was designed to determine whether the *Anopheles arabiensis* populations in Western Kenya partition their larval breeding habitats. The types of habitats considered in this study were, the standing permanent/semi-permanent water collections such as rice fields and large water puddles (> 1m diameter) and transient temporary water collections in footprints, hoof prints and car tyre ruts (< 30cm diameter). The analysis was done using eight microsatellite loci.

1.2 LITERATURE REVIEW

1.2.1 Malarious Regions In Kenya.

Malaria endemicity is largely dependent on three factors, namely: (1) type of mosquito vector in an area, (2) parasite species and (3) climate. The malarious regions in Kenya have been classified into various zones based on the transmission intensity. These zones are: (1) endemic zones, (2) epidemic zones and (3) unstable endemic zones. The endemic zones are those zones that experience high intensities of malaria transmission and the transmission is continuous throughout the year. The endemic zones primarily exist in tropical Africa except in highland areas. The high transmission intensity in endemic zones is due to high number of vectors and favourable climate. The epidemic zones are zones that experience irregular rapid increases in malaria incidences, usually related to the season and population movements. And lastly, the unstable endemic malaria zones are zones that experience moderate intensities of transmission with seasonal and year-to-year fluctuations. In malaria endemic zones, children are the most vulnerable to attack, as other adults acquire a degree of immunity through continued exposure. Whereas in the zones of less intense transmission, particularly in epidemic areas, most of the population are likely to be non-immune and all age groups are at risk of infection. Figure 1 shows various malaria zones in Kenya.

Fig. 1: **Endemicity of Malaria in Kenya (Ministry of Health, Kenya, 1998)**



1.2.2 The *Anopheles gambiae* Complex Mosquitoes.

The *Anopheles gambiae* complex mosquitoes were initially regarded as a single species under the name *Anopheles gambiae* Giles (Coluzzi *et al.*, 1979). Crossing experiments (White, 1974) and introduction of residual insecticides application for control of malaria (Service *et al.*, 1978) revealed that the once thought to be a single species was in fact a collection of six morphologically similar species. This is because in some groups of organisms including *Anopheles* mosquitoes, the morphological divergence has been shown to occur more slowly than reproductive isolation (White, 1984). The acquisition of reproductive isolation without morphological divergence results into sibling species. Sibling species are defined as morphologically similar, naturally occurring populations that are reproductively isolated and genetically different (White, 1984).

The *Anopheles gambiae* complex mosquitoes consist of six morphologically similar species with pronounced ecological and behavioural diversity (White, 1974). The species include *Anopheles gambiae sensu stricto* (s.s), *An. arabiensis, An. quadriannulatus, An. merus, An. melas* and *An. bwambae.* These six species widely differ in their epidemiological importance. The main human malaria vectors within the complex in sub-Saharan Africa are *An. gambiae* s.s and *An. arabiensis.* These two species have high vectorial potential. This is because of their ability to acquire high rates of *Plasmodium* infection (Petrarca & Beier, 1992), their close association with human settlements and their anthropophilic behaviour (Coluzzi *et al.*, 1979). Other species belonging to the complex i.e. *An. merus, An. melas* and *An. bwabwae* have localized roles in

malaria transmission (White, 1974) while *An. quadriannulatus* is mostly zoophilic (Coluzzi, 1992) and generally of little importance in malaria transmission.

1.2.3 Distribution Of *Anopheles gambiae* Mosquitoes.

The six sibling species of the *Anopheles gambiae* complex are native to sub-Saharan Africa (Powell, *et al.*, 1999). The two most important species in terms of malaria transmission, i.e. *An. gambiae* s.s and *An. arabiensis* are widely distributed in the region. They are found from the southern limits of Sahara desert, southwards through most of continental Africa including Madagascar (Powell, *et al.*, 1999), Mauritius and along the Coasts of Africa (White, 1974). They occur separately or in sympatry (White, 1974). Where they occur singly, the distributional limits are determined by climatic factors. *An. gambiae* s.s predominates in humid situations whereas *An. arabiensis* is more predominant in arid zones than in humid zones. *An. gambiae* s.s can be found singly in regions like rain forest belts of West Africa and the Congo basin, *An. arabiensis* can be found singly in horn of Africa and southern Arabia (White, 1974). *An. bwabwae, An. merus* and *An. melas* have narrower distributional limits. *An. merus* and *An. melas* are brackish/salt-water breeders confined to East and West Coasts of Africa, respectively. *An. bwabwae* breeds in mineral springs and it has been found only from Semliki forest in Uganda (White, 1974; Powell *et al.*, 1999). *An. quadriannulatus* has been found in Zanzibar (White, 1974), highlands of Ethiopia (White, 1974; Powell *et al.*, 1999) and southern Africa along Limpompo and Zambezi valleys (White, 1974).

1.2.4 Biology And Behaviour Of *Anopheles gambiae* Mosquitoes.

The *Anopheles gambiae* complex sibling species show wide diversity in terms of their biology (Coluzzi *et al.,* 1977; Powell *et al.,* 1999). The diverse biology leads to dissimilar distribution patterns, different behaviours in terms of host preference, and choice of resting places by these mosquitoes. The diverse biology of these mosquitoes also leads to differences in susceptibility to parasite infection, which greatly affects their vectorial capacities and roles in malaria transmission (Petrarca & Beier, 1992). *Anopheles gambiae* s.s and *An. arabiensis* are extremely flexible as regards tolerance to a wide variety of macro- and micro-environments and they show distinct habitat diversification (Manguin *et al.,* 1996).

The polymorphism of paracentric chromosomal inversions is associated with ecological and behavioural flexibility in *Anopheles gambiae* complex (Lanzaro, *et al.,* 1998). The spatial distribution of certain gene arrangements has strong association with specific regional habitats among the *Anopheles gambiae* complex mosquitoes (Lanzaro *et al.,* 1998). For instance, in *An. gambiae* s.s, chromosomally distinct sub-populations that are adapted to different ecological zones have been identified (Toure *et al.,* 1998). These chromosomal sub-populations seem to be reproductively isolated even though they belong to a single species.

Among the *Anopheles gambiae* complex mosquitoes, *An. arabiensis* is the most widely spread species (Donnelly *et al.,* 1999; Simard *et al.,* 2000; Petrarca *et al.,* 2000). This species extends into less humid environments more than any other

sibling species of the complex, indicating its flexility in terms of ecological adaptation and it has a higher adaptive value (Coluzzi *et al.*, 1979). One adaptive method exhibited by *An. arabiensis* is adult diapause or aestivation (Coluzzi *et al.*, 1979: Simard *et al.*, 2000). This aestivating behaviour enables this species to survive harsh environmental conditions and become widespread compared to other sibling species of the *An. gambiae* complex. *Anopheles arabiensis* exhibits exophily and partial zoophily (Mnzava *et al.*, 1995; Donnelly *et al.*, 1999). Its exophilic behaviour is linked to its plastic biology, which enables it withstand less humidity and drier conditions outside houses. The ability to feed on animals (zoophily) ensures its existence and survival where human population is scarce or absent. The biological and behavioural plasticity of *An. arabiensis* make commonly used malaria control measures to be less effective when applied where this species exists. Its preference for outdoor resting limits its contact with insecticide-sprayed walls and enables it to bite people outdoors thus rendering bed nets less effective.

The complex biology of *Anopheles arabiensis* requires a thorough intra-species analysis of this species in relation to ecological parameters. *An. gambiae s.s* and *An. arabiensis* breed in temporary (3-5 weeks) stagnant waters that are turbid and lack in vegetation or surface film (Powell *et al.*, 1999; Gimnig *et al.*, 2001). *An. arabiensis* has also been shown to prefer permanent / semi-permanent artificial habitats such as rice fields (White *et al.*, 1972; Githeko *et al.*, 1996).

Marked differences in vector capacities of various species within a sibling species complex necessitate that each member is distinctively recognised

(Besansky, 1999). The distinct recognition and analysis of each sibling species within a sibling species complex is important in understanding the exact role a vector species has in disease transmission. The sympatric existence of sibling species has often complicated malaria control programmes and measures (Mnzava, *et al.,* 1995). In places where *An. gambiae* s.s co-exists with *An. arabiensis* it has been less successful to prevent malaria transmission using insecticide control. This is because a proportion of vector population is constantly exophilic (Mnzava, *et al.,* 1995) hence it does not come into contact with 'insecticide-sprayed' walls.

1.2.5 Malaria Transmission And Vector Dynamics.

The *Anopheles gambiae* complex mosquitoes comprise the most flexible genetic systems in nature (Coluzzi *et al.,* 1979; Collins & Besansky, 1994). Their genetic flexibility is evidenced in their rapid adaptation to changing environments (Coluzzi *et al.,* 1979; Collins & Besansky, 1994). The rapid adaptation to a new environment results in genetic divergence within a species leading to possible reproductive isolation and speciation (Powell *et al.,* 1999). The rapid adaptation of *Anopheles gambiae* complex mosquitoes to new environments has increased the geographical area affected by malaria.

Some of factors that cause vector infiltration to new areas are changes in land use and global climatic changes (WHO, 1998; Githeko *et al.,* 2000) and population movements. Climatic changes and changes in land use affect the biology and ecology of mosquito vector species and consequently the risk of disease transmission (Githeko *et al.,* 2000). The breaking up of virgin land for

agricultural activities encourages proliferation of mosquito-breeding habitats. The global warming due to the increase of greenhouse gases leads to increased temperatures of the atmosphere and the Earth's surface. This leads to a rise in temperatures of water, making the mosquito larvae mature faster, which results into increase in vector populations during transmission period. At a raised temperature, the female mosquito digests blood faster than usual and feeds more frequently therefore the transmission intensity is increased (Githeko *et al.*, 2000). In addition, at raised temperatures the malaria parasites take shorter incubation period in the mosquito vector. The consequence of this is that the proportion of infective vectors and risk of infection is higher during the transmission season.

1.2.6 Vector Identification And Population Genetics.

Population genetics is defined as the study of the genetic basis of naturally occurring variation within populations. Thus, it is the study of genetic composition of populations and of effects of factors such as selection, population size, mutation, migration and genetic drift on the frequencies of various genotypes and phenotypes. It involves analysis of population structure, which is defined as a relationship between collections of genes belonging to a population (Harlt and Clark, 1989), with an aim of evaluating the evolutionary forces that create variation within and between species. It involves the description of genetic variation in populations, and experimental and theoretical determination of how that variation changes in space and time.

The study and description of genetic population structure of vector species is of significance to malaria transmission dynamics and control. This is because the control programmes based on genetic modification of malaria vectors (Crampton *et al.*, 1994; Gwadz, 1994), will depend on understanding how genes associated with refractoriness to the parasite will move within and between populations. The spread of insecticide resistance genes in vector populations can also be predicted based on the population genetic structure of a population. The description of population structure of vector species starts with identification and characterisation of vector species. The characterisation is necessary because population genetic structure can only be meaningful when single panmictic unit is considered. Various techniques and methods are used for identification and description of population structure of disease vectors. These include crossing experiments, cytogenetic analysis, protein characterisation and molecular studies.

1.2.6.1 Crossing Experiments.

The crossing experiments (White, 1974) were used for the most part in identifying sibling species in the *Anopheles gambiae* complex. This technique involves mating laboratory strains of known identity with wild caught mosquitoes to determine with which member they interbreed successfully and hence be conspecific. The crossing experiments between the different members of the *Anopheles gambiae* complex mosquitoes result in sterile F_1 males (White, 1974) showing reproductive isolation between the involved mosquitoes. This method is not a reliable method of showing reproduction isolation since the normal behaviour

of the anopheline mosquitoes is grossly disturbed in laboratory cages (White, 1974). This results in ethological or behavioural mechanisms that isolate different species in nature breaking down and allowing mating to occur between species that never hybridise in nature. The inter-fertility cannot also be used conclusively to decide con-specificity. This is because between some *An. gambiae* complex mosquitoes inter-specific crosses produce fertile female hybrids (Coluzzi, *et al.*, 1979; White, 1974). Crossing experiments are good only in demonstrating the post-mating barriers to gene flow through hybrid sterility or inviability.

1.2.6.2 Cytogenetic Analysis.

Cytogenetic analysis identifies each mosquito species in a sibling species complex based on its characteristic chromosome-banding pattern (White, 1974; Coluzzi *et al.*, 1985). The identification depends on recognition of fixed chromosomal inversions in squash preparations of polytene chromosomes (Coluzzi *et al.*, 1985). The possession of different fixed chromosomal inversions is indicative of reproductive isolation between different species (Coluzzi *et al.*, 1985) and within a species (Toure *et al.*, 1998). In the population genetic analyses, the polymorphic chromosomal inversions are applied by comparing the frequencies of the different inversions. The cytogenetic studies are more informative in terms of ecological adaptation of various mosquito species (Toure *et al.*, 1998; Alessandra della Torre *et al.*, 2001). However, some inversions have shown strong association with specific ecological zones (Coluzzi *et al.*, 1979) hence they are not favourable for population genetics studies.

Cytogenetic studies have also been used to analyse and identify species in the *Anopheles gambiae* complex (White, 1974; Coluzzi *et al.,* 1985). Anopheline mosquitoes have low chromosome number, one pair of heteromorphic sex chromosome X and Y and two pairs of autosomes numbered 2 and 3. These chromosomes are capable of undergoing endoreplication process, which involves repeated replication without occurrence of any mitosis-like event (Clements, 1992). The results of endoreplication are polytene chromosomes, which appear as bundles of extended interphase chromatids characterized by peculiar banding patterns (Alessadra della Torre, 1997).

The use of polytene chromosomes to characterise *Anopheles* mosquitoes utilizes the paracentric inversions (Coluzzi *et al.,* 1985) that occur along the chromosomal arms. Two kinds of paracentric inversions exist, namely, fixed and polymorphic inversions. An inversion is said to be polymorphic when both forms of the inversion, the standard and inverted forms are present in a species. The fixed inversions occur at homozygous state only (Coluzzi *et al.,* 1985) and are used for cytotaxonomic identification of the six species in *An. gambiae* complex (Coluzzi *et al.,* 1985; Coluzzi, 1988) because each species carries a different fixed chromosomal inversion. The polymorphic inversions are used in vector analysis at intra-species level. Distribution of the polymorphic inversions among the six sibling species of *Anopheles gambiae* complex mosquitoes is not uniform. The chromosomal inversion polymorphism is more common in *An. gambiae* s.s and *An. arabiensis* than in other members of the complex (Coluzzi *et al.,* 1985).

1.2.6.3 Protein Characterisation.

Protein characterisation (Coluzzi, 1988) has also been applied to vector analysis. The technique is based on the differential migration of proteins within a matrix, which is detected in an electric field. The migration is a function of electric charge and molecular weight of the protein. The differential migration of enzymes can be translated as the presence of different alleles coding for the target enzyme in the population. Protein characterisation is applied to study gene flow and genetic differentiation by evaluating allele frequencies using the population genetics tools (Coluzzi, 1988). The technique has the advantage that it can be applied to all adult males and females regardless of the stage in the life cycle. The limitation of this technique is that it is only more effective when used on distantly related species. This is because recent divergence may not provide scorable polymorphism within the enzyme variants (Favia *et al.,* 1999). This technique is also limited by the need for fresh or frozen specimens for analysis and this makes it generally expensive as require a sophisticated and well-equipped laboratory, which was not available for this work.

1.2.6.4 Molecular Studies.

Molecular studies on the *Anopheles* species was revolutionised by the advent of PCR. The molecular techniques rely on PCR and DNA based markers in analysing different individuals. PCR generates multiple copies of the DNA from initial small amount of sample DNA. Markers applied in *Anopheles* mosquitoes molecular studies include, ribosomal DNA (rDNA) (Cornel & Collins, 1996), microsatellites (Budowle *et al.,* 1991; Collins, et al 2000), restriction fragment length polymorphism (RFLP) markers and randomly amplified polymorphic DNA (RAPD) markers (Favia *et al.,* 1997).

Two methods are commonly used for identification of *Anopheles gambiae* complex mosquitoes. These methods are: (a) the sequence variation in the ribosomal DNA (rDNA) intergenic spacer (IGS) region (Cornel & Collins, 1996) and (b) chromosomal paracentric inversion analysis (White, 1974; Coluzzi *et al.,* 1985). The sequence variation in the ribosomal DNA (rDNA) intergenic spacer region (IGS) (Cornell & Collins, 1996) has been commonly used for identification of *An. gambiae* complex mosquitoes.

The rDNA-PCR technique utilizes primers designed on the basis of fixed sequence differences in rDNA of the mosquitoes. In the genome of *Anopheles gambiae* complex mosquitoes, the rDNA exists in many repeat units that have very high degree of species-specificity and conservation (Cornel & Collins, 1996). This

clearly identifies and separates members of *Anopheles gambiae* complex. The technique utilizes five PCR primers, one universal plus strand primer designed from a conserved 28S gene sequence and four minus strand primers designed from the intergenic spacer region (IGS) of rDNA of individual species. The minus strand primers are species specific and anneal only with rDNA of the species it was designed to amplify (Cornel & Collins, 1996). The fragments generated by this PCR technique are species specific. They vary in base pair lengths as follows: 153 bp for *An. quandriannulatus,* 315 bp for *An. arabiensis,* 390 bp for *An. gambiae* s.s, two bands of 390 and 690 bp for *An. bwabwae,* 464 bp for *An. melas* and 466 bp for *An. merus* (Cornel & Collins, 1996). The bands are visualised using UV transilluminator after electrophoresis of PCR products in ethidium bromide stained agarose gels.

The microsatellite are trimeric and tetrameric short tandem repeats (STRs) of DNA with high variability among individuals (Budowle *et al.,* 1991). Microsatellites loci are used in population studies due to sufficient variation in the number of tandem repeats between individuals caused by high mutation rates at these loci. These mutations result in new alleles that differ from parental alleles by the number of copies of the simple repeat sequence. The microsatellite loci are neutral markers (Anderson *et al.,* 1999) i.e. they are not under selection. Markers under selection are not suitable for population genetic studies because they give a distorted view rather than a true population genetic structure. This is because it is selection that determines patterns of distribution of alleles for such markers within populations (Anderson *et al.,* 1999). Multiple microsatellite loci are preferred in

population genetic studies because population structure characteristically shows high levels of variance among loci and also gene flow varies among regions of mosquito genome (Lanzaro *et al.,* 1998).

The microsatellite loci have been used to make inferences into population structure of *Anopheles gambiae* ((Lehmann *et al.,* 1996; Walton *et al.,* 1998; Kamau *et al.,* 1999). For example in a study by Walton *et al.,* (1998) microsatellite analysis showed substantial genetic divergence between chromosomal forms of *An. gambiae* s.s from different ecological zones. This demonstrated the power of the microsatellites to detect barriers to gene flow within and between closely related species hence making them valuable tools for population structure studies as was done in this thesis work.

1.2.7 Barriers To Gene Flow Between Species.

An analysis of population structure involves defining factors that create genetic variation within a population (Harlt and Clark, 1989). Therefore, it is important to define types of barriers to gene flow that can cause the genetic variations and divergence, which lead to reproductive isolation within a population.

In sexually reproducing organisms, individuals belonging to same species interbreed and contribute to the species gene pool's variation in space and time (Coluzzi *et al.,* 1977). Members of different species do not usually interbreed due to the barriers to gene flow, which reproductively isolate them. These barriers can either be biological or physical. The biological barriers to reproduction are of two general categories namely pre-zygotic and post-zygotic. Pre-zygotic isolating

mechanisms prevent the formation of hybrids between members of different populations through two main ways, ethological and ecological isolation. The post- zygotic isolating mechanisms reduce viability or fertility of hybrids or the progeny.

1.2.7.1 Ecological Isolation.

Ecological isolation results when species occupy habitats in different ecological zones. This limits the contact between members of different populations or even same population hence there is no mating or gene flow between them. The populations of *Anopheles merus* and *Anopheles melas* exhibit this type of isolation. They are found along the East and West Africa coasts, respectively. Ecological isolating mechanism acts as a barrier between species of different populations or even a single species whose various members occupy different ecological zones. For example, forest and savanna chromosomal forms of *An. gambiae* s.s were found to carry different inversions on chromosome 2 that showed absence of gene flow between the two forms and correlated with clines of climatic conditions and vegetation zones (Coluzzi *et al.*, 1985).

1.2.7.2 Ethological (Behavioural) Isolation.

Populations of different species of organisms have different courtship rituals that lead to mating and fertilisation. When these rituals are not in harmony between two species, mating cannot occur. The disharmonious courtship rituals

make sexual attraction between males and females of different populations weak or absent. The courtship rituals are specific to a species and play a significant role in species identification. The ethological reproductive isolation mechanisms (Coluzzi, *et al.,* 1977) are very potent and they can be strong even among closely related species. It is the main isolating mechanism between sympatric species. This form usually isolates individuals of a species that is undergoing incipient speciation process.

1.2.7.3 Hybrid Inviability And Sterility.

In the occasions where the pre-zygotic isolating mechanisms are absent or break down like when mosquitoes are put into cages (White, 1974), inter-specific zygotes are formed. Hybrid inviability and sterility prevent the subsequent gene flow in such circumstances. The interspecific zygotes often fail to develop into mature individual adults capable of reproducing, i.e. inviable. However, when they develop into mature adults they fail to develop functional gametes, i.e. sterile. This form of barrier to gene flow has been seen in *Anopheles gambiae* mosquitoes where crosses between some sibling species produce sterile males (White, 1974). Although crosses between some sibling species in the *An. gambiae* complex can produce viable females capable of reproducing, this is prevented in nature by the interplay of ethological and ecological variations.

1.2.7.4 Intra-Species Isolating Mechanisms.

The geographic and ethological isolating mechanisms are main barriers to gene flow within a species. Geographic barriers to gene flow include physical landmarks like the rift valley, which has been shown to play a role as a barrier to gene flow in the *An. gambiae* mosquitoes (Kamau *et al.,* 1999). Large water bodies like the oceans and lakes, deserts and harsh terrain can also prevent the gene flow within a species. Ethological/behavioural isolating mechanisms develop between members of a single species as they diverge. As members of a species diverge, they fail to recognise each other's unique mating rituals and therefore fail to interbreed.

1.3 THE JUSTIFICATION OF THIS STUDY.

Anopheles arabiensis Patton is one of the important malaria vectors in sub-Saharan Africa. It has been found in larval breeding habitats of varying characteristics (Gillies and De Meillon, 1968; McCrae, 1984; Gimnig, *et al.,* 2001). To understand the fluctuations in adult mosquito populations and transmission intensity of malaria, factors that influence breeding site choice by the vector populations should be understood. Studies carried out on *Anopheles* breeding habitats have involved determination of characteristics such as presence of algae and vegetation cover of the breeding habitats (Gimnig *et al.,* 2001) and of biotic and abiotic factors that determine presence of mosquito larvae in breeding habitats

(McCrae, 1984). It is hypothesised that mosquitoes select breeding habitat depending on habitat size and permanence and that different breeding habitat types have different populations of the same species.

An. arabiensis is known to exploit both temporary water collections (Gimnig *et al.,* 2001) and semi-permanent/permanent standing water (Githeko *et al.,* 1996) as larval breeding habitats. This study was designed to determine whether *An. arabiensis* populations spatially partition breeding habitats. Two types of breeding habitats were considered, namely: (a) permanent/semi permanent water bodies (e.g. rice fields) and (b) transient water collections (e.g. footprints, hoof prints and car tyre ruts).

The hypothesis of this study was that *An. arabiensis* mosquito populations select breeding habitats based on size and permanence, such that some populations of this mosquito will deposit eggs in large water bodies (permanent) while others deposit their eggs in transient water pools. And it is possible that different larval breeding habitats may be carrying reproductively isolated individuals of this species. Also, it may be possible that breeding habitats of same type carry species of a panmictic unit. Information on this can be got through undertaking a population genetic study of a target vector species to understand population structure of its members found in different types of breeding habitats, as this study proposed.

The knowledge on whether there is partitioning of larval breeding habitats within malaria vectors is important in understanding the vector dynamics and how different land use patterns affect the population structure of *An. arabiensis*. The

presence of several sub-populations in one type of breeding habitat may suggest that such habitat is favoured most by the species, while the presence of one sub-population predominantly in one habitat type would suggest breeding habitat partitioning or competitive exclusion. In relation to malaria control, results of this study, which was designed to determine the presence of gene flow in *An. arabiensis* mosquitoes collected from different breeding habitats could be useful in predicting the outcome of the use of genetically modified mosquitoes. Because it can be used to predict the refractory genes would spread throughout the entire *An. arabiensis* population. The information would also be useful in prediction of spread of insecticide-resistance genes in vector populations and in understanding the role that diverse land use can have on the genetic diversity of malaria vector *Anophèles arabiensis*.

CHAPTER TWO: OBJECTIVES OF THIS STUDY.

2.1 General Objective.

To determine whether *Anopheles arabiensis* populations from two larval habitat types, namely, (a) permanent/semi-permanent standing water bodies, and (b) small transient fresh water pools in Ahero and Nyakach areas of western Kenya form separate panmictic units by analysing variability at micro-satellite loci.

2.1.2 Specific Objectives.

2.1.2.1 To identify the similarities and differences in population structure of *An. arabiensis* populations from Ahero and Nyakach found in permanent/semi-permanent and transient breeding habitats.

2.1.2.2 To determine levels of gene flow between the *An. arabiensis* populations from permanent/semi-permanent breeding habitats and transient breeding habitats at two sites in Ahero.

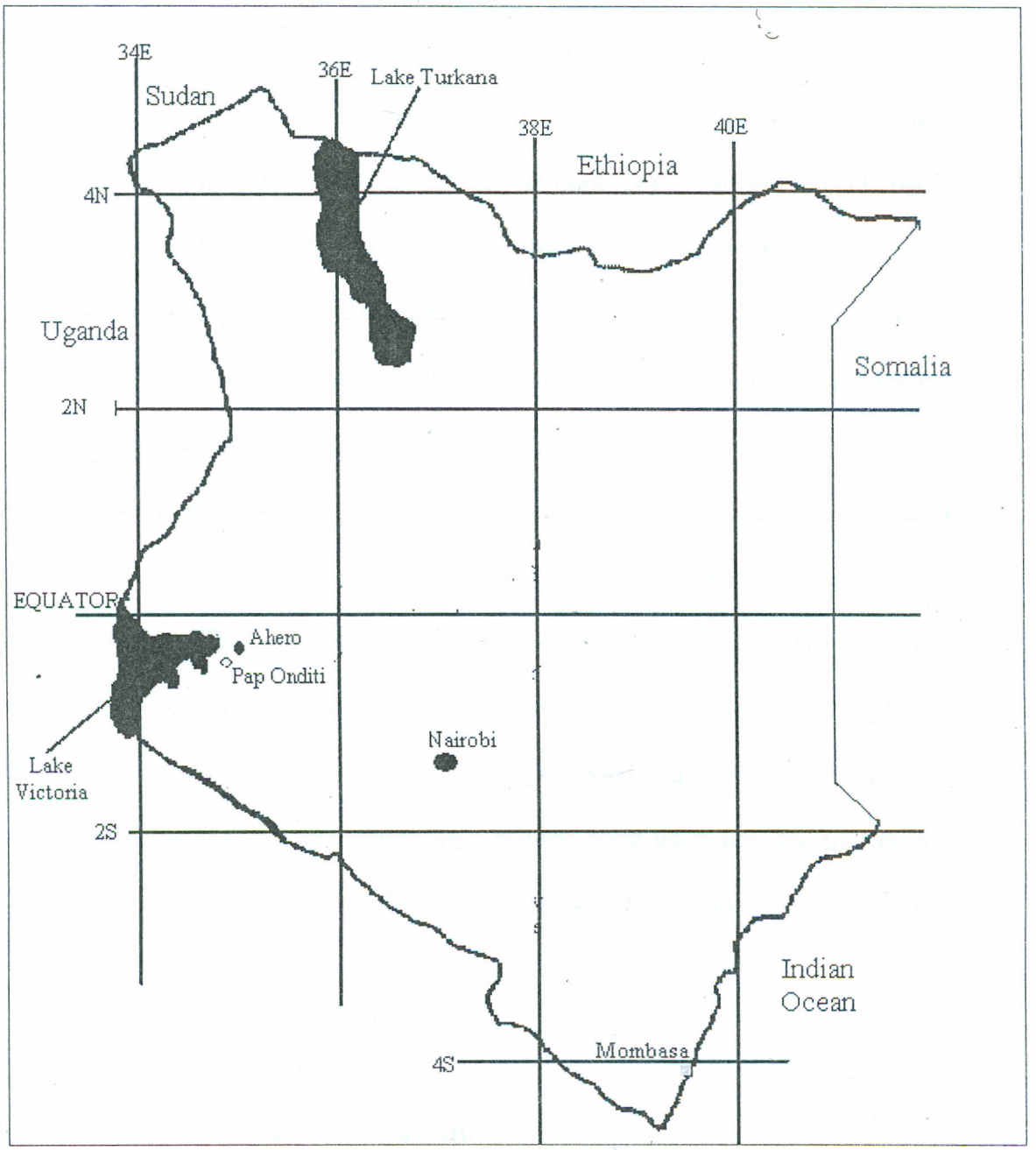
CHAPTER THREE: MATERIALS AND METHODS

3.1 Study Sites.

This study was carried out in Ahero and Nyakach (Pap Onditi) areas of Nyanza province in western part of Kenya, see Figure 2. Ahero (0°10'17. 16"S 34°55'07. 73"E 1,138m) is in Nyando district, it is a rice growing area. The permanent and semi-permanent water bodies mostly rice fields are the abundant mosquito breeding habitats in this area. Malaria vector reported in Ahero is *An. arabiensis* (Githeko *et al* 1996). The Nyakach study site, (Pap Onditi) (0° 17' 46.55" S 34° 59' 18.92"E 1,193m) is not a rice growing area as Ahero site and the only large water body in this area is a seasonal stream. The most abundant mosquito breeding habitats in the area are transient fresh water collections in, footprints, tyre ruts, cattle hoof-prints, etc. Both study sites experience rainy and dry seasons. The rainy season is experienced between months of March and July while dry spell is between October and February (Githeko per communication). The average temperature is 23 degrees C and average relative humidity is 64.9%.

Sampling was carried out in permanent/semi-permanent water bodies (>1m diameter) and transient water collections (<30cm diameter). Two sites were sampled in Ahero for both habitat types and one site was sampled in Nyakach study site for transient habitat type.

Fig. 2: **Map of Kenya showing the two study sites: Ahero site and Pap Onditi, Nyakach site.**



3.2 Mosquito Larvae Collection.

The specimens were randomly collected for a period of five days at each site. This decision was aimed at reducing the chances of sampling larvae from different egg batches laid by same female mosquito since egg-laying cycle takes between two to four days. The collection was done using standard World Health Organization (1975) sampling techniques, which included dipping and pipetting. The ladle (dipper), pipettes and turkey basters were used to scoop and pick the larvae from the breeding habitats. After scooping, one larva was picked per habitat to avoid sampling siblings and kept in plastic containers (poly pots) for transportation to the laboratory. Fourth instar larvae were preferred because of their larger size, which made processing easier than would have been with either first or second instars. They were identified based on their orientation in water. The anophelines lie parallel to the water surface; this characteristic was used to strictly pick anophelines only. All anophelines were processed and assayed by PCR to get the targeted species, *An. arabiensis*. Two primers were used, one specific for *An. arabiensis* and the other specific for *An. gambiae* s.s. Primers for other anophelines were not available.

3.3 Larvae Desiccation.

The fourth instar larvae identified as anophelines based on their orientation on water surface were put individually into 1.5ml vials then dried in anhydrous Calcium Sulphate before storage. The desiccation process was aimed at preventing

the decay of specimens, which causes the degradation of DNA. During desiccation, uncorked vials containing the specimens were put in a container with anhydrous calcium sulphate. The container was then sealed and kept for three days for water to be absorbed from the specimens.

3.4 Mosquito Larvae DNA Extraction.

An extraction method described by Cornel and Collins (1996) was used. An individual mosquito, (in this case, the fourth instar larvae,) was put into a 1.5ml vial and 100µl of grind buffer (0.08M NaCl, 0.16M sucrose, 0.06M EDTA, 0.1M Tris-HCl and 0.05% SDS) added. The pestle was used to grind the larvae until homogenous lysate was formed. The lysate was then incubated for 30 minutes in a water bath at 65^0 C then 14µl of 8M Potassium acetate was added and the mixture vortexed. The mixture was then incubated on ice for 30 minutes and then centrifuged for 10 minutes at 10,000 rpm. The supernatant was transferred into a new 1.5ml vial and 200µl of 95% ethanol added. This mixture was then kept at -20^0C for at least 20 minutes after which it was centrifuged at 10,000 rpm for 20 minutes. The 95% ethanol was then discarded and replaced with 70% ethanol, which was then poured off and replaced with 95% ethanol again. The 95% ethanol was then poured off and the tubes dried (inverted on blotting paper) overnight at room temperature. Following day, the pellet DNA was suspended in 100µl sterile distilled water and kept at -20^0C until needed for analysis.

3.5 *Anopheles gambiae* Species Identification.

The *Anopheles gambiae* and *Anopheles arabiensis* sibling species were identified by rDNA Polymerase Chain Reaction (rDNA-PCR) and electrophoresis technique of Cornel and Collins (1996). The PCR was done using PERKIN ELMER^{TM} Gene Amp PCR system 9600. During the rDNA-PCR, each reaction mixture of 15µl contained 10X PCR buffer (10mM Tris [pH 8.3], 50mM KCl, 1.5mM $MgCl_2$), 25mM $MgCl_2$, 10mM dNTPs, (N= adenine, guanine, cytosine, and thymine), 0.5U *Taq* polymerase, 10pmol of each primer and 1µl DNA template. The master mix protocol is given in *Table 1* below.

Table1: **The reaction mixture for species identification.**

Component	Volume
Distilled sterile water	8.8µl
10X PCR buffer	1.5µl
dNTP mix (10mM)	1.14µl
Primers (GA, AR, UN)	0.7µl
$MgCl_2$ (25mM)	1.8µl
Taq polymerase	0.06µl
DNA template	1µl
Total	15µl

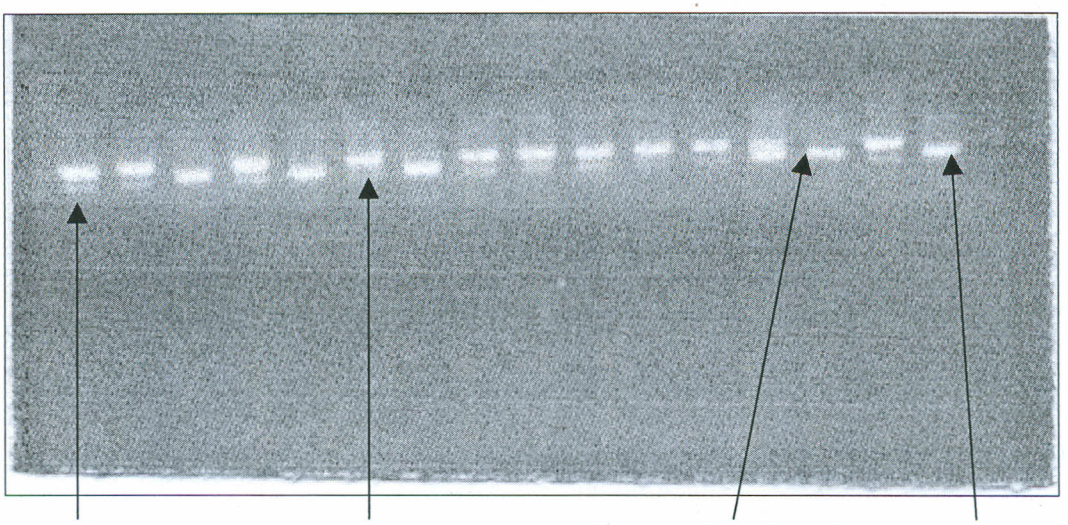
The primer sequences used were as indicated below:

Universal primer (UN)	GTG TGC CCC TTC CTC GAT GT
An. gambiae primer (GA)	CTG GTT TGG TCG GCA CGT TT
An. arabiensis primer (AR)	AAG TGT CCT TCT CCA TCC TA

The cycling conditions were as follows: 20 s at 95^0 C, 30 s at 55^0 C, and 30 s at 72^0C. The cycle was repeated 30 times and a final extension at 72^0C for 5 minutes included.

After the polymerase chain reaction (PCR), the amplified DNA was loaded onto a 3% Agarose gel in the electrophoresis tank (E-C Apparatus Corporation, St. Petersburg, Florida) and electric field applied. The 3% agarose gel contained 3µl of ethidium bromide, which enabled the separated bands to be visualized by UV illumination. The bands were photographed for future reference using Polaroid film (Polaroid Corporation Cambridge, MA 02139 USA). See Figure 3 for the schematic representation of agarose gel electrophoresis of PCR products for *An. arabiensis* and *An. gambiae* specimens from Ahero.

Fig. 3: Schematic representation of agarose gel electrophoresis of PCR products for *An. arabiensis* and *An. gambiae* specimens from Ahero.



To get the controls, PCR was run for 30 specimens using primers specific for *An. gambiae* s.s and *An. arabiensis.* The specimens which gave PCR products with primer specific to *An, gambiae* were taken as controls for *An. gambiae* in the subsequent reactions and those that gave PCR products with primers specific for *An. arabiensis* were taken as controls for subsequent reactions.

3.6 Microsatellite Analysis.

The microsatellite loci used in this study were initially designed for *An. gambiae* s.s (Zheng *et al.,* 1993) studies. They were applied in this study due to close relatedness of the two species (Besansky *et al.,* 1994; Kamau *et al.,* 1998). Table 2 shows the names, the cytological location, length and primer sequence of each locus used in this study.

Table 2: **The microsatellite loci used in the study.**

Locus	Cytological location	Length (bp)	Primer sequence (5'-3')
AG2H143	II	160	CGTACGAGTGAGTGAGTTGG
			CAAAAATAGCATCACGGCCG
AG2H161	II	92	TTCACCTGTCCCGTGTGGTC
			GGAACTTTCGGTGCTGTAGG
AG3H83	III	70	TTTGTTCCAAACGGGGTGCG
			CCAGCGCGAACTATGGGC
29C	III	148	TGTTGCCGGTTTGTTGCTGA
			ATGTTCCAGAGACGACCCAT
AG3H577	III	113	TTCAGCTTCAGGTTGGTCTC
			GGGTTTTTTGGCTGCGACTG
AG3H746	III	105	TGGGTTCGAAATTCGCCAAC
			GACGTGTGCACCCGTTGTG
ID1	X	135	GTTATCCACTGCGCATCATG
			TAATGGTCCCAAATCGTTGC
AGXH99	X	127	CGGGAATTTGTTGCTTCCTG
			TCGCCCTCTTTCTCCATCTC

3.6.1 Amplification Conditions.

The method used for the amplification of DNA for microsatellite analysis was that of Kamau, *et al.,* (1998). In this method, 1µl of the *An. arabiensis* DNA was amplified in 9µl PCR mixture containing primer concentration of 0.04µl/ml, 25mM $MgCl_2$, 10mM deoxynucleotides (dNTPs) and 0.5U Taq polymerase plus 10x buffer in Perkin Elmer PCR machine. Different reaction mixtures containing primers specific for particular loci were amplified in 0.2ml thin-walled micro-tubes. Thermal cycling conditions included initial denaturation step at 94^0C for 5 minutes followed by 30 cycles of 94^0 C for 15 s, 56^0C for 20 s, and 72^0C for 30s and final extension at 72^0C for 5 min. The master mix protocol for microsatellite amplification is given in Table 3 below:

Table 3: **The reaction mix protocol for microsatellite analysis (one specimen).**

Component	Volume
Distilled sterile water	1.69µl
10X PCR buffer	0.9µl
dNTP mix (10mM)	0.54µl
MgCl2 (25mM)	0.27µl
Primers (0.04µl/ml) @	2.25µl
Taq polymerase	0.1µl
DNA template	1µl
Total	9µl

3.6.2 Allelic Ladder Preparation.

To construct the allelic ladders, the method of Kamau *et al.*, (1998) was used. Briefly, PCR was performed for 30 specimens for a given locus. Then specimens that contained a whole spectrum of alleles seen at a particular locus were selected and amplified in 70µl reaction mixtures. Electrophoresis was then performed and the DNA was then recovered from the gel matrix after silver staining. Briefly, the extracted gel portion containing amplified DNA for each allele was scooped and placed in separate well-labelled 1.5ml vials. 400µl TE buffer (1M Tris pH 7.4, 0.5M EDTA pH 8.0) was then added and the tubes vortexed and then incubated at 37^0 C overnight. On the following day, tubes were centrifuged for 3 minutes at 10000 rpm and the supernatant carefully picked by a pipette. The supernatant was then placed in a 15ml screw-cap tube and 100µl of 7M Ammonium acetate (NH_4Ac) and 1ml absolute ethanol added. The mixture was vortexed and frozen at -70^0C for 15 minutes. The tubes were then centrifuged and the supernatant discarded. The resultant pellet was washed with 500µl of 70% ethanol by centrifuging for 5 minutes. The pellet was then left to dry in the air for three hours or until the 70% ethanol has dried off completely. Then 50µl of TE buffer was added and the tube incubated overnight at 56^0 C for DNA to be resolubilized. Following day, the tubes were centrifuged and electrophoresis performed to check for the presence of the extracted alleles. Obtained alleles for a particular locus were then mixed, serially diluted and amplified to form allelic ladders. The alleles were then numbered consecutively from bottom to top depending on their sizes. The allelic band with few base

pairs, which migrated faster in the gel, was assigned number 1. The immediate band behind it was assigned number 2. The third band in the series was given number 3 e.t.c. All specimens amplified at a particular locus were scored against the allelic ladder of that locus.

3.6.3 Electrophoresis.

The amplified DNA products were analysed by polyacrylamide gel electrophoresis (PAGE). Acrylamide gel containing 8% acrylamide, 5% cross-linker (PDA: Piperizine diacrylamide) and 60mM Formate on Gel Bond Page™ was prepared and allowed to set for about 30 minutes. The gel was then transferred onto an electrophoretic plater (E-C Corporation, St. Petersburg, Florida) maintained at 15^0C. Sample applicator tabs (Amersham Pharmacia Biotech, SE-751 84 Uppsala Sweden) cut into thirds were laid onto the gel at one end and 7µl of amplified DNA was placed onto each tab. Two strips of filter paper (Whatman USA) soaked in 0.28M Tris borate containing 0.002% Bromophenol blue was placed at each end on top of the gel and then the electrodes were connected and electrophoresis carried out for approximately two hours.

3.6.4 Silver Staining.

After the electrophoresis, the separated bands were visualized by silver staining, according to Cairns and Murray (1994) with modifications, Cairns and Murray (1994) used 0.1% formaldehyde this study used 0.27% formaldehyde.

Briefly, 6% polyacrylamide gels were soaked in a solution containing 10% ethanol and 0.5% acetic acid (fixing solution) for 3 min and then placed in freshly prepared 0.2% silver nitrate in fixing solution for 5 min. The gel was then washed three times with distilled water and then incubated in distilled water for 2 minutes. The gel was then developed in a reducing solution containing 3% sodium hydroxide and 0.27% formaldehyde. After the bands had developed the gel was washed in tap water and then placed in 10% acetic acid for 5 min to stop further development. The gel was washed again in tap water and then placed in 5% glycerol for at least 5 minutes to prevent hardening and cracking of the gel. After the gel had dried well, a transparency film was laid onto it to prevent the damage of the gel and facilitate easy storage and handling when scoring the alleles.

3.6.5 Scoring Of Alleles.

The specimens were scored for the genotypes according to the number of the allelic band that was at the same level with it after electrophoresis. For instance, if a given specimen had two alleles (heterozygous), one co-migrating with band 1 and the other with band 2 of the allelic ladder and they be at the same level after the electrophoresis. Then its genotype was scored as 1,2. When a specimen had only one allele, meaning it was homozygous at that locus, it was then scored as either 1,1, 2,2 or 3,3 depending on the number of the allelic ladder band lying at same level with it. Figure 4 and 5 below show how the specimen microsatellite bands and allelic ladder bands appeared after electrophoresis.

Fig. 4. **Schematic representation of polyacrylamide gel electrophoresis of PCR products for locus AG2H143 for *An. arabiensis* specimens from Ahero.**

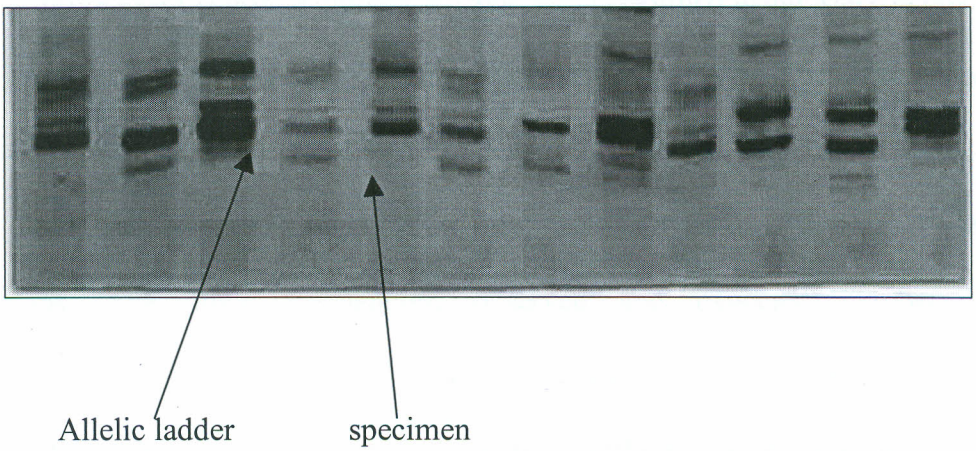
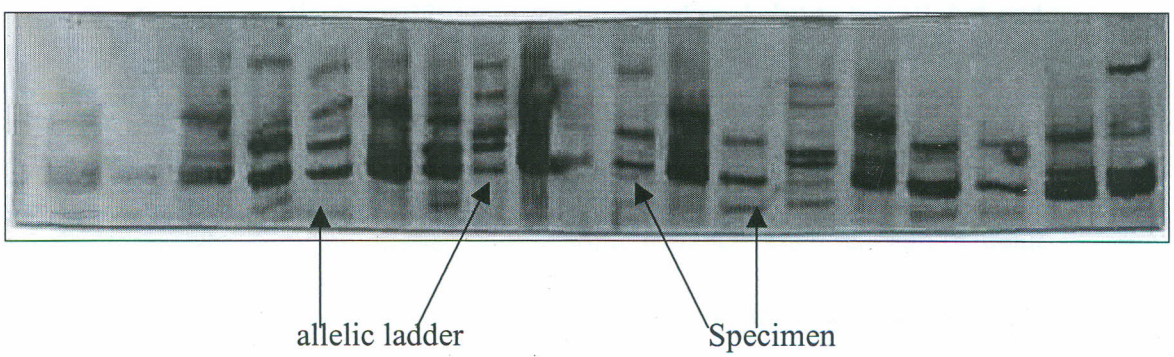


Fig. 5: **Schematic representation of polyacrylamide gel electrophoresis of PCR products for locus AG3H83 for *An. arabiensis* specimens from Nyakach study site.**



3.6.6 Data Analysis.

The data were analysed by GENEPOP version 3.2a (Raymond & Rousset, 1995a). The Hardy-Weinberg test was performed for each locus in each population. There were more than four alleles at some loci, therefore an unbiased estimate of the exact Hardy-Weinberg *P*-value was performed using Markov chain method available in GENEPOP. The chain was set at 2000 de-memorization steps, 2000 batches and 10000 iterations in order to get standard error of less than 0.005.

The Hardy-Weinberg equilibrium was tested for pooled and un-pooled populations. The allele frequencies, observed and expected number of homozygotes and heterozygotes were also tested across all populations. Difference between the observed and expected genotype distribution at each locus in all habitat populations were tested using genotype-based exact test of population differentiation available in GENEPOP (Raymond and Rousset, 1995a). The genotype frequency homogeneity was determined using an unbiased estimate of the *P* value of a log likelihood based exact test (*G*-like test), with Markov chain set as above. The significance of multiple tests was evaluated using the sequential rejective Bonferroni test (Holm, 1979) and the binomial probability test (Sokal and Rohlf, 1981).

Testing the significance of multiple tests is necessary when there are a number of tests to be done simultaneously. In some instances each test can be dealt with separately without connection to each other, however in most cases tests are connected to each other and totality of all tests is used to get a general picture (Holm, 1979). Testing significance of multiple tests gives protection against error

of type one i.e. rejecting true hypothesis. In this study, variation at six microsatellite loci was tested simultaneously using GENEPOP whereby *P* value of the test corresponds to the sum of the probabilities of all tables with same or lower probability (Raymond and Rousset, 1995a). There was a need to protect against first type error by requiring that tests have probability equal or less than the set threshold P value (0.05).

CHAPTER FOUR: RESULTS

A total of 500 anopheline mosquito larvae were collected from all the study sites. Out of these, one hundred and three, were from transient habitats in Ahero site1, ninety-five were from permanent/semi-permanent habitats in Ahero site 1, one hundred and two were from transient habitats in Ahero site 2, ninety were from permanent/semi-permanent habitats in Ahero site 2 and one hundred and ten were from transient habitats in Nyakach. After the polymerase chain reaction (PCR) the number of *Anopheles gambiae* s.s from all study sites was six while that of *An. arabiensis* from all study sites was two hundred and fifty. Number of mosquito specimens that did not yield PCR products was two hundred and forty four. Table 4 shows only those specimens that gave PCR products using primers specific to *An. arabiensis* and *An. gambiae.* At both study sites *An. arabiensis* occurred in higher proportions than *An. gambiae.*

Table 4: **The number of *An. gambiae* and *An. arabiensis* larvae collected from each site.**

Site	*An. gambiae* s.s	*An. arabiensis*	Total
Transient Ahero 1	0	56	56
Permanent Ahero 1	0	44	44
Transient Ahero 2	1	52	53
Permanent Ahero 2	3	52	55
Nyakach transient	2	46	48
Total	6	250	256

There was polymorphism at six of the eight loci used in this study. The polymorphic loci were: 29C, AG3H83, AG2H161, AG2H143, AG3H577 and AG3H746. Two loci ID1 AGXH99 were monomorphic i.e. each was fixed for a single allele and they were therefore discarded from further analysis. Table 5 below show number of alleles at each locus. Hardy-Weinberg equilibrium (HWE) test was also done for all studied populations and *P* values for each locus in each population are indicated in Table 5 below. The expected and observed number of heterozygotes and F_{IS} values in each population at each locus are also indicated in Table 5.

Table 5: **The genetic variability at all eight loci for the *An. arabiensis* populations studied.**

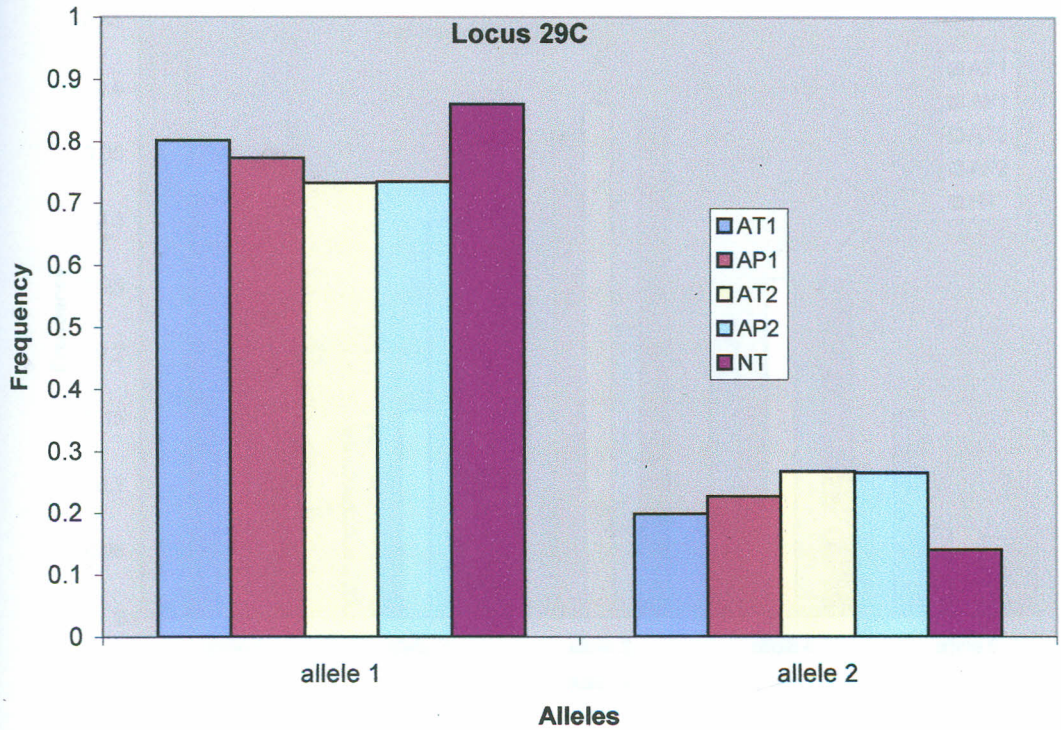
Population		No. of alleles,	Heterozygotes Obs.	Expt.	Hardy Weinberg F_{IS}	*P*
AT1 (*n*= 56)						
Locus	ID1	1	00	00	-	-
	AGX99 1	00	00		-	-
	AG2H161	3	27	26	-.019	0.44
	AG2H143	5	39	35	-.113	0.96
	AG3H83	5	29	44	0.34	0.000
	29C	2	7	17	0.59	0.000
	AG3H577	3	21	30	0.32	0.01
	AG3H746	3	55	28	-1	1
AP1 (*n*= 44)						
Locus	ID1	1	00	00	-	-
	AGX99 1	00	00		-	-
	AG2H161	3	7	20	0.65	.0.012
	AG2H143	5	28	26	-.086	0.67
	AG3H83	5	33	32	-.021	0.007
	29C	2	5	15	0.67	0.0001
	AG3H577	3	14	23	0.402	0.024
	AG3H746	3	44	22	-.96	1
AT2 (*n*= 52)						
Locus	ID1	1	00	00	-	-
	AGX99 1	00	00		-	-
	AG2H161	3	8	23	0.65	0.000
	AG2H143	5	38	31	-.22	0.99
	AG3H83	5	18	34	0.46	0.000
	29C	2	10	17	0.4	0.005
	AG3H577	3	20	29	.322	0.005
	AG3H746	3	44	25	-.92	1
AP2 (*n*= 51)						
Locus	ID1	1	00	00	-	-
	AGX99 1	00	00		-	-
	AG2H161	3	7	16	0.57	0.000
	AG2H143	5	34	30	-.15	0.96
	AG3H83	5	34	39	0.12	0.06
	29C	2	11	20	0.45	0.02
	AG3H577	3	32	29	-.077	0.7
	AG3H746	3	50	25	-1	1
NT (*n* = 46)						
Locus	ID1	1	00	00	-	-
	AGX99 1	00	00		-	-
	AG2H161	3	6	10	0.4	0.0001
	AG2H143	5	23	25	0.18	0.02
	AG3H83	5	17	30	0.4	0.000
	29C	2	6	10	0.42	0.02
	AG3H577	3	22	25	0.14	0.08
	AG3H746	3	20	16	-.201	0.05

AT= Ahero transient habitat, AP= Ahero permanent habitats and NT= Nyakach transient habitats.
obs= observed, exp= expected

All populations deviated from Hardy-Weinberg equilibrium at locus 29C as shown by significant P value ($P < 0.05$). Two loci, AG3H83 and AG2H161 showed deviations from Hardy-Weinberg equilibrium in four out of six populations. Locus AG3H577 showed deviations from Hardy-Weinberg equilibrium in three out of six populations ($P < 0.05$). Five out of six populations conformed to Hardy-Weinberg expectations at locus AG2H143. All populations conformed to Hardy-Weinberg expectations at locus AG3H746 ($P > 0.05$) as as already shown in Table 5.

The allelic frequencies for each locus were determined for each population using GENEPOP (Version 3.2a) (Raymond & Rousset, 1995). The allele frequency distribution for each of the six polymorphic loci is indicated in Figures 6 to 11. Each cluster of bars represents frequency of one allele in all habitat populations sampled at all study sites.

Fig. 6: **Allele frequency at locus 29C in different habitat populations**



Key:

AT1= Ahero transient habitats at site 1

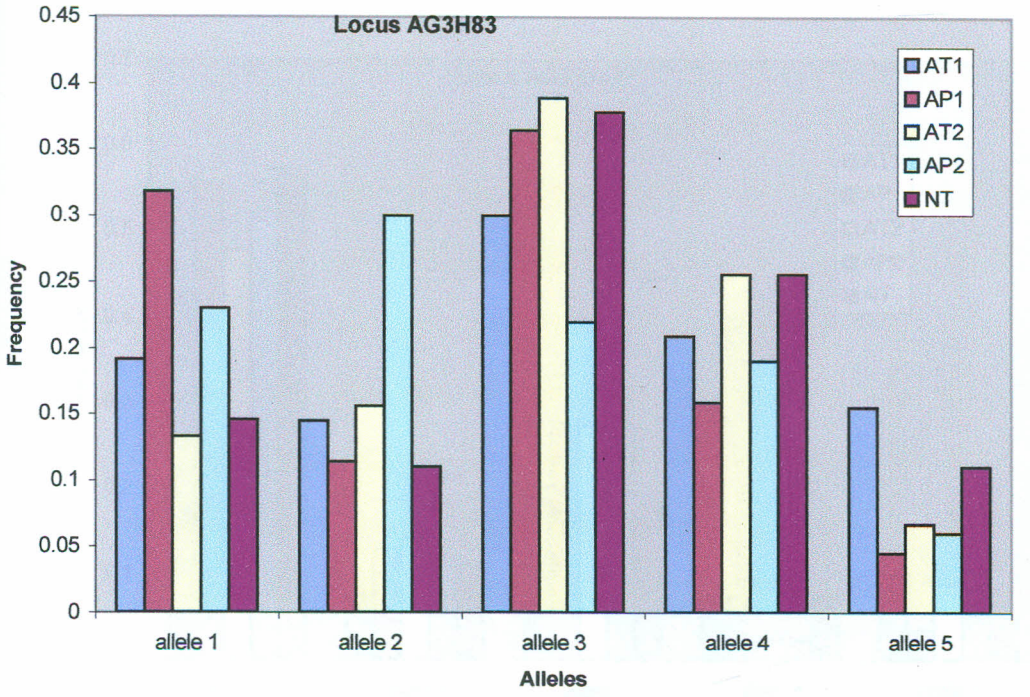
AP1= Ahero permanent habitats at site 1

AT2= Ahero transient habitats at site 2

AP2= Ahero permanent habitats at site 2

NT= Nyakach transient habitats.

Fig. 7: **Allele frequency at locus AG3H83 in different habitat populations**



Key:

AT1= Ahero transient habitats at site 1

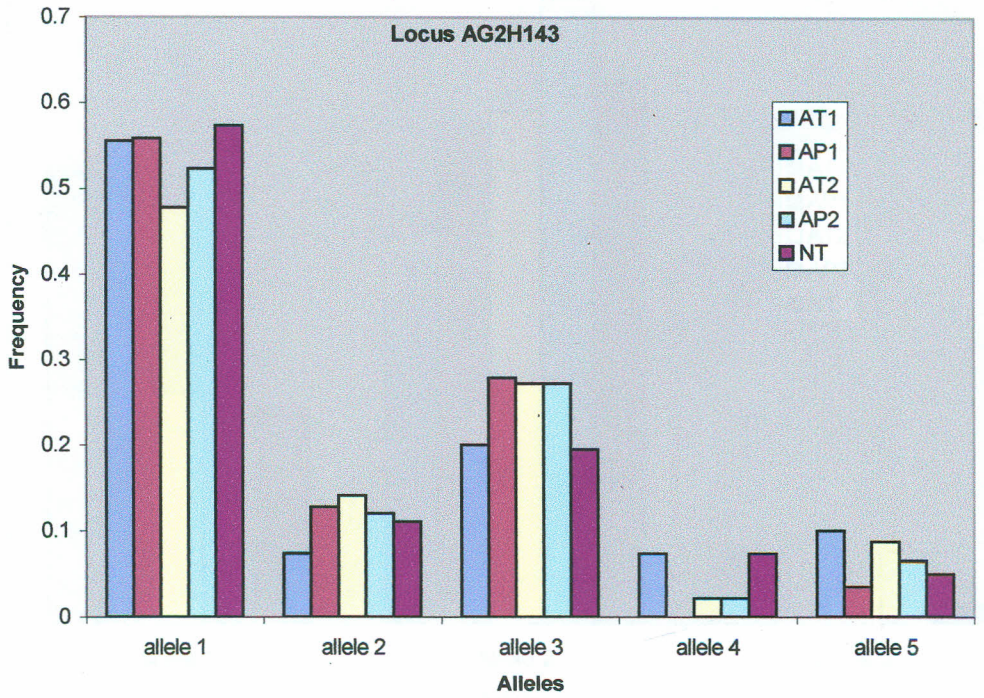
AP1= Ahero permanent habitats at site 1

AT2= Ahero transient habitats at site 2

AP2= Ahero permanent habitats at site 2

NT= Nyakach transient habitats.

Fig. 8: **Allele frequency at locus AG2H143 in different habitat populations.**



Key:

AT1= Ahero transient habitats at site 1

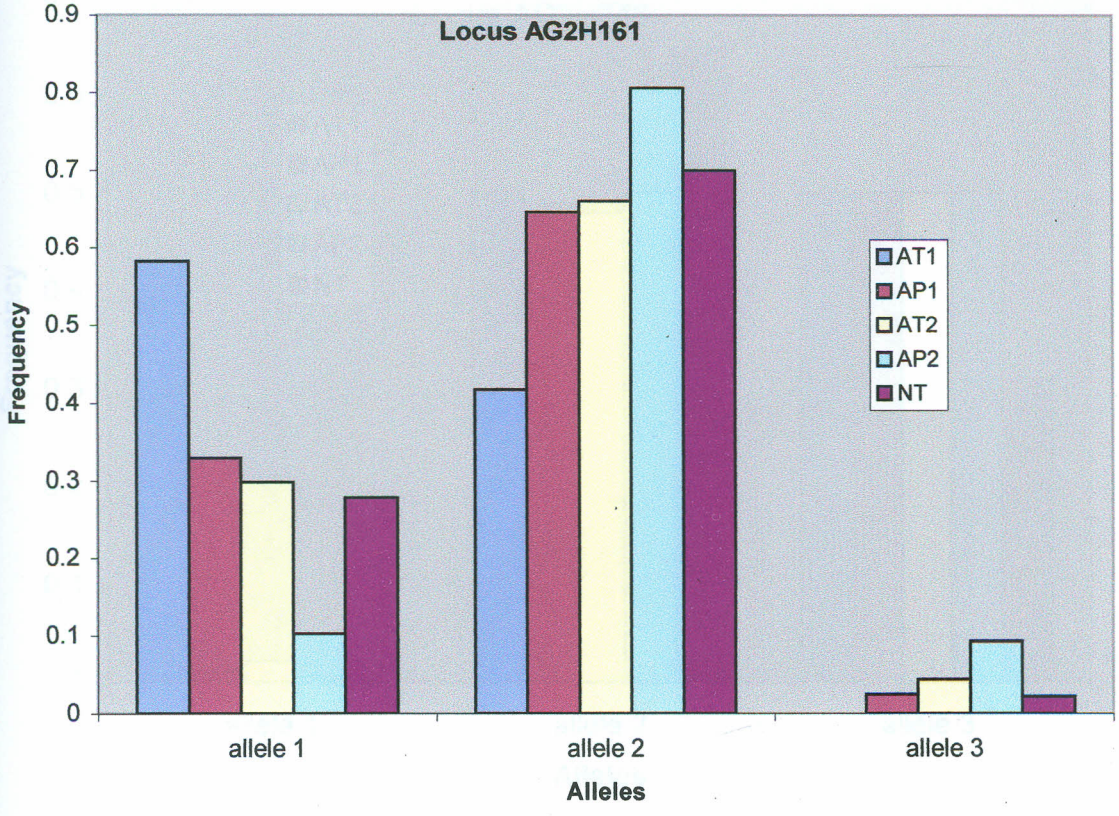
AP1= Ahero permanent habitats at site 1

AT2= Ahero transient habitats at site 2

AP2= Ahero permanent habitats at site 2

NT= Nyakach transient habitats.

Fig. 9: **Allele frequency at locus AG2H161 in different habitat populations.**



Key:

AT1= Ahero transient habitats at site 1

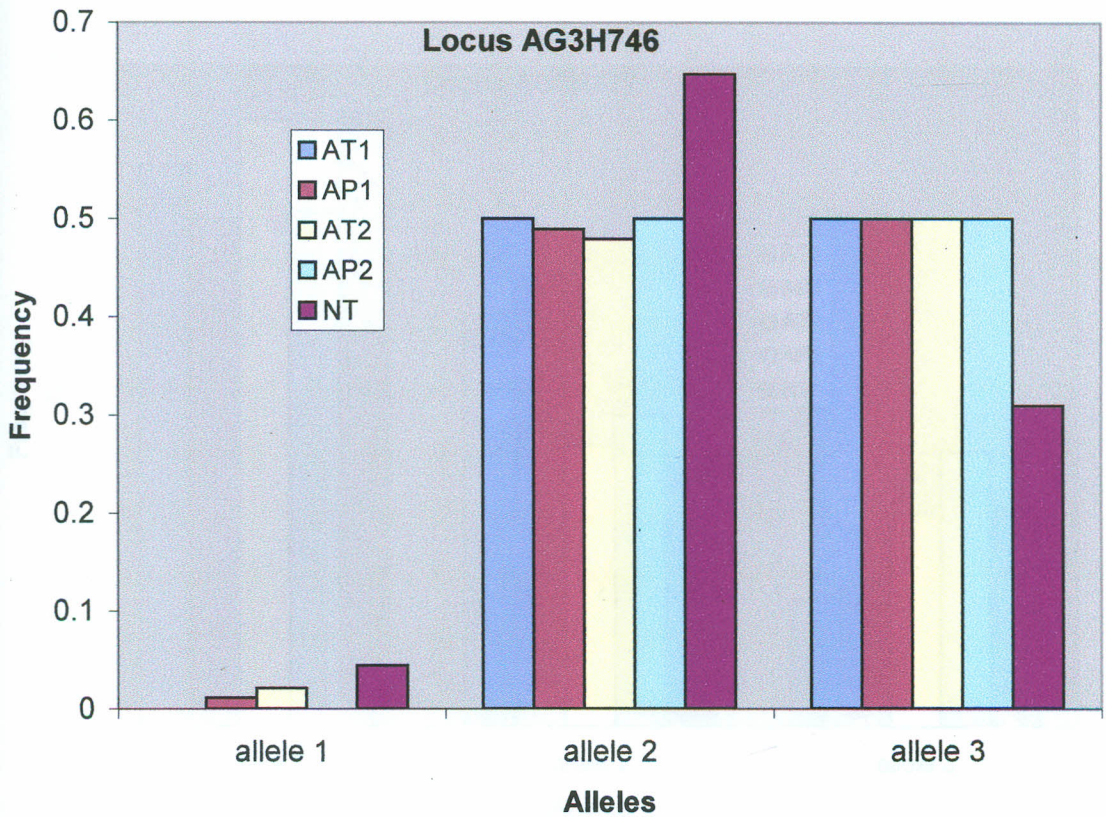
AP1= Ahero permanent habitats at site 1

AT2= Ahero transient habitats at site 2

AP2= Ahero permanent habitats at site 2

NT= Nyakach transient habitats.

Fig. 10: **Allele frequency at locus AG3H746 in different habitat populations.**



Key:

AT1= Ahero transient habitats at site 1

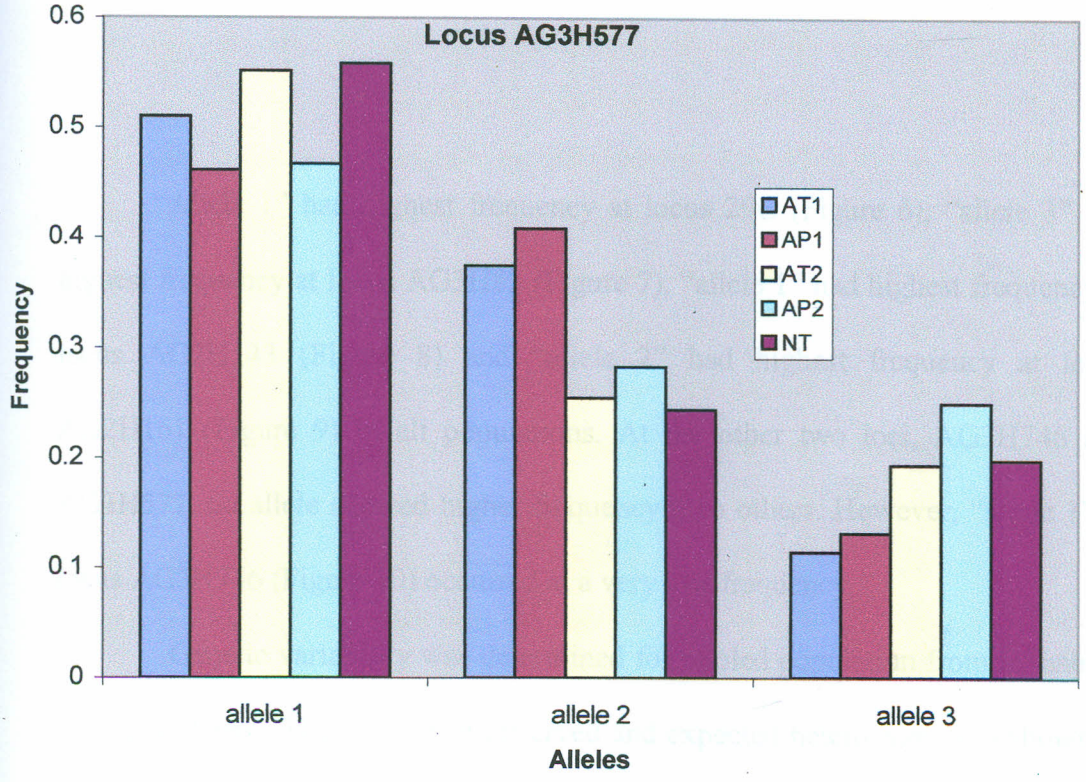
AP1= Ahero permanent habitats at site 1

AT2= Ahero transient habitats at site 2

AP2= Ahero permanent habitats at site 2

NT= Nyakach transient habitats.

Fig. 11: **Allele frequency at locus AG3H577 in different habitat populations.**



Key:

AT1= Ahero transient habitats at site 1

AP1= Ahero permanent habitats at site 1

AT2= Ahero transient habitats at site 2

AP2= Ahero permanent habitats at site 2

NT= Nyakach transient habitats.

"Allele 1" had highest frequency at locus 29C (Figure 6), "allele 3" had highest frequency at locus AG3H83 (Figure 7), "allele 1" had highest frequency at locus AG2H143 (Figure 8) and "allele 2" had highest frequency at locus AG2H161 (Figure 9) in all populations. At the other two loci, AG3H746 and AG3H577, no allele showed higher frequency than others. However, "allele 1" at locus AG3H746 (Figure 10) occurred at a very low frequency.

Genetic variability was determined for pooled population from all habitat types in Ahero. The number of observed and expected heterozygotes is shown in Table 6 below. Hardy-Weinberg equilibrium test was also done for the pooled Ahero populations and the *P* value for each locus in each population shown in Table 6.

Table 6: **The genetic variability in pooled *An. arabiensis* populations from Ahero sites.**

Locus	N	F_{IS}	H_E	H_O	*P* value (HWE)
29C	191	0.053	33	70	0.000
AG3H83	194	0.24	114	150	0.000
AG2H161	191	0.48	49	95	0.000
AG2H143	190	-0.15	139	121	0.0024
AG3h577	185	0.24	87	134	0.000
AG3H746	197	-.97	197	100	0.000

H_E: expected heterozygotes H_O: observed heterozygotes

There were significant deviations from Hardy-Weinberg expectations at all loci as shown by significant *P* values ($P < 0.05$). The observed heterozygotes were lower than expected in all the populations for loci, 29C, AG3H83, AG2H143, AG2H161, and AG3H577. At locus AG3H746 the deviation from Hardy-Weinberg expectation was due to higher number of observed heterozygotes than expected. These results suggest the presence of restricted gene flow within *An. arabiensis* populations collected from all types of breeding habitat in Ahero.

Genotypic differentiation was also assessed at all loci across all studied populations at the two sites. The observed genotypes, Hardy-Weinberg equilibrium *P* value and standard error at all loci are shown in Table 7. The overall genotype distribution was approximated using G-like test available in GENEPOP version 3.2a. This was used to determine whether the observed genotype frequency at each locus across all populations was in agreement with expected genotype frequencies. The sample size was large (> 200) and therefore the G statistic was approximated by the χ^2- distribution.

The approximated χ^2 value was 63.574 with 12 degrees of freedom and *P* value 0.0000 for the *G* statistic. The *P* value was significant and showed that observed genotypes were not at equilibrium and could not agree with the expected genotype distribution. The observed genotype frequencies deviated from a random mating situation, which is expected for panmictic populations. Suggesting non-random association of alleles in formation of different genotypes.

Table 7: **The genotypic differentiation at each locus across all studied populations.**

Locus	Observed genotypes	*P* values	S.E
29C	234	0.4068	0.0008
AG3H83	235	0.0146	0.0003
AG2H143	231	0.1039	0.0008
AG2H161	236	0.0000	0.00000
AG3H577	228	0.1699	0.0009
AG3H746	231	0.0030	0.0001

* S.E – standard error

The observed genotype frequencies at three out of six loci were not significant ($P > 0.05$), Table 7. However, when the overall genotype frequency distribution was determined for all the loci across all populations using the G-like test the results were significant with $P = 0.000$ as shown by the approximated χ^2 value above. Thus, the overall observed genotype distribution deviated significantly from the expected and this suggested the populations studied were not exchanging genes at random i.e. alleles were not associating randomly in formation of genotypes as expected in panmictic situation. The deviations from Hardy-Weinberg expectation were also observed after significance of multiple tests was determined using sequential rejective Bonferroni and binomial probability tests ($P < 0.05$).

CHAPTER FIVE: DISCUSSIONS

Anopheles arabiensis, Patton, was the most abundant member of the *An. gambiae* complex in Ahero and Nyakach areas. Out of the positively identified species, 97.66% was *An. arabiensis.* This species has been recorded as abundant member of the *Anopheles gambiae* complex in some parts of tropical Africa (Donnelly *et al.,* 1999; Petrarca *et al.,* 2000). *An. arabiensis* is known to be anthropophilic (Coluzzi *et al.,* 1979; Githeko *et al.,* 1996). It is able to acquire high *Plasmodium falciparum* infection rates (Petrarca & Beier, 1992) and it is endophagic and rests in houses (Githeko *et al.,* 1996). These factors coupled with its abundance in study areas provide it a major role in malaria transmission during malaria transmission season.

Anopheles gambiae s.s comprised 2.34% of the positively identified mosquitoes. A total of 244 specimens, representing 48.8% of the total collected specimens failed to amplify after three PCR trials. This suggested that other anophelines besides *An. arabiensis* and *An. gambiae* existed in the study area. Only *An. gambiae* s.s and *An. arabiensis* were assayed for in this study. The other anophelines could have been *An. funestus* Giles, *Anopheles pharoensis* Theobald, *An. rufipes* Gough or *An. coustani* Laveran. Githeko *et al.,* (1996) have found *An. funestus* present in Ahero. Ijumba *et al.,* (1990) have found *Anopheles pharoensis* Theobald, *An. rufipes* Gough or *An. coustani* Laveran to be associated with rice growing schemes. Ahero is rice-growing area and study site in Nyakach is approximately 30 km from Ahero site, hence it is possible that these anophelines exist in the study sites. No alternative identification was done for these other

species. This is because an alternative identification could have been done using polytene chromosomes (White, 1974; Coluzzi *et al.,* 1985), which are found in larval salivary glands. However, the specimens had been already crushed for DNA extraction and therefore it was impossible to get polytene chromosomes at that stage. Also primers for other anophelines besides *An. arabiensis* and *An. gambiae* were not available. It is suggested that future studies should consider identifying all species that belong to same genus or a sibling species complex in a study area. This is because factors like global warming, vector migration and human population movements may introduce of new species, which may establish and become stable in regions where once they were non-existent.

There was considerable allelic diversity in populations from Ahero study sites. The number of alleles per locus was not consistent with previously reported data. This study reported a generally lower number of alleles per locus compared to those reported by Kamau *et al.,* (1999) in *An. arabiensis* specimens from Ahero. Kamau *et al.,* (1999) reported 3 alleles at locus AGXH99 compared to one allele reported in this study. They also recorded 10 alleles at locus AG3H577 compared to 3 alleles reported in this study. These differences can be attributed to temporal genetic variations within the populations and suggests that genetic structure of *An. arabiensis* is not stable but experiences temporal variations. This is because in the year 1999 Kamau *et al* got three alleles at locus AGXH99 while in this study done in year 2001 only one allele was got at the same locus.

There were significant departures ($P < 0.05$) from the Hardy-Weinberg expectations at most loci. The departures from Hardy-Weinberg equilibrium arises

due to several factors, such as, null alleles, selection, or grouping of gene pools (Wahlund's effect). In addition, inbreeding or non-random mating may also cause lack of conformance.

The null alleles (Callen *et al.,* 1993) are those alleles that fail to amplify because of mutations or sequence variation at the primer annealing sites. In heterozygote specimen where only one of the alleles is a null allele, an individual is scored as homozygote and this reduces the number of observed heterozygotes. The frequency of null alleles is expected to be higher when microsatellite loci applied in the study had been initially isolated from a species different from the one under study (Walton *et al.,* 1998). The microsatellite loci used in this study were initially isolated from *An. gambiae* s.s (Zheng *et al.,* 1993). It is therefore likely that the seen deviations from Hardy-Weinberg expectations could have been due to presence of null alleles. The number of specimens that did not yield PCR products and the absence of some alleles at some loci in some populations support this suggestion. It is possible that such situation could have occurred in some specimens at loci AG2H161, 29C, AG3H83 and AG3H577 where the observed heterozygotes were fewer than expected, and at loci AG2H143 and AG3H746 where some alleles were missing in some populations.

Wahlund's effect could also have caused the departures from Hardy-Weinberg equilibrium. The Wahlund's principle is whereby a study sample consists of several sub-samples each with a different allelic frequency (Harlt and Clark, 1989), each sub-sample could be in Hardy-Weinberg equilibrium on its own but when they are mixed the resulting sample would no longer be in Hardy-

Weinberg equilibrium. In Wahlund's principle situation, the actual frequency of heterozygotes is smaller than Hardy-Weinberg expectation (Harlt and Clark, 1989; Petrarca *et al.,* 2000). There were lower numbers of observed heterozygotes than expected in both individual and the pooled populations. It is possible that specimens from different *An. arabiensis* sub-populations were mixed and analysed as a single population. This suggests that breeding habitats of similar type did not carry panmictic populations of *An. arabiensis.*

The departure from Hardy-Weinberg equilibrium ($P < 0.05$) at locus AG3H746 in pooled Ahero population may have been due to selection for heterozygotes. The number of observed heterozygotes at this locus was higher than expected. Such an occurrence is likely to be due to heterozygous advantage, which tends to increase fitness of heterozygote genotype in the population and results in stable polymorphism and avoids fixation of an allele. It is possible that AG3H746 locus is linked to a gene that has heterozygous advantage. Heterozygous advantage has been seen in sickle cell anaemia where individuals with heterozygous genotype HbS are more resistant to certain types of malaria infection.

Non-random mating could have also contributed to departures from Hardy-Weinberg equilibrium. Speciation has been demonstrated to be an ongoing process within the *Anopheles gambiae* complex mosquitoes (Toure *et al.,* 1998). So far, chromosomally distinct subpopulations that seem to be reproductively isolated have been recognized within *Anopheles gambiae* s.s (Toure *et al.,* 1998). Available evidence suggests close relatedness between *An. arabiensis* and *An. gambiae* (Besansky *et al.,* 1994). Therefore it is likely that *Anopheles arabiensis* may also

contain reproductively isolated sub-populations that have not been identified. It is possible that sub-populations are reproductively isolated based on feeding and resting preferences i.e. isolated by ethological means but share larval breeding habitats when sympatric.

Anopheles arabiensis has diverse behaviour in terms of feeding (Mnzava *et al.,* 1995) and resting preferences (Coluzzi *et al.,* 1979) and it is possible that divergence within this species has occurred based on the preferences. If this is the case then the seen departures from Hardy-Weinberg equilibrium can be attributed to non-random mating within the *An. arabiensis* sub-populations. Previous studies reported on *An. arabiensis* analysed specimens mainly collected resting indoor (Coluzzi *et al.,* 1985; Kamau *et al.,* 1999; Donnelly *et al.,* 1999). Coluzzi and others (1985) showed that endophilic *An. arabiensis* karyotype frequencies were in substantial agreement with expectations of Hardy-Weinberg equilibrium. Using microsatellite loci, Kamau *et al.,* (1999) showed that there was considerable gene flow between *An. arabiensis* populations. Specimens used in this study were collected from larval breeding habitats and it is likely that specimens from several sub-populations were sampled.

The egg depositing mechanism of the female *Anopheles* mosquitoes may have also contributed to departures from Hardy-Weinberg expectations. The female *Anopheles* mosquitoes lay eggs in batches of 50 to 500 eggs (Clements, 1992). The eggs are dropped individually to float on the water surface. The deposition is sometimes done when the female mosquito is in flight (Clements, 1992). This means that there are high chances that wind could have caused the

female mosquito to move over several habitats while dropping the eggs. This is most likely to happen when laying in small habitats that are mostly close to each other. Although one larva was picked per habitat, this laying mechanism can result in eggs from a single egg batch being spread to various habitats and therefore being picked and lumped together. The sampling of relatives leads to inflation of homozygosity and reduction of heterozygosity.

The overall difference in allelic frequency and genotypic distribution was significant ($P = 0.000$) in both Ahero and Nyakach populations. This suggested that overall allele and genotype distribution was heterogeneous across all habitat populations at both sites, and suggests that the populations in various habitats types at both sites are not similar. Most loci showed differing allelic patterns between habitat populations at both sites, which implies that there is extensive differentiation between studied populations. The results show that *An. arabiensis* populations at both sites are not in Hardy-Weinberg equilibrium ($P < 0.05$) at most loci and departure from Hardy-Weinberg equilibrium may be due to several factors, which include; null alleles, selection, grouping of gene pools (Wahlund's effect) and non-random mating which act to cause lack of conformance to Hardy-Weinberg equilibrium as already discussed.

These findings are in line with those of Donnelly and Townson (2000), which showed that there was significant differentiation within *An. arabiensis* samples only 25km apart, which suggested recent separation and range expansion by the species. This study suggests that the differentiation is due to process of incipient speciation and fixation in feeding and resting preferences in several sub-

populations. It also suggests that the breeding habitat type has little to do with the population differentiation in *An. arabiensis*.

This study was designed to analyse whether there was partitioning of larval breeding habitats in *Anopheles arabiensis.* The results suggested high levels of sub-division within *An. arabiensis* populations found in a single habitat type. The results have shown that populations found in one habitat type do not fit the Hardy-Weinberg equilibrium. The genotype distributions in various habitat types did not fit a random mating situation and there was significant genotypic differentiation. Therefore, *An. arabiensis* population found in similar type of larval breeding habitat in the study area do not form one interbreed unit, but consists of sub-populations that seem to be reproductively isolated. These results are consistent with other studies that have shown that there is population differentiation and sub-division in *An. arabiensis* populations separated by 25km (Donnelly & Townson, 2000). Further sampling is required to determine the consistency of population sub-division. The designing of control measures based on mosquitoes that are refractory to malaria parasite infection (Crampton *et al.,* 1994; Gwadz, 1994;) can considerably benefit from such studies.

A thorough analysis of *An. arabiensis* population genetic structure will tell whether intra-specific variants exist. This study suggests that the proposed use of transgenic mosquitoes as an option in malaria control may encounter difficulties if all *An. arabiensis* intra-specific variants are not genetically modified. This is because of there is a restriction to gene flow within *An. arabiensis* populations. Thus, for an effective vector population replacement to occur when the transgenic

mosquitoes will be released into the wild, the identified variants should also be genetically modified. Otherwise genetic modification of one or a few variants will not be effective in replacing all the wild vector populations of this species.

CHAPTER SIX: CONCLUSIONS AND SUGGETIONS FOR FUTURE RESEARCH WORK

6.1 Conclusions

This study was designed to analyse whether there was partitioning of larval breeding habitats within *Anopheles arabiensis* based on size or permanence. The hypothesis was that similar breeding habitats harboured interbreeding mosquito populations. The results suggest that there is no partitioning of larval breeding habitats by *An. arabiensis* based on permanence or size of the habitat. This suggests that gene flow is not influenced by the habitat type, and also that larval breeding habitat of different *An. arabiensis* populations overlap. The conclusions are that different *Anopheles arabiensis* sub-populations do not select egg deposition sites based on habitat size or water permanence and *Anopheles arabiensis* mosquitoes found in similar habitat type do not form panmictic unit and there no random gene flow between members of this species in Ahero and Nyakach.

6.2 Suggestions For Future Research Work.

Future studies should combine markers such as paracentric chromosomal inversions, which may show association with some habitats, and microsatellite loci in order to understand whether there is association between some variants with specific larval breeding habitats. The finding of such a study will be relevant to land management in relation to vector species diversity. The arising hypothesis is

that *Anopheles arabiensis,* which has been held as a single species, may consist of reproductively isolated forms. Further investigations are necessary so as to understand population structuring of this species.

REFERENCES

ALESSADRA Della TORRE, (1997). Polytene chromosome preparation from anopheline mosquitoes. *Molecular Biology of Insect Disease Vectors: A methods manual.* Edited by J.M Crampton, C.B Beard and C. Louis

ALESSADRA Della TORRE, Fanello C, Akogbeto M, Dossou-yovo J, Favia G, Petrarca V and Coluzzi M, (2001). Molecular evidence of incipient speciation within *Anopheles gambiae* s.s in West Africa. *Insect Molecular Biology* 10 (1), 9-18.

ANDERSON T.J.C, Xin-Zhuan S.U, Bockarie M, Lagog M and Day K.P, (1999). Twelve microsatellite markers for characterization of *Plasmodium falciparum* from finger-prick blood samples. *Parasitology.* (Pt 2): 113-25.

BESANSKY N, (1999). Complexities in the analysis of cryptic taxa within the genus *Anopheles. Parassitologia* 41(1-3) 97-100.

BESANSKY N, Powell J.R, Caccone A, Hamm D.M, Scott J.A and Collins F.H, (1994). Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between principal malaria vectors *Proc. Natl. Acad. Sci. USA Evolution* Vol. 91, 6885-6888.

BUDOWLE B, Charkraborty R, Giusti A.W, Eisenberg A.J and Allen R.C, (1991). Analysis of the variable number of tandem repeats locus D1S80 by the polymerase chain reaction followed by high-resolution polyacrylamide gel electrophoresis. *Am. J. Hum. Genet.* 48: 137-144.

CAIRNS M and Murray V, (1994). Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques* 17(5): 915-919.

CALLEN D.F, Thompson A.D, Shen Y, Philips H.A, Richards R. I, Mulley J.C and Sutherland G.R, (1993). Incidence and Origin of "Null" Alleles in the (AC)n Microsatellite Markers. *Am. J. Hum. Genet.* 52:922-927.

CLEMENTS A.N, (1992). *The Biology of Mosquitoes, Development, Nutrition and Reproduction.* Vol.1 Chapman and Hall

COLLINS F.H, Kamau L, Ranson H.A and Vulule J.M, (2000). Molecular Entomology and Prospects for Malaria Control. *Bull. WHO.* 78 (12): 1412-1423.

COLLINS F.H, Petrarca V, Mpofu S, Brandling-Bennett A.D, Were J.B, Rasmussen M.O and Finnerty V, (1988). Comparison of DNA probe and cytogenetic methods for identifying field collected *Anopheles gambiae* complex mosquitoes. *Am. J. Trop. Med. and Hgy.* 39(3): 545-500.

COLLINS F.H and Besansky N, (1994). Vector Biology and the Control of Malaria in Africa. *Science* Vol. 264. 1874-1875

COLUZZI M, (1988). Anopheline mosquitoes: genetic methods for species differentiation. *Malaria: Principles and Practice of malariology.* Vol.1 411-430 edited by Wernsdorfer WH & Sir I. McGregor.

COLUZZI M, (1992). Malaria vector analysis and control. *Parasitology Today* Vol.8, No4

COLUZZI M, Petrarca V and Di Deco M.A, (1985). Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae. Boll. Zool.* 52: 45-63

COLUZZI M, Sabatini A, Petrarca V and Di Deco M.A, (1979). Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Tran Royal Soc Trop Med and Hyg.* Vol. 73 No 5 483-497.

COLUZZI M, Sabatini A, Petrarca V, Di Deco M.A, (1977). Behavioural divergences between mosquitoes with different inversion karyotypes in polymorphic populations of the *Anopheles gambiae* complex. *Nature* 266: 832-833.

CORNEL J and Collins F, (1996). PCR of the ribosomal DNA Intergenic Spacer Regions in the *Anopheles gambiae* Complex. In *Methods in Molecular Biology* Vol. 50: 312-329. Edited by J.P Clapp Humana Press Inc, Totowa, NJ

CRAMPTON J.M, Warren A, Lycett G.J, Hughes M.A, Comley I.P an Eggleston P, (1994). Genetic Manipulation of Insect Vectors as a Strategy to Control Vector Borne Diseases. *Ann. Trop. Med. Parasitol.* 88(1): 3-12.

DONNELLY M.J and Townson H, (2000). Evidence for extensive genetic differentiation among populations of malaria vector *Anopheles arabiensis* in eastern Africa. *Insect Mol Biol.* 9(4): 357-67

DONNELLY M.J, Cuamba N, Charlwood J.D, Collins F.H & Townson H, (1999). Population structure in the malaria vector *Anopheles arabiensis* Patton, in East Africa. *Heredity* 83: 408-417

FAVIA G and Louis C, (1999). Molecular identification of chromosomal forms of *Anopheles gambiae* sensu stricto. *Parassitologia.* 41: 115-118.

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 Mol. Biol.* 6: 377-383.

GILLIES M.T and B. De Meillon, (1968). The Anopheline of Africa South of the Sahara (Ethiopian Zoogeographical region) 2nd ed. *Publications of The South Africa Institute for Medical research* No. 54

GILLIES M.T and Coetzee M, (1987). A Supplement to the Anophelinae of Africa south of the Sahara (Afro tropical Region) Publications of the South African Institute for Medical Research. No. 55 1-143.

GIMNIG J, Ombok M, Kamau, L, and Hawley W.A, (2001). Characteristics of Larval Anopheline (Diptera: Culicidae Habitats in Western Kenya. *J. Med. Entomology* 38(2) 282-288.

GITHEKO A.K, Lindsay S.W, Confalonieri, U.E & Patz J.A, (2000). Climate change and vector-borne diseases: a regional analysis. *Bull. World Health Organ.* 78(9): 1136-47

GITHEKO A.K, Service M.W, Mbogo C.M and Atieli F.K, (1996). Resting behaviour, ecology and genetics of malaria vector in a large-scale agricultural area of western Kenya. *Parassitologia.* 58: 307-316.

GWADZ R.W, (1994). Genetic Approaches to Malaria Control: How Long the Road? *Am. J. Trop. Med. Hyg.* 50(6), 116-125.

HARLT D. L and Clark G.A, (1989). *The Principles and Practice of Population Genetics.* 3rd Edition, Sinauer Associates, Inc. Publishers Sunderland, Massachusetts

HOLM S, (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scand. J. Statist.* 6: 60-70.

IJUMBA J.N, Mwangi R.W Beier J.C, (1990). Malaria transmission potential of *Anopheles* mosquitoes in the Mwea-Tebere irrigation scheme, Kenya. *Med. Vet. Entomol.* 4(4): 424-32

KAMAU L, Hawley W.A, Lehmann T, Orago A.S.S, Cornel A, Zhaoxi K and Collins F.H, (1998). Use of short tandem repeats for the analysis of genetic variability of *Anopheles gambiae* and *Anopheles arabiensis. Heredity* 80: 675-682.

KAMAU L, Lehmann T, Hawley W.A, Orago A.S.S Collins F.H, (1998). Microgeographic genetic differentiation of *Anopheles gambiae* mosquitoes from Asembo Bay, western Kenya: A comparison within Kilifi in coastal Kenya. *Am .J. Trop. Med. Hyg.* 58(1), 64-69.

KAMAU L, Mukabana W.R, Hawley W.A, Lehmann T, Irungu L.W, Orago A.A, Collins F.H, (1999). Analysis of genetic variability in Anopheles arabiensis and Anopheles gambiae using microsatellite loci. Insect Mol Biol 8(2): 287-97

KING W.V and Bradley G.H, (1941). General Morphology of *Anopheles* and Classification of the Nearctic Species. *The American Association for the Advancement of Science.* Publication No. 5 63 –87.

LANZARO G.C, Toure T.Y, Carnahan J, Zheng L, Dolo G, Traore S, Petrarca V, Vernick D.K and Taylor C.E, (1998). Complexities in the genetic structure of *Anopheles gambiae* populations in West Africa as revealed by microsatellite DNA analysis. *Evolution.* Vol. 95, 14260-14265.

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

MANGUIN S, Roberts D.R Andre R.G Rejmankova E and Hakre S, (1996). Characterization of *Anopheles darlingi* (Diptera: Culicidae) larval habitats in Belize, Central America. *J. Med. Entomol.* 33: 205-211.

McCRAE A.W.R, (1984). Oviposition by African malaria vector mosquitoes II. Effects of site tone, water type and conspecific immatures on target selection by fresh water *Anopheles gambiae* Giles senso lato. *Ann. Trop. Med. Parasitol.* 78: 307-318.

MINISTRY of HEALTH, KENYA, (1998). National Guidelines for Diagnosis, Treatment and Prevention of Malaria for Health Workers.

MNZAVA A.E.P, Rwegoshora R.T, Wilkes T.J, Tanner M and Curtis C.F, (1995). *Anopheles arabiensis* and *Anopheles gambiae* chromosomal inversion polymorphism, feeding and resting behaviour in relation to insecticide house spraying in Tanzania. *Medical and Veterinary entomology* 9, 316-324.

PETRARCA V and Beier J.C, (1992). Intraspecific chromosomal polymorphism in the *Anopheles gambiae* complex as a factor affecting malaria transmission in the Kisumu area of Kenya. *Am. J. Trop. Med.* 46(2). 229-237.

PETRARCA V, Beier J.C, Onyango F, Koros J, Asiago C, Koech D.K and Roberts C.R, (1991). Species composition of *Anopheles gambiae* complex (Diptera: Culicidae) at two sites in Western Kenya. *J. Med. Ento.* 28(3): 307 –313.

PETRARCA V, Nugud A.D, Elkarim Ahmed, Haridi A.M, Di Deco M.A & Coluzzi M, (2000). Cytogenetics of the *Anopheles gambiae* Complex in Sudan, with special reference to *An. arabiensis*: relationships with East and West African populations. *Med. Vet. Entomol.* 14, 149-164.

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

RAYMOND M and Rousset F, (1995a). GENEPOP (version 3.2a): population genetics software for exact tests and ecumenism. *Journal of Heredity* 86: 248-249.

ROJAS W, Penaranda F and Echavarria M, (1992). Strategies for malaria control in Columbia. *Parasitology Today* Vol. 8 no 4 141-144.

SERVICE 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* complex in the Kisumu area of Kenya following insecticidal spraying with OMS-43 (fenitrothion*). Ann. Trop. Med. Parasitol.* 72: 377-386.

SIMARD F, Lehmann T, Lemason J.J and Fontenille D, (2000). Persistence of *Anopheles arabiensis* during the severe dry season conditions in Senegal: an indirect approach using microsatellite loci. *Insect Mol Biol.* 9(5): 467-79.

SNOW R.W, Gouws E, Omumbo J, Rapuoda B, Craig M.A, Tanser F.C, le Seur D and Ouma J, (1998). Models to predict the intensity of *Plasmodium* transmission: applications to the disease in Kenya. *Trans. R. Soc. Trop. Med and Hyg.* 92: 601-606.

TAYLOR K.A, Koros J.K, Nduati J, Copeland R.S, Collins F.H, and Brandling-Bennett A.D, (1990). *Plasmodium falciparum* infection rates in *Anopheles gambiae, An .arabiensis* and *An. Funestus*, in Western Kenya. *American J. Trop. Med. Hyg.* 43(2); 124-129.

TOURE Y.T, Petrarca V, Traore S.F, Coulibay A, Maiga H.M, Sankare O, Sow M, Di Deco M.A and Coluzzi M, (1998). Distribution and inversion polymorphism of chromosomally recognised taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia.* 40: 477-511

VULULE J.M, Beach R.F, Atieli F.K, Roberts J.M and Mwangi R.W, (1994). Reduced susceptibility of *Anopheles gambiae* to permethrin associated with the use of permethrin-impregnated bed nets and curtains. *Med. Vet. Entomology* 8(1): 71-5.

WALTON C, Thelwell N.J, Preistman A, Butlin R.K, (1998). The use of Micro satellites to study gene flow in natural populations of *Anopheles* malaria vectors in Africa: potential and pitfalls. *J. Am Mosquito Assoc.* 266-72.

WHITE G. B, (1984). Needs and progress in the application of new techniques of mosquito identification. In Newton BN and Michael F (eds.) New approaches to the identification of parasites and their vectors. *Proceedings of a symposium on application of Biochemical and Molecular techniques to problems of parasite and vector identification.* UNDP/WHO special programme; Geneva. 293-332.

WHITE G.B, 1974. *Anopheles gambiae* complex and disease transmission in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene.* Vol. 68.No.4

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): bionomics and vectorial activity of species A and B at Segera, Tanzania. *Bull. Entomol. Res. 62: 295-317.*

WORLD HEALTH ORGANIZATION, (2002). Emergency and Humanitarian Action, Health Intelligence Network for Advanced Planning v. 1.5.

WORLD HEALTH ORGANIZATION, (1975). Manual on Practical Entomology in Malaria.

WORLD HEALTH ORGANIZATION, (2000) The Abuja Declaration and the Plan of Action: An Extract from the African Summit on Roll Back Malaria, Abuja, (WHO/CDS/RBM/2000.17)

ZAKERI S, Najafabadi S, Zare A, Djadid N, (2002). Detection of malaria parasites by nested PCR in south-eastern, Iran: Evidence of highly mixed infections in Chahbahar district *Malaria Journal* 1:2

ZHENG L, Benedict M, Cornel A.J Collins F.H and Kafatos F.C, (1993). An Integrated Genetic Map of the African Human malaria Vector Mosquito, *Anopheles gambiae. Genetics* 143: 941-952.