EVALUATION OF GENETIC DIVERSITY AMONG *Glossina fuscipes fuscipes* SUBPOPULATIONS FROM SLEEPING SICKNESS FOCI IN UGANDA USING MICROSATELLITE DNA MARKERS

By

Patrick P'Odyek Abila

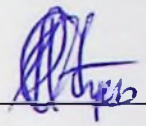
(2003/HD17/2107 U)

[January 2008]

A DISSERTATION SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES, IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN MOLECULAR BIOLOGY OF MAKERERE UNIVERSITY

DECLARATION

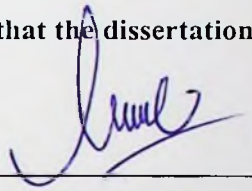
I **Patrick P'Odyek ABILA**, hereby declare that the work described in this dissertation is original and that it has not been presented to any other University as partial or full requirements for the award of any academic qualification.



Abila Patrick P'Odyek

CANDIDATE

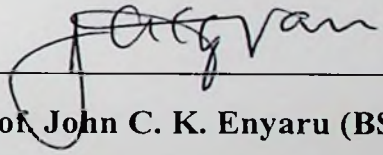
I am satisfied that the dissertation is ready for Examination.



Dr. Vincent B. Muwanika (BSc., MSc., Ph.D)

Makerere University Institute of Environment and Natural Resources, Kampala, Uganda

SUPERVISOR



Ass. Prof. John C. K. Enyaru (BSc., MSc., Ph.D)

Department of Biochemistry, Faculty of Science, Makerere University, Kampala, Uganda

SUPERVISOR

DEDICATION

This work is dedicated to my wife Barbara and my daughter Paula who endured my long absence from the family because of the field and laboratory work that culminated in putting together the results presented herein.

ACKNOWLEDGEMENTS

I am glad to acknowledge the assistance that I received from Dr. Allan Robinson of the International Atomic Energy Agency (IAEA), for introducing me to insect molecular genetics and providing the laboratory reagents that were used in the preliminary assays. Special thanks to Prof. Serap Aksoy and Dr. Gisella Caccone of Yale University for supporting the microsatellite genotyping using the automatic DNA sequencing machine. Needless to mention are my supervisors; Associate Professor John Enyaru and Dr. Vincent Muwanika for their guidance and time in following the study to its logical conclusion. I am indebted to the Board of Management Eastern African Network of Tsetse and Trypanomiasis (EANETT) and Dr. Loyce Okedi for their support that catered for the tuition, travel and stipend for the genotyping at the Yale Institute of Biospheric Studies. The Director and Staff of National Livestock Resources Research Institute (NaLIRRI) for their assistance in executing the study whenever it was required and the Director General, National Agricultural Research Organization (NARO) for granting me study leave that enabled me to participate in the coursework and do the genotyping components outside NaLIRRI.

TABLE OF CONTENTS

INTRODUCTION 1

1.1: STATEMENT OF THE PROBLEM 3

1.2: GENERAL OBJECTIVE 4

1.3: SPECIFIC OBJECTIVES 4

1.4: HYPOTHESIS 4

1.5: JUSTIFICATION 5

LITERATURE REVIEW 6

2.1: APPROACHES TO POPULATION GENETICS 8

2.1.1: Cytogenetics 8

2.1.2: Enzyme Electrophoresis 10

2.1.3: Molecular Biology Techniques 13

2.3: CHOICE OF GENETIC MARKER 23

MATERIALS AND METHODS 24

3.1: STUDY AREA 24

3.2: FLY SAMPLING AND PRESERVATION 24

3.3: GENOMIC DNA EXTRACTION 26

3.4: MICROSATELLITE GENOTYPING 26

3.5: DATA ANALYSIS 29

3.5.1: Hardy-Weinberg and Genotypic disequilibrium 29

3.5.2: Genetic variation 29

3.5.3: Population differentiation 29

RESULTS 31

4.1: PATTERNS OF GENETIC VARIATION 31

4.1.1: Microsatellite Genotyping 31

4.2.2: Genetic diversity within populations 33

4.2.3: Population Differentiation 37

DISCUSSIONS 41

5.1: PATTERNS OF GENETIC DIVERSITY 41

5.2: GENETIC DIFFERENTIATION 43

5.3: IMPLICATIONS FOR DISEASE CONTROL 44

5.4: CONCLUSIONS 45

5.5: RECOMMENDATIONS 45

REFERENCES 47

LIST OF TABLES

Table 1: Characteristics of the eleven microsatellite loci used in this study 28

Table 2: Summary of microsatellite primers screened by population 31

Table 3: Allele frequencies for seven *G. f. fuscipes* populations by locus 33

Table 4: Summary Statistics of genetic diversity at 5 loci in 7 populations of *G. f. fuscipes* ... 35

Table 5: Results of tests of genetic bottleneck by locus in each *G. f. fuscipes* population 36

Table 6: Pair-wise FST estimates across all loci 37

LIST OF FIGURES

Figure 1: Map of Uganda showing the locations of sampling sites 25

Figure 2: Electropherograph of locus D12 showing no polymorphism in samples from Namungalwe 32

Figure 3: Multiplex electropherograph of loci D101 and D12 showing polymorphism in samples from Ogur 32

Figure 4: Assignment of individuals to a subgroup based on Structure analysis 39

Figure 5: Un-rooted Neighbor Joining Tree based on pair-wise F_{ST} values 40

LIST OF ACRONYMS

AFLP	Amplified Fragment Length Polymorphism
DNA	Deoxyribonucleic acid
EANETT	Eastern African Network for Tsetse and Trypanosmiasis
FAO	Food and Agricultural Organization
HAT	Human African Trypanosomiais
HWE	Hardy-Weinberg Equilibrium
IBD	Isolation by distance
IAEA	International Atomic Energy Agency
mtDNA	Mitochondrial Deoxyribonucleic acid
NaLIRRI	National Livestock Resources Research Institute
NARO	National Agricultural Research Organization
PCR	Polymerase Chain Reaction
RAPD	Rapid Amplified Polymorphic DNA
rDNA	ribosomal Deoxyribonucleic acid
RFLP	Restriction Fragment Length Polymorphism
SIT	Sterile Insect Technique
TDR	Tropical Diseases Research
WHO	World Health Organization
YIBS	Yale Institute of Biospheric Studies

SUMMARY

Vector population genetic structure is rarely analyzed yet it is important in the understanding of the epidemiology of vector borne diseases and their control. In this study, patterns of genetic variation in the tsetse fly *Glossina fuscipes fuscipes*, the main vector of Human African Trypanosomosis (HAT) and Animal African Trypanosomosis (AAT) were assessed at five microsatellite loci. A total of 286 individuals were sampled from seven putative subpopulations in Uganda. Six of the subpopulations (Lumino, Kwapa, Osukuru, Namungalwe, Kaliro, and Serere) are sleeping sickness foci where the disease is endemic and the numbers of cases sometimes escalate. In one subpopulation (Ogur), *G. f. fuscipes* exists but active transmission is rarely observed. The mean number of alleles per locus for each of the subpopulations was relatively moderate ranging from 2 to 7 with generally low gene diversity (H_E ranging from 0.2455 to 0.4376). Analysis of population structure with 21 possible pair wise comparisons revealed 12 significantly divergent pairs of which, Kaliro and Lumino populations had the least divergence (F_{ST} =0.0443, P>0.05) while Kaliro and Serere represented the highest divergence (F_{ST} =0.2288, P<0.0000). Analysis of molecular variance (AMOVA) indicated significant differentiation at all hierarchical levels between populations (F_{ST} = 0.1474, P<0.001), among populations within each group (F_{SC} = 0.0313, P<0.001) and between the two groups (F_{CT} = 0.1199, P<0.001). These data show that *G. f. fuscipes* populations are highly structured, with clearly defined northern and southern cluster that are separated by Lake Kyoga. These results are interpreted in terms of their implications for an area-wide integrated tsetse and trypanosomosis control/elimination strategy.

CHAPTER 1

INTRODUCTION

Tsetse flies are the vectors of human African and animal trypanosomoses, diseases which have a considerable economic impact and constrain development in sub Saharan Africa where the diseases are either endemic or epidemic (WHO, 1998; Bhalla, 2002; Lewis, 2002). Current losses due to trypanosomosis in cattle are valued at 4.5 billion dollars (U.S.) per year, with about 3 million cattle deaths (Budd, 1999). In Uganda, major epidemics of human trypanosomosis continue to affect people from south eastern, mid north and north western Uganda.

Tsetse flies are vectors of several trypanosome species, including *Trypanosoma brucei brucei*, *T. congolense*, *T. simiae* and *T. vivax* that affect livestock. The fly also transmits *T. b. rhodesiense* and *T. b. gambiense* which cause Rhodesiense and Gambiense forms of human African Trypanosomosis (HAT) respectively. Rhodesiense type trypanosomosis can be contracted in parts of eastern and central Africa, including Uganda, Kenya, Tanzania, Ethiopia, Malawi, Zimbabwe, and Botswana. The Gambiense form is mainly in West Africa and extends to the Democratic Republic of Congo, Sudan, Angola, and the north western Uganda. In Uganda, both forms of the human disease exist with *Glossina fuscipes fuscipes* as the main vector (Okoth, 1999; Fevre *et al.*, 2001).

Human African Trypanosomosis occurs in localized micro-ecological zones known as sleeping sickness foci. The disease tends to reoccur and sometimes persist in such foci. Tsetse population in Uganda is estimated to inhabit approximately two-thirds of the total land area (Chizyuka 1998) with *G. f. fuscipes* having a very wide and patchy distribution throughout

much of Uganda and nearby countries. The known range includes the primary forest area as well as secondary forests and thickets in Uganda, western Tanzania, western Kenya, western Rwanda, western Burundi, and southern Sudan (Jordan, 1986). Some of the *G. f. fuscipes* sub-populations coincide with active sleeping sickness foci while others do not. In such a fragmented population, local selection regimes may cause one or more subpopulations to differ biologically in ways that could make a control strategy such as the Sterile Insect Technique (SIT) less or more effective due to mating barriers. Such barriers fall under ecological, temporal, or behavioral categories, and their strengths vary.

The bio-ecology of population isolation can lead to differentiation and possible genetic isolation depending on the extent of gene flow. Genetic differentiation and isolation may be among candidate factors to explain the persistence of sleeping sickness in some foci while other areas that harbor tsetse flies remain disease-free. It is therefore essential to accurately identify, characterize and detect variation within and among vector sub-populations in order to understand the epidemiology of African trypanosomosis (Maudlin, 1980).

Several laboratory based and field research on the life cycle and bionomics of tsetse flies have demonstrated differences in behavior, susceptibility to infection and vector competence among subpopulations (Jordan, 1974; Maudlin, 1982, 1985; Solano *et al.*, 1997, 2000). It is not well known whether differences in susceptibility to infection, vector competence as well as localization are a result of limited gene flow between them. The aim of the present study was to investigate the genetic variation between and among sub populations of *G. f. fuscipes* and relate this to the characteristics of the disease foci.

Population genetic studies in entomology are facilitated by methods that reveal detectable genetic variation using molecular markers (Loxdale and Lushai, 1998; Behura, 2006). Allozyme markers provide one method to study genetic variation in insects (Szalanski, 1995),

but studies on *Nhemorhina* species using this method have revealed little polymorphism (Krafsur and Griffiths, 1997; Leak, 1999; Gooding and Krafsur, 2005). Alternative approaches are polymerase chain reaction (PCR) based analysis. These involve the use of DNA markers such as random amplified polymorphic DNA (RAPD), restriction fragment length analysis, sequencing and microsatellite fragment analysis (Krafsur 2005). Of recent, among the PCR based analyses, microsatellite analyses have been the most frequently used (Krafsur and Wohlford, 1999; Solano *et al*, 1999, Ouma *et al*., 2005, 2006; Ravael *et al*., 2007). This is because they exhibit considerable variation that allows for assessment of population genetic structure.

1.1: STATEMENT OF THE PROBLEM

The ability of tsetse flies to move between patches of suitable habitat determines its dispersal potential into areas where the target population is under suppression or has already been eliminated. Dispersal characteristics can be determined using ecological approaches such as mark, release and recapture experiments, but this is expensive, time consuming and uncertain. Genetic methods can be used instead as indirect measures of dispersal (Krafsur, 2003; 2005) through estimation of genetic variability and levels of population differentiation within and between sleeping sickness foci. This information is vital in the development and planning of tsetse intervention and disease control strategies. This is particularly important for Uganda where the disease has persisted despite concerted medical, veterinary and entomological efforts to eradicate it. This study therefore seeks to assess the patterns of genetic variation within and among different *G. f. fuscipes* sub-populations in Uganda.

1.2: GENERAL OBJECTIVE

The general objective of the study was to investigate the genetic structure of *G. f. fuscipes* sub-populations within and among sleeping sickness foci in Uganda

1.3: SPECIFIC OBJECTIVES

i) To determine the level of genetic variation within the different *G. f. fuscipes* subpopulations in different sleeping sickness foci,

ii) To determine the extent of genetic differentiation among *G. f. fuscipes* subpopulations from different sleeping sickness foci,

iii) To detect any recent population bottlenecks in the sampled *G. f. fuscipes* populations.

1.4: HYPOTHESIS

There are no significant levels of genetic structure among the *G. f. fuscipes* subpopulations from sleeping sickness foci in Uganda.

1.5: JUSTIFICATION

Studies on vector genetics are basic and essential to the understanding of the epidemiology of vector borne diseases. In the study of the epidemiology of any vector borne disease, it is essential to accurately identify the characteristics of the vector subpopulations. This study documents the patterns of genetic variation and differentiation among *G. f. fuscipes* from different sleeping sickness foci. Data on levels of genetic variation, population differentiation, and gene flow can be constructively applied in the assessment of areas at risk of trypanosomosis and to prioritize decisions on vector control strategies (Marquez *et al.*, 2004, Gooding and Krafsur, 2005 and Krafsur, 2005) targeting tsetse populations in an area-wide integrated vector control approach.

In Uganda, no study has been done on the patterns of genetic variation among *G. f. fuscipes* subpopulations from different infested/affected areas. This study, therefore, aimed at assessing the levels of genetic variation, population differentiation and rates of gene flow for subsequent use in determining the areas at risk of African trypanosomosis.

CHAPTER 2

LITERATURE REVIEW

Population genetics is the study of gene frequencies in and among populations. For microsatellites, it encompasses estimations of variation in terms of allelic and genotypic frequencies. Given appropriate data, the application of population genetic theory allows estimates of gene flow within and among populations, estimates of effective (i.e. reproductive) population sizes, tests for past bottlenecks and rapid expansions of population size, and pair-wise genetic distances. Due to the fact that all evolutionary change is accompanied by change in gene frequencies, while population genetics is hinged on analyses of gene frequencies, population and evolutionary genetics overlap (Krafsur, 2005). The natural theory of evolutionary genetics proposes that the majority of mutational changes are either: (i) selectively neutral, (ii) slightly deleterious, or (iii) strongly deleterious with only a small minority being, (iv) selectively advantageous. Because mutants of type (iii) are removed from the population and type (iv) are rare, most molecular evolution in populations proceeds with mutants of type (i) and (ii) (Allendorf and Luikart, 2007).

Before the advent of biochemical techniques, population genetics was divided into two schools of thought. The classical school postulated that genetic variability was minimal within a species, and most conspecific individuals had the same "wild type" allele at each locus (Avise, 1994). The opposing view, the balancing school, believed that most loci were polymorphic and heterozygosity was high. In the 1960's allozyme studies (Lewontin and Hubby, 1966) revealed a wealth of genetic variation that supported the balancing school and was designated as the neutralist doctrine (Kimura, 1991). The high frequency of alternate alleles observed suggested

that most polymorphisms are essentially neutral (Avise, 1994). Variation is maintained by a balance between mutation and extinction by random genetic drift.

Despite different opinions on genetic variability, evolution remains a common element. Evolution can be defined as the change in genetic composition of populations through time (Dobzhansky, 1937). Evolution is dependent on three components: a source of genetic variation (mutation); forces that change the frequency of alleles (selection, genetic drift); and historical components in time and space (Dobzhansky, 1937).

Mutations are "errors" that occur during DNA replication resulting in changes in nucleotide sequences (Futumya, 1986). Examples include substitutions, duplications, deletions, translocations, and inversions. Frequently, mutations are lethal, and the probability of their establishment in a population is low (Hartl, 1988). However, over millions of years, mutations are incorporated into the species, resulting in genetic variation (Avise, 1994).

Partitioning of genetic variation within and among populations is dependent on several factors including environmental, genetic drift and population bottlenecks. Localized environmental forces can reduce migration (gene flow) among populations (Hartl, 1988) and therefore result in population differentiation. Environmental forces may include geographical barriers or selection pressures (e.g. temperature tolerance and mating preference). Genetic drift can partition variation if immigration is absent. Finally, bottlenecks during colonization can reduce genetic variation or change gene frequencies significantly (Hartl, 1988) and lead to differentiation as well.

2.1: APPROACHES TO POPULATION GENETICS

Population genetic studies in veterinary and medical entomology have been facilitated by cytogenetic, biochemical and molecular methods that reveal detectable genetic variation using appropriate markers (Loxdale and Lushai, 1998, Behura, 2006).

2.1.1: Cytogenetics

Chromosome morphologies and structural rearrangements may vary within species and provide markers that distinguish populations (Robinson, 2005). Examples include paracentric inversions, easily detected in stained squashes of polytene chromosomes found in *Drosophila*, *Anopheles* mosquitoes, and the supernumerary chromosomes of some *Glossina morsitans* group of tsetse flies (Gooding and Krafsur, 2005).

A study of chromosomal polymorphism within natural *Anopheles funestus* populations from West Africa revealed high levels of genetic structuring. In Burkina Faso, two distinct chromosomal forms were identified. This was interpreted as evidence for incipient speciation, and two chromosomal forms were described as "Folonzo" and "Kiribina" (Cohuet *et al.*, 2004). Although, incipient speciation was not ruled out by this analysis, the data suggests that differential environmental selection pressure acting on the different inversions should be considered a major factor in shaping patterns of genetic variation in *An. funestus* populations (Cohuet *et al.*, 2004). It has been postulated that such geographical variability within a relatively limited area, suggests a restricted gene flow between the populations, which may be due to discontinuities in the *An. funestus* distribution and to its bioecology (Dia *et al.*, 2000).

Polytene chromosomes in *An. gambiae sensu stricto* from West Africa represent a complex of paracentric inversions, with the majority occurring on the right arm of the second chromosome (2R). Extensive field studies of *An. gambiae* s.s. in areas of sympatry indicate that 2R inversion genotypes are seldom in Hardy–Weinberg equilibrium due to a deficiency or absence of

inversion heterozygotes. This result suggests barriers to gene flow among mosquito subpopulations with different 2R inversion genotypes (Black and Lanzaro, 2001).

Several instances of chromosome polymorphisms have been reported in tsetse flies. Chromatin (C) banding polymorphism in the Y chromosome of *G. m. submorsitans* attributed to a paracentric inversion has been reported and another inversion on the left arm of autosome L1 was identified in a colony of *G. pallidipes* originating from Uganda (Willhoeft, 1997). *Glossina morsitans morsitans* in a colony originating from Handeni, Tanzania, differed by a large paracentric inversion on the X chromosome from those in a colony from Kariba, Zimbabwe (Jordan, 1974). More recently, one inversion has been identified on the L1R chromosome arm of *G. austeni* (Gariou-Papalexiou *et al.*, 1998) and another on L2L chromosome arm of *G. pallidipes* in laboratory colonies (Willhoeft, 1997). In a study of polytene chromosome maps of *G. m. submorsitans* sex ratio distortion, an important bionomic characteristic, was associated with an X chromosome characterized by a complex inversion in both arms (Gariou-Papalexiou *et al.*, 2002).

Two colonies of *G. p. gambiensis*, both of which were established with flies collected from the same locality in Burkina Faso, but with different colonization histories, differed in susceptibility to and maturation time of *T. b. gambiense*. The two colonies also presented different allele frequencies at two X chromosome loci (Hk and Pgm) and one autosomal locus (Pgi) (Gooding, 1996).

Sex chromosome aneuploidy and sex ratio distortion in wild populations of *G. p. palpalis* from Nigeria was attributed to non disjunction of the X chromosome during female meiosis. Although data on sex chromosome aneuploidy was not available for *Glossina*, the occurrence of a high proportion of females (8.4%) in a sample shown to be XXY by producing XXY,

XYY, XX and XY pupae suggests that they were maintained in the population as a polymorphism rather than by spontaneous non disjunction (Maudlin, 1979).

Unfortunately, these kinds of variation rarely provide the high degree of resolution afforded by molecular variation in DNA or its transcription and translation products such as allozymes. Most, if not all chromosomal variation, seem to respond to selective forces, and therefore do not provide unbiased estimators of gene flow. They may nevertheless provide important information about reproductive isolation and the operation of natural selection, but the analysis is time consuming and requires much skill, and therefore adequate and representative sampling is a problem (Krafsur, 2005).

2.1.2: Enzyme Electrophoresis

Protein electrophoresis, a powerful and simple technique for detecting and quantifying genetic variation, has been used to establish the amount of genetic variation in most natural populations of tsetse flies. Allozymes have been used to measure genetic diversity in several species of *Muscidae*, including house fly (Black and Krafsur, 1985a, 1985b, 1986), face fly (Krafsur and Black, 1992), stable fly (Szalanski, 1995), horn fly (McDonald *et al.*, 1987) and the tsetse fly, *G. m. centralis* (Gooding, 1989). Mean heterozygosity in Muscid species range from 0.059 in tsetse to 0.161 in face flies. Within the taxonomic family, Stable flies are less heterozygous than house flies, face flies and horn flies. Genetic distances between stable fly populations are therefore lower than those reported for the horn fly (McDonald *et al.*, 1987).

Based on multi-locus enzyme electrophoresis, analysis of *Triatoma infestans*, samples from twenty two Andean localities in Bolivia and Peru exhibited panmixia, while genetic differentiation of *T. infestans* between villages was correlated with the distance between them. The genetic structure of *T. infestans* natural populations followed an "isolation by distance" model, involving a series of founder effects followed by genetic drift, rather than adaptation in

response to differential selection pressures. This conforms to circumstantial evidence that *T. infestans* spread, mainly in association with recent human migrations, from a source, probably in southern Bolivia. Isoenzyme characterization of populations of *T. infestans* could be used to infer sources of re-infestation during the surveillance phase of control programs (Dujjardin *et al.*, 1998).

In an effort to characterize stable fly populations from North America, Szalanski (1995) reported no evidence of reproductive isolation between sampled populations from Canada, Lincoln and Grand Island. Based on allozyme variation the average numbers of migrants between these populations indicated high levels of gene flow between populations and were in agreement with conclusions in previous studies on Stable flies (Black and Krafsur, 1985a).

Laboratory maintained tsetse colonies have shown genetic changes due to the effect of colonization. Allele frequencies at the allozyme locus for midgut alkaline phosphates changed in a large colony of *G. m. morsitans* at about the time that the colony was switched from an in vivo to in vitro feeding system (Gooding, 1992).

In wild populations, comparisons of intraspecific genetic diversity between tsetse populations using allozyme markers have on the other hand revealed significant differences in the mean heterozygosities of *G. pallidipes* populations at Lambwe and Nguruman, Kenya (Krafsur, 2002). These populations also differed significantly in the frequency of alleles at loci controlling an esterase, octanol dehydrogenase and phosphoglucose isomerase (Tarimo-Nesbitt *et al.*, 1990). Similarly, *G. m. centralis* at Keembe, Zambia differed significantly from those in the main fly belt at Nalusanga, Zambia at the loci for xanthine oxidase, aldehyde oxidase and an esterase (Gooding, 1989). These differences indicate lack of significant amounts of migration between these pairs of populations. However, it was difficult to relate the amount of heterozygosity to the history of a population. It might be expected that populations that had

been subjected to repeated control measures may have gone through genetic bottlenecks. Populations of *G. m. centralis* at Keembe, Zambia and *G. pallidipes* at Lambwe, Kenya have been subjected to insecticidal campaigns, although the former population showed significantly less heterozygosity (Gooding, 1989) than is typical of most invertebrates; the latter appears to be unchanged (Tarimo-Nesbitt *et al.*, 1990).

There was no information available on the size of the Keembe population before or after the insecticidal campaign but the Lambwe population was apparently so large that even a 99.9% reduction in its size (Turner and Brightwell, 1986) did not reduce it to the point where significant changes occurred due to genetic drift (Tarimo-Nesbitt *et al.*, 1990). Allele frequencies in samples of *G. m. centralis* collected by three different trapping systems did not produce any evidence of division of the population at Nalusanga, even though the different collecting systems returned samples with significantly different sex ratios, age structures and infection rates (Gooding, 1992).

An investigation of gene diversity in four taxa of tsetse flies including *Glossina morsitans morsitans*, *G. m. centralis*, *G. swynnertoni*, and *G.pallidipes* using histochemical techniques on 35-46 isozymes proved that allozyme gene diversities were considerably less than those reported for many Diptera. The low gene diversities are probably related to small effective population sizes (Krafsur and Griffiths, 1997). In general, studies of tsetse genetic variability using allozymes have revealed little polymorphism within and between populations.

Additionally, there are two major problems associated with the use of allozyme loci; first, freshly killed or frozen tissues are required in order to preserve enzyme activity. Second, population genetics studies assume that the genetic markers employed are selectively neutral, and evidence has accumulated that balancing selection acts upon some allozyme (Gooding and Krafsur, 2005), including those in *G. pallidipes* (Krafsur, 2002). Tsetse flies also show low

levels of diversity, averaging 6%, compared with other dipterans such as house flies, *Drosophila*, or stable flies, among which diversities vary from 10% to 19%. However, average heterozygosities at polymorphic loci was 27% in field-collected specimens of *G. morsitans s. l.* and 19.5% among colonies of *G. p. palpalis* and *G. p. gambiensis* (Gooding and Krafsur, 2005).

2.1.3: Molecular Biology Techniques

These are a family of methods that have tremendously increased the speed and accuracy of field of population genetics applications with the ever increasing capabilities of different laboratories to handle large amounts of generated data (Caterino *et al.*, 1998; Loxdale and Laishi, 1998; Behura, 2006). They are DNA based and apply different techniques to detect the differences in the DNA of the individual organisms within populations, code and analyse the data using specific models (Wang *et al.*, 2001).

2.1.3.1: Restriction Fragment Length Polymorphism (RFLP)

Restriction fragment length polymorphisms (RFLPs) are generated following the digestion of DNA with restriction endonucleases. In general these recognition sites are four or six bases in length (hence the enzymes are termed tetra or hexanucleases), but various shorter and longer length base-cutters also exist, but are less frequently employed. These enzymes cut non-methylated double-stranded DNA at specific locations, thereby generating a reproducible array of fragments that may be separated by size using gel electrophoresis (Hoy, 2003). The variations in RFLPs represent changes in the lengths of DNA between the specific enzyme cutting sites using one or more restriction endonucleases. The cut DNA fragments may be visualized under UV-light using ethidium bromide stained agarose gels or with more resolution in 6% polyacrylamide gels.

Restriction fragments segregate as codominant alleles, and can therefore be used to test the Hardy-Weinberg Equilibrium (HWE) assumptions. However, the main disadvantage is that for

higher resolutions (increased allele number), radioactive end-labeling of DNA probes is required with nucleotides tagged with ^{32}P or ^{35}S, followed by exposure to X-ray film, which show the alleles. Many of restriction enzymes (Res) are required with different recognition sequences, which make RFLPs an expensive technique to use, and are not ideal for large interpopulation surveys. The amount of genomic DNA required is high and may not be feasible when studying tiny samples such as insects.

Because of the limited amount of genomic DNA that can be obtained from a single insect, the power of RFLP analysis may be increased by pre-amplifying the genomic DNA fragments with PCR. In this case, sequence information flanking the restriction site is required. This approach has been applied to the inter-transcribed spacer 2 between the 5S and 28S rDNA sequences and produced excellent markers for species systematics of the *An. punctatus* (Beebee and Saul, 1995) and different snow pool *Aedes* species (West *et al.*, 1997).

Contributing to the debate on the dynamics of the migration/invasion processes exhibited by the Mediterranean fruit fly (medfly), *Ceratitis capitata*, a major agricultural pest, He and Haymer (1999) used the PCR–RFLP technique, based on the variation in intron sequence of the glucose-6-phosphate dehydrogenase (*Zw*) gene to assess genetic variability. The relationships among the populations of *C. capitata* within geographical groupings were analyzed using the AMOVA and showed that variances differed considerably (He and Haymer, 1999).

2.1.3.2: Amplified Fragment Length Polymorphism (AFLP)

Amplified fragment length polymorphism (AFLP) is the selective amplification of DNA restriction fragments from a digest of total genomic DNA using the polymerase chain reaction (PCR). With AFLP, genetic polymorphisms are identified by the presence or absence of DNA fragments following restriction and amplification of genomic DNA (Blears *et al.*, 1998). The

procedures involve the initial digestion of genomic DNA by two or more restriction enzymes. Double-stranded oligonucleotide adapters, homologous to one 5'- or 3'-end generated during restriction digestion, are ligated to the DNA fragments. The ligated DNA fragments are then amplified by PCR using primers complementary to the adapter and restriction site sequence with additional selective nucleotides at their 3'-end. The use of selective primers reduces the complexity of the mixture. Only those fragments with complementary nucleotides extending beyond the restriction site will be amplified by the selective primers under stringent annealing conditions (Behura, 2006).

The AFLP method produces bands (DNA fragments) that are separated based on differences in length (polymorphism) using denaturing polyacrylamide gel electrophoresis or more recently sequencing robots and the patterns generated for each sample compared. A band of a particular length represents an allele (scored 1) at such an AFLP locus. Individuals not having a band of that length are scored 0. The presence and absence character of the data cannot provide complete genotypic information for diploid organisms. This is because individuals with the band can either have two (1/1) or one (1/0) copy of the allele. Polymorphic AFLP loci (within the sample of study) are example of dominant genetic markers, and for each AFLP locus individuals with a band (the presence allele) are either homozygous (1/1) or heterozygous (1/0) and those without the band homozygous for the absence allele (0/0). Most population genetic analyses based on AFLP data assume that the absent allele is really non-existent from the data (Blears *et al.*, 1998). However, other types of mutations may result in a DNA fragment of a different length and hence a band at a different position in the gel. In this way two alleles at the same AFLP locus will mistakenly be scored as different alleles at two different AFLP loci.

Indel variation (insertions or deletions) between the primers, such as when the primers span a polymorphic microsatellite region, may result in many spurious AFLP loci. Also, a substitution

that creates a new cut site for any of the two restriction sites between the primers may make the absent allele for one AFLP locus to be scored as a presence allele at another AFLP locus (Wong *et al.*, 2001). Because AFLP gels are typically complex containing many polymorphic sites, it is rarely possible to find the alternative allele.

Such alternative alleles are not independent and thus violate an important assumption in analyses of population structure and estimates of population genetic diversity. The introduced bias is however assumed to be negligible as long as AFLP-length co-dominance is rare (< 10%) and a large number of informative bands (> 100) have been studied (Parsons and Shaw, 2001).

Another problem is size homoplasy, i.e. that bands of the same length are not homologous and thus representing two or more different AFLP loci, and this is of particular concern in studies of genetic diversity and phylogenetic reconstructions (Bensch and Akesson, 2005). Size homoplasy has been found to increase with the density of amplified fragments and decrease with the length of the amplified fragments (Vekemans *et al.*, 2002)

Because AFLP markers have to be analyzed as dominant loci in order to estimate population genetic diversity and genetic structure parameters, one must assume that dominant (amplified) alleles are identical in state, recessive (unamplified) alleles are identical in state, AFLP fragments segregate according to Mendelian expectations and that the genotypes of an AFLP locus are in Hardy-Weinberg equilibrium (HWE). The HWE assumption is not testable for natural populations using dominant markers.

2.1.3.3: Randomly Amplified Polymorphic DNA

Randomly Amplified Polymorphic DNA (RAPD) is a PCR based technique in which banding profiles are created using small oligonucleotide primers (around 10 base pairs in length) of arbitrary sequence. These primers bind to homologous sequence along the genome, and PCR

amplification occurs where opposing primer sites are at least about 3,000 bp apart (Robinson and Gooding, 1997). PCR products are separated by agarose gel electrophoresis and the bands are visualized under UV light after staining with ethidium bromide. Within a population sample, mutations influence the base sequence of primer binding sites, allowing polymorphism to be detected (Wilkerson *et al.*, 1993).

A RAPD-PCR based genetic analysis of samples of *G. austeni* collected from a laboratory colony maintained at Seibersdorf, Austria (established with material collected in Zanzibar in 1969), wild flies from Zanzibar and Tanga showed very little genetic differences among the three populations. However, analysis including *G. austeni* from the Hellsgate area, South Africa, a population which has been given a separate taxonomic status (*G. austeni mossurizensis*), and thus would be expected to show significant genetic differences from *G. a. austeni*, the northernmost subspecies indeed indicated some divergence but the level of similarity remained high (Robinson and Gooding, 1997). Similarly, in a study aimed at generating comparative genetic data between *G. f. fuscipes* from Buvuma Islands, and mainland Mukono district in Uganda, only a single RAPD primer pair produced distinct diagnostic banding profile (Gidudu *et al.*, 1999).

At its primacy, developments in molecular genetics, including the use of random amplified polymorphic DNA by the polymerase chain reaction (RAPD-PCR) greatly increased the opportunities for, and the rate of analysis of poorly known genomes. It represented a major step forward in the analysis of important population parameters related to the Sterile Insect Technique (SIT) and other methods of control (Robinson and Gooding, 1997). Due to inherent constraints and development of other methods it has not been extensively applied in tsetse population genetics. The bands visualized, show dominance in heterozygotes, and it is not possible to distinguish between a homozygote and heterozygote. A major weakness of RAPDs

is lack of reproducibility. It is also difficult to establish that RAPD fragments observed on the gel are primed from the same locus, and achieving reproducibility between reactions and laboratories is difficult (Krafsur, 2005).

2.1.3.4: Single Strand Conformation Polymorphism

Single-strand conformational polymorphism (SSCP) analysis is a simple and sensitive technique for mutation detection and genotyping. The principle of SSCP analysis is based on the fact that single-stranded DNA has a defined conformation. Altered conformation due to a single base change in the sequence can cause single-stranded DNA to migrate differently under non-denaturing electrophoresis conditions (Hodgkinson *et al.*, 2002). Therefore wild-type and mutant DNA samples display different band patterns. Analysis using SSCP involves four steps; PCR to amplify DNA sequence of interest; denaturation of double-stranded PCR products; cooling of the denatured DNA (single-stranded) to maximize self-annealing; detection of mobility difference of the single-stranded DNAs by electrophoresis under non-denaturing conditions.

Several methods have been developed to visualize the SSCP mobility shifts. These include the incorporation of radioisotope labeling, silver staining, fluorescent dye-labeled PCR primers, and more recently, capillary-based electrophoresis. The use of SSCP analysis to discover and genotype single-nucleotide polymorphisms (SNPs) in population genetic studies of insect disease vectors was demonstrated (Sunnucks *et al.*, 2000). The SSCP technique was further used to study variable loci of the mitochondrial genome in tsetse populations (Marquez *et al.*, 2004, Krafsur, 2002).

2.1.3.5: DNA Sequencing

DNA Sequencing involves the use of nucleotides as character states in the analysis of DNA sequences. Silent and neutral substitutions undetected by allozyme analysis are detectable by sequencing. Selected regions of the genome are sequenced for each individual and aligned.

Nucleotide site changes are converted into quantitative measurements of genetic distance (Loxdale and Lushai, 1998). DNA sequencing was originally performed by using radioactive labels for detection of the reaction products (Avise, 1994). However, current sequencing protocols employ fluorescent nucleotides to label the DNA whose sequence is then read with an automated instrument (Behura, 2006). Sequencing of DNA begins with PCR amplification of a sample DNA directed at regions of interest, followed by sequencing reactions of the PCR products.

Although sequencing provides a complete dataset, it is not practical as the sole method for sampling of the entire genome in population genetic analysis (Loxdale and Lushai, 1998; Behura, 2006). It also examines only a very small portion of the sites which can potentially vary between individuals and populations. The structure of the region of DNA selected must consist of a variable sequence flanked by highly conserved regions. This enables PCR amplification and typing of individuals. The variability within the selected sequence must be sufficient to differentiate individuals and populations of particular species.

Probably the most frequently used genes in population studies are those from the mitochondrial genome. The Mitochondrial genome is haploid, self-replicating, and is maternally transmitted as a circular molecule of about 16 kilobases (Loxdale and Lushai, 1998). It encodes 22 transfer RNA genes and 13 proteins, including the cytochrome oxidases, NADH, RNAase, and ribosomal RNAs. There are many highly conserved regions from which universal oligonucleotide primers have been designed that are commercially available in kits (Loxdale and Lushai, 1998). These primers allow the amplification of regions of the mitochondrial genome by the polymerase chain reaction (PCR). Mitochondrial variation can be assessed by a number of methods. Nucleotide sequencing of amplified regions can be followed by the comparison of sequences among individuals. Variants in mitochondrial DNA (mtDNA) do not

recombine and, with rare exceptions, inheritance is strictly matrilineal. Such diversity is particularly useful for examining maternal lineages and testing for bottlenecks in population size (Krafsur and Wohlford, 1999).

For instance, in western Kenya, the genetic relatedness of mosquitoes (*Anopheles gambiae*) caught within houses in villages spanning up to 50 km distance and comparing the haplotypes of catches between houses revealed no differentiation. The indices of kinship based on mtDNA haplotypes of samples caught in houses were even lower than in the pooled samples. Likewise, estimates of haplotype divergence of mtDNA between populations were not higher than the within population estimates, representing a lack of genetic isolation from one another (Lehmann et al. 1997).

On the other hand, abundant variation occurs at mitochondrial loci in *morsitans* group flies but much less variation was found in *G. palpalis gambiensis* populations (Gooding and Krafsur, 2005). Wohlford *et al.*, (1999) detected twenty-seven mitochondrial haplotypes at *12S*, *16S1*, *16S2* and *CO1* loci in *Glossina morsitans morsitans* sampled from Zimbabwe and Zambia. Mitochondrial diversities were greatest in *G. m. morsitans* samples from Mana Angwa and Kakumbi, riverside locations, northern Zimbabwe.

Nevertheless, application of mtDNA in population genetics of animals is not without problems (Zhang and Hewitt, 2003). The recent demonstration of the presence of mitochondrial pseudogenes in the nuclear genome of a wide range of organisms is, for population studies, an unwanted reality (Bensasson *et al.*, 2001). Although there exists methods and techniques to reduce the interference of mtDNA pseudogenes in sample preparation and data analysis, they have only limited resolving power (Bensasson *et al.*, 2001). The effectiveness of using mtDNA in population-genetic studies has been greatly weakened by this. It has been shown that mitochondrial-like sequences can exist in high copy number with little difference among

members in many organisms, sometimes group-wide as in the acridid grasshoppers and locusts, making mtDNA of little practical value for population-genetic studies in these groups (Bensasson *et al.*, 2001). The deployment of mtDNA marker in organisms without first being aware of this phenomenon can be a real headache. Ironically, the great value of mitochondrial DNA is nowhere better appreciated than in these situations where it can not be employed, as in cases where numerous nuclear pseudogenes exist very similar to the authentic mitochondrial sequences (Zhang and Hewitt, 2003).

2.1.3.6: Microsatellite DNA Markers

Microsatellite DNA sequences are short core (2 to 10 base pair) tandem repeats up to 500 base pairs long scattered throughout the genome. They are unique in that many are closely associated with conserved loci that contain coding regions. Changes in the motif repeat number occur at a higher rate than in the coding region and polymorphism due to variation in motif repeat number is recordable as either fragment length, repeat number or both (Bennett, 2000).

Microsatellite DNA markers are amenable for high throughput genotyping by nonradioactive labeling and scoring by automated machines (Behura, 2006). Although the cost involved in the initial setup is higher than that of the AFLP, this is often compromised in population and ecological studies where more number of genotypes are required for meaningful statistical analyses. However, recent findings have shown that in some cases microsatellites are not neutral markers as they are found involved in gene regulation, genetic hitchhiking, and sex-specific differential selection, and hence, such loci may not be useful for evolutionary and phylogenetic studies (Behura, 2006).

Their generally high mutation rate ensures high levels of polymorphism (Beebee and Rowe, 2004; Behura, 2006) and they have therefore been useful for examining relationships among individuals and breeding groups within insect populations. Repeat sequences of different

lengths at a given locus are referred to as alleles and co-dominance is observed. Development of microsatellite loci for a particular organism is rather labor intensive, but there is indication that loci may be more broadly applicable than originally thought. In addition, methods have been developed recently to screen for microsatellite loci without creating a complete genomic library (Caterino *et al.*, 2000). Once primers have been designed for a target organism they can sometimes be used in closely related taxa including different genera although this is dependent on the loci identified (Solano *et al.*, 1997, Luna *et al.*, 2001).

Microsatellites have come into widespread use for genetic analysis at or below the species level (Queller *et al.*, 1993). The major Advantage of microsatellite DNA method is the ability to detect greater levels of genetic variability. Many tsetse microsatellite loci are available often with numerous alleles (Solano *et al.*, 1997, Luna *et al.*, 2001, Ouma *et al.*, 2003) and can be screened for use in molecular epidemiology and ecology of tsetse flies. Individual alleles can be scored at particular loci and provide good Mendelian markers. Di, tri and tetra-nucleotide repeats provide good signals for allele calling. Appropriate PCR conditions can help reduce stutter bands caused by miss-priming that result in templates of smaller size to the original band. Microsatellite DNA fragments are more easily recorded for size variation either as absolute base pair sizes or repeat numbers (Bennett, 2000).

Several microsatellite loci in the *palpalis* and *morsitans* groups of tsetse flies have been characterized. Three polymorphic loci have been isolated from *G. p. gambiense* (Solano *et al.*, 1997). Another thirteen microsatellite markers were characterized in detail for *G. p. palpalis* and four other tsetse species (Luna *et al.*, 2001), fourteen in *G. morsitans sensu lato* (Baker and Krafsur, 2001), and eleven loci from *G. pallidipes* (Ouma *et al.* 2003, 2006). Robinson (Unpublished data) reports 39 loci in *G. f. fuscipes*.

2.3: CHOICE OF GENETIC MARKER

Among codominant markers, usually the most powerful for population genetics work, microsatellites currently dominate the field (Beebee and Rowe, 2004). This is because they are both highly polymorphic and relatively easy to score. Their main disadvantages include the time and expense of development and to a lesser degree of recurrent use. Theoretically too much polymorphism can also be problematic in population genetic studies but this is rarely a problem in practice (Beebee and Rowe, 2004).

Whichever marker system is chosen, there is always a need to develop a method for any new study to make sure there is an appropriate level of polymorphism for the task in hand. Although microsatellites are currently the best for most population genetic work, characterizing loci for a new species is relatively time consuming and expensive task. To get around this problem, we obtained *G. f. fuscipes* sequences of fragments containing microsatellites with good flanking regions from Robinson (Unpublished data).

CHAPTER 3

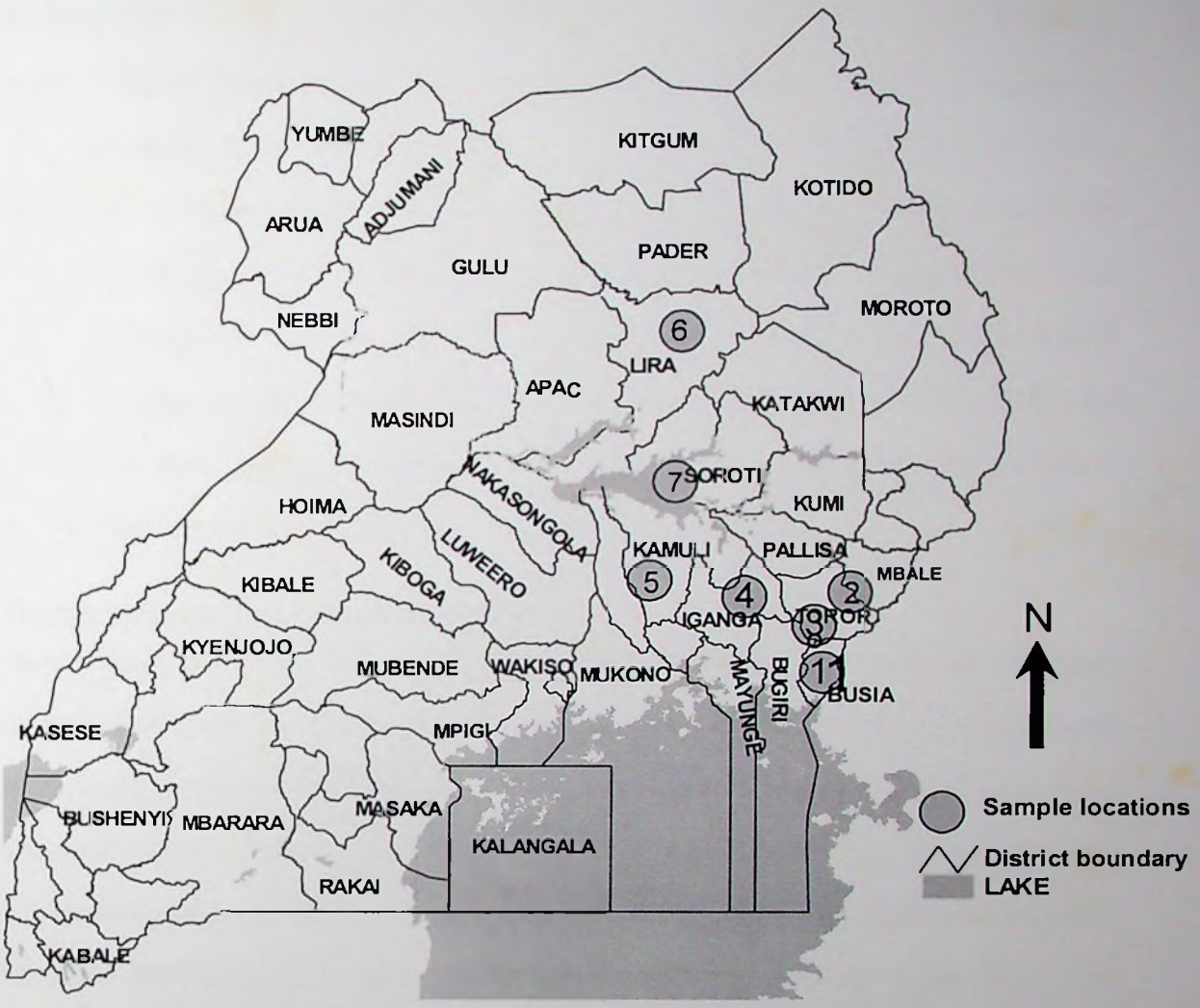
MATERIALS AND METHODS

3.1: STUDY AREA

Seven putative populations from sleeping sickness foci were purposively selected on the basis of their epidemiological, geographical and ecologically distinguishable characteristics. The samples were drawn from the Serere (Soroti district), Kaliro (Kamuli district), Namungalwe (Iganga district), Lumino (Busia district), Osukuru and Kwapa (Tororo district), where the disease is endemic and occasionally the numbers of cases escalate. Samples were also collected from Ogur (Lira district) where *G. f. fuscipes* exists but active transmission was last reported by Shore (1960) and only recently (Abilá *et al.*, 2006). *G. f. fuscipes* from Kamuli and Iganga have been reported to inhabit peridomestic vegetation (Okoth, 1986) while the rest are known to dwell in typical riparian environments. Lira and Soroti districts are located north of Lake Kyoga as opposed to the rest of the districts which are south of the Lake. The districts from which samples were obtained are shown in Figure 1.

3.2: FLY SAMPLING AND PRESERVATION

For each of the foci, *G. f. fuscipes* were sampled from either riparian or peridomestic tsetse habitats. Sampling was done using non-impregnated biconical traps using standard procedures (Pollock, 1982). The legs of trapped flies were excised and preserved in dry 1.5ml cryovials (no preservative used) until DNA extraction.



Key to populations: 1=Lumino, 2=Kwapa, 3=Osukuru, 4= Namungalwe, 5=Kaliro, 6=Ogur, 7=Serere

Figure 1: Map of Uganda showing the locations of sampling sites

3.3: GENOMIC DNA EXTRACTION

Extraction of genomic DNA was done using standard procedures (Aljanabi and Martinez 1997): the excised fly legs were homogenized in 300µl of extraction buffer [0.1M NaCl, 0.2M Sucrose, 0.1M Tris pH 9.1, 0.5% SDS]; the homogenates were incubated for 30 to 60 minutes at 60°C. 35µl of 8M Potassium Acetate was added to the homogenate while still warm and mixed by tapping gently followed by dipping in ice crystals for at least 30 minutes in order to precipitate SDS and proteins. The homogenate was centrifuged for 15 minutes at 14,000 revolutions per minute, and the supernatant transferred to a new vial. DNA was precipitated by adding 600µl of 100% ethanol followed by centrifuging for 15 minutes at 14,000 revolutions per min to produce a DNA pellet. The pellet was then washed using 600µl of 70% ethanol and air dried at room temperature. The DNA was reconstituted in solution by adding 50 µl of double distilled water and then stored at -4°C until PCR.

3.4: MICROSATELLITE GENOTYPING

Eleven previously cloned but unpublished sequences containing microsatellite loci from *G. f. fuscipes* were obtained from Alan Robinson of the Joint FAO/IAEA Agriculture and Biotechnology Laboratories, Seibersdorf, Austria. Primers were designed using the web based computer program; Primer3 (available at http://primer3.sourceforge.net/). The primers were synthesized, optimized and used in genotyping the samples at the following loci; A03, A10, B03, B05, B08, B101, C107, D05, D06, D12 and D101. The details and characteristics of the different microsatellite DNA loci are summarized in table 1. These microsatellite loci were then amplified using polymerase chain reaction (PCR). The forward primer of each pair was synthesized with a fluorescent dye group — 6-FAM, PET, or VIC (Applied Biosystems) on the 5' end to allow detection and sizing of fragments on an ABI Prism 3730 DNA sequencer.

All loci were amplified separately in 12.5 µl PCR volumes performed containing final amounts of 1 ng of template DNA, 2.5 µM of dNTPs, 12.5 µM of $MgCl_2$, 0.10 unit *Taq* DNA polymerase (AmpliTaq Gold) and 2.5 µM of each of the forward and reverse primers.

Reactions were performed in a GeneAmp® PCR System 9700 (Applied Biosystems) thermocyler according to the following profile; an initial denaturation for 10 min at 94 °C followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 58.5 °C for 30 s (55.0 °C for loci D06 and D12), extension at 72 °C for 30 s. All reactions were followed by a final extension step for 20 min at 72 °C.

To maximize throughput, a multiplex of three or four PCR products were combined in 96 well PCR plates with the following combinations of loci: (1) A03, B05, D06; (2) A10, B03, B08, D12, D101; and (3) D05, B101, C107. The multiplex of PCR products were subjected to capillary electrophoresis with Liz 600 internal marker. The loading mix (0.5 µl of each PCR Product, 0.5 µl of Size Standard and 14 µl of Hi-Di Formamide in 15 µl total reaction volume) was denatured at 92 °C for 3 min and loaded on POP-7TM performance optimized polymer of ABI Prism 3730 DNA sequencer (Hitachi - Applied Biosystems) and electrophoresed at 1.5kV for 3.8 hours at constant temperature of 18.7 °C. The microsatellite fragment lengths estimation and allele calling were then performed using GeneMapper version 3.7 software package (Applied Biosystems 2000a).

Table 1: Characteristics of the eleven microsatellite loci used in this study

Loci	Repeat	Repeat Characteristics	Product length	Primer sequence	
				Forward	Reverse
A03	(AC)12	Perfect	236	VIC-CGAAGCAGGTAGATAGAATCC	CTGCATTGGAAGTTACTTATCC
A10	(CA)13	Perfect, but flanked by TC repeats	190	PET-GCAACGCCAAGTGAAATAAAG	TACTGGGCTCGCGTACATAAT
B03	(AG)17	Perfect	236	6-FAM-GGAGGCTATGCTGATGAATG	TGATGCGAAAAAGAGAAACAG
B05	(GA)12	Perfect	212	6-FAM-CGCGCTTAGCTAGGAAACTC	AACGATTTGCTGTCCTCGAT
B08	(TC)20	Perfect	204	VIC-TGACCACTAAATCGTCCTCAC	TCCCGAGCTTAGAGAGAGTG
B101	(AG)19	Perfect	285	VIC-CGACTTTTTAGAGCCAGTGC	TTTGGTTGTTCATTCCTCAAC
C107	(CAT)12	Perfect	296	PET-CCCATTTAGACGATATTCTGC	CGTATAGCCAGACGTTGATATG
D05	(CAG)10	Perfect	111	6-FAM-TTTCCTTCCAGACGAACTG	CTTGGTATGGTCGTACATGG
D06	(GCT)9	Perfect	268	PET-TAACGGGGAGCTAAAAGAAG	AAATCATCAGCAGCATCATC
D12	(TGC)7	Perfect	135	6-FAM-GTTGATGGTCACACAACATAAG	TCAATGAGGAAAACTGAACTG
D101	(GCT)7	Perfect	176	6-FAM-TGCCTTTACACTGCATACTACC	AAAAAGAGGAGCAATGATGTG

3.5: DATA ANALYSIS

Microsatellite data were analyzed for three major aspects; (i) genetic diversity within populations, (ii) population differentiation and (iii) phylogenetic relationships between populations. Due to missing data resulting from non amplification of some loci, only those with 30% or less missing data were subjected to statistical analysis. As a result, only five of the eight loci were analyzed.

3.5.1: Hardy-Weinberg and Genotypic disequilibrium

GENEPOP 3.2 (Raymond and Rousset, 1995) was used to test each locus in each population for departure from Hardy-Weinberg proportions using Guo and Thompson (1992) method and to test all pairs of loci for genotypic disequilibrium. For the Hardy-Weinberg test, significance levels were adjusted using the sequential Bonferroni method to take into account multiple comparisons (Rice, 1989)

3.5.2: Genetic variation

Measures of genetic diversity for each population were assessed by calculating mean number of alleles per locus (N_A), total number of alleles per population, observed heterozygosity per locus (H_o) and mean expected heterozygosity (H_E) per locus under the Hardy-Weinberg expectation (Nei, 1987). F_{IS} as a measure of heterozygote deficiency or excess (Wright 1978) was estimated for each locus in each population

3.5.3: Population differentiation

Population differentiation was assessed using Weir and Cockerham's (1984) analog of Wright's F_{ST} computed using ARLEQUIN (Excoffier *et al.*, 2005). The null hypothesis of identical allelic and genotypic frequency distributions was rejected when the probability (*P*-value) of the exact test was < 0.05. Two approaches were used to evaluate if any population structure can be deduced from localities with different ecological conditions and epidemiological histories using; i) the assignment test as implemented in the programme

STRUCTURE ver.2.1 (Prichard *et al.*, 2000) and ii) Analysis of Molecular Variance (AMOVA) using ARLEQUIN (Excoffier *et al.*, 2005). Using these approaches, it was examined if; i) flies sampled from localities south of Lake kyoga and those from north of Lake Kyoga would cluster into separate groups and ii) flies sampled from peridomestic habitats and those from riparian habitats would cluster into separate groups. Phylogenetic relationships between populations was assessed using Neighbor-joining population trees, computed using MEGA 2.1 (Kumar *et al.*, 2004) based on matrices of F_{ST} values. A correlation between genetic distances as measured by F_{ST} values (computed with the pair-wise distance method) and geographic distances was investigated using the Mantel test. Approximate straight line geographical distances were calculated using Google Earth (http://earth.google.com).

BOTTLENECK Version 1.2.02 (Cornuet and Luikart, 1996) was used to detect any recent population bottleneck. Calculations were performed for each population and locus under the Two Phase Model (TPM). The Wilcoxon sign-rank test was conducted to determine whether populations exhibit significant number of loci with gene diversity excess. The test provides a relatively high power of resolution and can be used with a few loci and individuals (Lowe *et al.*, 2004).

CHAPTER 4

RESULTS

4.1: PATTERNS OF GENETIC VARIATION

4.1.1: Microsatellite Genotyping

In total, 274 *G. f. fuscipes* individuals originating from 7 subpopulations within sleeping sickness foci in Uganda were genotyped at 11 microsatellite loci. Loci B101 and C107 were not amplified after modification of the forward primers with dye labels. Locus D06 was monomorphic in all sub-populations except Kaliro and Ogur. Samples from Namungalwe and Kaliro were not polymorphic at locus D12, while B03 and B08 did not show any polymorphism in samples from Serere. Therefore, only five loci: D05, B03, D101, B05 and D12 met the criterion of analysis allowing only 30% missing data, and therefore only five loci were subsequently analyzed. The results are summarized in Table 2 and representative eletropherographs in Figures 2 and 3.

Table 2: Summary of microsatellite primers screened by population

District	Population	Loci										
		A03	A10	B03	B05	B08	B101	C107	D05	D06	D12	D101
Busia	Lumino	P	P	P*	P*	P	-	-	P*	M	P*	P*
Tororo	Kwapa	P	P	P*	P*	P	-	-	P*	M	P*	P*
	Osukuru	P	P	P*	P*	P	-	-	P*	M	P*	P*
Iganga	Namungalwe	P	P	P*	P*	P	-	-	P*	M	M	P*
Kamuli	Kaliro	P	P	P*	P*	P	-	-	P*	P	M	P*
Lira	Ogur	P	P	P*	P*	P	-	-	P*	P	P*	P*
Soroti	Serere	P	P	M	M	M	-	-	P*	M	P*	P*

- = not amplified at all, M = monomorphic, P = polymorphic, * = met criteria for analysis

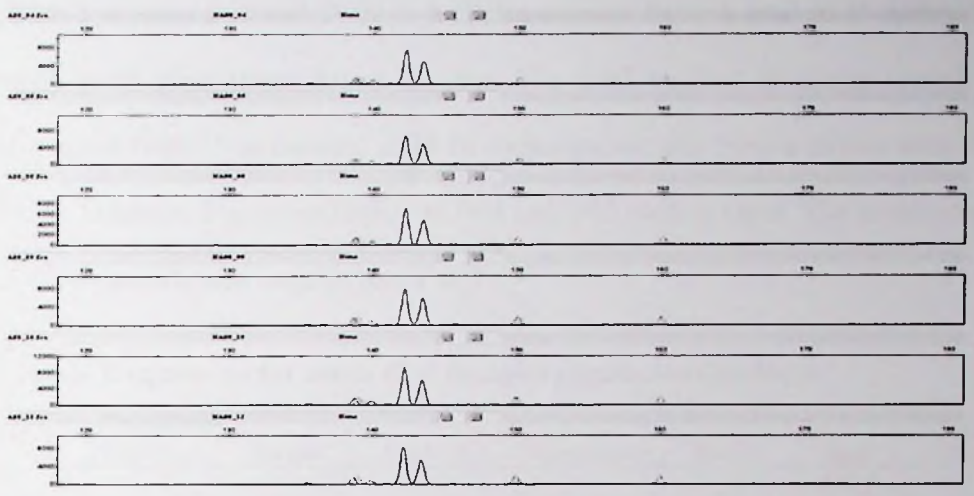


Figure 2: Electropherograph of locus D12 showing no polymorphism in samples from Namungalwe

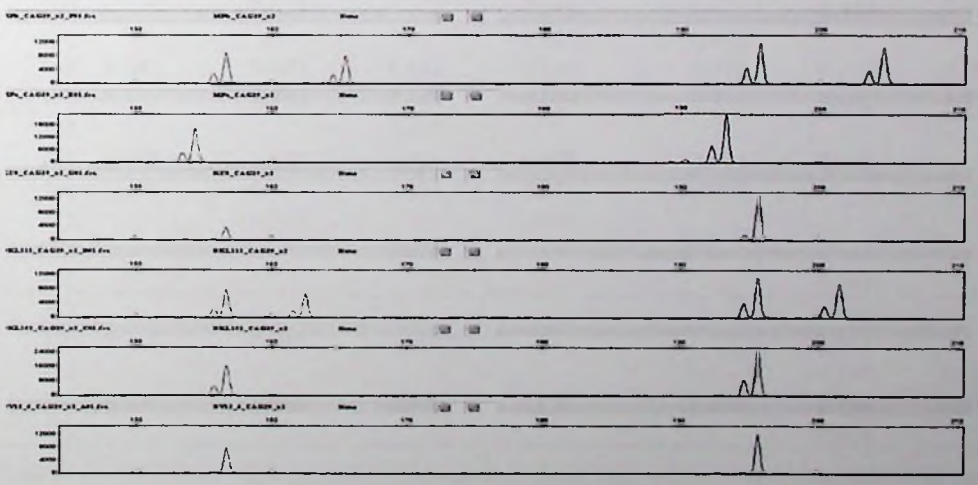


Figure 3: Multiplex electropherograph of loci D101 and D12 showing polymorphism in samples from Ogur

4.2.2: Genetic diversity within populations

4.2.2.1: Allele frequencies

The five loci analyzed in this study were polymorphic with a total number of alleles per locus ranging from 3 at locus B05 and D101 to 8 alleles at locus D05. A total of 25 different alleles were scored in all populations across all loci. The total number of alleles scored in each population varied from 10 in Lumino to 14 in Namungalwe. Six private alleles were detected; two at B03 in Osukuru, 2 at locus D05, 1 at B03 and D12 each in Ogur. The number of alleles per locus within populations ranged from 1 to 7.

Table 3: Allele frequencies for seven *G. f. fuscipes* populations by locus

Locus	Allele	Populations						
		Lumino	Kwapa	Osukuru	Namungalwe	Kaliro	Ogur	Serere
D05	93						*0.011*	
	96		0.053		0.022		0.245	0.063
	99				0.022	0.014		
	102	**0.583**	**0.632**	**0.827**	**0.630**	**0.771**	0.096	**0.281**
	105	0.417	0.289	0.173	0.326	0.214	0.245	0.219
	108						**0.309**	0.125
	111						*0.021*	
	114		0.026				0.074	0.313
B05	205	0.458	**0.548**	**0.688**	0.313	0.387	0.163	0.344
	207	**0.542**	0.429	0.313	**0.646**	**0.613**	0.388	**0.531**
	209		0.024		0.042		**0.449**	0.125
D101	156				0.125	0.008		
	165	**0.583**	0.447	**0.531**	**0.667**	**0.762**	0.139	0.267
	171	0.417	**0.553**	0.469	0.208	0.230	**0.861**	**0.733**
B03	222			*0.011*				
	224	**0.692**	**0.833**	**0.924**	**0.857**	**0.992**	**0.941**	
	226	0.308	0.111	0.054	0.071	0.008		1.000
	228		0.056		0.071			
	232			*0.011*				
	250						*0.059*	
D12	267						*0.076*	
	270	0.115	0.045	0.133			0.011	
	273						0.130	0.143
	276	**0.885**	**0.955**	**0.867**	1.000	1.000	**0.750**	**0.857**
	279						*0.033*	

Italicized values denote private allele (s) in a population; bold faced values denote most frequent alleles in each population

4.2.2.2: Hardy-Weinberg Expectations

There were five significant departures from Hardy-Weinberg Expectations (HWE) after applying a Bonferroni correction on the P-values. Samples from Kwapa showed significant departure from HWE at loci D05 and B03. Osukuru, Namungalwe and Ogur populations were not in HWE at loci D05, B03 and B05 respectively. All the observed deviations from H-W proportions were a result of heterozygote deficiency as indicated by the positive F_{IS} values. All the five loci were in linkage disequilibrium ($P \geq 0.05$). As a result, all the five loci were considered as independent markers. The details of the analyses are summarized in Table 4.

4.2.2.3: Fixation indices

All estimates of fixation indices (F_{IS}) at different loci returned positive values indicating variable extents of homozygote excess except for Kamuli, Tororo and Lira at loci D101, B03, and D12 respectively where negative F_{IS} values were detected. The average F_{IS} value was negative only for samples from Kaliro population. The results are presented in detail in Table 5. The majority of the F_{IS} values although positive, were not significantly positive (Table 4).

4.2.2.4: Gene Diversity

The gene diversity indices as measured by mean observed heterozygosity [H_O (average)] and expected heterozygosity [H_E (average)] in the sampled populations were generally low. Single locus expected heterozygosities (H_E) varied from 0.00 at loci D12 (Namungalwe and Kaliro) and B03 (Serere) to 0.7802 at B05 in Serere.

Table 4: Summary Statistics of genetic diversity at 5 loci in 7 populations of *G. f. fuscipes*

Locus	Genetic parameter	Populations						
		Lumino	Kwapa	Osukuru	Namungalwe	Kaliro	Ogur	Serere
D05	N	12	19	24	17	35	41	16
	A	2	4	2	4	3	7	5
	Ho	0.3333	0.1579	0.0385	0.4348	0.3429	0.7021	0.6250
	He	0.5072	0.5277	0.2919	0.5063	0.3640	0.7781	0.7802
	F_{is}	0.353	0.707	0.870	0.144	0.059	0.099	0.204
	P(HW)	0.2548	**0.000**	**0.0002**	0.2902	0.4820	0.0505	0.1692
B05	N	12	21	46	18	62	43	16
	A	2	3	2	3	2	3	3
	Ho	0.5833	0.5238	0.4167	0.4583	0.4194	0.4490	0.7500
	He	0.5181	0.5285	0.4342	0.4938	0.4784	0.6278	0.6028
	F_{is}	-0.132	0.009	0.073	0.124	0.041	0.287	-0.254
	P(HW)	0.8553	0.5846	0.4007	0.2390	0.5157	**0.0012**	0.9291
D101	N	12	12	32	10	61	18	15
	A	2	2	2	3	3	2	2
	Ho	0.3333	0.2632	0.3750	0.5000	0.4426	0.1667	0.1333
	He	0.5072	0.5078	0.5060	0.5181	0.3692	0.2460	0.4046
	F_{is}	0.353	0.489	0.262	0.036	-0.201	0.329	0.678
	P(HW)	0.2550	0.0414	0.1258	0.4776	0.9822	0.2732	0.0209
B03	N	13	18	45	12	61	16	2
	A	2	3	4	3	2	2	1
	Ho	0.1538	0.1111	0.1522	0.0000	0.0161	0.0000	0.0000
	He	0.4431	0.2984	0.1448	0.2646	0.0161	0.1141	0.0000
	F_{is}	0.662	0.634	-0.052	1.0000	-	0.634	-
	P(HW)	0.0331	**0.0051**	1.0000	**0.0014**	-	0.0307	-
D12	N	13	11	30	15	41	41	7
	A	2	2	2	1	1	5	2
	Ho	0.0769	0.0909	0.2000	0.0000	0.0000	0.4565	0.0000
	He	0.2123	0.0909	0.2350	0.0000	0.0000	0.4181	0.2637
	F_{is}	0.647	-	0.151	-	-	-0.093	1.000
	P(HW)	0.1196	-	0.4143	-	-	0.3595	0.0775
A (Total)		10	14	12	14	11	19	13
H_o (Ave)		0.2962	0.2294	0.2365	0.2786	0.2442	0.3549	0.3017
H_e (Ave)		0.4376	0.3907	0.3224	0.3565	0.2455	0.4103	0.4103
F_{IS}(Ave)		0.377	0.350	0.261	0.326	-0.034	0.084	0.407

N=number of individuals genotyped; A=number of alleles detected in each population; Ho=observed heterozygosity; He =expected heterozygosity; F_{IS} =Weir and Cockerham's (1984) analog of Wright's fixation index; P(HW)=single locus P-Values for Hardy-Weinberg tests; H_o (Ave)=average observed heterozygosity across all loci; H_e (Ave) average expected heterozygosity across all loci; F_{IS} (Ave)=average Wright's fixation index across all loci. Bold denotes significant departure from HWE after Boferroni's correction.

4.2.2.5: Population Bottleneck

The expected gene diversity H_E was significantly greater than expected equilibrium gene diversity (*Heq*) at loci B03 and B05 in Osukuru and Ogur respectively (P<0.05). The Wilcoxon test, which tests for bottlenecks across all loci, did not detect significant effects of recent bottlenecks except in the Lumino population (P<0.05). Details of the population bottleneck analysis are presented in Table 5.

Table 5: Results of tests of genetic bottleneck by locus in each *G. f. fuscipes* population

Population	Locus	H_E	H_{eq}	Probability (Single locus)	Probability (Wilcoxon test)
Lumino	D05	0.507	0.284	0.1251	
	B05	0.518	0.285	0.0726	
	D101	0.507	0.286	0.1321	***0.03125***
	B03	0.443	0.279	0.2633	
	D12	0.212	0.278	0.4641	
Kwapa	D05	0.528	0.568	0.2977	
	B05	0.528	0.438	0.3653	
	D101	0.508	0.256	0.0669	0.59375
	B03	0.298	0.446	0.2178	
	D12	0.091	0.289	0.2533	
Namungalwe	D05	0.506	0.552	0.2957	
	B05	0.494	0.428	0.4433	
	D101	0.518	0.481	0.5220	0.84375
	B03	0.265	0.467	0.1543	
	D12				
Kaliro	D05	0.364	0.404	0.3810	
	B05	0.478	0.201	0.0934	
	D101	0.369	0.374	0.4422	0.84375
	B03	0.016	0.204	0.1439	
	D12				
Ogur	D05	0.778	0.706	0.2008	
	B05	0.628	0.383	***0.0470***	
	D101	0.246	0.261	0.5230	0.68750
	B03	0.114	0.263	0.3302	
	D12	0.418	0.598	0.0942	
Serere	D05	0.780	0.666	0.0631	
	B05	0.603	0.457	0.1558	
	D101	0.405	0.272	0.3167	0.06250
	B03	0.000			
	D12	0.264	0.330	0.4774	
Osukuru	D05	0.292	0.238	0.3981	
	B05	0.434	0.210	0.1752	
	D101	0.506	0.229	0.1752	0.0.3125
	B03	0.145	0.510	***0.0213***	
	D12	0.235	0.232	0.4661	

Probability is calculated to determine whether the expected gene diversity (H_E) is higher than the expected equilibrium gene diversity (H_{eq}) which is calculated from observed number of alleles for each locus in each locus under the assumption of a mutation-drift equilibrium

4.2.3: Population Differentiation

4.2.3.1: Pair-wise F_{ST} estimates

Using F_{ST} to quantify the proportion of genetic differences between populations, 12 out of 21 pair wise comparisons showed significant population differentiation. The significant differentiation revealed by F_{ST} values ranged from 0.0443 ($P<0.05$) between Kaliro and Lumino to 0.2288 ($P<0.001$) between Serere and Kaliro. Details of all pair wise comparisons between populations are summarized in Table 6.

Table 6: Pair-wise FST estimates across all loci

Population	Lumino	Kwapa	Namungalwe	Kaliro	Ogur	Serere
Kwapa	0.0255					
Namungalwe	0.0847	0.0091				
Kaliro	0.0443*	0.0862***	0.0417			
Ogur	0.0896***	0.0773***	0.1678***	0.1567***		
Serere	0.0069	0.0542	0.1273***	0.2288***	0.0083	
Osukuru	0.0097	0.0476	0.0723**	0.1186***	0.1428***	0.0953**

* denotes significance at 0.05 level
** denotes significance at 0.01 level
*** denotes significance at 0.001

4.2.3.2: Isolation by geographical distance

Approximate straight line geographical distances ranged from 18.3km (between Osukuru and Kwapa) to 228.78 km (between Ogur and Lumino). According to this test, the correlation between geographic and log transformed genetic distances (log F_{ST}) was statistically significant. The correlation coefficient was estimated to be $rY_1 = 0.4110$ ($P < 0.05$).

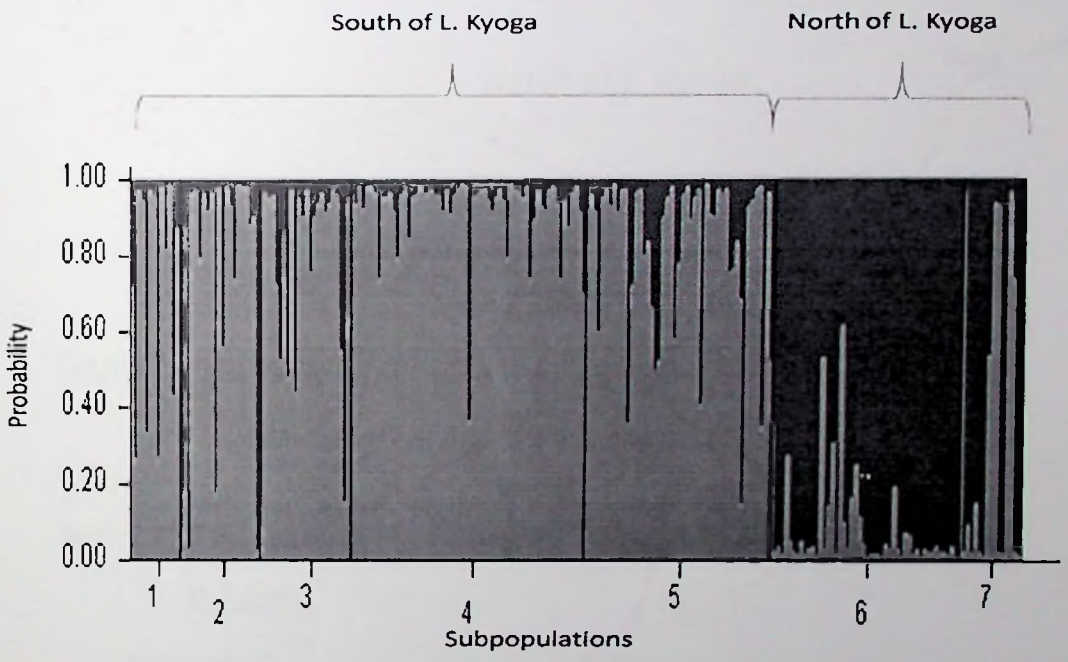
4.2.3.3: Analysis of Molecular Variance (AMOVA)

To examine the degree of population structure exhibited by the seven *G. f. fuscipes* populations, they were grouped according to geographical origin. Two groups were formed, the first group included all populations south of Lake Kyoga (Namungalwe, Kaliro, Lumino, Osukuru and Kwapa), and the second group included samples from north of the Lake (Ogur and Serere). The AMOVA indicated significant differentiation at all hierarchical levels between populations (F_{ST} = 0.1474, $P < 0.001$), between populations within each group (F_{SC} = 0.0313, $P < 0.001$) and among the two groups (F_{CT} = 0.1199, $P < 0.001$). Relatively little of the variation was explained by differences between populations within groups (2.75%); a substantial portion was explained by differences between northern and southern populations (11.99%), while differences between individuals accounted for 85.26% of the variations observed.

In order to explore the genetic basis of the observed differences in *G. f. fuscipes* peridomestic habitat preference in Kaliro and Namungalwe populations, and the rest of the populations sampled from riparian habitats, three groups were created. One group comprised samples from north of Lake Kyoga (Ogur and Serere), a southern but peridomestic group (Lumino, Namungalwe, Kaliro) and a southeastern-riverine group (Osukuru, Kwapa). The AMOVA indicated significant differentiation at all hierarchical levels between populations (F_{ST} = 0.1143, $P < 0.001$), between populations within each group (F_{SC} = 0.0178, $P < 0.001$) and among the two groups (F_{CT} = 0.1298, $P < 0.05$). This is corroborated by the pair-wise F_{ST} values that indicate differentiation between populations belonging to the same geographical group, as well as between populations belonging to different groups.

4.2.3.4: Assignment of individuals to populations

Analysis of microsatellite data using Structure (Pritchard *et al.*, 2000) returned strong support for subdivision of the sample into two groups (see figure 2). A large percentage of individuals in the populations from the south of Lake Kyoga (Lumino [77.6%], Kwapa [73.1%], Osukuru [84.5], Namungalwe [82.6%] and Kaliro [93.5%]) were assigned to one subgroup. Most individuals from populations north of Lake Kyoga (Ogur [89.7%] and Serere [63.7%]) were assigned to the other subgroup (Figure 4).



Key to populations: 1=Lumino, 2=Kwapa, 3=Osukuru, 4= Namungalwe, 5=Kaliro, 6=Ogur, 7=Serere

Figure 4: Assignment of individuals to a subgroup based on Structure analysis

4.2.3.5: Phylogenetic Analysis

Two major clades were detected after construction of phylogeographic tree using the Neighbor-Joining method. The clades were consistent with the geographic origins of the *G. f. fuscipes* samples. Samples from the Busoga sub-region (Namungalwe and Kaliro) cluster together while those from north of Lake Kyoga form another clade. The remainder, from the greater Tororo (Lumino, Osukuru and Kwapa) formed a more complex branching as shown in Figure 5.

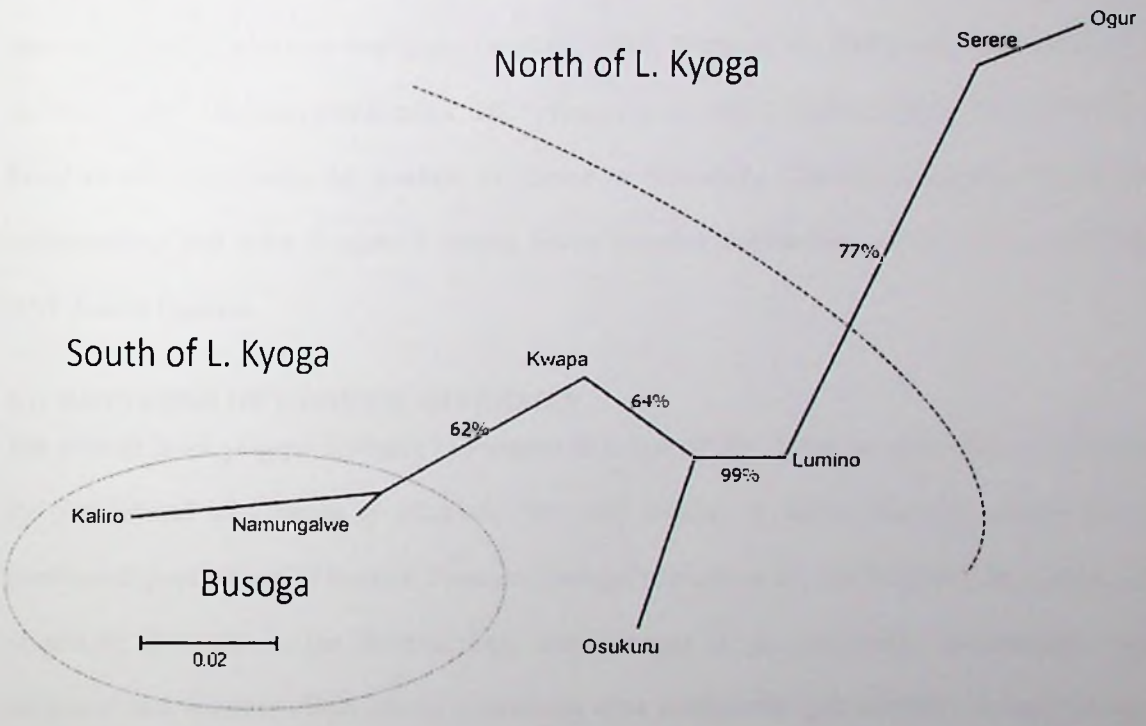


Figure 5: Un-rooted Neighbor Joining Tree based on pair-wise F_{ST} values

CHAPTER 5

DISCUSSIONS

This study was undertaken to evaluate the population genetic diversity of *G. f. fuscipes* the main tsetse species found in active HAT foci of Uganda. Knowledge of the genetic structure of a vector population is useful for understanding epidemics and will contribute to the design of a rational disease control operation (Krafsur 2005). Microsatellite DNA (Bennett, 2000; Loxdale and Lushai, 1998), which has proven successful for population studies of many arthropod species, including insect vectors (Lanzaro *et al.*, 1995; Wang *et al.*, 2001) and tsetse (Solano *et al.*, 1999, 2000; Gooding and Krafsur, 2005; Ouma *et al.*, 2005, 2006; Camara *et al.*, 2006 and Ravel *et al.*, 2007) were the markers of choice in this study. The allele frequencies at five microsatellite loci were compared among seven putative populations of *G. f. fuscipes* from HAT foci in Uganda.

5.1: PATTERNS OF GENETIC DIVERSITY

The overall level of gene diversity (Measured in terms of H_E, Table 4) observed in the tsetse fly populations of Uganda is relatively low and similar to those reported among *G. p. gambiensis* populations of Burkina Faso and Senegal (Solano *et al.*, 2000, 1999). The low level of genetic diversity in the Burkina Faso and Senegal *G. p. gambiensis* populations was attributed to a founder effect due to re-invasion after eradication and control efforts (Politzar and Cuisance, 1984). In contrast however, studies of *G. pallidipes* from Kenya revealed relatively higher levels of gene diversity (Krafsur, 2003; Ouma *et al.*, 2005).

The low levels of gene diversity in *G. f. fuscipes* samples from Uganda can be interpreted in two ways: as a natural characteristic of *G. f. fuscipes* and/or as a result of population bottlenecks following control campaigns (Okoth, 1999; Okoth *et al.*, 1991).

Low gene diversity being a natural characteristic feature of *G. f. fuscipes* microsatellite loci can not be the reason to explain the observed diversity among the putative populations analyzed. This is because relatively high gene diversities have been reported in separate studies involving samples from a laboratory colony of *G. f. fuscipes* (Solano *et al.*, 1997; Luna *et al.*, 2001, Ouma *et al.*, 2003).

The most plausible explanation for the observed gene diversity in this study is population bottleneck (s). There have been extensive tsetse control campaigns in Uganda to eliminate local populations of tsetse to control both animal and human trypanosomiasis. Studies by Okoth *et al.* (1991) and Lancien (1991) indicated drastic reduction in the population density of tsetse flies in Uganda. The genetic test for bottleneck in this study however, did not reveal bottleneck events except for a single population – Lumino.

The failure to detect a bottleneck in tsetse fly populations may be ascribed to the long generation time. Tsetse have approximately eight generations per year, and their effective population sizes are relatively small, so four years is probably the maximum extent that detectable disequilibrium occurs between allelic diversity and heterozygosity (Cornuet and Luikart, 1996; Ouma *et al.*, 2005). They reasonably infer from such observations that tsetse populations recover full allelic diversities in the intervals between population minima and the current genetic sampling. *Glossina f. fuscipes* populations from Lumino have been subjected to a sustained deterioration of the habitat in the recent past (Rutto *et al.*, 2007). This probably explains why the genetic bottleneck was detectable as opposed to the rest of the samples.

5.2: GENETIC DIFFERENTIATION

Strong degree of genetic structuring of *G. f. fuscipes* populations was detected by all estimates of genetic differentiation. Pair-wise F_{ST} estimates disclosed highly significant genetic differentiation at microsatellite loci within and between the samples from the south and north of Lake Kyoga. Lake Kyoga appears to be the main factor in the genetic structuring of the putative *G. f. fuscipes* populations, separating clearly distinct northern and southern population clusters. In the Kaliro Vs Kwapa, and the Ogur Vs Serere population pairs which are only approximately 100km from each other, significant divergences were detected, while putative populations from Tororo district (Osukuru and Kwapa) showed no evidence of genetic differentiation within the district suggesting that the two putative populations are in a state of panmixia and probably represent a single subpopulation.

These observed genetic differentiation between and within the geographical clusters can be interpreted in light of the dispersal capacity of *G. f. fuscipes* and physical barriers:

(i) Dispersal capacity: *G. f. fuscipes*, like other *Palpalis* group tsetse is not very vagile as some ecological studies have suggested. Its ability to disperse and reproduce successfully over long distances is limited, effectively rendering gene flow insufficient to neutralize genetic drift. It has been shown that tsetse may not use their maximal capacities of dispersal if they do not need to: very recent data on mark-release-recaptures of *G. palpalis* in Burkina Faso showed that when tsetse are released in very favorable area (i.e. an area with humidity and shade typical of *G. palpalis*), a statistically significant number stays within this area, whereas when they are released in a degraded area, they tend to look for a favorable area and then diffuse over much longer distances (Ravel *et al.*, 2007). The expectation from organisms with limited dispersal capacity is high genetic differentiation, even across short distances. Data presented in this study clearly conforms to this expectation.

(ii) Physical barriers: The limited gene flow between populations within the two clusters suggests that the patchy distribution of *G. f. fuscipes* populations is another factor shaping the population structure of these vectors despite their ability to disperse (338 m/day in male *G. fuscipes* (Rogers, 1977). In this, although being a riverine species, *G. f. fuscipes* is more similar to savannah species such as *G. morsitans* and *G. pallidipes* than to the other riparian tsetse species; *G. p. gambiensis*. Consequently, forces of genetic drift in *G. f. fuscipes* are much stronger than gene flow (Ouma *et al.*, 2007).

5.3: IMPLICATIONS FOR DISEASE CONTROL

In this study, the observed strong differentiation of flies and the apparent lack of gene flow have implications for the vector based control of both animal and human African trypanosomiasis:

(i) Disease transmission between foci is unlikely to be aggravated by vector movements since there is evidence of limited *G. f. fuscipes* (major vector) dispersal between the different foci.

(ii) It is possible to eliminate tsetse from either side of Lake Kyoga without fear of reinvasion from another side. The lake can therefore be used as an effective landmark when delimiting tsetse control/elimination units.

5.4: CONCLUSIONS

Thorough understanding of the factors influencing population structure and bionomics of *G. f. fuscipes* is necessary to develop adequate regional and or area-wide vector control strategies. These results suggest that;

(i) *G. f. fuscipes* subpopulations from sleeping sickness foci in Uganda are highly structured.

(ii) Physical barriers, particularly Lake Kyoga can be effective land marks in delimiting control units when designing area-control strategies against tsetse (*G. f. fuscipes*).

(iii) *G f. fuscipes*, as other *Palpalis* group tsetse fly populations are not very vagile as some ecological studies have suggested (Ogwal and Kangwajye, 1988; Cuisance *et al.*, 1985; Bouyer *et al.*, 2007) and this could have important economic considerations when planning area-wide vector control programs, which may prove to be more effective than are now anticipated by knowledgeable tsetse fly ecologists.

5.5: RECOMMENDATIONS

The design of this study (sample sizes, geographic and genome sampling) is limited by at least three factors;

(i) Samples were collected only once from each putative population. Extended periods of sample collection should facilitate collection of sufficient numbers of samples per site for both temporal and spatial analyses.

(ii) One caveat of microsatellite analyses is the potential presence of null alleles. This could bias F_{ST} estimates as well as the structure analysis. To minimize this effect, primer pairs can be redesigned to eliminate null alleles.

(iii) Only five microsatellite loci were included so far, but additional microsatellite loci should be tested in similar studies in order to even increase the resolution of the analyses. An analysis with increased number of loci will provide a more robust picture of the structuring of these populations.

REFERENCES

Abila, P. P., Nakayima J., Okedi L.M., Akol M., Alioni V., and Otim, C.P. (2006). Assessment of the prevalence of sleeping sickness in Lango and Teso Sub regions, Uganda. 8th EANETT Scientific Conference (18th - 21st September, 2006). Hotel Africana, Kampala.

Aljanabi, S.M., and I. Martinez. (1997). Universal and rapid salt-extraction of high quality. Nucleic Acids Research, 25, no. 22: 4692–4693.

Allendorf, F. W., and G. Luikart (2007). Conservation and the Genetics of Populations. Malden, MA, USA.: Blackwell Publishing.

Applied Biosystems (2000a). ABI PRISM® GeneMapper high-performance genotyping software for linkage mapping applications. Foster City, CA, USA: Applied Biosystems.

Applied Biosystems (2000b). Automated DNA Sequencing: Chemistry Guide. Applied Biosystems

Avise, J.C. (1994). Molecular Markers, Natural History and Evolution. New York: Chapman and Hall.

Baker M.D., and Krafsur, E.S. (2001). Identification and properties of microsatellite markers in tsetse flies *Glossina morsitans sensu lato* (Diptera: Glossinidae). Molecular Ecology Notes, 1: 234-236.

Beebee, T., and G. Rowe. (2004). An Introduction to Molecular Ecology. New York: Oxford University Press.

Beebee, N. W., and A. Saul (1995). Discrimination of all members of the *Anopheles punctatus* complex by polymerase chain reactionrestriction fragment analysis. American Journal of Tropcal Medicine and Hygiene, 53: 478-481.

Behura, S. K. (2006). Molecular marker systems in insects: current trends and future avenues. Molecular Ecology, 15: 3087-3113.

Bennett, P. (2000). Demystified ...Microsatellites. Molecular Pathology, 53: 177–183.

Bensasson, D., D. X. Zhang, D.L. Hartl, and G. M. Hewitt (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology and Evolution, 16: 314–321.

Bensch, S., and M. Akesson (2005). Ten years of AFLP in ecology and evolution: why so few animals? Molecular Ecology 14: 2899–2914.

Bhalla, N. (2002). Pan African group takes lead against the tsetse fly. The Lancet, 359: 686.

Black, W.C. and E.S. Krafsur (1985a). Seasonal breeding structure in house fly, *Musca domestica* L., populations. Heredity, 56, no. 33: 289-298.

Black, W.C. and E.S. Krafsur (1985b). Use of sticky traps to investigate seasonal trends in the spatial distribution of houseflies and stable flies (Diptera: Muscidae). Journal of Medical Entomology, 22 no. 5: 550-557 .

Black, W.C., and E.S. Krafsur (1986). Temporal and spatial trends in allozyme frequencies in house fly populations, *Musca domestica* L. Theoretical and Applied Genetics, 71 no. 5.

Black, W. C., and G. C. Lanzaro (2001). Distribution of genetic variation among chromosomal forms of *Anopheles gambiae s.s.*: introgressive hybridization, adaptive inversions, or recent reproductive isolation? Insect Molecular Biology, 10: 3-7.

Blears, M. J., De Grandis S. A., H. Lee, and J. T. Trevors (1998). Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. Journal of Industrial Microbiology and Biotechnology, 21: 99-114.

Bouyer, J., A. Sibert, M. Desquesnes, D. Cuisance, and S. de La Rocque (2007). A model of diffusion of *Glossina palpalis gambiensis* (Diptera: Glossinidae) in Burkina Faso. *In* M.J.B. Vreysen, A. S. Robinson, and J. Hendrichs [eds.], Area-wide Control of Insect Pests. From Research to Field Implementation. Springer, Dordrecht, The Netherlands

Budd, L. (1999). DFID-Funded Tsetse and Trypanosomiasis research and Development since 1980. Economic Analysis. Department of International Development, London.

Camara, M. C.O. Harling, S. Ravel, J. Dujardin, J. Hervouet, T. De Meeu, M.S. Kabadouno, J. Bouyer,and P. Solano (2006). Genetic and Morphometric Evidence for Population Isolation of *Glossina palpalis gambiensis* (Diptera: Glossinidae) on the Loos Islands, Guinea. Journal of Medical Entomology, 43 no. 5: 853-860.

Caterino, M. S., Cho, S., and Sperling, F.A.H. (1998). The Current State of Insect Molecular Systematics: A thriving tower of Babel. Annual Review of Entomology, 45 (2000): 1-54.

Chizyuka, G. (1998). FAO Liaison Officers Summary Report. Harare: Food and Agricultural Organisation,.

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

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

Cuisance, D., J. Fevrier, J. Dejardin, and J. Filledier (1985). Dispersion line′ aire de *Glossina palpalis gambiensis* et *G. tachinoides* dans une galerie forestiere en zone soudano-guine′enne. Rev. Elev. Med. Vet. Pays Trop, 38: 153 -172

Dia, I., Lochouarn L, Boccolini D, Costantini C, and F. D. (2000). Spatial and temporal variations of the chromosomal inversion polymorphism of *Anopheles funestus* in Senegal. Parasite, 7: 179-184.

Dobzhansky, T. (1937). *Genetics and the origin of species.* New York: Columbia University Press.

Dujjardin, Schofield, and Tibayrenc (1998). Population structure of Andean Triatoma infestans: allozyme frequencies and their epidemiological relevance. *Medical and Veterinary Entomology* 12: 20-29.

Excoffier, L., G. Laval, and S. Schneider (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online, 1: 47-50.

Fevre, E., P. G. Coleman, M. Odiit, S. C. Welburn, M. E. J. Woolhouse (2001). The origins of a new *Trypanosoma brucei rhodesiense* sleeping sickness outbreak in eastern Uganda. *Lancet* 358: 625 - 628.

Futumya, D. J. (1986). Evolutionary biology. Sunderland, MA: Sinauer Assoc. Inc.

Gariou-Papalexiou, A., G. Yannopoulos, A. S. Robinson and A. Zacharopoulou (1998). Polytene Chromosome Maps and RAPD Polymorphisms in *Glossina auteni. In* K.H. Tan [Eds]. International Conference on Area-Wide Control of Insect Pests, May 28 - June 2, 1998 and the Fifth International Symposium on fruit Flies of Economic Importance, June 1 - 5, 1998. Penang, Malaysia: Penerbit Universiti Sains. pp 239 - 250.

Gariou-Papalexiou, A., G. Yannopoulos, A. Zacharopoulou, R.H. Gooding (2002). Photographic polytene chromosome maps for *Glossina morsitans submorsitans* (Diptera: Glossinidae): cytogenetic analysis of a colony with sex-ratio distortion. Genome, 45: 871-880.

Gidudu, A. M., Robinson, A.S. and Stamenova, A. (1999). Initial Studies on RAPD Polymorphism in *Glossina fuscipes fuscipes*. 25th International Scientific Conference on Trzpanosomiasis Research and Control, Mombassa, Kenya: OAU/IBAR, 214-215.

Gooding, R. H. (1996). Genetic Variation in Arthropod Vectors of Disease-Causing Organisms: Obstacles and Opportunities. Clinical Microbiology Reviews, 9: 301–320.

Gooding, R. H. (1992). Genetic Variation in tsetse flies and implications for trypanosomiasis. Parasitology Today, 8: 92-95.

Gooding, R. H. (1989). Genetics of two Populations of *Glossina morsitans centralis* (Diptera: Glossinidae) from Zambia. Acta Tropica, 46: 17-22.

Gooding, R. H., and E. S. Krafsur. (2005). TSETSE GENETICS: Contributions to Biology, Systematics, and Control of Tsetse Flies. Annual Review of Entomology, 50: 101-123.

Guo S.W. and E. A. Thompson (1992): Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. Biometrics, 48, 361-372

Hartl, D. L. (1988). A primer of population genetics. Sunderland, MA: Sinauer Associates.

He, M., and D. S. Haymer (1999). Genetic relationships of populations and the origins of new infestations of the Mediterranean fruit fly. Molecular Ecology, 8: 1247-1257.

Hodgkinson, V. H., Birungi J., Haghpanah M., Joshi S., Munstermann L.E. (2002). Rapid Identification of Mitochondrial Cytochrome B Haplotypes by Single Strand Conformation Polymorphism in *Lutzomyia longipalpis* (Diptera: Psychodidae) Populations. Journal of Medical Entomology, 39, no. 4: 689-694.

Hoy, M. A. (2003). Insect molecular genetics: An introduction to principles and applications. 2nd Edition. San Diego, CA: Academic Press.

Jordan, A. M. (1974). Recent Developments in the ecology and methods of control of tsetse flies (*Glossina* ssp) (Diptera: Glossinidae) - A Review. Bulletin of Entomological Research, 63: 361–399.

Jordan A. M. (1986). Trypanosomiasis Control and African Rural Development. Essex: Longman Group Limited.

Kimura, M. (1991). Recent developments of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. Proceedings of the National Academy of Science of America, 88: 5969-5973.

Krafsur, E. (2005). Role of Population Genetics in the Sterile Insect Technique. *In* V.A., J. Hendriches and A.S. Robinson Dyck [Eds] Sterile Insect Technique: Principles And Practice in Area-wide Integrated Pest Managemen. Dordrecht: Springer, 2005.

Krafsur, E. S. (2002). Population structure of the tsetse fly *Glossina pallidipes* estimated by allozyme, microsatellite and mitochondrial gene diversities. Insect Molecular Biology, 11: 37-45.

Krafsur, E. S. (2003). Tsetse fly population genetics: an indirect approach to dispersal. Trends in Parasitology, 19: 162–66.

Krafsur, E. S., and D.L. Wohlford (1999). Breeding structure of *Glossina pallidipes* populations evaluated by mitochondrial variation. Journal of Heredity, 90: 635–42.

Krafsur, E. S., and N. Griffiths (1997). Genetic Variation at Structural Loci in the *Glossina morsitans* Species Group. Biochemical Genetics, 35: 1-11.

Krafsur, E. S., and W. C. Black (1992). Analysis of isozyme loci in the face fly, *Musca autumnalis* DeGeer. Biochemical Genetics, 30: 625-634.

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

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

Lancien, J. (1991). Campaign against sleeping sickness in South-West Uganda by trapping tsetse flies. Ann Soc Belg Med Trop, 71 no. 1: 35-47.

Leak, S. G. A. (1999). Tsetse Biology and Ecology: their role in the epidemiology and control of Trypanosomiasis. New York: CABI Publishing.

Lewis, R. (2002). African sleeping sickness: A Recurring Epidemic. The Scientist, 16: 26.

Lewontin, R. C., and J. L. Hubby (1966). A molecular approach to the study of genetic heterozygosity in natural populations. II. Amount of variation and the degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics, 54: 595-609.

Lowe, A., S. Harris, and P. Ashton (2004). Ecological Genetics. Design, Analysis, and Application. Malden, MA: Blackwell Publishing.

Loxdale, H. D., and G. Lushai (1998). Molecular Markers in Entomology (Review Article). Bulletin of Entomological Research, 88: 577-600.

Luna, C., M. Bonizzoni, Q. Cheng, A. S. Robinson, S. Aksoy, and L. Zheng (2001). Microsatellite Polymorphism in Tsetse Flies (Diptera: Glossinidae). Journal of Medical Entomology, 38: 376-381.

Magona, J.W, J. Walubengo, M. Odiit, L.A. Okedi, P.P. Abila, B.K. Katabazi, A.M. Gidudu and W. Olaho-Mukani (2005). Implications of the re-invasion of Southeast Uganda by *Glossina pallidipes* on the epidemiology of bovine trypanosomosis. Veterinary Parasitology, 128, no. 1: 1-9.

Marquez, J. G., M. J. B. Vreysen, A. S. Robison, S. Bado, and E. S. Krafsur (2004). Mitochondrial diversity analysis of *Glossina palpalis* gambiensis from Mali and Senegal. Medical and Veterinary Entomology,18: 288–295.

Maudlin, I. (1979). Chromosome polymorphism and sex determination in a wild population of tsetse. Nature, 277: 300-301.

Maudlin, I. (1982). Inheritance of susceptibility to *Trypanosoma congolense* infection in *Glossina morsitans*. Parasitology Today, 1: 59-60.

Maudlin, I. (1985). Inheritance of susceptibility to trypanosomes in tsetse flies. Annals of Tropical Medicine and Parasitology,76: 225-227.

Maudlin, I. (1980). Population Genetics of tsetse flies and its relevance to trypanosomiasis research. Insect Science and its Application,1: 35 - 38.

McDonald, P. T., L. R. Hilburn and S. E. Kunz (1987). Genetic similarities among natural populations of the hornfly (Diptera: Muscidae). Annals of the Entomology Society of America, 80: 288-292.

Nei, M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Ogwal LM, Kangwagye TN. (1988). Population Dynamics of Glossina fuscipes fuscipes on Buvuma Islands, Lake Victoria, Uganda: Proceedings of the SIT Research Coordination Meeting, Vom. Plateau state Nigeria 6–10 June 1988 pp.67–78.

Okoth, J.O. (1986). Peridomestic breeding sites of Glossina fuscipesfuscipes Newst. in Busoga, Uganda and epidemiological implications for trypanosomiasis. Acta Tropica, 43: 283–6.

Okoth, J.O. (1999). Tsetse and trypanosomiasis control problems in South-East Uganda:past, present and alternative strategies. Schweiz Med Wochenschr, 129: 1091–8.

Okoth, J.O., V. Okethi, A. Ogola (1991). Control of tsetse and trypanosomiasis transmission in Uganda by applications of lambda-cyhalothrin. Medical and Veterinary Entomology, 5, no. 1: 121-128.

Ouma, J. O., J.G. Marquez, and E. S. Krafsur (2005). Macrogeographic population structure of the tsetse fly, *Glossina pallidipes* (Diptera: Glossinidae). Bulletin of Entomological Research, 95: 437–447.

Ouma, J. O., M. A. Cummings, K. C. Jones, E. S. Krafsur (2003). Characterization of microsatellite markers in the tsetse fly, *Glossina pallidipes* (Diptera: Glossinidae). Molecular Ecology Notes, 3: 450-453.

Ouma, J. O., Marquez,J. G., Krafsur,E. S. (2006). Microgeographical breeding structure of the tsetse fly, *Glossina pallidipes* in south-western Kenya. Medical and Veterinary Entomology, 20: 138-149.

Parsons, Y. M., and K. L. Shaw (2001). Species boundaries and genetic diversity among Hawaiian crickets of the genus *Laupala* identified using amplified fragment length polymorphism. Molecular Ecology, 10: 1765–1772.

Politzar, H. and D. Cuisance (1984). An integrated campaign against riverine tsetse *Glossina palpalis gambiensis* and *G. tachinoides* by trapping and the release of sterile males. Insect Science and its Application, 5: 439-442.

Pollock, J. N. (1982). Tsetse biology, Sytematics and Distribution, Techniques. Vol. 1. Rome: Food and Agricultural Organisation of the United Nationa.

Pritchard, J. K., M. Stephens, and P. Donelly (2000). Inference of population structure using multilocus genotype data. Genetics, 155: 945-959.

Queller, D. C., Strassmann, J.E., Hughes, C.R. (1993). Microsatellites and Kinship. Trends Ecological Evolution, 8: 258-88.

Ravel, S., T. de Meeus, J. P. Dujardin, D. G. Zeze, R. H. Gooding, I. Dusfour, B. Sane, G. Cuny, and P. Solano (2007). The tsetse fly *Glossina palpalis palpalis* is composed of several genetically differentiated small populations in the sleeping sickness focus of Bonon, Cote d'Ivoire. Infection Genetics and Evolution, 7: 116-125.

Raymond, M., and F. Rousset (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity, 86: 248–249.

Rice, W.R. (1989). Analyzing tables of statistical tests. Evolution, 43: 223-225.

Robinson, A. S. (2005). Genetic basis of the Sterile Insect Technique. *In* J. Hendrichs and A.S. Robinson Dyck V.A. [Eds] Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management, 95-114. Dordrecht: Springer.

Robinson, A. S., and R. H. Gooding. (1997). RAPD-PCR analysis of the genome of *Glossina austeni*. 23rd International Scientific Conference on Trypanosomiasis Research and Control. Banjul: OAU/IBAR, 264-265.

Rogers, D.J. (1977). Study of a natural population of tsetse flies and a model for fly movements. Journal of Animal Ecology, 46: 309–330.

Rutto, J.J., L. Okedi, O. Odipo and V. Odenyo. (2007). Remote Sensing and Geographical Information Systems Technologies in Human African Trypanosomosis Epidemiology. 9th Annual EANETT Conference. Dar es Salaam, Tanzania.

Shore, H. (1960). An Epidemic of sleeping sickness in the Lango District of Uganda. Part I. East African Medical Journal, 37, no. 1: 9-21.

Solano, P., G. Duvallet, V. Dumas, D. Cuisance, G. Cuny (1997). Microsatellite markers for genetic population studies in *Glossina palpalis* (Diptera: Glossinidae). Acta Tropica, 65: 175-180.

Solano, P., S. De La Rocque, D. Cuisance, B. Geoffroy, T. De Meeus, G. Cuny, G. Duvallet (1999). Intraspecific variability in natural populations of *Glossina palpalis gambiensis* from West Africa, revealed by genetic and morphometric analyses. Medical and Veterinary Entomology, 13: 401-407.

Solano, P., S. de La Rocque, T. de Meeus, G. Cuny, G. Duvallet and D. Cuisance. (2000). Microsatellite DNA markers reveal genetic differentiation among populations of *Glossina palpalis gambiensis* collected in the agro-pastoral zone of Sideradougou, Burkina Faso. Insect Molecular Biology, 9: 433 - 439.

Szalanski, A. (1995). Genetic Characterization and Population Genetics of Stable fly (Diptera: Muscidae). PhD Thesis, Faculty of the Graduate College of the University of Nebraska, University of Nebraska, Nebraska.

Sunnucks, P., A. C. C. Wilson, L. B. Beheregaray, K. Zenger, J. French, and A. C. Taylor. (2000). SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. Molecular Ecology 9: 1699-1710.

Tarimo-Nesbitt, S. A., R. H. Gooding, and B. M. Rolseth (1990). Genetic variation in two field populations and a laboratory colony of *Glossina pallidipes* (Diptera: Glossinidae). Journal of Medical Entomology, 27: 586–591.

Turner, D. A., and R. Brightwell (1986). An evaluation of a sequential aerial spraying operation against *Glossina pallidipes* Austen (Diptera: Glossinidae) in the Lambwe Valley of Kenya: Aspects of post-spray recovery and evidence of natural population regulation. Bulletin of Entomological Research,76: 331-349.

Vekemans, X., T. Beauwens, M. Lemaire, and I. Roldán-Ruiz (2002). Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. Molecular Ecology, 11: 139–151.

Wang, R., L. Zheng, Y.T. Toure, T. Dandekar, and F.C. Kafatos (2001). When genetic distance matters: Measuring genetic differentiation at microsatellite loci in whole-genome scans of recent and incipient mosquito species. PNAS, 98, no. 19: 10769–10774.

Weir, B. S., and C. C. Cockerham (1984). Estimating F-statistics for the analysis of population structure. Evolution, 38: 1358–1370.

West, D. F., T. Payette, T. Mundy, and W. C. Black (1997). Regional molecular genetic key for of thirteen snow pool *Aedes* species (Diptera: Culicidae) in Northern Colorado. Journal of Medical Entomology, 34: 404-410.

WHO. (1998). Control and surveillance of African Trypanosomiasis: Report of a WHO Expert Committee on the control and surveillance of African Trypanosomiasis. WHO Thechnical Report, World Health Organization.

Wilkerson, R. C., T. J. Parsons, D. G. Albrightt, T. A. Klein, and M. J. Braun (1993). Random amplified polymorphic DNA (RAPD) markers readily distinguish cryptic mosquito species (Diptera: Culicidae: Anopheles). Insect Molecular Biology, 1: 205-211.

Willhoeft, U. (1997). Fluorescence in situ hybridization of ribosomal DNA to mitotic chromosomes of tsetse flies (Diptera: Glossinidae: *Glossina*). Chromosome Research, 5: 262-267.

Wohlford, D. L., E. S. Krafsur, N. T. Griffiths, J. G. Marquez, and M. D. Baker (1999). Genetic differentiation of some *Glossina morsitans morsitans* populations. Medical and Veterinary Entomology, 13: 377-385.

Wong, A., Forbes M.R., and S. M.L. (2001). Characterization of AFLP markers in damselflies: prevalence of codominant markers and implications for population genetic applications. Genome, 44: 677–684.

Wright, S. (1978). Evolution and the genetics of populations: variability within and among natural populations. Chicago: University of Chicago Press.

Zhang, D.-X., and G. M. Hewitt (2003). Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Molecular Ecology, 12: 563-584.