MOLECULAR CHARACTERIZATION OF *Trypanosoma brucei brucei* AND *Trypanosoma brucei rhodesiense* ISOLATES AND RESPONSES OF *Glossina pallidipes* TO *Trypanosoma brucei brucei* CHALLENGE

BATETA ROSEMARY WANYONA

A Thesis Submitted to the Graduate School in Partial Fulfillment of the Requirements for the Award of Doctor of Philosophy Degree in Biochemistry of Egerton University

EGERTON UNIVERSITY

April, 2016

DECLARATION AND RECOMMENDATION

DECLARATION

This thesis is my original work and has not been submitted in part or whole for an award in any other institution.

Rosemary Wanyona Bateta

SD14/315/2011

Signature: ……. Date:……..

RECOMMENDATION

This thesis has been submitted for examination with our approval as university supervisors

Dr. Paul O. Mireji,
Egerton University, Kenya

Signature:..…… Date: …….

Dr. Grace Murilla
Kenya Agricultural and Livestock Research Organisation - Biotechnology Research Institute

Signature: …… Date: …….

Professor Serap Aksoy
Yale University

Signature: …….. Date: ……

COPY RIGHT

DEDICATION

This thesis is dedicated to my spouse Mr Mumasi Wanyama and children; Laura, Triza, Mitchelle and Nicole for their support and understanding while I was undertaking my studies.

ACKNOWLEDGEMENTS

I thank God for having given me sufficient grace and strength during the course of my study. I would like to thank my supervisors Dr Paul Mireji of Egerton University and KALRO–BioRI, Dr Grace Murilla of KALRO–BioRI and Professor Serap Aksoy of Yale School of Public health, Department of Epidemiology of Microbial diseases, for their support, constructive discussions, encouragement and guidance throughtout my Phd studies. Many thanks to my colleagues at KALRO-BioRI Biochemistry department for their assistance in animal experiments and sample generation during the initial phase of my research work. I am also grateful to Prof. Serap Aksoy and Dr Adalgisa Caccone of Yale University Department of Ecology & Evolutionary Biology for giving an opportunity to carry out part of my research work in their Laboratories and the great support that I received from staff in the two laboratories. Special thanks to my beloved husband Mr Mumasi Wanyama and daughters Laura, Triza, Mitchelle and Nicole for their love, patience, support and encouragement throught my studies. I would also like to thank my mother Selina, brothers and sisters especially Rebecca for their prayers and support. I am also grateful to Global Infectious Disease (GID) Research training programme funded by NIH and managed by Prof. Serap Aksoy and Evidence based control strategies of sleeping sickness NIH Grand Number 5R03RTW8413-2 by Dr Johnson Ouma for supporting my studentship. Lastly I thank the Director KALRO for granting me time to pursue my studies.

ABSTRACT

African trypanosomiases are a group of related diseases that affect humans (Human African Trypanosomiasis - HAT) and their livestock (Animal African Trypanosomiasis), with devastating medical and economic consequences for Africa. The diseases are caused by trypanosome parasites cyclically transmitted by tsetse flies. Recent reports indicate that HAT is on the decline in sub Sahara Africa. However, the disease still remains a major health problem in some parts of Africa such as South Sudan, Democratic Republic of Congo and Uganda. Reports also indicate repeated irregular *T. b. rhodesiense* out breaks in traditionally endemic areas in south-east Uganda, with continued spread to previously unaffected areas in central Uganda. Although the disease has been reported to spread to new areas in Uganda, observations in the field indicate that the infection rates in tsetse flies are very low. Competence of the flies to transmit the parasite is determined by host midgut responses that either leads to clearance (self-cure) or establishment of the parasite infections. Investigations were conducted to determine evolutionary dynamics behind the origin of new foci and the impact of host species on parasite genetic diversity in Uganda. *Trypanosoma brucei* isolates {N = 269, n= 58(for Samples isolated from cattle)} were collected from different areas in Uganda and western Kenya and genotyped at 17 genetic marker loci (microsatellite). Analysis was carried out using Bayesian clustering and Discriminant Analysis of Principal Componens. Presence of serum resistance associated (SRA) gene in the isolates was determined using PCR. Results showed that the genotyped trypansosome isolates partitioned into three distinct genetic clusters. Clusters 1 and 3 included isolates from central and southern Uganda, whereas cluster 2 was composed of mainly isolates from western Kenya. F_{ST} values between sampling sites ranged from 0 to 0.67 while between the three genetic structures it ranged from0.24 to 0.46 with most values being statistically significant $P<0.01$. AMOVA results at $p<0.05$ indicated that the genetic variation was apportioned within (71.8%) rather than among the three clusters. These analyses revealed genetic admixture among the three genetic clusters and long-range dispersal, suggesting recent and possibly on-going gene flow between them and the new foci of HAT in central Uganda is as a result of northward movement of *Trypanosoma brucei rhodesiense* from the traditionally endemic foci. Therefore disease control efforts need to be enhanced to prevent continued spread to new foci.

To determine molecular responses in *Glossina pallidipes* challenged with trypanosomes, teneral female *Glossina pallidipes* flies were challenged with *Trypanosoma brucei brucei* and

dissected at 24 or 48 hours post challenge (hpc). Tissues were collected and analysed to establish key molecular responses mediating initial phase of establishment of the parasite in the fly. Transcriptomes of midguts and respective carcasses from the challenged and unchallenged flies were sequenced on illumina RNA-Seq platform, and analyzed for differentially expressed transcripts by mapping the RNA-Seq reads on *G. pallidipes* gene models. The transcripts were annotated and examined for enrichment of gene categories using, heat maps, BLAST2GO and R spider network software. Results from this study indicated that most of the differentially expressed transcripts at 24 hpc were associated with lipid remodeling/lipogenesis, proteolysis, urea cycle, carnitine trafficking, collagen metabolism, apoptosis, and cell growth/differentiation. Transcripts associated with 48 hpc included those linked to embryonic growth and development, muscle/motility, suppression of tumor, serine endopeptidase and related proteosomal degradation of target protein, enhanced translation of mRNA and neuronal development. There was pronounced expression of immune responsive transcripts 48 relative to 24 hpc, indicative of gradual maturity of immune responses in the fly or institution of vector-parasite endemic stability in the guts to facilitate the establishment of infection. Overall there was a systematic suppression of immunity in the *G. pallidipes* midgut in the initial phase of *T. b. brucei* challenge, which potentially facilitated initial establishment of the infection. Gradual and sequential immunological responses subsequently emerge contingent with the durations of challenge.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMNS

AAT	Animal African Trypanosomiasis
AMOVA	Analysis of Molecular Variance
AMP	Adenosine Monophospahet
ATP	Adenosine Triphospahte
BLAST	Basic Local Alignment Search Tool
cDNA	Complementary Deoxyriboxynucleic Acid
DAPC	Discriminant Analysis of Principle Component
DNA	Deoxyriboxynucleic Acid
dNTPs	Deoxyribonucleotide Triphosphates
ds RNA	Double stranded Riboxynucleic Acid
DSCAM	Down Syndrome Cell Adhesion Molecule
dsRNA	Double Stranded Riboxynucleic Acid
EP	glu-pro
FDR	False Detection Rate
GAPDH	Glyceraldehyde 3- Phosphate Dehydrogenase
gDNA	Genomic Deoxyriboxynucleic Acid
GO	Gene Ontology
GoK	Government of Kenya
GPEET	gly-pro-glu-glu-thr
GPI	Glycophosphatidylinositol
GPS	Global Positioning System
HAT	Human African Trypanosomiasis
HDPs	Host Defense peptides
Hpc	Hours Post Challenge
IACUC	Institutional Animal Care and Use Committee
IAEA	International Atomic Energy Agency
IGGI	International Glossina Genome Initiative
IMD	Immune Deficiency Pathway
ITS	Internal Transcribed Spacer

KALRO	Kenya Agricultural and Livestock Organisation
KARI	Kenya Agricultural Research Institute
KETRI	Kenya Trypanosomiasis Research Institute
LD	Linkage disequilibrium
LPS	Lipopolysaccharide
$MgCl_2$	Magnesium Chloride
mRNA	Messanger Riboxynucleic Acid
NCBI	National Centre for Biotechnology Infromation
PCF	Procyclic Forms
PCR	Polymerase Chain Reaction
PKAC	cAMP Dependent Protein Kinase
PSG	Phosphate Saline Glucose
qPCR	Quantitative Polymerase Chain Reaction
RFLP	Restriction Fragment Length Polymorphism
RLO	Rickettsia Like Organisms
RNA	Riboxynucleic Acid
RNAi	Riboxynucleic Acid Interference
ROS	Reactive Oxygen Species
RPKM	Reads per Kilobase of Transcript per Million Mapped
RT-PCR	Reverse Transriptase Polymerase Chain Reaction
SIT	Sterile Insect Technique
SNPs	Single Nucleotide Polymorphisms
SPSS	Statistical Package for Social Sciences
SRA	Serum Resistance Associated
Tbb	*Trypanosoma brucei brucei*
TBE	Tris Borate Ethylenediaminetetraacetic Acid
Tbg	*Trybanosoma brucei gambiense*
TBI	Transmission Blocking Vaccine
TRC	Trypanosomiasis Research Centre
WHO	World Health Organization
YSPH	Yale School of Public Health

LIST OF APPENDICES

CHAPTER ONE
GENERAL INTRODUCTION

1.1 Background information

African trypanosomias refers to a group of related diseases that cause Human African Trypanosomiasis (HAT) (sleeping sickness) in humans and African Animal Trypanosomiasis (AAT) (nagana) in cattle (Vickerman, 1985). Wild animals are reservoirs of trypanosomes, the causative agents of the disease (Mehlitz *et al.,* 1982; Njiokou *et al.,* 2006; Cordon-Obras *et al.,* 2009). The trypanosomes causing HAT and AAT are extracellular protozoan parasites in *Trypanosoma brucei* species complex. The HAT occurred on the African continent in the early 20th century and by early 1960's the disease was almost eradicated (Steverding, 2008). However, HAT remerged as a result of relaxed surveillance and control measures to about 450,000 cases in 1997 (Barrett, 2006). The HAT cases ranged from 50,000 to 70,000 until 2009 when new cases dropped to 9878 (Simarro *et al.,* 2010) and 7931 in 2010, representing a 28% decrease of a year (WHO, 2011). The cases had dropped to 6743 by 2011 (Simarro *et al.,* 2012) due to WHO and non-governmental intervention in disease control (Aksoy, 2011). Vector control and treatment of infections are the main disease control strategies that have been employed successfully in short term, although a long term intervention has not been found.

Trypanosoma brucei brucei has a complex life cycle that involves multiple differentiation steps in mammalian and invertebrate host (tsetse fly of *Glossina* species). All tsetse species are susceptible to some degree at least, to trypanosome infections. The susceptibility of the flies to trypanosomes under laboratory-controlled conditions is low, and significantly decreases with increased blood-meal intakes by the fly (Distelmans *et al.*, 1982) due to enhancement of the immune system of the fly (Hao *et al.*, 2001; Gibson and Bailey, 2003). There are many barriers that the parasite must overcome in order to survive and develop a mature infection. The initial immune response is able to clear parasites in over 95% of challenged flies (Aksoy *et al.*, 2003). In mammalian hosts, the parasite surface is covered with a variant surface glycoprotein (VSG) molecule, which shields the underlying membrane proteins from the host immune responses. The parasite evades the adaptive immune response of the host by changing these surface exposed molecules in a random order (Aksoy *et al.,* 2003). Over the last three decades, population genetic research has provided important insights into the biology of *T. brucei* and the epidemiology of sleeping sickness (Hide and Tait, 2009), but the fine genetic processes/factors underlying disease

dynamics and especially those distinguishing phenotypes in human infective parasite forms are poorly understood. Further understanding on comparative *T. brucei* parasite genomes and factors responsible parasite for clearance in the fly, can be determined using more advanced illumina sequencing technology hence opening novel long term and short term disease control strategies.

Development of *T. brucei* in tsetse involves initial establishment in the midgut and its successive maturation into infective metacyclic forms in the salivary glands of the fly (Aksoy *et al.*, 2003). In the midgut, most of the ingested parasites are cleared probably through the actions of antimicrobial peptides, lectins and reactive oxygen species produced by the fly (Aksoy *et al*., 2003). In only a small proportion of challenged flies, the parasites remain in the midgut and establish an infection this froms a bottleneck that reduces successful transmission of *T. brucei* in the fly. Many of the thirty-one species of *Glossina* have been noted to transmit trypanosomes but significant differences have been noted between these species in terms of their vector competence (Harley and Wilson, 1968; Maudlin *et al.*, 1986; Moloo and Kutuza, 1988a; Peacock *et al.*, 2012). To date, much of the molecular studies that focus on tsetse's vector competence have been performed on the species *Glossina morsitans morsitans* given the availability of molecular resources and therefore this information can be used for genetic study in other *Glossina* species. The expansion of genetic information to understand barriers to parasite transmission in the *G. pallidipes,* a major trypanosome vector in Kenya, will lead to more insight on development of transmission blocking strategy and hence vector control.

1.2 Statement of the problem

Underlying genetic and evolutionary factors supporting the observed differences in clinical presentations of infection by *Trypanosome brucei gambiense* and *Trypanosome brucei rhodesiense* in humans have been characterized. However, variations have been observed in clinical manifestations among and between isolates of the same subspecies of the trypanosomes which have in turn complicated treatment outcomes. Therefore to better understand these variations, genetic diversity, population structure and whole genome sequencing of the various trypanosome isolates was carried out in this study. Similarly, *G. pallidipes* tsetse flies are more refractory to infection by trypanosomes than other related *Glossina* species including the more extensively investigated *G. m. morsitans*. It was therefore necessary to identify molecular

factors/genes mediating this resistance phenomenon at transcriptome level that can be targeted for parasite transmission blocking in the fly.

1.3 Objectives

1.3.1 General Objective

To establish genetic polymorphisms conferring differences in phenotypes of medical importance among *T. brucei* field isolates from various parts of East Africa, and determine immune responsive genes influencing vector competence of *G. pallidipes* challenged with *T. brucei* parasite.

1.3.2 Specific objectives

1. To determine genetic differences between and among field isolates of *T. brucei* of medical importance in East Africa.
2. To determine immune responsive genes in *G. pallidipes* challenged with *T. b. brucei*.
3. To determine functional roles of immune responsive genes in *G. pallidipes* challenged with *T. b. brucei.*

1.4 Hypotheses

1. There are no genetic differences among *T. brucei* field isolates of medical importance in East Africa.
2. There are no differences in expressions of immune responsive genes in *G. pallidipes* challenged with *T. b. brucei*.
3. Immune responsive genes in *G. pallidipes* challenged have no functional roles when challenged by *T. b. brucei.*

1.5 Justification

Currently there are no vaccines against African Trypansomiasis (AAT and HAT), and chemotherapeutic drugs available are inefficient due to drug toxicity and complications related to their administration. In addition, available diagnostic tests are unreliable. Therefore advances in DNA and RNA sequencing technology using Next Generation Sequencing (NGS) platform allow the study of *T. brucei* genetic variations at genome level and use of microsatellites and SNPs to

study genetic diversity of the parasites with the aim of understanding their pathogenicity and genome evolution. The findings can help improve drug, vaccine and diagnostic tests design. Since trypanosome must complete part of its life cycle in the tsetse biological vector, disease transmission can be reduced through vector control. While current vector control methods are effective in reducing tsetse populations, the methods are unsustainable with some requiring trained personnel or community participation. In the current study microsatellite analysis and RNA-Seq approach was used to determine population structure and genetic diversity of *T. b. brucei* and *T. b. rhodesiense* and to identify early molecular responses that manipulate *G. pallidipes* challenged by *T. b. brucei*. The information gained will inform design of novel genetic based tools for management of the parasite and the vector in integrated trypanosomiasis control initiatives.

CHAPTER TWO
GENERAL LITERATURE REVIEW

2.1 History of African Trypanosomiasis

African sleeping sickness is endemic to parts of Africa. First historical note of the disease came from Ibn Khaldun, historian from Arabia, who described how in 1374 A.D Sultan Mari Djata of Mali died of an illness, which was compatible with this disease (Dobson, 2008). The disease was later observed and described in 1803 by Thomas Masterman Winterbottom, an English doctor working in Sierra Leone with David Livingstone attributing nagana to tsetse fly bite in 1813-1873 and reporting disease outbreak of *T. b. rhodesiense* in the valleys of Limpompo, Zambezi River, lakes Nyasa and Tanganyika from which cattle died after being bitten by tsetse flies in 1952 (Winkle, 2005). In 1881 Griffith Evans, a veterinary surgeon in India, found trypanosomes in the blood of sick camels and horses. These animals died of a serious disease, known locally as "surra".with the parasite given the name *"Trypanosoma evansi*". In 1895 Dr. David Bruce a Scottish pathologist and microbiologist discovered that trypanosomes were responsible for a cattle disease referred to as nagana and hence the name *Trypanosoma brucei brucei*. In 1901 Dr. Forde discovered a motile parasite in the blood of a sick captain of a river boat in Gambia; the parasite was identified as *T. gambiense* by Dutton in Liverpool. An Italian physian and pathologist Aldo Castellani found trypanosomes in cerebral spinal fluid of sleeping sickness patients and suggested that they cause the disease (Cox, 2004). Cyclical transmission of *T. brucei* in tsetse fly was first described by a German military surgeon Friedrich Karl Kleine in 1909, this prompted Bruce to change his initial opinion of mechanical transmission of trypanosomes and later describe full developmental cycle of the trypanosome in the insect vector (Cox, 2004). The *T. congolense* and *Trypanosoma vivax,* animal pathogenic trypanosome species, were discovered in 1904 and 1905 by Belgian physician Alphonse Broden and the German naval doctor Hans, respectively. *T. b. rhodesiense* the second human pathogenic trypanosome was described by John William Watson Stephens and Harold Benjamin Fantham parasitologists in 1910.

In Africa there were three severe epidemics in the 20th century the first one occurred in Uganda and Congo between 1896-1906 (WHO, 2006), in which an estimated 300,000 and 500,000 people died in the Congo basin and Busoga focus of Kenya and Uganda, respectively

(Hide 1999; de Raadt, 2005). The devastating effect of the epidemic prompted colonial administration to send out scientists to investigate the disease and find cure (Winkle 2005; de Raadt 2005). They started drug development for chemotherapy of trypanosomiasis which helped to fight the second epidemic which occurred in 1920 to 1940's in a number of African countries. Other control measures that were applied include vector control (use of traps, spraying using DDT and bush clearing), host reservoir control and game destruction some of which are still in use upto today (de Raadt, 2005). Combined application of the above control measures led to a dramatic reduction of sleeping sickness incidences in the early 1960s. After a decade of low endemicity, the control of trypanosomiasis was no longer a priority and control programmes were stopped (WHO/CDS/CSR/ISR/2000) and use of DDT in vector control was stopped due to environmental concerns in the 1970s. The outcome of all this was an increase in the number of reported cases of sleeping sickness in mid 1970s and the beginning the third sleeping sickness epidemic in the 20th century affecting Angola, Congo, Southern Sudan and the West Nile district of Uganda (WHO, 2006; de Raadt, 2005). The WHO is currently reporting very few new disease cases due to renewed control initiatives (Simarro *et al.,* 2008).

2.2 Trypanosomes as disease agents

Trypanosomes are protozoan parasites that cause trypanosomiasis. The species *T. brucei* consist of three sub-species: *T. b. rhodesiense* and *T. b. gambiense* which cause disease in humans, and *T. b. brucei* that causes infection in domestic and wild animals. Also *T. b. rhodesiense* can infect animals but not cause disease and rather act as a zoonosis for Human African Trypanosomiasis (HAT). All the three *T. brucei* sub-species are transmitted exclusively by tsetse flies of *Glossina* species. *T. evansi* is regarded as a mutant form of *T. b. brucei* with defects in its mitochondrial DNA (Jackson *et al*., 2012). It has extended its range beyond the tsetse belt of Africa through transmission by other biting flies and is an important livestock pathogen causing surra across subtropical regions of the world. Other species of Trypanosomes that cause Animal African Trypanosomiasis AAT include *T. congolense* and *T. vivax*. The severity of the diseases depends on pathogenicity of the parasite strain and the genetics of the mammalian host (Courtin *et al.,* 2008). The *T. b. rhodesiense* is resistant to lysis by human serum since it contains a serum resistance associated gene (SRA) (Gibson *et al.,* 2005), which enables *T. b. rhodesiense* to evade lysis by Apolipoprotein L-1 (ApoL-1) in human serum (Van

Xong *et al.,* 1998). The SRA is lacking in *T. b. brucei* (de Greef *et al.,* 1989) and *T. b. gambiense.* The mechanism that allows *T. b. gambiense* to survive in human serum is unknown.

The *T. b. gambiense* infection, present throughout western and central Africa, is responsible for 90% of cases of HAT (Simarro *et al.,* 2008). The infection causes chronic form of disease with progression taking approximately two years and clinical signs appear when the patient has already progressed to advanced stages of the disease (Simarro *et al.,* 2008). The *T. b. rhodesiense* causes acute form of HAT with progressing to death occurring in a matter of months. Clinical manifestations of these infections vary within different foci (Mcleon *et al.,* 2004). Infections with *T. b. rhodesinse* are found in eastern Africa. Historically the two diseases have occurred in geographically distinct ranges but currently in Uganda it is feared that the parasites will ultimately merge raising concerns about the implication for disease transmission. Previous studies have shown that despite the considerable variation in life history traits and clinical disease caused by the African *T. brucei* subspecies, genomic variation is extremely limited. This lack of genomic variation suggests that the functional differences are as a result of features shared between subspecies, and vary in either structure or expression (Jackson *et al.,* 2010). Therefore there is a need to determine the level and patterns of genomic variation within the *T. b. rhodesiense* subspecies.

2.3 Disease control strategies

Control of African trypanosomiasis relies on diagnosis, therapeutics, vaccine development, control of the tsetse fly vector, and reducing the disease in reservoirs. Accurate diagnosis is important in disease control and the available diagnostic tests include microscopy which is unreliable especially when parasitaemia is low (Magez and Radwanska, 2009). In a case where the parasite has been visualized microscopically, determination of infection stage, critical in providing appropriate treatment, is difficult since there is limited clinical distinction (Kennedy, 2008). Lumber puncture, the only diagnostic method that can be used in disease staging, is a painful procedure requiring specialized tools and trained personnel.

Drugs developed for trypanosomiasis control are few and toxic. Suramin was the first commercially available drug developed in 1916 while Pentamidine was introduced in 1936 (Cox, 2004). Pentamidine is used in early stage *T. b. gambiense* infection with Suramin as an alternative, while Suramin is administered in *T. b. rhodesiense* infection without any alternative

(Kennedy, 2008). Melarsoprol is used against late stage infections of both *T. b. rhodesiense and T. b. gambiense* (Cox 2004, Kennedy, 2008). Eflornithine (DFMO), registered in 1981, is used singly or in combination with nifurtimox as an alternative to melarsoprol resistant *T. b. gambiense* infection (Kennedy, 2006; Kennedy, 2008). Drug resistance, which is of great concern, has been reported in AAT (Anene *et al.,* 2001) while in humans reports concerning drug resistance is contradictory (Cox, 2004; Delespaux and de Koning, 2007).

Current control methods employed against tsetse such as trapping while effective are not sustainable especially when trapping activities are organized by local communities. When the fly densities drop, there is little adherence to the trapping schedule permitting reinvasion of flies (Baumgartner *et al.,* 2008). Pesticides have also been used in vector control but are expensive. Sterile insect technique (SIT) (release of sterile males) has been used to eradicate tsetse from Zanzibar Island (Vreysen *et al.,* 2000) and island of Pinciple (De Raadt, 2005). The challenge of SIT outside islands is the reinvasion of flies to previously cleared areas making eradication difficult.

2.4 Trypanosome life cycle

Trypanosomes undergo a complex life cycle within a vertebrate host and the tsetse fly. Within the vertebrate bloodstream they are long slender (which replicates by asexual division) and short stumpy (non-replicating) forms of the parasite. Tsetse is infected with bloodstream trypomastigotes when taking a blood meal on an infected mammalian host. In the tsetse fly, trypanosome undergoes a two stage differentiation: establishment in the midgut and maturation in mouthparts/ salivary glands. In the midgut, the mammalian short-stumpy forms of parasites rapidly differentiate to procyclic forms and begin to replicate (Roditi, *et al.,* 1989). Once established in the midgut, trypanosomes migrate forwards, cross the peritrophic matrix (Ellis and Evans, 1977) and enter the proventiculus where they proliferate and re-enter the midgut and migrate to the foregut and then enter the salivary glands (Van Den Abbeele *et al.,* 1999).

In the mammalian host, the bloodstream form (BSF) parasites are covered with an immunogenic surface coat composed of 10^7 identicle variant surface glycoprotein (VSG) molecules which protects the essential membrane proteins from host immune responses. The VSG coat displays antigenic variation. The trypanosome genome data suggests that there are many potential VSG proteins (Vickerman, 1969; Barry and McCulloch, 2001; Barry *et al.,* 2005). The presence of a large number of VSG genes allows expansion of antigenically distinct

trypanosome populations within the host. After activation of host immune responses (in reaction to high parasitaemia), the majority of the parasite population is killed but a small population that express an antigenically different VSG coat proceed to expand in numbers. The continuous cycles of trypanosome replication and destruction result in waves of fluctuating parasitaemia. The differentiation of the long slender bloodstream forms (BSF) into the non-dividing stumpy BSF occurs in high density populations of long slender BSFs (Vassella *et al.,* 1997; Seed, 2003). Short stumpy BSF are pre-adapted for survival within the insect midgut due to the presence of a functional mitochondrion.

Once ingested by a tsetse fly, the BSF differentiate by replacing the trypanosome surface VSGs by procyclins (Acosta-Serrano *et al.,* 2001; Vassella *et al.,* 2001; Gibson and Bailey, 2003). The VSG transcripts are undetectable 6h post blood meal while transcript for procyclin are upregulated 2 h post blood meal (Van Deursen *et al.,* 2001) with the protein detectable by 6hrs (Kabani *et al.,* 2009). Parasite replication can be followed in the midgut for approximately three days after acquasition, after which time majority of flies eliminate parasites excepts for the small proportion (~10%) where parasites multiply in number (Gibson and Bailey, 2003). When analyzed at eight days post acquastion, the flies either clear all the ingested parasites or have an established midgut infection. Microscopic investigations suggest that trypanosomes in an established infection migrate to the ectoperitrophic space by three to five days post infection (Gibson and Bailey, 2003) and this is assumed to occur by direct penetration through the PM (Ellis and Evans 1977; Gibson and Bailey, 2003). Typically the midgut population in an established infection reaches approximately 5 x 10^5 trypanosomes (Van den Abbeele *et al.,* 1999; Gibson and Bailey, 2003). From six to eight days post infection, large numbers of trypanosomes congregate within the proventriculus (Van den Abbeele *et al.,* 1999; Gibson and Bailey, 2003; Sharma *et al.,* 2008) where they begin to differentiate into epimastigotes and eventually colonize the proboscis or salivary glands, depending on the parasite species (Abbeele *et al.,* 1999). Differentiation (metacyclogenesis) involves appearance of a VSG coat in nascent and mature metacyclics and no dividing forms of these stages are observed (Tetley and Vickerman, 1984). Metacyclic forms are infective to mammals (maturation) and can be transmitted to the next host during blood feeding by the fly (Vickerman *et al.,* 1988) where they multiply for a few days at the site of fly bite before invading the blood stream. Inside the mammalian host, the

trypanosomes transform into BSF, and are carried to other sites throughout the body and continue to replicate by binary fission.

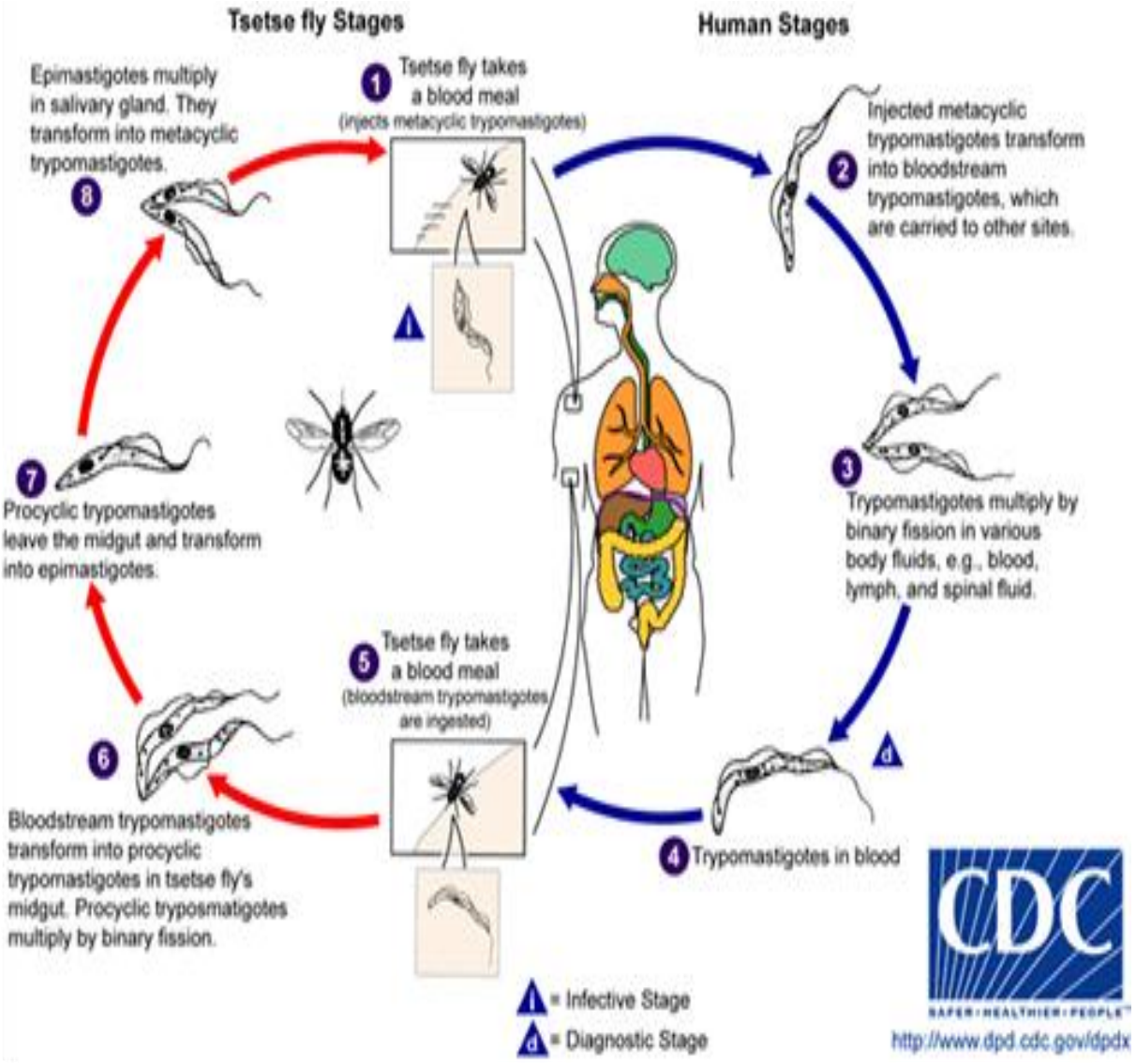


Figure 1: Life cycle of *T. b. rhodesiense* parasite of trypanosomosis; (Image source: Centres for Disease Control and Prevention. http://www.cdc.gov/parasites/sleepingsickness/biology.html 1989).

2.5 Trypanosoma. *brucei. brucei* and Trypanosoma *brucei. rhodesiense* population genetics

Disease causing organisms that easily adjust to environmental changes are a challenge to disease control. Changes in these organisms can lead to development of harmful traits that can rapidly spread through a population. The population structure and recombination within these organisms determines the mode and dynamics by which the trait spreads through a population. Therefore understanding the population structure and genetic diversity of an organism is important in risk assessment and designing of disease control strategies.

Earlier isoenzyme studies on *T. brucei* isolates in East Africa revealed a randomly mating population structure (Tait, 1980). The lack of agreement with Hardy-Weinberg principle, high levels of linkage disequilibrium and presence of high frequencies of similar genotypes suggested a clonal population structure with rare genetic exchange (Mathieu-Daude, 1995) or epidemic population structure with frequent mating and clonal expansion of a few genotypes.

Genetic studies have shown that *T. b. rhodesiense* is heterogeneous with its isolates varying both within and between foci (Gibson, 2001). Molecular characterization studies carried out on human isolates Collected from Busoga, Uganda, the Lambwe Valley, Kenya, north-west Tanzania and the Luangwa Valley in Zambia revealed many trypanosome genotypes circulating in each focus (Gibson *et al.,* 1980; Gibson and Gashumba, 1983; Gibson and Wellde, 1985; Mihok *et al.,* 1990; Hide *et al.,* 1991; Enyaru *et al*., 1993a; Komba *et al.,* 1997). Genetic exchange between *T. b. rhodesiense and T. b .brucei* could be the reason behind the observed heterogeneity in *T. b. rhodesiense but* it is unlikely that this variability could be as a result of the different hosts involved in the zoonotic cycle (Gibson, 1990; Mihok *et al.,* 1990; Cibulskis, 1992; Hide *et al.,* 1994). It has also been shown that *T. b. rhodesience* is clonal since one genotype has been collected from one area over a period of time and an example is a zymodyme from Busoga strain in Uganda that has been collected for over 40 years from Busoga and other neighbouring endemic areas in Uganda and Kenya (Gibson *et al.,* 1980; Gibson and Gashumba, 1983; Enyaru *et al.,* 1993a). Homogeneous isolates were also observed in North West Tanzania from 1959 to1994 (Gashumba *et al.,* 1994; Komba *et al.,* 1997). Duffy *et al.* (2013) showed that Ugandan *T. b. rhodesiense* isolates are clonal with few highly related genotypes and considerable linkage disequilibrium between pairs of loci but these populations were not stable and hence not conforming to strict clonality. They also demonstrated that Malawi *T. b. rhodesiense* is diverse and undergoes frequent genetic exchange and this shows that genetic diversity of the parasite is

complex. McLeod *et al.* (2000; 2001a) showed that *T. b. rhodesiense* is genetically isolated from *T. b. brucei* and can be identified by its multilocus genotype. Additionally *T. b. brucei* has an epidemic population structure. Studies have also shown that *T. b. brucei* is genetically isolated by geographic location with limited or no gene flow between them and they are also isolated by host specificity (McLeod *et al.*, 2001b). Comparison of *T. b. rhodesiense* from Uganda and Zambia foci showed different genotypes in each focus (McLeod *et al.*, 2000; McLeod *et al.*, 2001a). Studies from Tanzania also show a distinct focus (Komba *et al.*, 1997).

2.6 Tsetse trypanosome interactions

Tsetse flies, vectors of African trypanosomes are strictly hematophagous (Gooding and Krafsur, 2005), and they are widely distributed in Africa. In Kenya by 1996, it was estimated that 34% (202,774 km^2) of total land surface was infested with tsetse flies (KETRI, 1996; Figure 2) and eight species of tsetse have been described. The *Morsitans* subgenus is the most widely distributed which includes; *Glossina morsitans morsitans, Glossina pallidipes, Glossina swynnertoni and Glossina austeni*. The Fusca subgenus comprises of *Glossina longipennis, Glossina fuscipleuris and Glossina brevipalpis* while the *Palpalis* subgenus consists of *Glossina fuscipes fuscipes*, (Bourn *et al.*, 2001). The flies are distributed across seven tsetse belts represented as zones 1-7 in figure 2 (KETRI, 1996). *G. pallidipes* is a widely distributed and is a major vector of animal trypanosomes; *T. b. brucei, Trypanosoma vivax and Trypanosoma congolense*; it has also been implicated in the transmission of *T. b. rhodesiense* (*T. b. rhodesiense)* from its traditional focus of south eastern Uganda to western Kenya in 1950s (Wellde *et al.*, 1989). It is found in five fly belts separated by climate and geographical features.

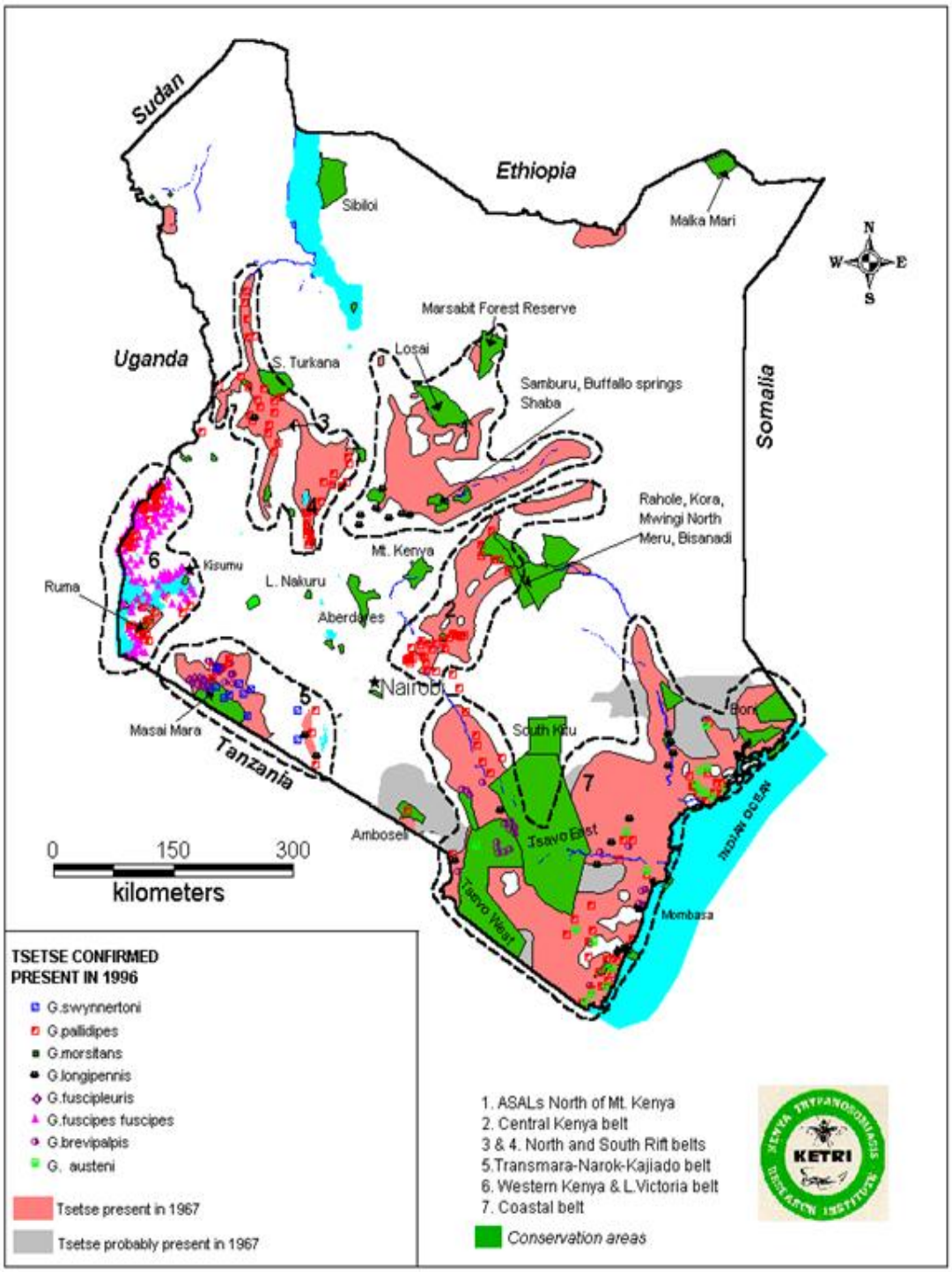


Figure 2. Tsetse habitats (Zones 1–7) and *Glossina species* in Kenya (Source: Kenya Trypanosomiasis Research Institute, Kenya 1996).

2.7 Tsetse vector competence

All tsetse species are capable of transmitting trypanosomes but differ in their vector competence (Aksoy *et al*., 2005). Flies in the *Palpali*s subgroup are most refractory to trypanosome infection (Nayduch and Aksoy 2007) as compared to other fly species, while those in the *Morsitans* subgroup are more susceptible (Moloo *et al.,* 1994). Within the *Morsitans* subgroup, there are two closely related species, *Glossina morsitans morsitans* and *G. pallidipes*, which show differential susceptibility with *G. pallidipes* being more refractory to trypanosome infection (Peacock *et al.,* 2012). Adult flies are also highly refractory to trypanosome infection and far less than 10% of the flies that are given an infected bloodmeal establish midgut infections and far less than 1% of trapped wild flies will have salivary gland infection (Aksoy *et al.,* 2003). Laboratory transmission of *T. vivax, T. congolense and T. b. brucei* varies between 1-20% depending on the fly species and parasite strain (Moloo and Kutunza, 1988; Moloo *et al.,* 1992; Moloo *et al.,* 1994) and infection rates reported from the salivary glands and mouth parts of field flies also varies from 1-5% (Woolhouse, 1994; Msangi *et al.,* 1998; Lehane *et al.,* 2000). This refractorieness is maintained throughout the life time of the fly, although it can be reversed by long periods of starvation and sub optimal environmental conditions (Haines, 2013). For transmission to occur, the parasite must first establish in the fly midgut and then mature in salivary gland or mouth parts depending on the species (Van Den Abbeele *et al.,* 1999 and Vickerman *et al.,* 1988). Less than half of the established midgut infections mature into salivary gland infections in *G. m. morsitans* due to immune challenge from the fly (Hao *et al.,* 2001; Gibson and Bailey, 2003). While the midgut infection rates are lower in *G.pallidipes*, all midgut infections give rise to salivary gland infections (Peacock *et al.,* 2012).

There are barriers to trypanosome establishment and maturation processes in the fly which includes; lectins in the gut at the time of infection, species, sex, age, and symbiotic associations in tsetse flies, these barriers determine success or failure of parasite infections (Welburn, 1999). Immune genes such as defensin, attacin and diptericin, lectins and reactive oxygen intermediates also influence trypanosome establishment in tsetse. Additionally, systemic stimulation of the immune system of the fly before receiving an infectious blood meal severely blocks establishment of midgut infections (Hao *et al.,* 2001; Hao *et al.,* 2003; Hu and Aksoy, 2006; Abubakar *et al.,* 2006). The initial responses in the gut provide the first protective barrier to

invading microorganism and thus are important targets to parasite transmission blocking methods.

2.8 Tsetse immune system

Insects possess a complex immune system comprising of physical barriers such as the cuticle and the peritrophic matrix (PM), production of host defense peptides (antimicrobial peptides) reactive oxygen species (ROS), melanisation and coagulation, cellular responses such as encapsulation and phagocytosis and humoral responses such as production of host defense peptides (Lemaitre and Hoffmann, 2007). Toll and immune deficiency pathways (Imd) are the two main humoral immune pathways in insects including tsetse fly (Hoffmann and Reichhart 2002). The Imd pathway acts both in systemic and epithelial immunity. In tsetse trypanosome interactions, trypanosomes are exposed only to epithelial surfaces throughout the parasite life cycle though they have also been reported in the hemolymph which is not a major route of invasion (Mshelbwala, 1972). Those that cross the midgut epithelium are rapidly killed by an unknown systemic immune response (Croft *et al.,* 1982).

2.8.1 Effector molecules and tsetse genomic resourses available

Invertebrates lack memory and adaptive immune responses and therefore rely on recognition of conserved molecules specific to a particular pathogen. These molecules include proteins and peptides that have antimicrobial activity which are synthesized and induce systemic immune response that is initiated in the fat body. In *Glossina morsitan morsitan* antimicrobial peptides (AMP) cecrospin, attacin (an effector in Imd pathway), diptericin and defensin are produced in the midgut (Hao *et al.,* 2001; Hao *et al.,* 2003 and Hu and Aksoy, 2006). Attacin occurs in three different forms where attacin A (attA) and attacin B (attB) have amino acid sequence almost identical and attacin D (attD), which is 69% identicle to attA/attB. These genes are differentially regulated (Wang *et al.,* 2008). The AMPs are differentially regulated in the haemolymp, midgut fatbody and the proventiculus (Hao *et al.,* 2001; Hao *et al.,* 2003; Hu and Aksoy, 2006). Attacin (attA/B and attD) and defensin transcript levels in tsetse fat body increase three days post trypanosome infection, and at six day post infection Attacin, defensing and cecrospin are detected in the haemolymph (Hao *et al.,* 2001; Hao *et al.,* 2003; Wang *et al.,* 2008). This indicates that the fly is able to detect the procyclic form of the parasite and not the

blood stream form of the parasite (Baulanger *et al.,* 2002). In self-cured flies there is a decrease in transcript expression levels of AMPs but in flies with established infections the levels remain high in the fat body and the proventiculus (Hao *et al.,* 2001; Hao *et al.,* 2003). Haines *et al*. (2003) observed that procyclics exhibit a high resistance to trypanocidal activity of AMP than the BSF trypanosomes. Peptides synthesized in the fat body and circulating in the haemolymph may fail to reach the parasite. Studies have shown that *Glossina* host defense peptides are able to kill trypanosomes (Hu and Aksoy, 2005; Hu and Aksoy, 2006; Nayduch and Aksoy, 2007). Recombinant attacin is able to inhibit growth of BSF and PCF of trypanosomes *in vitro* (Hu and Aksoy, 2005). *In vivo* gene knockdown of attacin or its Imd pathway transcriptional regulator Relish prior to administration of trypanosome infected blood meal lead to a significant increase in midgut and salivary gland infection rates in *G. m. morsitans* (Hu and Aksoy, 2006). Constitutive systemic (fat body) expression of attacin has been observed in *G. pallidipes* and *G. palpalis* species that are refractory to trypanosome infection and not in the susceptible *G. m. morsitans.* Refractory tsetse species had a higher attacin transcript expression in the fat body and proventiculus and upregulated attacin tissues as compared to susceptible flies both in blood fed and teneral states (Nayduch and Aksoy, 2007).

Tsetse EP protein contains a glutamic acid proline repeats (EP) that are found also on the C-terminal of the *T. brucei* EP procyclins (Chandra *et al.,* 2004). The protein is expressed in the midgut of *Glossina* and it is upregulated in response to stimulation by Gram negative bacteria (Haines *et al.,* 2005). This protein may also be involved in tsetse immune modulation since gene knock down by RNAi increases the susceptibility of the fly to trypanososme infection (Haines *et al.,* 2010) but the interaction between tsetse EP protein and trypanosome is still under investigation.

The genome of *G. m. morsitans* has also been published (International *Glossina* genome initiative, 2014). Analysis of the 366 megabase genome has led to discovery of many genes that could be targeted for fly control. These available resources can now be used to investigate tsetse trypanosome interaction at molecular level in related species so as to develop disease control strategies. Complete and annotated *G. pallidipes* genome and associated transcripts are also available in Vectorbase (Giraldo-Calderón. *et al.,* 2015) and the extensive literature on immune responses in *G. m. morsitans* presents a unique resource and opportunity that has facilitated identification of novel *G. pallidipes* immune associated molecule(s) and pathways responsible

for the differential resistance to trypanosome infection observed in the fly. Other available tsetse genomic resourses in vectorbase include; genome sequences and gene sets for *G. austeni, G. brevipalpis*, *G .fuscipes* and *G. palpalis*.

2.8.2 Peritrophic matrix

The insect midgut epithelium is protected physically by the PM a highly organized glucosaminoglycan rich layer supported with chitin (Lehane, 1997). Tsetse PM (type II PM) is constitutively expressed forming a protective sheath along the entire length of the midgut. The PM separates the midgut into two compartments, the endoperitrophic space which contains a blood meal and the ectoperitrophic space the region between the PM and the gut epithelium. The tsetse PM has a pore size of 9 nm and therefore permeable to molecules less than 150kDa (Miller and Lehane, 1990). Trypanosomes developing in the midgut move to the ectoperitrophic space 3-5 days post infection (Gibson and Bailey, 2003) and whether they penetrate the PM or circumnavigate via the hind gut are unknown and no chitinase gene has been found in the trypanosome genome.

Insect epithelium is also protected by reactive oxygen species (ROS) which has been implicated in protecting the gut from trypanosome infection (Lehane and Msangi 1991; Hao *et al.,* 2003; Hao *et al.,* 2005). This is supported by experimental evidence that involved feeding a variety of antioxidants to tsetse which significantly increased midgut tryapanosome establishment and maturations (Munks *et al.,* 2005; Macleod *et al.,* 2007).

2.8.3 Fly sex, age and starvation

Fly sex influences susceptibility to infection, most laboratories but not all have suggested that male flies are more susceptible to infection than female flies (Peacock *et al.,* 2012). Female *G. palpalis* are more resistant to developing mature *T. congolense* infection (Distelmans *et al.,* 1982) than females. In *G. m. morsitans* a mated female is twice refractory to *T .b. brucei* infection (Macleod *et al.,* 2007). Both male and female flies are resistant to trypanosome infection and this resistance increases with age. Natural refractoriness of the fly to trypanosome infection has been shown both in laboratory and field situations and it is called teneral phenomenon (Distelmans *et al.,* 1982; Welburn and Maudlin, 1992; Leak, 1999). Susceptibility of tsetse to trypanosome midgut infection decreases during the first 48 hrs after emergence. This

change in susceptibility could be due to immaturity of the PM, variation in the concentrations of antioxidant and lectins, PH gradient differences and variation in symbiont numbers. Research needs to be done to determine which of these factors are responsible for the teneral phenomenon.

Starvation influences the host choice as fly's hunger intensifies (Bouyer *et al.,* 2007) and environmental conditions such as temperature and relative humidity also influences fly starvation with *Glossina* species becoming hungrier during the dry season. Starvation plays a role in tsetse trypanosome interaction as the nutritional status of a fly at the time of an infective blood meal determines the susceptibility of the fly to trypanosome infection (Gingrich *et al.,* 1982; Mwangelwa *et al.,* 1987; Gooding, 1988; Kubi *et al.,* 2006). In the field a higher proportion of adult tsetse flies than predicted develop mature infections (Woolhouse *et al.,* 1993; Msangi *et al.,* 1998; Lehane *et al.,* 2000). Kubi *et al.* (2006) demonstrated that a starvation period of 3-4 days for teneral or 7 days for adult flies increases fly susceptibility to *T .b. brucei or T. congolense* infection resulting in a high number of mature infections in the salivary gland for *T. b. brucei* and midgut infections in *T. congolense*.

2.8.4 Stage specific parasite surface coat

Differentiation of *T. brucei* bloodstream (BSF) form to procyclic form involves release of VSG coat and its replacement by procyclins (Roditi and Pearson, 1990; Beecroft *et al.,* 1993), a family of (Glycophosphatidylinositol) GPI anchored proteins possessing internal C-terminal glu-pro (EP) or gly-pro-glu-glu-thr (GPEET) repeats (Roditi and Clayton, 1999). There are three isoforms of EP procyclins (EP1, EP2 and EP3) and the four procyclins are present in equal amounts within a few hours of differentiation from BSF to procyclins. The GPEET procyclic coat is expressed three days post infection and from day seven post infection GPEET procyclin coat disappears and expression switched to glycosylated isoforms EP1 and EP3. EP2 isoform is found abundantly in fly derived procyclic culture forms but its expression *in vivo* has not been confirmed (Urwyler *et al.,* 2005). The N-terminal in all procyclins is removed during proteolysis in the fly but the acidic amino acids are resistant to proteolysis (Acosta-Serrano *et al.,* 2001) hence help in parasite development and influence ligand associated parasite vector signaling (Roditi and Pearson, 1990; Ruepp *et al.,* 1997). Parasites expressing procyclins with truncated N-termini can still establish midgut infections to similar levels as wild type parasites under laboratory conditions, and form mature metacyclics within the salivary glands (Liniger *et al.,* 2004). Thus, procyclins are not crucial for migration of procyclic forms from the tsetse midgut to

the salivary glands (Vassella *et al.,* 2009). However, in co-infection experiments, procyclin null mutants were rapidly outgrown in the midgut by wild type parasites. This suggests that under field conditions; where mixed infections in flies are common surface procyclins probably play a significant role in trypanosome fitness (Lehane *et al.,* 2000; Vassella *et al.,* 2009). As trypanosome epimastigotes within the tsetse salivary glands typically lack a procyclin coat (Urwyler *et al.,* 2005), the function of the procyclin coat molecules probably occurs earlier in trypanosome development in the fly.

2.9 Symbiont tsetse trypanosome interaction

Tsetse flies being single diet insects, they harbour microbial symbionts that provide additional nutrients that are not found in their limited diet and which they are unable to synthesize. These microorganisms have been shown to interact with tsetse physiology and have been implicated in vector competence of the fly. Tsetse fly harbor three symbionts and these are: *Wiggleswothia, Wolbachia* and *Sodalis. Wiggleswothia glossinidus* is an obligate mutualist primary gram-negative endosymbiont (Aksoy, 1995b; Aksoy, 1995) that is found in all *Glossina* species. The bacteria reside within the cytoplasm of specialized epithelial cells, bacteriocytes which form an organ called bacteriome found in the anterior midgut of the fly. The bacterium is thought to produce metabolites to compensate for nutritional deficit and metabolism of B-complex vitamins that are necessary for tsetse survival (Nogge, 1981; Akman *et al.,* 2002). Removal of *Wiggleswothia* reduces fly longevity, rate of bloodmeal digestion and fecundity by making female sterile (Nogge, 1976; Welburn *et al.,* 2008). Feeding of aposymbiotic flies on a bloodmeal supplemented with B-complex vitamins partially restore fecundity (Nogge, 1981). Wiggleswothia is vertically transmitted from the female fly to her offsprings through milk gland secretions (Denlinger and Ma, 1975). Studies have shown that elimination of *Wiggleswothia* by ampicillin treatment increases susceptibility of non-teneral flies to trypanosome infection (Pais *et al.,* 2008). This indicates that symbiont clearance may lead to an increase in basal immunity and therefore affecting the host immune responses to trypanosome infection. Wiggleswothia cleared flies also have a compromised ability to digest their blood meal (Kubi *et al.,* 2006).

Sodalis glosinidus is a gram negative non motile bacterium that was first described as a rickettsia like organisms (RLO), separated from the cytoplasm of midgut cells (Reinhardt *et al.,* 1972). The bacterium resides in the midgut, haemolymph, muscle, fatbody, milk glands and in

certain species the salivary glands (Cheng and Aksoy, 1999; Attardo *et al.,* 2008). *Sodalis* can exist both intracellularly and as free living forms in the gut lumen. The density of *Sodalis* increases in the fly tissues with increase in fly age and differs with species and fly sex (Cheng *et al.,* 2000; Nogge and Ritz, 1982). It has been suggested that *Sodalis* may influence tsetse vector competence. Laboratory and wild tsetse flies that have high densities of RLO have been shown to have increased trypanosome infections (Maudlin and Ellis, 1985; Maudlin *et al.,* 1990). It has been observed that susceptible lines of *G. m. morsitans* have high symbiont densities than the refractory lines and that susceptibility is a maternally inherited factor (Maulin 1982; Moloo and Kutuza, 1988). There is a direct correlation between trypanosome and RLO densities in wild caught flies (Maudlin *et al* 1980). However there are other populations of wild flies that do not reflect such a direct relationship between bacterial densities and tsetse refractoriness or susceptibility (Moloo and Shaw, 1989; Geiger *et al.,* 2005b). *Glossina* species have a highly variable vectorial competence (Harley and Wilson, 1968; Moloo and Kutuza, 1988a; Reifenberg *et al.,* 1997; Kazadi *et al.,* 2000). Using this fact as a basis, Geiger *et al.* (2005) demonstrated no correlation between *Sodalis* prevalence and maturation of a *T. congolense* infection in two distinct species of tsetse. The genetic diversity of *Sodalis* was carried out by screening symbionts isolated from trypanosome infected and uninfected flies (Geiger *et al.,* 2007). The ability of a specific parasite species to establish in the insect midgut is statistically linked to the *Sodalis* genotype present. To address host-specificity among *Sodalis* Weiss *et al.* (2006) trans-infected two species of tsetse (previously cleared of native *Sodalis* with ampicillin) with reciprocal *Sodalis* strains. Equal symbiont densities were obtained in surrogate hosts without seriously deleterious effects on the fly.

Wolbachia pipientis is a gram negative intracellular alpha-proteobacteria secondary tsetse symbiont that is also maternally inherited and resides in the female reproductive organs (O'Neil *et al.,* 1993). Survey has shown that more than 70% of all insects are infected with *Wolbachia* (Werren and Windsor, 2000) and it has been shown to manipulate the host's biological processes which includes parthenogenesis, feminization, male-killing and cytoplasmic incompatibility (Siozios *et al.,* 2008). It has been reported that 100% of laboratory-reared tsetse colonies are infected with *Wolbachia,* while infections in wild populations vary significantly (Cheng *et al.,* 2000). *Wolbachia* colonizes several tissues in infected tsetse, depending on the species of *Glossina*. The role of *Wolbachia* plays in the tsetse trypanosome interaction is unknown. .

2.10 References

Abubakar L.U., Bulimo W.D., Mulaa F.J. and Osir E.O (2006). Molecular characterization of a tsetse fly midgut proteolytic lectin that mediates differentiation of African trypanosomes. *Insect Biochemisrty and Molecular Biology*. **36**:344-52.

Aksoy S., Gibson W.C. and Lehane M.J. (2003). Interactions between tsetse and trypanosomes with implications for the control of trypanosomiasis. *Advances in Parasitology* **53**:1–83.

Berriman M., Ghedin E., Hertz-Fowler C., Blandin G., Renauld H., Bartholomeu D.C., Lennard N.J., Caler E., Hamlin N.E., Haas B., Böhme U., Hannick L., Aslett M.A.,Shallom J., Marcello L., Hou L., Wickstead B., Alsmark U.C., Arrowsmith C., Atkin R.J., Barron A.J., Bringaud F., Brooks K., Carrington M., Cherevach I., Chillingworth T.J.,Churcher C., Clark L.N., Corton C.H., Cronin A., Davies R.M., Doggett J., Djikeng A., Feldblyum T., Field M.C., Fraser A., Goodhead I., Hance Z., Harper D., Harris B.R.,.Hauser H., Hostetler J., Ivens A., Jagels K., Johnson D., Johnson J., Jones K., Kerhornou A.X., Koo H., Larke N., Landfear S., Larkin C., Leech V., Line A., Lord A., Macleod A., Mooney P.J., Moule S., Martin D.M., Morgan G.W., Mungall K., Norbertczak H., Ormond D., Pai G., Peacock C.S., Peterson J., Quail M.A., Rabbinowitsch E,.Rajandream M.A., Reitter C., Salzberg S.L., Sanders M., Schobel S., Sharp S., Simmonds M., Simpson A.J., Tallon L., Turner C.M., Tait A., Tivey A.R., Van Aken S., Walker D., Wanless D., Wang S., White B., White O., Whitehead S., Woodward J.,W ortman J., Adams MD., Embley T.M., Gull K., Ullu E., Barry JD, Fairlamb A.H., Opperdoes F., Barrell B.G., Donelson J.E., Hall N., Fraser C.M., Melville S.E. and El-Sayed N.M. (2005). The genome of the African trypanosome *Trypanosoma brucei*.*Science* **309**: 416-422.

Bourn D., Reid R., Rogres D., Snow W. and Wind W. (2001). Environmental change and the autonomous control of tsetse and trypanosomosis in sub-Saharan Africa: case histories from Ethiopia, Gambia, Kenya, Nigeria and Zimbabwe. Oxford: Environmental Research Group Oxford Limited. ISBN: 1-898028-05-2

Cox F.E.G (2004). History of sleeping sickness (African trypanosomiasis). *Infectious Disease Clinics of North America* **18:**231-245.

Dobson M. J. (2008). "African Trypanosomiasis." Disease: The Extraordinary Stories Behind History's Deadliest Killers. *Book Sales Print* 94-101.

Ellis D.S. and Evans, D.A. (1977). Passage of *Trypanosoma brucei rhodesiense* through the peritrophic membrane of *Glossina morsitans morsitans*. Nature **267**:834–835

Ford J. (1971). The role of trypanosomes in African Ecology. A study of tsetse fly problem. London George Allen and Unwin

Geiger A., Ravel S., Mateille T., Janelle J., Patrel D., Cuny G. and Frutos R. (2007). Vector competence of *Glossina palpalis gambiensis* for *Trypanosoma brucei s.l.* and genetic diversity of the symbiont *Sodalis glossinidius*. *Molecular Biology and Evolution* **1**:102-109

Gibson W. and Bailey M. (2003). The development of Trypanosoma brucei within the tsetse fly midgut observed using green fluorescent trypanosomes. *Kinetoplastid Biology and Disease* **2**:1.

Gibson W.C. (2005). The SRA gene: the key to understanding the nature of *Trypanosoma brucei rhodesiense*. *Parasitology* **131**:143-150.

Haines L.R. (2013). Examining the tsetse teneral phenomenon and permissiveness to trypanosome infection. *Frontiers in Cellular and Infection Microbiology* **19:**3:84.

Hao Z., Kasumba I., Lehane M.J., Gibson W.C., Kwon J. and Aksoy S. (2001). Tsetse immune responses and trypanosome transmission: Implications for the development of tsetse-based strategies to reduce trypanosomiasis *Proceedings in National Academy of Sciences* **98**: 12648-12653.

Hao Z., Kasumba I. and Aksoy S. (2003). Proventriculus (cardia) plays a crucial role in immunity in tsetse fly (*Diptera :Glossinidiae*) . *Insect Biochemistry and Molecular Biology* **33**:1155-1164.

Hu C. and Aksoy S. (2006). Innate immune responses regulate trypanosome parasite infection of the tsetse fly *Glossina morsitans morsitans*. *Molecular Microbiology* **60**:1194-1204.

International *Glossina* Genome Initiative (2014). Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis *Science*. 344:380-386

Jackson A.P., Berry A., Aslett M., Allison H.C., Burton P., Vavrova-Anderson J., Brown R., Browne H., Corton N., Hauser H., Gamble J., Gilderthorp R., Marcello L., McQuillan J., Otto T.D., Quail M.A., Sanders M.J., van Tonder A., Ginger M.L., Field M.C., Barry J.D,.Hertz-Fowler C. and Berriman M., (2012). Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. *Proceedings in the National Academy of sciences* **109**:3416-3421.

Jackson A.P., Sanders M., Berry A., McQuillan J., Aslett M.A., Quail M.A., Chukualim B., Capewell P., MacLeod A., Melville S.E., Gibson W., Barry J.D., Berriman M. and Hertz-Fowler C., (2010). The Genome Sequence of *Trypanosoma brucei gambiense*, Causative Agent of Chronic Human African Trypanosomiasis. *Proceedings in the National Academy of sciences* **4**:e658

Kiss L. (2013). [David Livingstone (1813-1873)] - not just an Africa explorer, but also a contributor to the development of tropical medicine]. *Hungarian medical Journal* **154**:394-396.

Lehane M.J., Msangi A.R., Whitaker C.J. and Lehane S.M. (2000). Grouping of trypanosome species in mixed infections in *Glossina pallidipes*. *Parasitology* **120**: 583-592.

Moloo S.K. and Kutuza S.B. (1988). Comparative study on the susceptibility of different *Glossina* species to *Trypanosoma brucei brucei* infection. *Tropical Medical Parasitology* **39**: 211-213.

Moloo S.K. (1993). A comparison of susceptibility of two allopatric populations of *Glossina pallidipes* for stocks of *Trypanosoma congolense*. *Medical and Veterinary Entomology* **7:**369-372.

Moloo S.K., Zweygarth E and Sabwa C.L (1994). Comparative study on the susceptibility of different laboratory strains of *Glossina* species to *Trypanosoma simiae. Medical Veterinary Entomology* **8:**225-30

Moloo S. K. and Okumu I.O. (1995). A comparison of susceptibility to stocks of *Trypanosoma vivax* to *Glossina pallidipes* from allopatric populations in Kenya *Medicaland Veterinary Entomologv* **9**: 202-204.

Moloo S.K. and Gooding R.H. (1995). A comparison of *Glossina morsitans centralis* originating from Tanzania and Zambia, with respect to vectorial competence for pathogenic Trypanosoma species, genetic variation and inter-colony fertility. *Medical Veterinary Entomology* **4:** 365-371.

Moloo S.K., Zweygarth E. and Sabwa C.L. (1995). Comparative study on the susceptibility of different laboratory strains of *Glossina* species to *Trypanosoma simiae*. *Medical Veterinary Entomology* **8**: 225-230.

Msangi A.R., Whitaker C.J. and Lehane M.J. (1998). Factors influencing the prevalence of trypanosome infection of *Glossina pallidipes* on the Ruvu flood plain of Eastern Tanzania. *Acta Tropica* **70:** 143-155.

Nayduch D. and Aksoy S. (2007). Refractoriness in tsetse flies (*Diptera: Glossinidae*) may be a matter of timing. Journal of *Medical Entomology* **44**: 660-665.

Njiokou F., Laveissiere C., Simo G., Nkinin S. and Grebaut P. (2006). Wild fauna as a probable animal reservoir for *Trypanosoma brucei gambiense* in Cameroon. *Infection Genetics and Evolution* **6**: 147–153.

Peacock L., Ferris V., Bailey M. and Gibson W (2012). The influence of sex and fly species on the development of trypanosomes in tsetse flies. *PLoS Neglected Tropical Diseases.* **6:**e1515.

Roditi I, Schwarz H, Pearson TW, Beecroft RP, Liu MK, Richardson JP, Bühring HJ, Pleiss J, Bülow R, Williams RO, *et al.* (1989) Procyclin gene expression and the loss of the variant surface glycoprotein during differentiation of *Trypanosoma brucei*. *Journal of Cell Biology* 108:737-746.

Simarro P.P., Diarra A., Ruiz Postigo J.A., Franco J.R. and Jannin J.G. (2011). The Human African Trypanosomiasis Control and Surveillance Programme of the World Health Organization 2000–2009: The Way Forward. *PLoS Neglected Tropical Diseases* **5**: e1007.

Van Den Abbeele J., Claes Y., van Bockstaele D., Le Ray D. and Coosemans M. (1999). Trypanosoma brucei species development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology* **118** : 469-478

Vickerman K. (1985). Developmental cycles and biology of pathogenic trypanosomes. *British Medical Bulletin* **41**:105-114.

Vickerman K., Tetley L., Hendry K.A. and Turner C.M. (1988). Biology of African trypanosomes in the tsetse fly. *Biology of the Cell.* **64**:109-19.

Welburn S.C. and Maudlin I. (1999). Tsetse-trypanosome interactions: Rites of passage *Parasitology Today.* **15:** 399-403.

Wellde B.T., Chumo D.A., Reardon M.J., Waema D., Smith D.H., Gibson W.C., Wanyama L. and Siongok T.A. (1989). Epidemiology of Rhodesian sleeping sickness in the Lambwe Valley Kenya. *Annals of Tropical Medicine and Parasitology* **83** :43- 62.

Wijers D.J. (1974). The complex epidemiology of Rhodesian sleeping sickness in Kenya and Uganda. II. Observations in Samia (Kenya). *Tropical Geography and Medicine* **26**: 182-197.

Woolhouse M.E., Bealby K., McNamara J.J. and Silutongwe J. (1994). Trypanosome infections of the tsetse fly *Glossina pallidipes* in the Luangwa Valley, Zambia. *International Journal of Parasitology* **24:** 987-993.

World Health Organisation Report (2010). First WHO report on neglected tropical diseases. http://www.who.int/neglected_Diseses/2010report/en/

World Health Organization Report (2011). Neglected tropical diseases: New cases of Human African Trypanosomiasis continue to drop. http://www.who.int/neglected diseases/disease management/HAT cases drop/en/ex.Html

World Health Organization (2005). Control of Human African Trypanosomiasis: A strategy for the African region http://www.who.int/trypanosomiasis african/resources/afro tryps strategy.pdf.

World Health Organization (2005). The history of sleeping sickness: Fourth International ours on African Trypanosomoses Tunis, 11-28 October.

World Health Organization (2013). Trypanosomiasis, human African (sleeping sickness), factsheet 259 World Health Organization, Geneve Switzerland http://www.who.int/mediacentre/factsheets/fs259/en/.

World Health Organization Report (1998). Control and Surveillance of African trypanosomiasis. http://www.who.int/trs/WHO_TRS_881.pdf

Xong H.V., Vanhamme L., Chamekh M., Chimfwembe C.E., Van Den Abbeele J., Pays A., Van Meirvenne N., Hamers R., De Baetselier P. and Pays E. (1998). A VSG Expression Site–Associated Gene Confers Resistance to Human Serum in *Trypanosoma rhodesiense*. *Cell* **95**:839-846.

CHAPTER THREE

GENETIC DIFFERENCES BETWEEN *T. b. brucei* AND *T. b. rhodesiense* ISOLATES FROM WESTERN KENYA AND UGANDA

3.1 Abstract

Human African Trypanosomiasis (HAT) is a disease of economic importance in Sub Saharan Africa. Currently the disease is on the decline in affected countries, although in Uganda it is still has a major health impact. There are frequent irregular outbreaks of the disease in the traditionally endemic areas in south-eastern Uganda, and continued spread to new unaffected areas in central Uganda. This study evaluated the evolutionary dynamics behind the origin of new foci and the impact of host species on parasite genetic diversity in Uganda and western Kenya. *Trypanosoma brucei brucei* and *Trypanosoma brucei rhodesiense* isolates (n=269) collected from different regions in Uganda and western Kenya, were checked for the presence of the SRA gene that confers human infectivity to *T. b. rhodesiense* by PCR and genotyped at 17 microsatellite loci. Both Bayesian clustering methods and Discriminant Analysis of Principal Components partitioned the isolates into three distinct genetic clusters. Clusters 1 and 3 include isolates from central and southern Uganda, while cluster 2 contains mostly isolates from western Kenya. The analyses also showed evidence of genetic admixture among the three genetic clusters and long-range dispersal, suggesting recent and possibly on-going gene flow between them. Results from this study also indicate that the emergence of the new disease foci in central Uganda is as a result of the northward spread of *T. b. rhodesiense* form the traditionally endemic foci. The results likewise confirm cattle as reservoirs of *T. b. rhodesiense* and their role in shaping HAT epidemiology in the region.

3.2 Introduction

Recent reports show that HAT is on the decline on the continent of Africa (WHO 2010). However, the disease still remains a major health problem in Uganda, characterized by recurrent sporadic outbreaks in the traditional endemic areas and spread to new unaffected areas in central Uganda (Fèvre, 2001). Uganda is currently the only country in sub-Saharan Africa known to harbor all three subspecies of *T. brucei.* The locations of districts affected by HAT are shown in Figure 2 (Fèvre, 2001; Enyaru *et al.,* 1999; Welburn *et al.,* 2001; Picozzi *et al.,* 2001 and Lwala Hospital Medical Reports 2010-2012). During most of the 20th century, *T. b. rhodesiense* was

limited to south-east Uganda in the old foci of Busoga (BS) and Bugiri (BG), and in areas bordering Tanzania and Kenya, such as Busia (BU). However, since the 1970's, *T. b. rhodesiense* has been spreading towards the northern and the eastern parts of the country and the affected areas have increased in size from 13,820 to 34,843 km2, doubling the human population at risk (Picozzi *et al.,* 2001). The *T. b. rhodesiense* and *T. b. gambiense* are now less than 120km apart (Figure 1). By the late 1980's HAT appeared in Tororo (TR) and by 1998, HAT cases began to be recorded in the Soroti (SR) district, north of Lake Kyoga in Central Uganda. From 2004 to date, all the districts in central Uganda - Kaberamaido (KA), Dokolo (DK), Lira (LR), Apac (AP) and Kole (KO) have reported HAT cases (Lwala Hospital Medical Reports, 2010-2012). These foci in central Uganda are reffered to as the new foci (Figure 3). The epidemic in the new foci has been attributed to import of cattle carrying *T. b. rhodesiense* from disease endemic areas in the south (Fèvre, 2001), although recent work on the tsetse vector, *Glossina fuscipes fuscipes*, suggests that movement of susceptible flies from south to north could also be implicated in the emergence of new foci (Abila *et al.,* 2008; Beadell *et al.,* 2010; Echodu *et al.,* 2013; Aksoy *et al.,* 2013). Analyses of microsatellite and mitochondrial haplotype data show that the populations of *G. f. fuscipes* north and south of Lake Kyoga are separate (Abila *et al.,* 2008, Beadell *et al.,* 2010).

Population genetics studies have been carried out on *T. brucei* isolates across Africa, including HAT foci in Uganda and western Kenya. Analysis of *T. b. rhodesiense* isolates from the old foci in southeastern Uganda (BS, BG, BU, Figure 3) by isoenzyme, RFLP, and microsatellite analyses show that they are relatively heterogeneous (Gibson and Gashumba, 1983; Enyaru *et al.,* 1993; Hide *et al.,* 1994; MacLeod *et al*., 2000;, Goodhead *et al.,* 2013; Duffy *et al.,*2013; Aksoy *et al.,* 2013). Genotype has been correlated with clinical presentation in patients and virulence in experimental mice (Duffy *et al.,* 2013). Although it is assumed that *T. b. rhodesiense* spread from the old to the new foci, *T. b. rhodesiense* isolates from Soroti and Tororo (SR and TR respectively, Figure 3) were genetically distinct from those in the old foci, but closely related to each other (Aksoy *et al.,* 2013). This confirms that *T. b. rhodesiense* was introduced into Soroti via cattle from Tororo (Welburn *et al.,* 2001). Microsatellite analysis (7 loci) of *T. b. rhodesiense* populations from Tororo/Soroti and Malawi showed that levels of genetic diversity were much higher in the Malawi focus, with evidence of recent genetic exchange between isolates (Aksoy *et al.,* 2013). The lack of genetic exchange, clonal expansion

and epidemic population structure of Tororo/Soroti Ugandan *T. b. rhodesiense* agrees with the conclusions of previous population genetics studies (Hide *et al.,* 1994; Goodhead *et al.,* 2013). Thus, the local population structure of *T. b. rhodesiense* seems to depend on the relative amounts of clonal versus sexual reproduction, driven by transmission dynamics specific to the local conditions.

In this study a set of 17 highly variable microsatellite loci markers were used (Balmer *et al.,* 2006, Sistrom *et al.,* 2013) to investigate the patterns of genetic variation among 269 *T .b. brucei* and *T. b. rhodesiense* isolates from Uganda and the neighbouring region of western Kenya. The aim was to understand the extent of genetic exchange both within and between *T. b. brucei* and *T. b. rhodesiense* and to investigate the origin and spread of HAT in Uganda. This is by far the most comprehensive study of genetic variation for Ugandan *T. brucei* yet documented. Understanding the population structure of *T. brucei* and the extent of genetic variation in both human infective and non-infective subspecies will reveal the potential for generation and spread of new human infective strains and is thus of critical relevance for disease control.

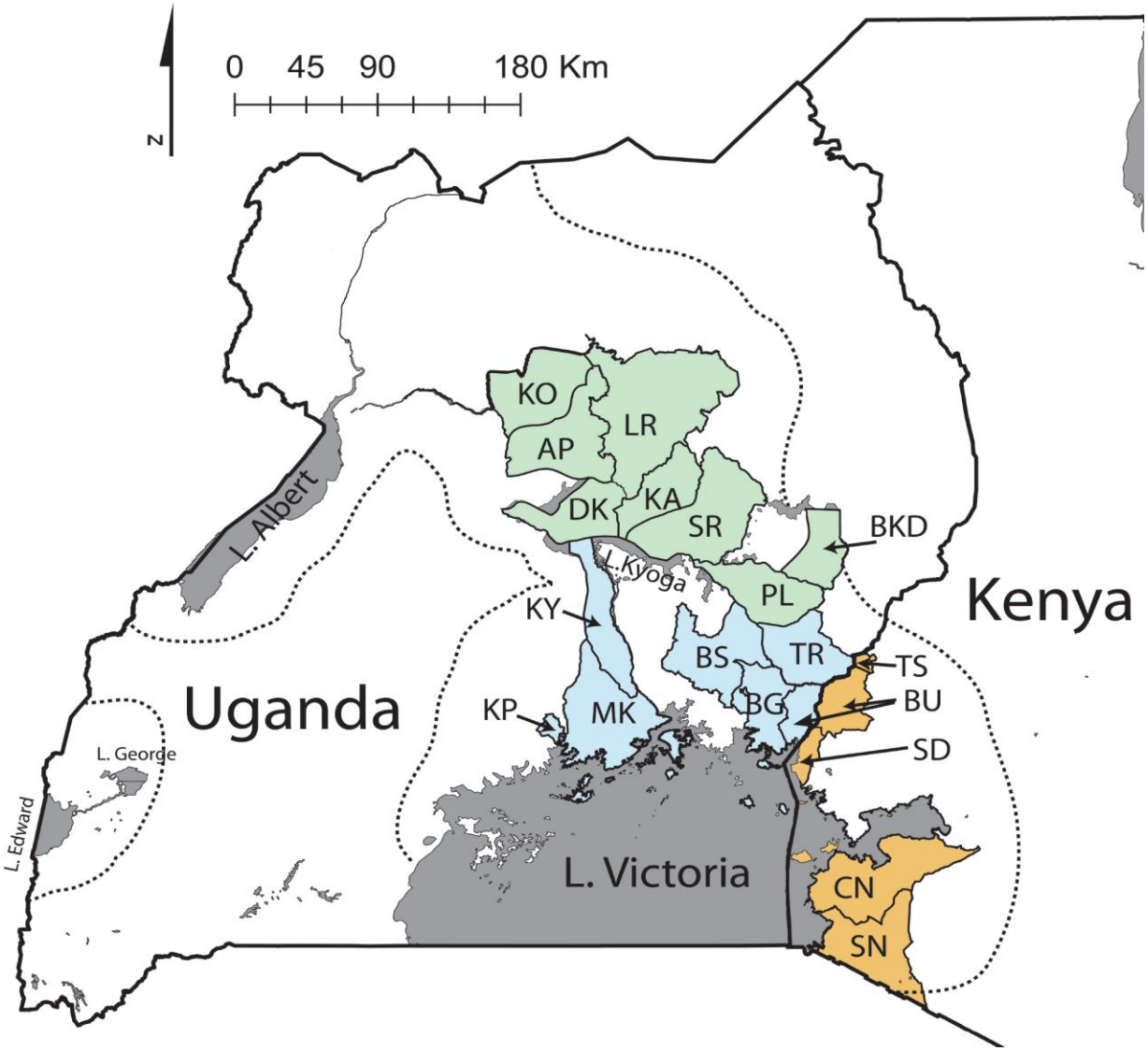


Figure 3. The 19 Ugandan and Kenyan districts from which *T. brucei* samples were collected. The dotted lines indicate the *G. f. fuscipes* distribution in the study region, and thus the distribution of *T. brucei*; there is a disjunct area of *G. f. fuscipes* around Lake George. Lakes (grey shading) are indicated by name. Districts are identified by two/three letter abbreviations Districts are color-coded as follows: green—new foci of *T. b. rhodesiense* in central Uganda; blue—old foci of *T. b. rhodesiense* in south eastern Uganda; orange—foci of *T. b. rhodesiense* in western Kenya. The blue and green shaded areas separated by Lake Kyoga also demarcate the genetically northern and southern *G. f. fuscipes* populations (Abila *et al.,* 2008; Beadell *et al.,* 2010).**Legend**: AP (Apach), BKD (Bukedae), BG (Bugiri), BU (Busia), BS (Busoga), DK (Dokolo), KA (Kaberamaido), KP.(Kampala), KY.(Kayunga), KO.(Kole), LR.(Lira), MK.(Mukono), PL.(Pallisa), SR.(Soroti), TR.(Tororo), TS.(Teso), CN.(Central Nyanza), SN.(South Nyanza) and SD.(Sidende).

3.3 Materials and methods

3.3.1 Source of trypanosome isolates for this study

The 269 *T. brucei* isolates (Appendix I) collected between 1959 and 2011 in 19 sites from the known parasite range in Uganda and western Kenya were analysed. The isolates were obtained from various hosts (180 from humans, 57 from cattle, 1 from sheep, 11 from pigs, 1 from dogs, 7 from wild animals and 12 from tsetse). Most of the samples (N= 194) were from archival cryopreserved collections obtained from KARI-Trypanosomiasis Research Centre (KARI-TRC, Kenya), Yale University School of Public Health (US) and National Livestock Resources Research Institute (NALIRRI, Uganda) cryo-banks. The isolates were expanded in cyclophosphamide immune-suppressed Swiss white mice or rats at KARI-TRC with approval from the Institutional Animal Care and Use Committee (IACUC). Other parasites were expanded in axenic culture at Yale University.

3.3.2 Isolation of trypanosome DNA and detection of SRA gene by PCR

Depending on the quality of the material, DNA was extracted using either Qiagen DNeasy Blood or Tissue kit (Qiagen, Valencia, CA) according to the manufacturer's instructions or by traditional phenol-chloroform DNA extraction method (Sambrook *et al.,* 1989). DNAs from field samples were screened by ITS-PCR test to separate *T. brucei* from other African trypanosomes (Njiru *et al.,* 2005). All *T. brucei* samples were tested for presence of the serum resistance associated (SRA) gene by polymerase chain reaction (PCR). The PCR detection was performed using SRA gene specific primers (SRA A: 5′-GACAACAAGTACCTTGGCGC and SRA E: 5′-TACTGTTGTTGTACCGCCGC), SRA H-SRA J using protocols developed by Gibson *et al.*, (2002). Briefly, 1µl DNA was amplified with 1 unit of Taq polymerase (Promega, Madison, MO) in the buffer supplied by the manufacturer which contained no $MgCl_2$ in the presence of the specific primers for SRA gene. Empirically determined concentration of $MgCl_2$ was used in each reaction to provide optimum yield and specificity. Reactions were carried out in the Applied Biosystems® 2720 Thermal Cycler (Life Technologies, USA) using the following cycling conditions: initial at 94^0C 1 minute, 30 cycles of denaturation at 92^0C for 30 seconds, annealing at 60^0C for 45 seconds, elongation at 72^0C for 45 seconds and a final extension step of 72^0C for 4 minutes. Each set of reactions included a no-DNA negative control. Another set of primers used were SRA-R-SRA-F (Radwanska *et al.,* 2002). The PCR products were loaded onto ethidium

bromide stained 1% agarose gels in Tris borate Ethylenediaminetetraacetic acid (TBE) buffer (Sambrook *et al.,* 1989). On every gel, a 100 base pair (bp) DNA molecular weight marker (Fermentas, Pennsylvania USA) was separated to confirm expected molecular weight of the amplification product. Gel images were captured with a Uvidoc HD2 gel imager (Cambridge, UK).

3.3.3 Multi locus microsatellite typing

Fluorescent labeled forward primers for 17 *T. brucei* loci (Appendix II) described by Balmer *et al*. (2006) and Sistrom *et al.* (2012) were used to amplify specific allele from each isolate by PCR. Amplification was carried out using Type-it microsatellite PCR kit (Qiagen, Germany). Briefly, 1µl of genomic DNA diluted to approximately 100ng/µl was amplified using 5µl of Type-it Master Mix and 1µl each of forward and reverse primers in a total reaction volume of 15µl. PCR reactions were carried out using Eppendoff Mastercycler Pro thermocycler (Eppendorf, Germany) under the following PCR cycling profile: initialization step of 95^0C for 4 minutes, followed by twelve touch-down cycles of 95^0C for 30 seconds, 60-50^0C for 25 seconds and 72^0C for 30 seconds, an additional 30 cycles of 95^0C for 30 seconds, 50^0C for 25 seconds and 72^0C for 30 seconds, and a final extension step of 72^0C for 20 minutes.

PCR products were multiplexed in groups of two or three before fragment analysis and sizing by capillary electrophoresis using an automatic 3730xl DNA Analyzer (Applied Biosystems Inc.). Allele sizes were determined using Genescan ROX-500 internal size standard for loci; TB1/8, TB5/2, TB6/7, TB9/6, TB10/5, TB11/13, Tryp51, Tryp67, Tryp55, Tryp53 and Tryp59 and Liz-500 internal size standard for loci; Tryp66, Tryp54, Tryp62, Tryp59 and Tryp53. In a 96-well microtitre plate, 1 µl of PCR product was added to 9 µl formamide and 0.5µl of either ROX500 or Liz500 size standard.

3.3.4 Genetic diversity

Allele size calling was performed using GeneMarker version 2.4.0 (SoftGenetics, USA) and manually edited. Raw alleles were exported from GeneMarker to TANDEM version 1.0.9 (Matschiner and Salzburger, 2009) for allele binning. Genepop version 4.2 (Rousset, 2008) was used to calculate number of alleles (N_a), observed level of heterozygosity (H_o) and expected (H_e) level of heterozygosity under Hardy-Weinberg equilibrium conditions, allele richness (Ar) and

inbreeding coefficient (Fis). Linkage disequilibrium (LD) was evaluated using the log likelihood ratio statistic (G statistic) implemented in Genepop v4.2 (Rousset, 2008).

3.3.5 Population structure and differentiation

Bayesian clustering implemented in STRUCTURE v2.3.3 (Pritchard *et al.,* 2000) was used to assign isolates to genetic clusters (K) according to the allele frequencies at each locus. Five independent runs for K = 1-10 were carried out. For all runs, an admixture model and independent allele frequencies were used with a burn-in value of 250,000 steps followed by 1,000,000 iterations. The optimal value of K was determined using STRUCTURE HARVESTER v0.6 (Earl and vonHoldt, 2011) to calculate ad hoc statistic "ΔK" (Evanno *et al.,* 2005).
Assignment of individual strains to a given cluster and levels of genetic admixture within each individual were assessed using STRUCTURE membership coefficients (*Q*-values), which represent the fraction of the sampled genome that has ancestry in a given cluster. Genetic clustering between *T. brucei* isolates was also determined using Discriminate Analysis of Principal Components (DAPC) implemented in the R (R Core Development Team, 2013) package Adegenet (Jombart, 2008) a method that makes no assumptions on HWE or LD and tends to perform better when hierarchical and clonal structure is present (Jombart *et al.,* 2010). DAPC comprises two steps: 1) a principal component step, where the dimensionality of the multilocus allelic data is reduced to 15 principal components based on a-scores; and 2) a discriminant analysis step, where two discriminants were used to identify the linear combination of principal components from the first step that best distinguished prior groupings (populations) of individuals.

To measure the amount of genetic divergence among sampling localities, and the inferred genetic clusters and sampling sites, pairwise F_{ST} values and associated P values were calculated using ARLEQUIN v3.5 (Excoffier and Lischer, 2010) calculations were performed for 10,000 permutations. The same software was used to carry out a hierarchical analysis of molecular variance (AMOVA) to analyze the partitioning of the genetic variance (a) among and within the genetic clusters detected using previously described methods, (b) among and between three pre-defined groups within each genetic cluster: host (human, cattle, sheep, pig, dog, wild animals and tsetse flies), decade of isolation, subspecies, and (c) among all samples based on date of collection. For this analysis, samples were grouped by the decade of collection to determine

whether observed genetic variation could be attributed to temporal turnover. Each AMOVA analysis was run for 10,000 permutations with an allowable missing data level of 40%. LD bias correction method (Waples, 2006) implemented in LDNe (Waples and Do CHI, 2008) was used to estimate the effective population size (*Ne*) of each genetic cluster. We ran the analysis using a lowest allele frequency of 0.01.

3.4 Results

3.4.1 Taxon identification and genetic diversity

Of the 269 *T. brucei* isolates analyzed, 210 (78%) were *T. b. rhodesiense*, as determined by the presence of the *SRA* gene. Majority of *SRA* positive samples were found among human isolates and 34% (21/69) of the isolates from non-human vertebrate hosts tested positive for the *SRA* gene (Appendix III). Cattle formed the largest percentage (16 of 21; 76%) of the SRA positive non human hosts indicating that *T. b. rhodesiense strains* are circulating in these animals.

The final dataset for analysis included samples from 19 districts in Uganda and Kenya (Figure 2), averaging 13 samples per district. The average amplification rate was 70.0% across the 17 microsatellite loci (S.E. 12.13%); the 2010/2011 field samples collected on FTA cards had variable template concentration, leading to non-amplification due to low template concentration (Aksoy *et al.*, 2013). As expected, due to clonal reproduction in *T. brucei*, all loci deviated from HWE in at least one district (Appendix IV). A_R ranged between 2.24 and 7.35 (districts for which a single sample was collected were excluded; Table 1). H_E ranged from 0.34 to 0.70, H_O from 0.27 to 0.57 and F_{IS} from -0.16 to 0.43 (Table 1). All genotypic data were submitted to Dryad (http://datadryad.org).

Table 1. Sample sizes and genetic diversity statistics for seventeen microsatellite loci across *Trypanosoma brucei* isolates from 19 districts

Sampling site	Site Symbol	Country	**N**	A_R	H_E	H_O	F_{IS}
Apach	AP	Uganda	1	1.2	N/A	N/A	N/A
Bukedae	BKD	Uganda	1	1.5	N/A	N/A	N/A
Bugiri	BG	Uganda	7	3.2	0.47	0.41	0.13
Busia	BU	Uganda/Kenya	32	3.4	0.35	0.28	0.2
Busoga	BS	Uganda	23	5.3	0.51	0.43	0.18
Dokolo	DK	Uganda	11	2.4	0.24	0.17	0.42
Kaberamaido	KA	Uganda	59	3.7	0.25	0.19	0.32
Kampala	KP	Uganda	1	1.7	N/A	N/A	N/A
Kayunga	KY	Uganda	2	2.4	0.75	0.75	-0.04
Kole	KO	Uganda	25	3.9	0.29	0.18	0.39
Lira	LR	Uganda	10	2.8	0.33	0.24	0.26
Mukono	MK	Uganda	3	2.2	0.6	0.64	-0.16
Pallisa	PL	Uganda	15	2.4	0.3	0.37	-0.17
Soroti	SR	Uganda	25	4.2	0.39	0.24	0.43
Tororo	TR	Uganda	31	4.7	0.47	0.45	0.11
Teso	TS	Kenya	1	1.4	N/A	N/A	N/A
Central Nyanza	CN	Kenya	9	2.1	0.41	0.55	-0.25
South Nyanza	SN	Kenya	10	2.3	0.36	0.28	0.24
Sidende	SD	Kenya	1	1.44	N/A	N/A	N/A

N = number of samples analyzed, AR = allele richness, HE = expected heterozygosity, HO = observed heterozygosity and FIS = Fisher's inbreeding coefficient. N/A = data not available because only a single sample was collected.

3.4.2 Population Structure, differentiation among groups, and Ne estimates

Bayesian clustering analyses as implemented in STRUCTURE; grouped the 269 isolates in 3 genetic clusters (Figure 4). The three genetic clusters; 1, 2 and 3 (blue, green and red represent clusters 1-3 respectively) are shown in Figure 5 which shows geographic origin and the assignment of each isolate to one of the three clusters in relation to its host and taxonomic designation (*T. b. rhodesiense* versus *T. b. brucei*) . Clusters 1and 3 includes isolates mainly from central and southeastern Uganda, while cluster 2 is composed of isolates from Kenya. Importantly, localities in the southeastern Ugandan foci (Busoga, Busia, Tororo), and central Uganda foci (Soroti, Kaberamaido, and Dokolo) share strains from both cluster 1 and 3 (only one strain from cluster 2). The *T. b. brucei* and *T. b. rhodesiense* samples are found together in clusters 1 and 3, indicating that *T. b. rhodesiense* strains are not genetically differentiated from the co-occurring *T. b. brucei* strains; most isolates in cluster 2 were *SRA* positive. The same analysis was done omitting all the Kenyan samples and the same two clusters were recovered for the *T. b. brucei* /*T. b. rhodesiense Ugandan* strains. Figure 6 shows the results of DAPC clustering approach which grouped isolates into the three genetic clusters as described above.

The western Kenyan samples belong mainly to cluster 2 although a few individuals with genetic assignment to cluster 1 (blue bars) are also found in this region. Similarly, a few individuals from cluster 2 (both pure and admixed) can be found in central and southeastern Uganda. F_{ST} values between sampling sites ranged from 0 to 0.67 (Appendix V), and F_{ST} values between the three DAPC and STRUCTURE inferred clusters ranged from 0.24 to 0.46 (Appendix VI), with most values being statistically significant $P<0.01$. This findings confirmed genetic differentiation not only among the three clusters but also among some of the isolates from the 19 sampling sites belonging to both the same or different genetic clusters.

AMOVA results show the level of genetic diversity explained by the genetic clusters and how much of the genetic variation is explained by collection date, species host, subspecies and sampling sites (Table 2). Most of the genetic variation was apportioned within (71.8%) rather than among the three STRUCTURE-defined clusters

Effective population size estimates (N_e) and confidence intervals calculated using LNDe (Waples, 2006) for the 3 clusters revealed smaller Ne values in clusters 1 and 2 (13.1 and 8.1, respectively) than in cluster 3 which had an Ne of 44.3 (Table 3). As the confidence intervals

around these estimates were relatively narrow, all clusters differed significantly ($p<0.05$) in effective population size estimates using LDNe analysis.

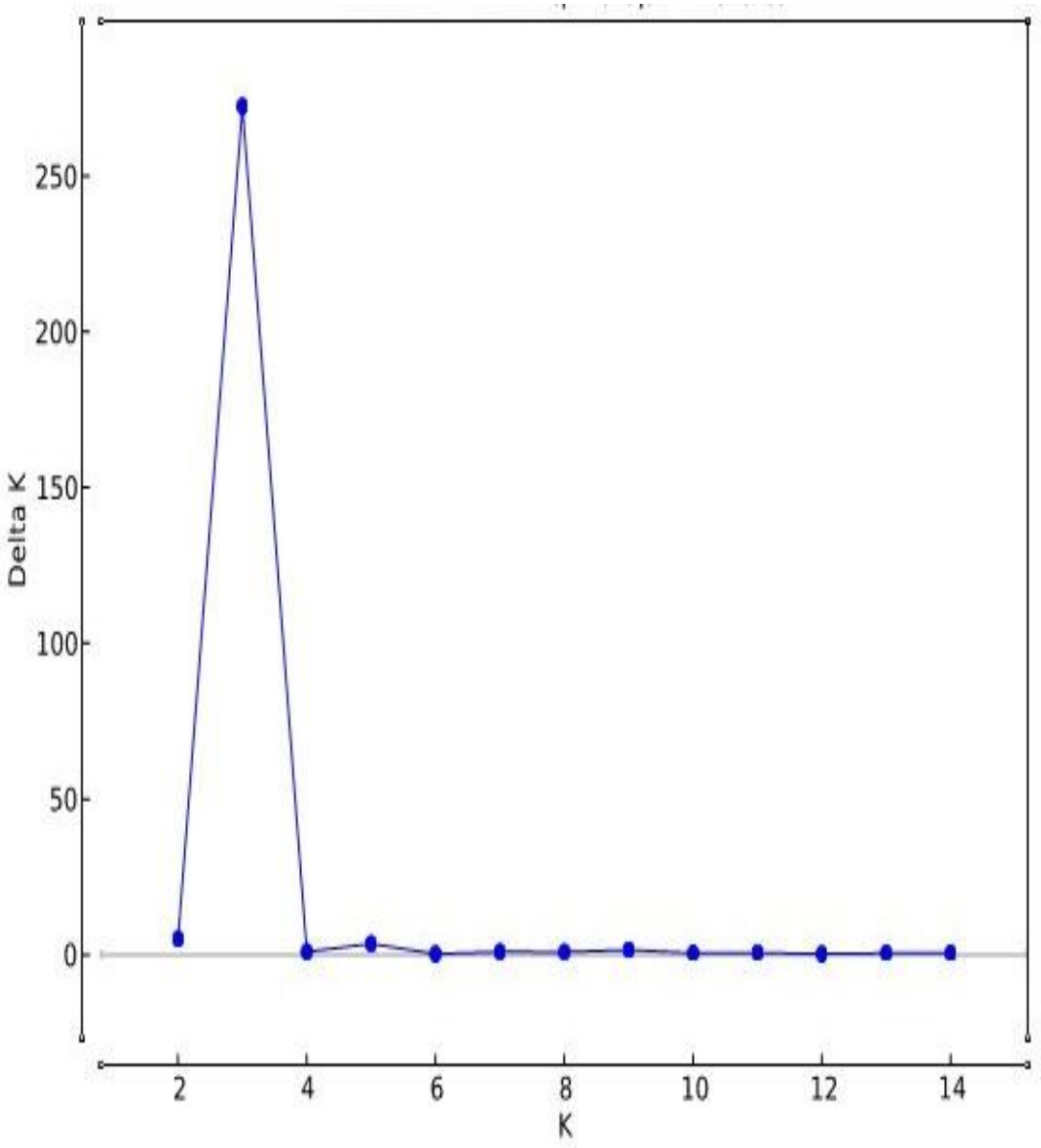


Figure 4. Estimation of population clustering level from *Trypanosoma brucei* microsatellite Genotypes. The highest peak at ΔK represents the most appropriate number of genetic clusters (K = 3).

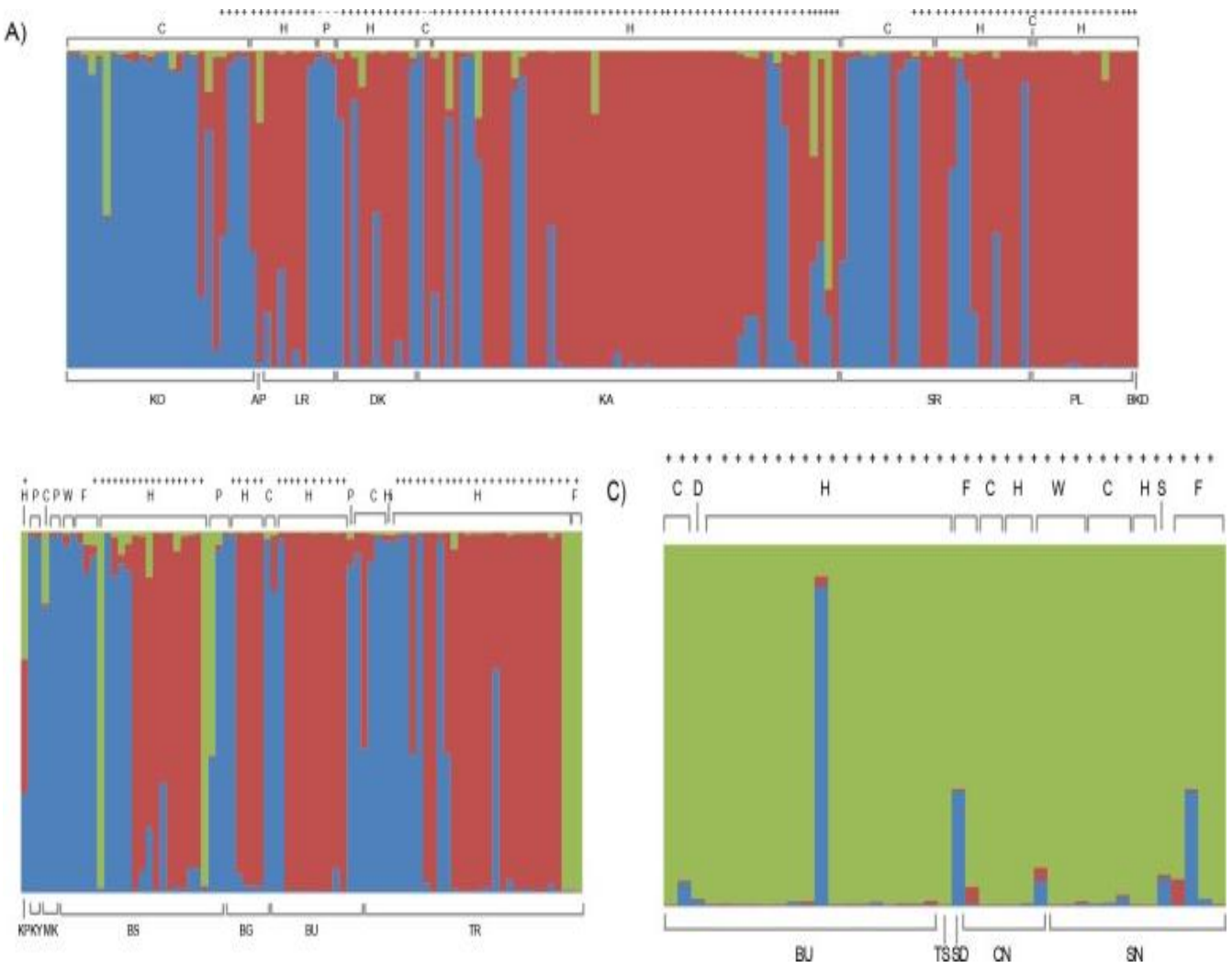


Figure 5. Population structure based on Bayesian clustering (ΔK = 3) for 269 samples of *T. b. brucei* and *T. b. rhodesiense* isolates from Uganda and Kenya, genotyped at 17 microsatellite loci. A) Central Uganda; B) Southern Uganda; C) Kenya. The district of origin of each sample is reported at the bottom of each panel (A-C). AP (Apach), BKD (Bukedae), BG (Bugiri), BU (Busia), BS (Busoga), DK (Dokolo), KA (Kaberamaido), KP.(Kampala), KY.(Kayunga), KO.(Kole), LR.(Lira), MK.(Mukono), PL.(Pallisa), SR.(Soroti), TR.(Tororo), TS.(Teso), CN.(Central Nyanza), SN.(South Nyanza) and SD.(Sidende). A bracket line group's samples from the same district. Within each panel (A-C), samples are organized by districts. The districts are shown below each A-C plot in a west-east direction—with abbreviations. Host is shown immediately above each plot (H = human, C = cattle, D = dog, P = pig, S = sheep, F = tsetse fly, W= Wildlife). Above the host information, + denotes samples with the SRA gene present. Each bar represents an isolate, the colors within the bar reflect the percent assignment (shown on the Y axis) of that individual to one of three genetic clusters (blue, green and red represent clusters 1–3, respectively). The proportion of each color in each individual represents the probability with which an individual is assigned to each of the three color-coded clusters.

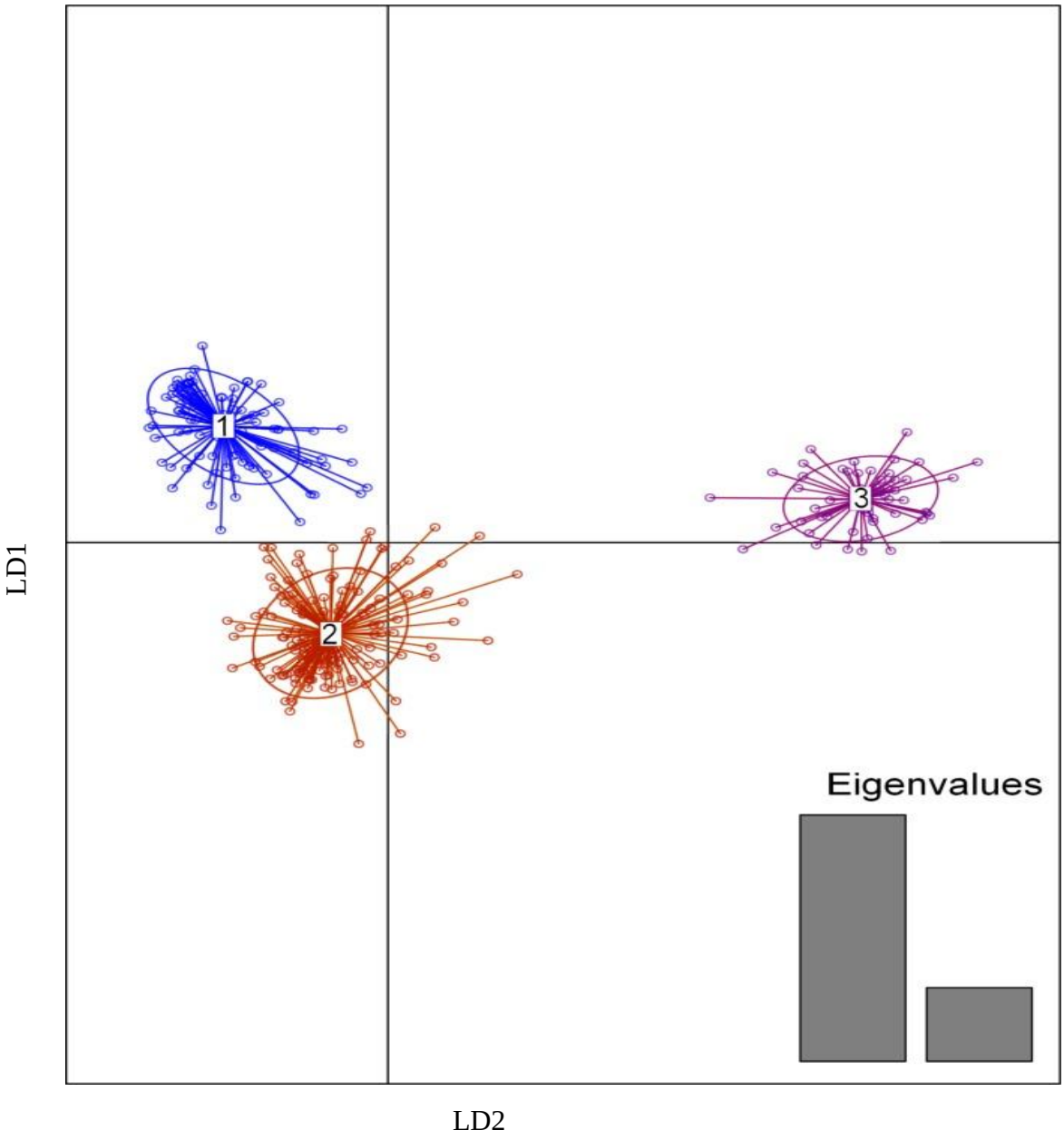


Figure 6. Discriminant analysis of principal components (DAPC). Two linear discriminants (LD1 and LD2) were used, following selection of principal components using a-score optimization, to plot *T. brucei* individual isolates. Dots represent individual genotypes connected by a line to the center of an ellipse with different colors representing the three clusters; blue (cluster 1), red (cluster 2), and purple (cluster 3).

Table 2. AMOVA analyses on seventeen microsatellite loci of *T. brucei* isolates partitioned into four groups

	Within Groups	Among Groups	P value
Clusters	79.21%	20.79%	<.01*
Sampling Dates	91.51%	8.49%	<0.01*
Cluster 1			
Host	83.87%	16.13%	<0.01*
Date	87.39%	12.61%	<0.01*
Subspecies	87.64%	12.36%	<0.01*
Cluster 2			
Host	95.15%	4.85%	0.01*
Date	92.57%	7.43%	0.03*
Subspecies	N/A	N/A	N/A
Cluster 3			
Host	97.37%	2.63%	0.07
Date	99.23%	0.77%	0.07
Subspecies	83.75%	16.25%	<0.01*

Host (human, cattle, sheep, pig, dog, wild animals and tsetse flies), year of isolation (decade), subspecies, and structure/dapc inferred genetic clusters. Asterisks denote comparisons with significant p values (<0.05)

Table 3. Estimates of effective population size (Ne) calculated using LNDe among the three structure/dapc genetic clusters

Cluster	Ne	Lower and Upper C.I
1	9.7	(8.1–11.6)
2	2.3	(1.7–3.0)
3	86.4	(28.1–170.4)

C.I = Confidential interval.

3.5 Discussion

This study examined the pattern of genetic differentiation of *T. b. brucei* and *T. b. rhodesiense* isolates from Uganda and western Kenya, and the results demonstrated that the expansion of the disease to the new foci in central Uganda occurred from the northward spread of *T. b. rhodesiense* and confirmed the emergence of the human infective strains from non-infective *T. b. brucei* strains of different genetic backgrounds. Results of genetic diversity showed that 34% (21/69) of isolates from non-human vertebrate hosts tested positive for the *SRA* gene, indicating that *T. b. rhodesiense* strains were circulating in the animals. Cattle formed the largest percentage (16 of 21; 76%). These results confirmed that cattle are an important reservoir for *T. b. rhodesiense* and they are likely to fuel the epidemiology of sleeping sickness in Uganda as 28.1% of the *T. brucei* isolates found in cattle (16/57) were SRA positive. In previous studies cattle was connected to earlier *T. b. rhodesiense* out breaks (Baldry, 1972; Gibson and Wellde, 1985; Fèvre, 2001). Additionally in the STRUCTURE analyses they clustered together with the human isolates from the same geographic regions suggesting on going genetic exchange between *T. brucei* isolates from cattle and humans in the same area (Hide *et al.,* 1994). The distance separating the *T. b. rhodesiense* and T. b. gambiensis foci in North western Uganda is less than 100 km (Fèvre, 2001; Picozzi *et al.,* 2005), understanding the role and impact of cattle in fuelling movement of *T. b. rhodesiense* strains is important as these results suggest that continued cattle movement from southern districts can accelerate the fusion of the two disease belts with unknown public health consequences. Similarly, increased livestock trade across south eastern Uganda and Western Kenya also poses a risk transferring *T. b. rhodesiense* from the old Uganda HAT foci in that region to western Kenya, which has been reporting low HAT prevalence in the last decade (Enyaru *et al.,* 2006; Rutto *et al.,* 2013).

During PCR analysis, average amplification rates were obtained, at 70.0% across the 17 microsatellite loci (S.E. 12.13%); the 2010/2011 field samples collected on FTA cards had variable template concentration, leading to non-amplification (Aksoy *et al* 2013). Data from this study show clear evidence of linkage disequilibrium at most loci this is due to clonal reproduction in *T. brucei*, this is in agreement with previous studies by Duffy *et al*. (2013).

Results of the analysis of population structure, differentiation among groups, and Ne estimates showed that individuals within sampling sites with varying degrees of assignment to each of the three genetic clusters co-occur. These clusters show evidence of recent long-range

dispersal, as demonstrated in (1) individuals with 100% assignment to a different genetic cluster than other samples from the same locality, and (2) genetically admixed individuals, likely the result of mating between local and immigrant strain in a given sampling locality. Localities in the southeastern Ugandan foci and central foci shared strains from both cluster 1 and 3 implying that the strains from the old and new foci are not genetically distinct.

Two of the three clusters contained a mixture of Ugandan *T. b. rhodesiense* isolates from the old foci in the southeast and from the new foci in central districts, while the third cluster groups *T. b. rhodesiense* isolates from western Kenya. Thus, despite their geographic proximity and the widespread view that the Kenyan focus was an extension of southeast Uganda (Baldry, 1972) the Ugandan and Kenyan *T. b. rhodesiense* populations seem to be genetically distinct. There is also evidence of genetic admixture likely via both long and short-range dispersal. From the earliest isoenzyme studies onwards, it has been clear that *T. b. rhodesiense* differs between geographically distant foci (Gibson *et al*., 1980; Gibson *et al.,* 1985; Hide *et al.,* 1994; Gibson and Stevens, 1999; Goodhead *et al.,* 2013 and Duffy *et al.,* 2013), but more overlap might have been expected between these neighbouring foci in Uganda and Kenya, which were in close contact via Lake Victoria (Baldry, 1972). One factor distinguishing HAT from the two areas is transmission by different tsetse species. HAT in the lakeshore region of southeast Uganda was originally transmitted by the *Morsitans* group fly *G. pallidipes* (Gibson and Gashumba, 1983; Baldry, 1972), and this fly was also the vector of HAT in South Nyanza, Kenya (Gibson and Wellde, 1985) and in Busia (Baldry, 1972); however, outbreak of *T. b. rhodesiense* in Alego Central Nyanza, Kenya, transmission was by the *palpalis* group fly *G. f. fuscipes* (Onyango *et al.,* 1966). In Uganda, transmission of *T. b. rhodesiense* also switched to *G. f. fuscipes* as *T. b. rhodesiense* extended northwards into areas infested with this species from the mid 1970's onwards (Gibson and Gashumba, 1983) and *G. f. fuscipes* is regarded as the main HAT vector in Uganda (Waiswa *et al.,* 2006; Aksoy *et al.,* 2013). Therefore, the factor that led to genetic isolation of cluster 2 could be adaptation to transmission by a different tsetse vector, *G. pallidipes*. Results clearly rule out the hypothesis that *T. b. rhodesiense* spread from its traditional focus in southeastern Uganda to western Kenya in the 1950's along with *G. pallidipes* (Baldry, 1972) and furthermore, *G. pallidipes* populations in Uganda and Kenya are genetically distinct (Ouma *et al.,* 2011).

Separate transmission cycles may also explain the partitioning of Ugandan *T. b. rhodesiense* isolates into two genetic clusters, despite the fact that they are now sympatric. *Glossina f. fuscipes* and *G. pallidipes* occupy different biomes; have different host-feeding preferences, and susceptibility to trypanosome infection. Therefore, a priori, divergence would be expected among the trypanosome populations adapted to transmission cycles involving either of these vectors. A switch from transmission by *G. pallidipes* to *G. f. fuscipes*, as occurred in south eastern Uganda, would be expected to select for certain genotypes, while allowing the two divergent trypanosome populations to mix. It may also be significant that the *G. f. fuscipes* populations to the north and south of Lake Kyoga are genetically distinct (Abila *et al.,* 2008; Beadell *et al.,* 2010), implying that transmission cycles in the old and new foci were separate until trypanosomes were transferred via movement of infected humans and livestock.

In earlier studies, two *T. b. rhodesiense* genotypes circulating in the old foci were defined by isoenzyme profiles (zymodemes) and correlated with clinical presentation; the Zambezi zymodeme was associated with more chronic progression of HAT than the Busoga zymodeme (Godfrey *et al.,* 1990). Although Goodhead *et al*. (2013) found no simple correlation between zymodeme designation and clade based on 11 microsatellite loci, some population sub-structuring was evident in their analysis, and perhaps inaccuracy in zymodeme classification, which is based on relatively few informative isoenzyme loci, has obscured the relationship. Goodhead *et al*. (2013) also compared the genome sequences of one representative Busoga and Zambezi isolate and found that, although the genomes were >99.8% identical, they showed extensive chromosome-wide SNP variation. Comparison with *T. b. brucei* or *T. b. gambiensis* genomes revealed that some chromosomes were mosaics of shared alleles, suggesting that the Ugandan *T. b. rhodesiense* strains might have originated through a hybridization event between *T. brucei* of East and West African origin. Historically it is known that *T. b. gambiensis* was present in the lakeshore region of southeast Uganda in the early 20th century, so it is indeed possible that introgression has occurred.

Previous studies showed that there is sub-structuring in trypanosome populations in relation to host and geography suggesting that both geography and host play a role in shaping the patterns of genetic differentiation among *T. b. brucei* and *T. b. rhodesiense* isolates (Hide *et al.,* 1994; MacLeod *et al.,* 2001; Goodhead *et al*., 2013;), this study does not support this. Although the estimates of genetic differentiation among sampling sites are statistically significant for a

number of pairwise comparisons the biological significance of this result is questionable, given the AMOVA results from Table 2 which show that within each of the three genetic clusters taxonomy, date of collection, and host explain less than 16% of the overall observed genetic variation. However, it should also be noted that the results from the AMOVA analyses are somewhat weakened by the fact that the representation of time and space points or hosts is not uniform.

The finding of individuals of two genetic clusters in both the old and new Ugandan foci challenges previous studies (Enyaru *et al.,* 1993; MacLeod 2001; Maclean *et al.,* 2007; Ouma *et al.,* 2011; Goodhead *et al.,* 2013; Duffy *et al.,* 2013) which suggested that *T. b. rhodesiense* isolates from the Ugandan old and new foci were genetically distinct. This study, was based on a much larger data set both in terms of loci and number of samples, and including both *T. b. brucei* and *T. b. rhodesiense* co-occurring strains, suggests that the expansion of the disease to the new foci in central and western Uganda occurred from *T. b. rhodesiense* isolates spreading from the old to the new foci. In addition, estimates of Ne show that clusters 1 and 2 have much lower effective population sizes than cluster 3, indicating that clusters 1 and 2 experienced recent clonal expansion, whereas cluster 3 had a higher rate of sexual reproduction. This may also explain the discord between these results and those of others. The identified linkage disequilibrium and differences in effective population size estimates could be an example of the potential for rapid population contractions and expansion of different genotypes due to clonal reproduction

Results from this study concur with previous studies that identified *T. b. rhodesiense* epidemics involving multiple lineages (Gibson and Gashumba, 1983; Gibson and Wellde, 1985) since *T. b. rhodesiense* strains with different genetic background co-occur in both the new and the old foci. There was no evidence for temporal structure in Ugandan *T. brucei*, whereas Duffy *et al*. (2013) found evidence of genetic shifts in allelic frequencies between samples collected in 1970 and 1990, as well as very low genetic similarity between samples from the old and new Ugandan foci. Here, temporal variation does not explain the partitioning of the observed genetic variation as shown by the AMOVA analyses and by the occurrence in the same genetic clusters of samples collected at different time points from the same or different sampling sites. Instead this study revealed evidence of geographic genetic structuring. In this sense this study parallels better the Duffy *et al.* (2013) result for the Malawi strains rather than Uganda ones, underscoring

the importance of using highly variable markers for studies such as this, where genetic differentiation levels are expected to be small, given the narrow spatial and temporal scale of the study. The other important difference between these studies that may play a role in explaining the different results is that the Duffy *et al.* (2013) study was entirely focused on *T. b. rhodesiense* strains from human patients, while this study looked at the genetic differentiation of co-occurring *T. b. brucei* and *T. b. rhodesiense* isolates and included 32% of *T. b. rhodesiense* strains from non-human isolates. Looking at the whole spectrum of circulating genotypes provides additional insights on the evolutionary origin of the strains and their level of genetic admixture, as this and other studies have clearly shown that *T. b. rhodesiense* strains originate from *T. b. brucei* strains, when they acquire the SRA gene (Balmer *et al.*, 2011).

In conclusion this study of genetic diversity and population structure of *T. b. rhodesiense* and *T. b. brucei* isolates from Uganda and western Kenya was done and the aim was examine the pattern of genetic differentiation of *T. b. brucei* and *T. b. rhodesiense* isolates from Uganda and western Kenya, to understand population structure and the modalities of parasite spread to help support sustainable control strategies for AAT and HAT in this region. Continent wide studies have already shown that *T. b. rhodesiense* and *T. b. brucei* strains should not be treated as reproductively isolated taxa, as some *T. b. brucei* strains are more closely related to *T. b. rhodesiense* strains than their conspecifics and vice versa. This study shows that there is genetic structuring within *T. brucei* populations from Uganda and Kenya, separating the isolates into three groups. We found clear evidence of on-going genetic admixture and long-range dispersal among *T. b. brucei* and *T. b. rhodesiense* strains. The use of a dense sampling scheme and highly variable loci enabled us to detect genetic exchange between the old and new Uganda disease foci, possibly mediated by cattle movements across the region as both *T. b. brucei* and *T. b. rhodesiense* strains were found circulating in cattle. These results have important implications for disease control, as they provide empirical evidence for the occurrence of genetic exchange between co-occurring human infective and non-infective strains, and the role of cattle in spreading the human disease. This study also emphasizes the importance of studying both *T. b. brucei* and *T. b. rhodesiense* strains when attempting to understand the population dynamics of *T. b. rhodesiense.*

3.6 References

Abila P.P., Slotman M.A., Parmakelis A., Dion K.B., Robinson A.S., Muwanika V.B., Enyaru J., Lokedi L.M, Aksoy S. and Caccone A. (2008). High levels of genetic differentiation between Ugandan *Glossina fuscipes fuscipes* populations separated by Lake Kyoga." *PLoS Neglected Tropical Diseases* **2**.5: e242.

Aksoy S., Caccone A., Galvani A.P. and Okedi L.M. (2013). "*Glossina fuscipes* populations provide insights for human African trypanosomiasis transmission in Uganda." *Trends in Parasitology* **29**.8: 394-406.

Aksoy S., Almeida-Val V.M., Azevedo V.C., Baucom R., Bazaga P., Beheregaray L.B., Bennetzen J.L., Brassaloti R.A., Burgess T.I., Caccone A., Chang S.M., Ciampi A.Y., Ciancaleoni S., Clímaco G.T., Clouet C., Coimbra M.R., Coutinho L.L., Dantas H.L., De Vega C., Echodu R., Enyaru J., Figueira A., Filho M.A., Foltz B., Fressigné L., Gadomski M., Gauthier N., Herrera C.M., Hyseni C., Jorge E.C., Kaczmarczyk D., Knott E., Kuester A., Lima A.P., Lima M.A., Lima M.P., Longo A.L., Lor G., Maggioni R., Marques T.S., Martins A.R., Matoso D.A., Medrano M., Mendonça M.A., Mettler R., Nascimento P.R., Negri V., Oliveira K.K., Oliveira L.O., Ovcarenko I., Paula-Silva M.N., Raggi L., Sandoval-Castillo J., Santos C.H., Martin Schaefer H., Segelbacher G., Seino M.M., Sistrom M., Taole M.M., Teske P.R., Tsagkarakou A., Verdade L.M., Villela P.M., Vinson C.C., Wingfield B.D., Wingfield M.J. (2013). Permanent Genetic Resources added to Molecular Ecology Resources Database 1 October 2012–30 November 2012. *Molecular Ecology Resources* **13**:341-343

Baldry D.A.T. (1972). A history of Rhodesian sleeping sickness in the Lambwe Valley. *Bulletin of the World Health Organisation* **47**: 699-718

Balmer O., Beadell J.S., Gibson W. and Caccone A. (2011). Phylogeography and taxonomy of *Trypanosoma brucei. PLoS Neglected Tropical Diseases* **5**:e961.

Balmer O., Palma C., MacLeod A. and Caccone A. (2006). Characterization of di-, tri and tetranucleotide microsatellite markers with perfect repeats for *Trypanosoma brucei* and related species. *Molecular Ecolology Notes* **6**: 508–510.

Beadell J.S., Hyseni C., Abila P.P., Azabo R., Enyaru J., Ouma J.O., Mohammed Y.O., Okedi L.M., Aksoy S. and Caccone A. (2010). Phylogeography and population structure of

Glossina fuscipes fuscipes in Uganda: implications for control of tsetse." *PLoS Neglected Tropical Diseases* **4**.3 (2010): e636.

Duffy C.W., MacLean L., Sweeney L., Cooper A., Turner C.M., Tait A., Sternberg J., Morrison L.J. and MacLeod A. (2013). Population Genetics of *Trypanosoma brucei rhodesiense*: Clonality and Diversity within and between Foci. *PLoS Neglected Tropical Diseases* **7**.11: e2526.

Earl D. and vonHoldt B. (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservational Genetic Resourses* **1**–3.

Echodu R., Sistrom M., Bateta R., Murilla G., Okedi L., Aksoy S., Enyioha C., Enyaru J., Opiyo E., Gibson W. and Caccone A. (2014). Data from: Genetic diversity and population structure of *Trypanosoma brucei* in Uganda: implications for the epidemiology of sleeping sickness and Nagana Dryad digital Repository. doi: 10.5061/dryad.m7q4c

Echodu R., Sistrom M., Hyseni C., Enyaru J., Okedi L.M., Aksoy S. and Caccone A (2013). "Genetically Distinct *Glossina fuscipes fuscipes* Populations in the Lake Kyoga Region of Uganda and Relevance for Human African Trypanosomiasis." *Biomedical Research International* Article ID 614721

Enyaru J.C., Matovu E., Nerima B., Akol M. and Sebikali C. (2006). Detection of *T. b. rhodesiense* Trypanosomes in Humans and Domestic Animals in South East Uganda by amplification of Serum Resistance-Associated Gene." *Annals of the NewYork Academy of Sciences* NY. **1081**: 311-319.

Enyaru J.C., Odiit M., Winyi-Kaboyo R., Sebikali C.G., Matovu E., Okitoi D. and Olaho-Mukani W. (1999). Evidence for the occurrence of *Trypanosoma brucei rhodesiense* sleeping sickness outside the traditional focus in south-eastern Uganda. *Annal of tropical medicine and Parasitology* **93**:817-822.

Enyaru J.C., Stevens J.R., Odiit M., Okuna N.M. and Carasco J.F. (1993). Isoenzyme comparison of *Trypanozoon* isolates from two sleeping sickness areas of south-eastern Uganda. *Acta Tropica* **55**:97-115.

Evanno G., Regnaut S. and Goudet J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620.

Excoffier L. and Lischer H.E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resourses* **10**: 564-567.

Fèvre E. (2001). More thoughts on the control of trypanosomes in cattle. *Trends in Parasitology* **17**:412-413.

Gibson W. and Stevens J. (1999). Genetic exchange in the *Trypanosomatidae*. *Advances in Parasitology* **43**: 1–46

Gibson W.C., de C Marshall T.F. and Godfrey D.G (1980). Numerical analysis of enzyme polymorphism: a new approach to the epidemiology and taxonomy of trypanosomes of the subgenus *Trypanozoon*. *Advances in Parasitology* **18**: 175–246.

Gibson W.C. (1989). Analysis of a genetic cross between *Trypanosoma brucei rhodesiense* and *T. b. brucei*. *Parasitology* **99**: 391–402.

Gibson W.C., Backhouse T. and Griffiths A. (2002). The human serum resistance associated gene is ubiquitous and conserved in *Trypanosoma brucei rhodesiense* throughout East Africa. *Infection Genetics and Evolution* **1**: 207–214.

Gibson W.C. and Gashumba J.K. (1983). Isoenzyme characterization of some *Trypanozoon* stocks from a recent trypanosomiasis epidemic in Uganda." *Transactions of the Royal Society of Tropical Medicine and Hygiene* **77**: 114-118.

Gibson W.C. and Wellde B.T. (1985). Characterization of *Trypanozoon* stocks from the South Nyanza sleeping sickness focus in Western Kenya. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **79**: 671–676.

Godfrey D.G., Baker R.D., Rickman L.R. and Mehlitz D. (1990). The distribution, relationships and identification of enzymic variants within the subgenus *Trypanozoon*. *Advances in Parasitology* **29**:1-74.

Goodhead I., Capewell P., Bailey J.W., Beament T., Chance M., Kay S., Forrester S., MacLeod A., Taylor M., Noyes H. and Hall N. (2013). Whole-genome sequencing of *Trypanosoma brucei* reveals introgression between subspecies that is associated with virulence." *MBio* pii: e00197-13.

Hide G., Tait A., Maudlin I. and Welburn S.C (1994). Epidemiological relationships of *Trypanosoma brucei* stocks from South East Uganda: Evidence for different population structures in human and non-human trypanosomes. *Parasitology* **109**: 95– 111.

Hoare C.A. (1972). The trypanosomes of mammals. Oxford, Blackwell Scientific Publications, pp. 1– 749.

Jamonneau V., Barnabé C., Koffi M., Sané B., Cuny G. and Solano P. (2003). Identification of *Trypanosoma brucei* circulating in a sleeping sickness focus in Cote d'Ivoire: assessment of genotype selection by the isolation method. Infection *Genetics and Evolution* **3**: 143-149.

Jombart T. (2008). Adegenet: an R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**:1403-1405.

Jombart T., Devillard S. and Balloux F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**: 94

Komba E.K., Kibona S.N., Ambwene A.K., Stevens J.R. and Gibson W.C. (1997). Genetic diversity among *Trypanosoma brucei rhodesiense* isolates from Tanzania. *Parasitology* **115:** 571–579

Lwala Hospital Medical Reports 2010–2012.

Maclean L., Odiit M., Macleod A., Morrison L., Sweeney L., Cooper A., Kennedy P.G. and Sternberg J.M. (2007). Spatially and genetically distinct African Trypanosome virulence variants defined by host interferon-γ response." *Journal of Infectious Diseases* **196**: 1620-1628.

MacLeod A., Tait A. and Turner C.M.R (2001). The population genetics of *Trypanosoma brucei* and the origin of human infectivity. *Philosophical Transactions of the Royal Society of London* **356**: 1035-1044.

MacLeod A., Tweedie A., Welburn S.C., Maudlin I., Turner C.M.R. and Tait A. (2000). Minisatellite marker analysis of *Trypanosoma brucei*: Reconciliation of clonal, panmictic, and epidemic population genetic structures. *Proceedings in the National Academy of Sciences* **97**:13442–13447.

Mathieu-Daude F., Stevens J., Welsh J., Tibayrenc M. and McClelland M. (1995). Genetic diversity and population structure of *Trypanosoma brucei*: clonality versus sexuality. *Molecular Biochemistry and Parasitology* **72**: 89–101

Matschiner M.and Salzburger W. (2009). TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* **25**:1982-1983.

Njiru Z.K., Constantine C.C., Guya S., Crowther J., Kiragu J.M., Thompson R.C.and Dávila A.M (2005). The use of ITS1 rDNA PCR in detecting pathogenic African trypanosomes. *Parasitological Research* **95**: 186–192.

Onyango R.J., Van Hoeve K. and De Raadt P. (1966). The epidemiology of *Trypanosoma rhodesiense* sleeping sickness in Alego location, Central Nyanza, Kenya. I. Evidence that cattle may act as reservoir hosts of trypanosomes infective to man. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **60**:175-182.

Ouma J.O., Beadell J.S., Hyseni C., Okedi L.M., Krafsur E.S., Aksoy S. and Caccone A. (2011). Genetic diversity and population structure of *Glossina pallidipes* in Uganda and western Kenya. *Parasite and Vectors* **204**:122

Picozzi K., Fèvre E.M., Odiit M., Carrington M., Eisler M.C., Maudlin I. and Welburn S.C. (2005). Sleeping sickness in Uganda: a thin line between two fatal diseases. *British Medical Journal* **331**:1238-1241.

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

R Core Development Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Radwanska M., Chamekh M., Vanhamme L., Claes F., Magez S., Magnus E., de Baetselier P., Büscher P. and Pays E. (2002). The serum resistance-associated gene as a diagnostic tool for the detection of *Trypanosoma brucei rhodesiense*. *American Journal of Tropical Medicine and Hygiene* **67**: 684-690.

Rousset F. (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resourses* **8:** 103-106.

Rutto J.J., Osano O., Thuranira E.G., Kurgat R.K. and Odenyo V.A. (2013). Socio-Economic and Cultural Determinants of Human African Trypanosomiasis at the Kenya–Uganda Transboundary." *PLoS Neglected Tropical Diseases* **7**: e2186.

Sistrom M., Echodu R., Hyseni C., Enyaru J., Aksoy A. and Caccone A. (2013). Taking advantage of genomic data to develop reliable microsatellite loci in *Trypanosoma brucei. Molecular Ecology Resources* **13**: 341-343.

Stevens J.R., and Godfrey D.G. (1992). Numerical taxonomy of *Trypanozoon* based on polymorphisms in a reduced range of enzymes." *Parasitology* **104**: 75-86.

Tait A. (1980). Evidence for diploidy and mating in trypanosomes. *Nature* **287**: 536–538

Tait A., Barry J.D., Wink R., Sanderson A. and Crowe J.S. (1985). Enzyme variation in *T. brucei* species. II. Evidence for *T. b. rhodesiense.*being a set of variants of *T. b. brucei. Parasitology* **90**: 89–100

Waiswa C., Picozzi K., Katunguka-Rwakishaya E., Olaho-Mukani W., Musoke R.A. and Welburn S.C. (2006). *Glossina fuscipes fuscipes* in the trypanosomiasis endemic areas of south eastern Uganda: Apparent density, trypanosome infection rates and host feeding preferences." *Acta Tropica* **99**: 23-29.

Waples R.S. (2006). A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservational Genetics* **7**:167-184

Waples R.S. and Do C. (2008). ldne: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecological Resourses* **8**: 753-756

Welburn S.C., Picozzi K., Fèvre E.M., Coleman P.G., Odiit M., Carrington M. and Maudlin I. (2001). Identification of human-infective trypanosomes in animal reservoir of sleeping sickness in Uganda by means of serum-resistance-associated (SRA) gene *Lancet* **358**: 2017-2019.

World Health Organization. "First WHO report on neglected tropical diseases: working to overcome the global impact of neglected tropical diseases." Geneva: WHO, 2010.

Wright S. (1978). *Evolution and the Genetics of Populations*, Vol. **4**. University of Chicago Press, Chicago.

Xong H.V., Vanhamme L., Chamekh M., Chimfwembe C.E., Van Den Abbeele J., Pays A., Van Meirvenne N., Hamers R., De Baetselier P. and Pays E. (1998) "A VSG Expression Site–Associated Gene Confers Resistance to Human Serum in *Trypanosoma rhodesiense*. *Cell* **95**: 839–846.

CHAPTER FOUR

MOLECULAR RESPONSES IN *G. pallidipes* to CHALLENGE WITH *T. b. brucei*

4.1 Abstract

African trypanosomiases are diseases of medical and economic consequences for sub Saharan Africa. The diseases are caused by trypanosome parasites cyclically transmitted by tsetse flies. *Glossina pallidipes* is an important vector species in Kenya, responsible for transmission of both human and animal diseases. Tsetse flies display strong resistance to infection with parasites, which are typically eliminated early in the infection process in the midgut. To determine key molecular responses of *G. pallidipes* to infection with *Trypanosoma brucei brucei*, we analyzed transcriptomes of midgut and carcasses from newly eclosed teneral females 24 and 48 hours post parasite challenge (hpc) using the illumina RNA-Seq platform. Most of the differentially expressed transcripts at 24 hpc were associated with lipid remodeling/lipogenesis, proteolysis, urea cycle, carnitine trafficking, collagen metabolism, apoptosis, and cell growth/differentiation. Transcripts associated with 48 hpc included those linked to embryonic growth and development, muscle/motility, suppression of tumor, serine endopeptidase and related proteosomal degradation of target protein, enhanced translation of mRNA and neuronal development. There was pronounced expression of immune responsive transcripts 48 relative to 24 hpc, indicative of gradual maturity of immune responses in the fly or institution of vector-parasite endemic stability in the guts to facilitate the establishment of infection. Overall there was a systematic suppression of immunity in the *G. pallidipes* midgut in the initial phase of *T. b. brucei* challenge, which potentially facilitated initial establishment of the infection. Gradual and sequential immunological responses subsequently emerge contingent with the durations of challenge

4.2 Introduction

African trypanosomiasis constitutes one of the most neglected diseases affecting humans and their livestock with devastating health and economic consequences in Africa (Brun *et al.,* 2010; Hotez *et al.,* 2009). There are two forms of the human disease (Human African Trypanosomiasis, HAT), also known as Sleeping Sickness. The chronic form in West and Central Africa is caused by *Trypanosoma brucei gambiense*, while the acute form in East and Southern Africa is caused by *Trypanosoma brucei rhodesiense*. The animal disease (Animal African Trypanosomiasis,

AAT), also known as Nagana, is caused by *Trypanosoma brucei brucei,* and closely related *Trypanosoma vivax,* and *Trypanosoma congolense* parasites. All trypanosomes are transmitted to the mammalian host through the tsetse fly (*Diptera: Glossinidia*) bite. The HAT and AAT remain as major public health threats in most of Africa due to the long adult life of tsetse, and exclusive haematophagy of both sexes. Brought under control in the 1960s, HAT re-emerged and resurged to epidemic proportions by the end of the 20th century due to decreased control and surveillance activities. Concerted and collaborative control efforts over the last decade reversed the epidemic trend, reducing the cases to just 6,228 in 2013 (Franco*, et al.,* 2014). Informed by the progress in HAT control, the WHO Strategic and Technical Advisory Group for Neglected Tropical Diseases (NTD) declared a target to eliminate *gambiense* HAT as a public health problem by 2020 and zero incidence of HAT by 2030 (WHO, 2012). Kenya is under continuous risk of re-emergence of *T. b. rhodesiense* associated HAT due to the presence of parasites in wild game which are not affected by the parasite and serve as reservoirs, but AAT is rampant in livestock inhabiting tsetse-infested areas.

There are no mammalian vaccines against HAT. There are few available drugs for chemotheraphy, but treatment is expensive with long administration regiment and has adverse effects (Anene *et al.,* 2001; Ferve *et al*., 2006; Baker *et al.,* 2013). Chemotherapy is problematic due to the widespread and increasing resistance detected in trypanosomes to existing drugs (Barrett *et al.,* 2011), high cost of treatment and sporadic availability of drugs in areas with high fly challenge (Jordan, 1986). Tsetse population control efforts therefore constitute the corner stone in initial suppression and ultimate eradication of the disease. Suppression of tsetse populations has been accomplished using insecticide-based technologies **(**Allsopp, 1984; Vale *et al.,* 1999; Vale, 1993). In addition, tsetse eradication campaigns require, integrate sterile insect release (SIT) approach to eliminate residual tsetse population, as demonstrated in Zanzibar. However, the irradiated male flies released in SIT applications are still capable of transmitting pathogenic trypanosomes, a challenge that can be surmounted by development of tsetse strains refractory to trypanosome infections (Moloo *et al.,* 1988; Medlock *et al.,* 2013). The generation of parasite-resistant strains requires better understanding of the molecular interactions that lead to establishment or elimination of parasite infections in the midgut.

Tsetse (Genus: *Glossinidiae*), consist of three species complexes (*Morsitans, Palpalis* and *Fusca* subgenera) with differential vector competence (Moloo *et al.,* 1988). Flies in the *Palpali*s

subgroup are most refractory to trypanosome infection (Nayduch and Aksoy 2007), while those in the *Morsitans* subgroup are more susceptible (Moloo *et al.*, 1994). Within the *Morsitans* subgroup, there are two closely related species, *Glossina morsitans morsitans* and *G. pallidipes*, which show differential susceptibility with *G. pallidipes* being more refractory to trypanosome infection (Peacock *et al.*, 2012). The *G. pallidipes* is widely distributed in Kenya, and is a vector of AAT, and historically HAT in Kenya (Onyango *et al.,* 1966; Willett, 1955).

The process of trypanosome transmission in tsetse is initiated in the vertebrate host by differentiation of the long slender bloodstream forms (BSF) into non-dividing stumpy forms (Vassella *et al.,* 1997; Seed and Wenck, 2003). Within 24 hrs of blood meal ingestion, the stumpy forms differentiate to midgut-adapted procyclic forms (PCFs) while the long slender BSF are readily lysed (Rico *et al.,* 2013). The PCF express a different and non-varying surface coat composed of procyclin proteins (Roditi *et al.,* 1989). In majority of flies, trypanosomes are eliminated from the gut by about three days post-acquisition, while in a few susceptible individuals parasites survive and establish gut infections (Ellis and Evans, 1977; Gibson and Bailey, 2003). The parasites eventually colonize the proventriculus (cardia) organ in susceptible flies and differentiate into long epimastigotes, translocate into gut lumen and ultimately reach the salivary glands through the mouthparts where they differentiate into metacyclic forms (Aksoy 2003). Mammalian infective metacylics are transmitted to the next host in saliva as the fly takes a bloodmeal and parasites differentiate to BSF that promotes disease (Barry *et al.,* 1998).

Newly eclosed adults (termed teneral) are more susceptible to trypanosome infection than older flies (reviewed in Haines, 2013). The phenomenon has been linked to the immature nature of the peritropic matrix (PM), which is a chitinous barrier that physically protects the midgut epithelium from abrasion by components of the bloodmeal including the pathogens it may contain (Lehane, 1997). Alternatively, the immature nature of the teneral fly immune system may contribute to higher susceptibility. Among the factors that have been shown to influence parasite transmission are the PM barrier integrity (Weiss *et al.,* 2014), midgut proteolytic lectin(s) that may induce transformation of BSF to PCF (Abubakar *et al.,* 2006), antimicrobial peptides (Hao *et al.,* 2001; Boulanger *et al.,* 2002, Hu and Aksoy 2006), peptidoglycan recognition protein LB (Wang *et al.,* 2009 and Weiss *et al.,* 2013), Tsetse EP protein (Haines *et al.,* 2005; Haines *et al.,* 2010) as well as reactive oxygen species (MacLeod *et al.,* 2007).

Much of the molecular and functional work on tsetse-trypanosome dynamics has been

performed with *G. m. morsitans*, while the molecular dynamics underpinning differential susceptibility in the important vector species *G. pallidipes* are poorly understood. The differential resistance to infection between these species is more pronounced in the gut than in the salivary glands, such that all *G. pallidipes* with gut infections give rise to infections in the salivary glands while only a proportion of gut infections mature in the case of *G. m. morsitans* (Peacock *et al*., 2012). The purpose of this study was to determine the molecular responses of *G. pallidipes* to *T. b. brucei* challenge early in the infection process. In the laboratory setting, typically, less than 1% of adult flies (48 hours post eclosion) exposed to a parasite-containing bloodmeal are parasitized, and infection prevalence in natural populations is also very low even in disease endemic areas (Walshe *et al.,* 2011; Aksoy *et al.,* 2003). Given the high resistance older adult flies express, newly eclosed flies less than 24 hours post eclosion were used in the study and evaluation of gut and carcass transcriptional responses to trypanosomes at 24 and 48 hrs post challenge (hpc) when BSF-PCF transformation typically occurs done.

4.3 Materials and methods

4.3.1 Tsetse flies and parasites

Teneral female *G. pallidipes* flies used in the study were obtained from pathogen free colony maintained at insectary of Yale School of Public Health (YSPH), New Haven, CT, USA. The colony was previously established using pupae from International Atomic Energy Agency (IAEA) seibersdorf laboratory, Vienna, Austria. The flies were maintained at 25^{o}C and 50-60% relative humidity, and fed on bovine blood meal every 48hrs using artificial membrane feeding method of Moloo, (1971). The *T. b. brucei* strain RUMP 503 used for tsetse challenges was originally isolated from bovine in Nyanza, Kenya and maintained at YSPH.

4.3.2 Tsetse fly challenges with trypanosomes

The BSF *T. b. brucei* RUMP 503 parasites were expanded in rats diluted in bovine blood at $2x10^6$ trypanosomes/ml and used to challenge flies. A similar group of flies was fed on unchallenged blood meal (controls). Pools of ten midguts (minus proventiculus) or carcasses (plus proventiculus) each were dissected in 1x phosphate-buffered saline (137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 2mM KH_2PO_4) 24 or 48hrs post challenge (hpc) and from their respective controls. The proventiculus was included with the carcass since at 48 hpc, the parasites have not

interacted with the proventiculus are still enclosed within the bloodmeal being digested (Gooding, 1974). Presence of trypanosomes in the infected midguts was confirmed by microscopy method of Herbert and Lumsden, (1976). Challenged midguts and their respective controls were immediately placed in TRIzol®Reagent (Invitrogen, Carlsbad, USA), and subsequently at (-80^{o}C) until when required. Overall, eight samples were collected 24 or 48 hpc (i.e. single replicates of challenged and control midgut or carcase 24 or 48 hpc)

4.3.3 Isolation and validation RNA from *G. pallidipes* for RNA-Sequencing

Total RNA was isolated from each of the frozen (-80^{o}C) midgut and carcass samples using TRIzol reagent (Invitrogen, Carlsbad, USA) following the manufacturer's protocol. Potential genomic DNA (gDNA) contaminants in the RNA extracts were digested using TURBO DNase™ (Ambion life technologies USA) following manufacturer's instructions. Removal of the gDNA contaminants was confirmed via PCR of the digest using *beta-tubulin* primers (Appendix VII). Briefly, potential gDNA in 1µl digest products were amplified with 1 unit of Taq polymerase (Promega, Madison, MO) in the buffer (provided by the manufacturer which contained $MgCl_2$) in the presence of the specific primers for β-tubulin genes (Appendix VII). The first cycle included a 5 minutes hot start at 95^{0}C, 45 seconds at 55°C, and 1 minute at 72°C. Subsequent cycles involved 45 seconds at 94°C, 45 seconds at 55°C, and 1 minute at 72 °C for 30 cycles. The final extension was conducted at 72^{o}C for 10 minutes. Reactions were carried out in Bio-RAD DNA Engine® peltier thermocycler (Bio-RAD, Hercules, CA). The PCR products were loaded onto Ethidium Bromide 3% agarose gels in a TBE buffer (Sambrook *et al.*, 1989). On every gel, a 50 bp DNA ladder molecular weight marker (Life Technologies, Rockville, MD) was run to confirm expected molecular weights of the amplification product contaminants.

4.3.4 RNA Sequencing of the *G. pallidipes* Transcriptome

Total Isolated RNA was reverse transcribed into cDNA for illumina RNA sequencing using *illumina* TruSeq RNA *Sample Preparation Kit (*illumina, Hayward, CA) and sequenced on illumina HiSeq 2000 (paired-end 100bp) (illumina, Hayward, CA) at Washington University Biotechnology Center, city, MO, USA, following a published protocol (Low quality reads, reads with less than 100 base pairs and adapter sequences were removed by Illumina build in sequence

clean up (illumina, Hayward, CA). The resultant raw RNA-Seq reads from each treatment were stored in bam file formats of interleaved fastq formatted sequences for downstream analysis.

4.3.5 Identification and validation differentially expressed *G. pallidipes* transcripts

Quality of the RNA sequence reads in each file was assessed using FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The quality data was used to clean respective reads using SamToFastq software (http://picard.sourceforge.net), where reads with poor quality were discarded. The reads were mapped onto *G. pallidipes* gene-sets from vectorbase (Giraldo-Calderón *et al.,* 2015) and differential expressions profiles of the transcripts established using RNA-Seq analysis module in the CLC genomic workbench var 8.0 (CLC Bio, Aarhus, Denmark) as previously described (Telleria *et al*., 2014). The profiles were normalized using Kal's test (Kal *et al*., 1999). In the analysis, expression profiles were compared between challenged and control midguts or carcass (24 or 48 hpc). A conservative selection regime was adopted to minimize false detections of differential expressions. In that regime, transcripts were considered differentially expressed between treatments only if they had 1) at least two-fold change, 2) false detection rate (FDR) corrected $p < 0.05$, 3) at least five Reads per Kilo base per Million reads (RPKM), a proxy of gene expression (Mortazavi *et al*., 2008) and 4) supported by at least 100 read mappings. Most predominantly expressed transcripts, were considered as those within the 90 percentile in this selection and supported by at least 5000 reads. The fold changes were determined as a ratio of RPKM values between treatments and respective controls, and normalized based on the number of reads from each library. A secondary analysis was conducted to determine enrichment of transcripts within midguts (temporal), and, between midguts and respective carcasses (spatial).

The expression profiles were validated by comparing expression profiles of nine randomly selected transcripts in the transcriptome to profiles of the same genes independently established by real-time qPCR under similar experimental conditions as previously described (Telleria *et al.,* 2014). Briefly, flies were challenged with the parasite, and the RNA obtained and processed for gDNA contaminates in a similar manner as described for those used for preparation of the transcriptome. Total RNA (1ug) was reverse transcribed using by iScript^{TM} cDNA synthesis kit (BIO-RAD, Hercules, USA) according to manufacturer's protocol. Transcript expressions were evaluated by quantitative PCR (qPCR) with gene specific primers (Appendix 7) at 95^{o}C for 5

min (1x), 95ºC for 10 sec, 55ºC for 10 sec, 72ºC for 30 sec (45x) for each gene. The expression levels were analyzed with CFX Manager software version 3.1 (Bio-Rad, Hercules, CA) and normalized to *G. pallidipes* GAPDH (VectorBase Accession number GPAI033271) gene, chosen on basis of its stability analysis for reference genes (Vandesompele *et al.,* 2002). Fold change in transcript expressions were established by comparing levels of expression of the transcripts in challenged (treatment) to those in unchallenged (control) midguts. Pearson correlation analysis was conducted between fold changes obtained from qPCR to those obtained from the RNA-seq data to determine the validity of the transcriptome.

Identification of differentially expressed gene sets were performed using BLASTx (Altschul *et al.,* 1990) to compare nucleotide sequence to the non-redundant protein data base at National Centre for biotechnology Information(NCBI) using Blast2GO^{TM} software (Conesa *et al*., 2005; Götz *et al.,* 2008).

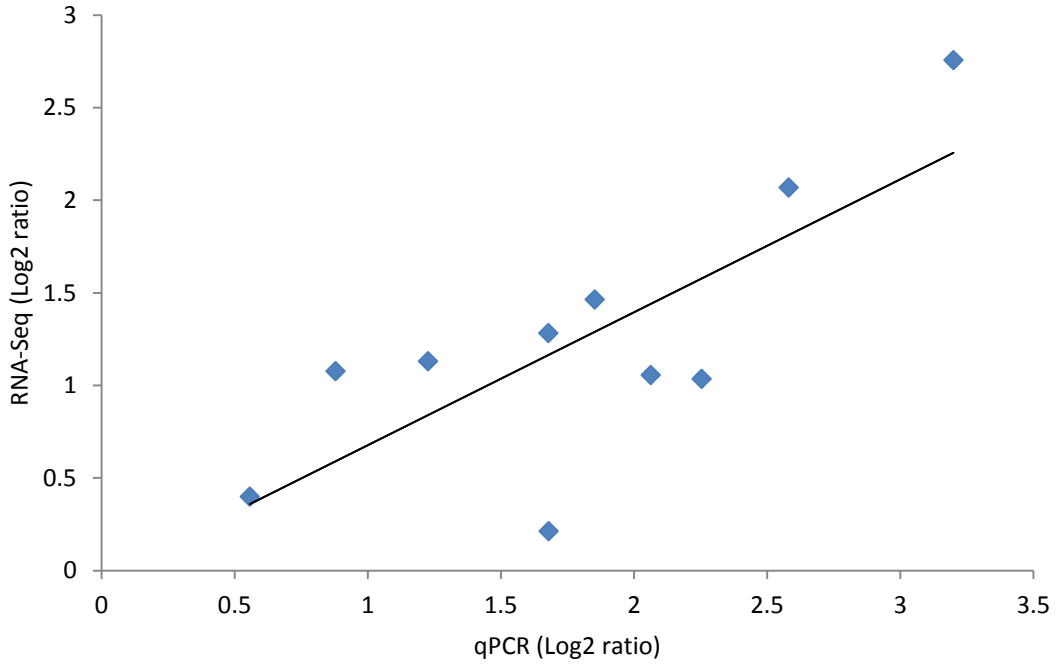
4.4 Results

4.4.1 Global expression profiles of transcripts in *G. pallidipes* challenged with *T. b. brucei*

Processing and mapping results of the RNA-Seq data yielded 43 to 92 million reads in the 24 and 48hpc midgut and carcass libraries (Figure. 7). Most (>99%) of the reads were of high quality and about 64.5 to 75.3% of the reads could be mapped to *G. pallidipes* genes (Figure. 7). At least 89% of the mapped reads were unique to specific genes. To validate the transcriptome data, the expression profiles of ten randomly selected transcripts were obtained using RT-qPCR from RNA extracted from independent biological samples of *G. pallidipes* challenged and control guts, respectively. The comparison revealed a Pearson correlation coefficient (R=0.766) and goodness of fit (R^2= 0.586; Table 4) for the ten genes evaluated, indicative of a valid transcriptome. When challenged guts and carcasses were compared to controls, more DE transcripts were found at 24 hpc than 48 hpc, respectively (Figure 8). When fold changes and significance (p-value) based global dispersion patterns of differentially expressed (DE) transcripts expressed in challenged and control guts and carcasses were analyzed, more transcripts were suppressed than induced at 24 hpc relative to 48 hpc, (Figure 8).

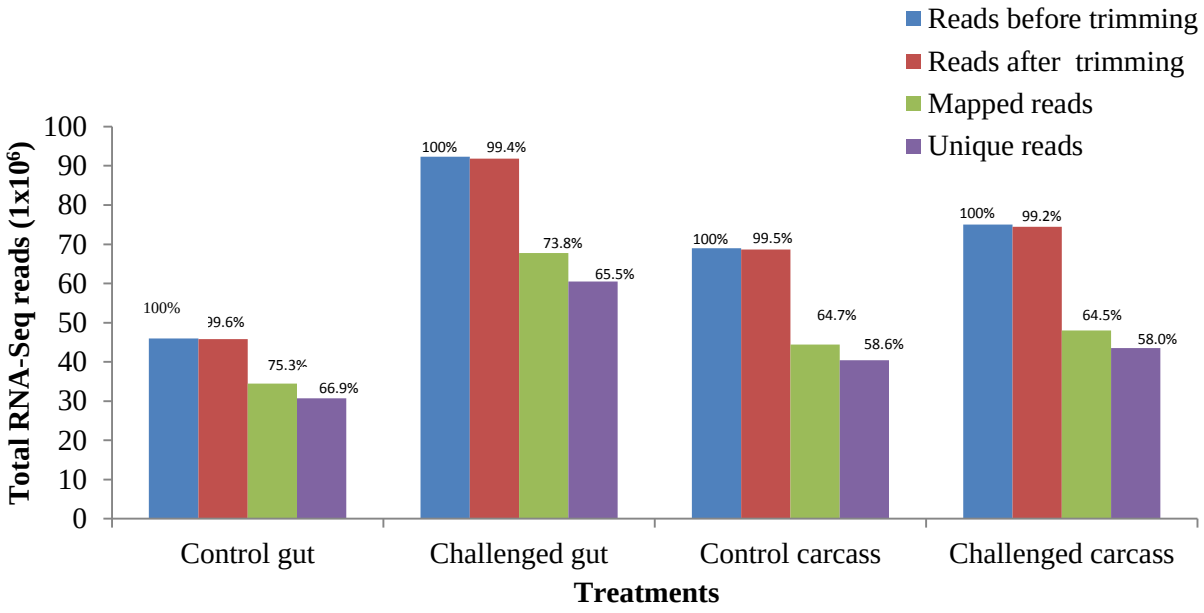
Table 4.Validation of *G. pallidipes* RNA Seq data with qPCR RNA-seq

Gene	Accession Number	qPCR Fold Change	Log2 of qPCR	RNA-seq Fold Change	Log2 of RNA-seq
Tep2	GPAI040205	3.61	1.85	2.76	1.46
Multicopper oxidase	GPAI025756	5.98	2.58	4.2	2.07
Glutamine synthase	GPAI006387	4.18	2.06	2.08	1.06
Heat shock protein 83	GPAI002368	4.77	2.25	2.05	1.04
Pyruvate carboxylase	GPAI003647	1.84	0.88	2.11	1.08
Transferrin	GPA1033230	2.34	1.23	2.19	1.13
Serpin 4	GPAI011576	3.20	1.68	2.44	1.28
Serpin 6	GPAI011576	3.20	1.68	1.16	0.21
PGRP-LB	GPAI047520	1.47	0.56	1.32	0.40
Chitinase	GPAI022616	9.19	3.20	6.76	2.76



Expression values (log2 ratios) for ten genes plotted against qPCR values (log2 ratios) with a Pearson correlation coefficient (R) of 0.766

24hrs (A)



48hrs (B)

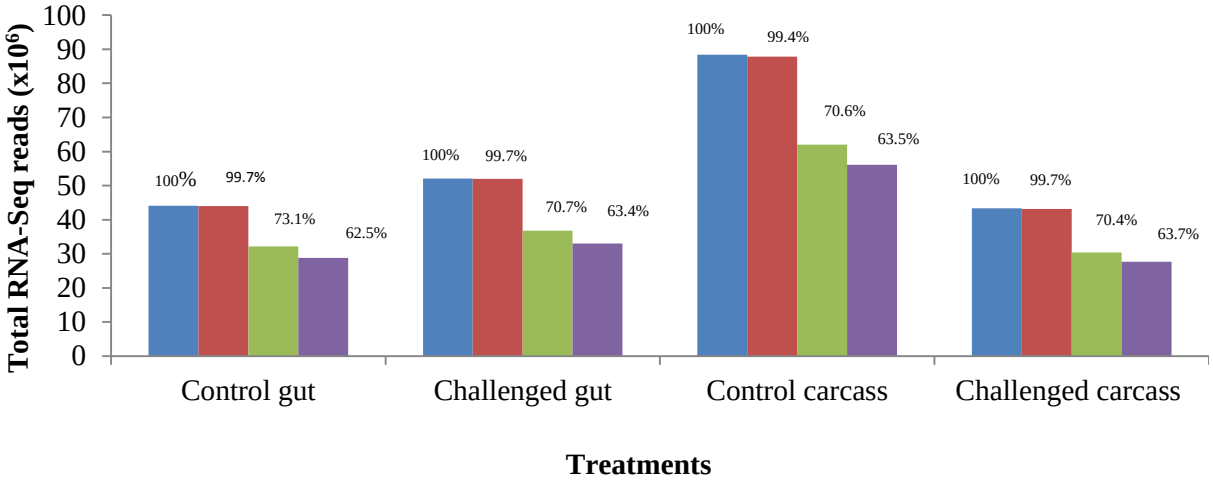


Figure 7. Summary of processing and mapping statistics of RNA-Seq reads from teneral female *G. pallidipes* gut and carcass 24 or 48 hours post challenge with *T. b. brucei*. The blue bars show the percentage raw reads after sequencing, red bars show reads after trimming (after quality control), Green bars reads mapped *G. pallidipes* geneset in vectorbase and the purple bars show reads that mapped uniquely on the geneset.

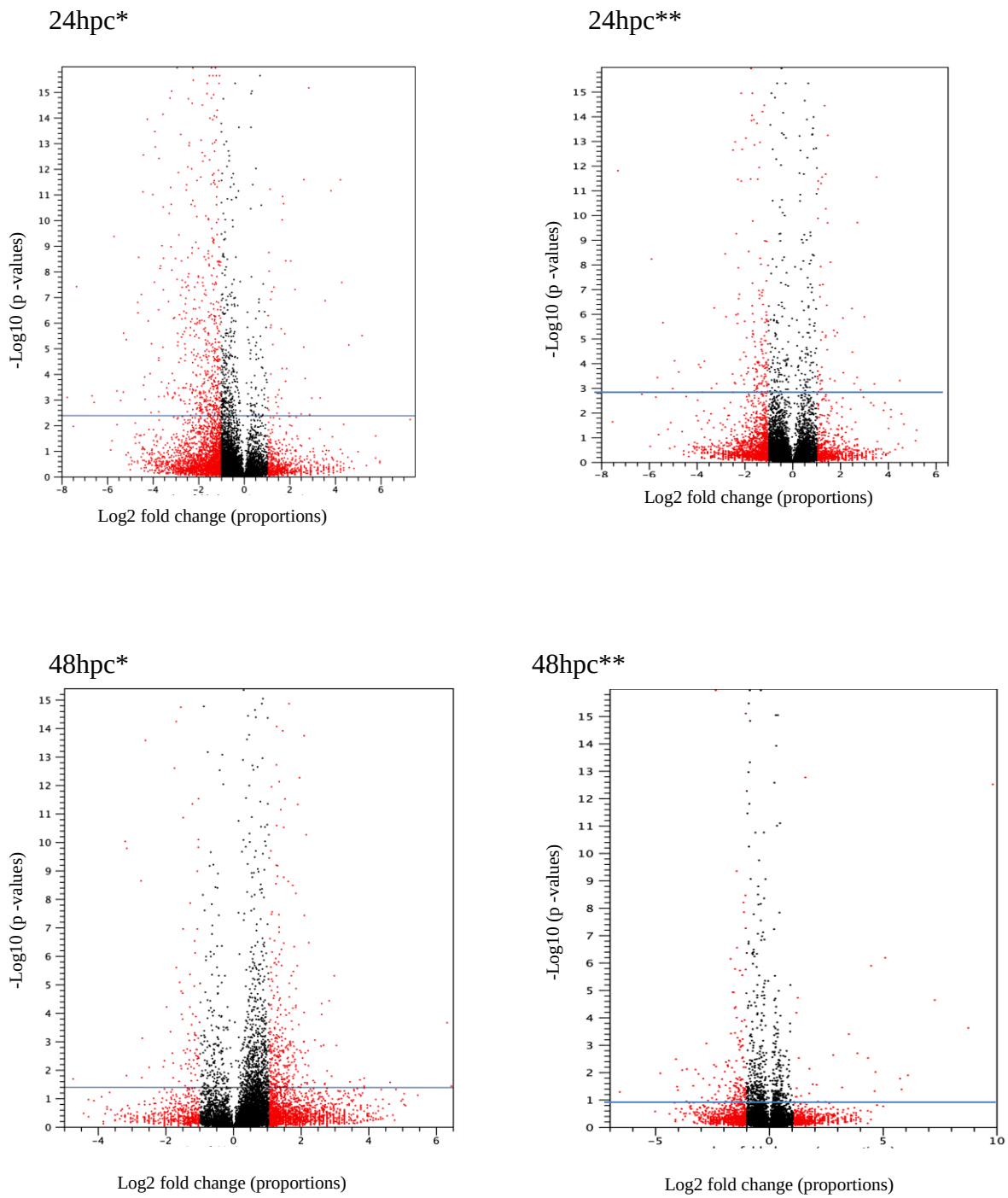
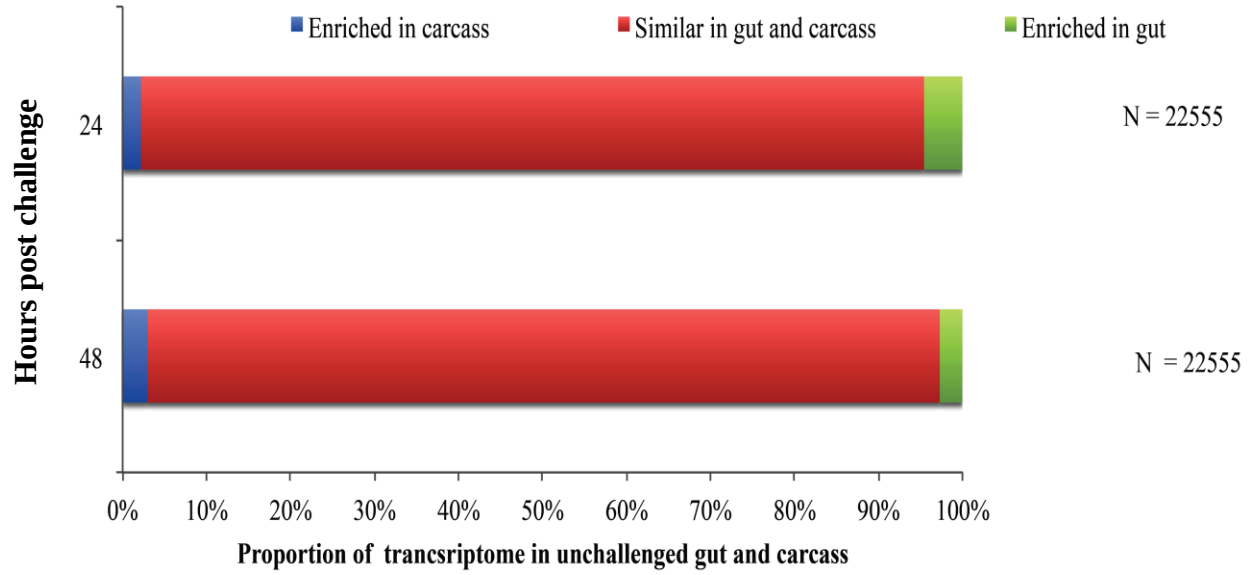


Figure 8. Volcano plots of expressed transcripts in *G. pallidipes* 24 and 48 hrs post challenges with *T. b. brucei*. Red spots indicate differentially expressed transcripts with two (Log2 = 1) or more fold-change (x-axis) and high statistical significance (-log10 of p-value, y-axis). Horizontal blue line shows where Bonferoni FDR corrected P- value = 0.05 with spots above the line having < 0.05 and spots below the line having > 0.05. Spots having a fold-change less than 2 are shown in black. Only the genes denoted with red spots above the blue line are considered significant. hpc= hours post challenge. * =challenged vs unchallenged gut, **= challenged vs unchallenged carcass.

4.4.2 Enriched transcripts and putative protein-protein-interactions in G. *pallidipes* challenged with *T. b. brucei*

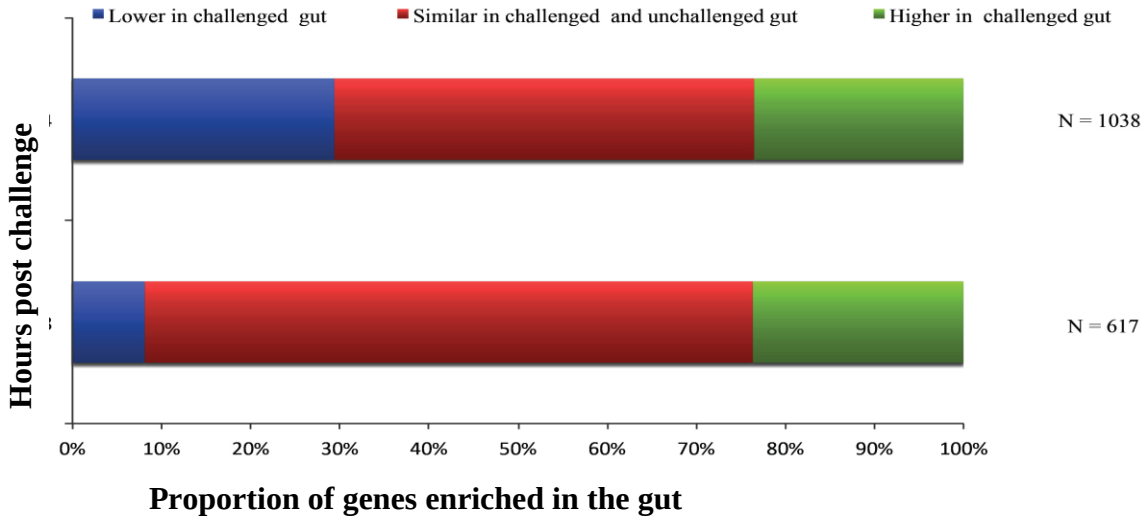
When preferential enrichment and temporal changes of transcripts in gut or carcass libraries under normal or parasite challenge physiological conditions were compared (Fig 9), most (>93%) of the transcripts were conserved between unchallenged gut and carcass. In the gut, 4.6 (1038 genes) or 2.7 (617 genes) % of the transcripts were enriched at 24 or 48 hpc datasets, respectively (Figure 9 A). The gut preferential transcripts were analyzed to understand physiological changes that occurred during the parasite challenge process. Temporal analysis of gut specific transcripts showed that more transcripts were suppressed at 24 hpc than at 48 hpc (Figure 9 B). Only 37 transcripts were shared among the 24 and 48 hpc datasets, and most of these were induced (Figure 9 C).

A



Transcripts in unchallenged guts and carcass

B



Transcripts in enriched guts

C

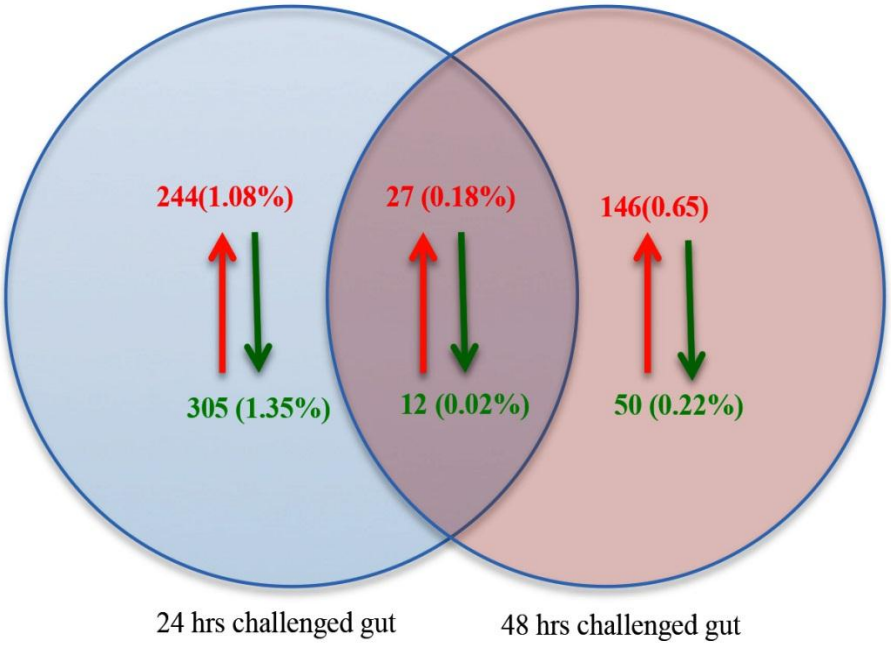


Figure 9 Summary of temporal differential enrichments of transcripts detected in *G. pallidipes* 24 or 48 hrs post challenge with *T. b. brucei*. **A**. Gut and carcass tissue species distribution of transcripts detected. **B.** Expression profiles of gut enriched trancripts in *G. pallidipes* challenged vs unchalleged with *T. b. brucei*. **C.** Temporal expression profile of gut specific *G. pallidipes* genes differentiallly expressed. hpc= hours post challenge

4.4.3 Immune associated transcripts in *G. palidipes* at 24 and 48hpc

*G.pallidipe*s immunity genes homologs described in *G. m. morsitans* and/or *D. melanogaster* were determined using BLAST (Altschul *et al.,* 1990) and 139 of these transcripts were differentially expressed. Among these transcripts 27.3 or 45.3% were induced or suppressed respectively 24 hpc, and 21.6 or 5.8% respectively 48 hpc (Appendix VIII). Predominant profiles included inductions of CD109 antigen and suppression of both trypsin epsilon and serine protease sp24d at 24 and 48 hpc (Appendix VIII). Expressions of Tld domain-containing protein 2, ejaculatory bulb-specific protein 3 and Endocuticle structural protein/glycoprotein were also suppressed 24 hpc. Similarly, expressions of Chymotrypsin-1, AP2-associated protein kinase 1, myosin heavy non-muscle, transferrin and protein croquemort were induced, while those of Lectin subunit alpha were suppressed 48 hpc. Serpin 3 and Toll transcripts associated with Toll immune pathway were induced and suppressed, respectively 24 hpc. Similarly, expressions of Mask and Peptidoglycan recognition protein LC (PGRP-LC) were induced and suppressed, respectively in the Immune Deficiency (IMD) pathway 24 hpc, while Notch was the only induced transcript in the IMD immune pathway 48 hpc.

4.4.4 Heatmap of transcripts 24 and 48 hpc

The heatmap of differentially expressed transcripts at 24 and 48 hpc revealed distinct time-specific gene expression profiles (Table 5). Relative inductions of Trypsin-1, 30S Ribosomal proteins II, chaperone protein, heat shock protein 83 and Glutamine synthetase at 24 hpc relative to 48 hpc were observed. Transcripts relatively induced at 48 hpc relative to 24 hpc included Estradiol 17-beta-dehydrogenase 11, faty acid synthase, Protein cronquemort and several hypothetical proteins. Among the temporal differentially expressed transcripts, only Trypsin-1 and Protein cronquemort were immune associated (Appendix 10)

Table 5. Heatmap of differentially expressed transcripts in *G. pallidipes* guts challenged with *T.b.brucei* for 24 or48 hrs in relation to their respective controls

Vectorbase Accesion No	24 hpc	48 hpc	Best BLAST Hit, None-Redundant (nr) NCBI database	E - Value
GPAI001482			---NA---	-
GPAI002755			---NA---	-
GPAI006903			Lim and calponin homology domains-containing protein 1	2.88257E-22
GPAI008345*			Trypsin-1	1.33501E-36
GPAI009122			---NA---	-
GPAI012773			Transmembrane protein 8b	9.88117E-52
GPAI016439			Apolipoprotein d	2.07837E-31
GPAI019336			---NA---	-
GPAI019955			Protease inhibitor	6.79549E-20
GPAI021719			---NA---	-
GPAI021775			---NA---	-
GPAI025756			l-ascorbate oxidase	4.72399E-51
GPAI033230*			Transferrin	0
GPAI038345			Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	3.13536E-70
GPAI040425*			Endocuticle structural glycoprotein bd-1	3.52053E-22
GPAI046500*			Serine protease sp24d	8.0528E-34
GPAI046602			Calponin homology domain-containing protein ddb_g0272472	0.00451322
GPAI000755			30s ribosomal protein s11	5.44256E-80
GPAI000789			Chaperone protein	0
GPAI002274			---NA---	-
GPAI002368*			Heat shock protein 83	0
GPAI005597			---NA---	-
GPAI006387			Glutamine synthetase mitochondrial	0
GPAI008893			---NA---	-
GPAI009123			---NA---	-
GPAI012803			Acyl- delta desaturase	4.1518E-107
GPAI014739			Filamin-a	0
GPAI015300			U4 small nuclear ribonucleoprotein 27 kda protein	1.27536E-24
GPAI021337			Estradiol 17-beta-dehydrogenase 11	1.79437E-71
GPAI022076			Proline-rich nuclear receptor coactivator 2	0.000284403
GPAI022900			---NA---	-
GPAI025714			40s ribosomal protein s23	2.69515E-14
GPAI028922			Protein fam13a	1.7218E-34
GPAI031242			---NA---	-
GPAI034237			Calcium calmodulin-dependent 3 -cyclic nucleotide phosphodiesterase 1c	2.6531E-172
GPAI035801			Protein lsm14 homolog a	3.4794E-48
GPAI039167			Fatty acid synthase	1.27508E-63
GPAI039186			Outer dense fiber protein 3-like protein 2	1.58607E-17
GPAI040411*			Myosin heavy non-muscle	0
GPAI040413			Myosin heavy non-muscle	6.99922E-47
GPAI041696			Spectrin beta non-erythrocytic 5	0
GPAI042995*			Ras-responsive element-binding protein 1	3.60274E-19
GPAI043434			Breast cancer anti-estrogen resistance protein 1	2.18348E-18
GPAI043644			---NA---	-
GPAI045204*			Protein croquemort	1.58E-125

***** Immunity associated genes
Blue: Non secreted proteins
Orange: Secreted proteins

4.5 Discussion

This study reports on the molecular responses of the tsetse species, *G. pallidipes* to the parasite *T. b. brucei* at a critical moment in infection process when most trypanosomes typically die, with few surviving to establish permanent infections in the fly midgut (Gibson and Bailey, 2003). Previous studies in *G. p. palpalis* established that teneral adults show greater susceptibility to *Trypanosome brucei gambiense* infection than older adults, a phenomenon that has also been observed in teneral *G. m. morsitans* infection by *T. b. brucei* (Reviewed in Haines, 2013). Infection susceptibility of teneral flies however varied with trypanosome species and parasite- vector pairings where teneral *G. m. morsitans* is more susceptible than *G. p. palpalis* to infection with *T. b. brucei* (Walshe *et al.,* 2011). Within the *Morsitans* group, *G. pallidipes* were shown to be generally more refractory to trypanosome infection than other species (Nayduch and Aksoy, 2007). Unlike in adult flies, susceptibility to trypanosome infection in teneral flies was not affected by the number of parasites ingested (*Walshe et al.,* 2011).

Analysis was performed with newly eclosed teneral adult *G. pallidipes* flies. Host responses from two different fly compartments were analyzed: the gut where epithelial immune responses can limit parasite survival and the carcass that is involved in systemic immunity and metabolic responses. These datasets were used to identify host responses that are preferentially enriched in the gut.

Induced transcripts, such serine protease sp24d involved in toll signalling pathway may suggest a process of systematic suppression of host immunity in the gut (Mika *et al.,* 2012a; Mika *et al.,* 2012b) and immune associated cytoskeletal reorganization (Baton and Ranford-Cartwright, 2005; Vlachou and Kafatos, 2005; Vlachou *et al.,* 2005). Suppression of oxidation-reduction processes at 24 hpc was unexpected. In another study in teneral *G. p. gambiensis* challenged with *T. b. gambiense,* proteome analysis revealed induction of these processes at 72 hpc (Geiger *et al.,* 2015), and has been suggested to reflect oxidative stress from heme present in the bloodmeal and/or invading parasites (Hu and Aksoy, 2006; Haines *et al.,* 2010). Oxidative stress responses have been reported in a variety of insects upon pathogen challenge, including tsetse (Hao *et al.,* 2003; Nappi *et al.,* 2000; Luckhart *et al.,* 1998), suggesting that suppression of these responses in this study may inevitably promote parasite survival or differentiation early in the gut infection process. A study that looked at 48 and 72 hpc responses of adult *G. m. morsitans* using the same parasite strain we studies here also noted induction of ROS responses

in the gut as part of the immune arsenal that may result in parasite refractoriness of tsetse. Results from this study may reflect the immature nature of the gut immune responses in teneral tsetse. Alternatively, lack of ROS responses at earlier times in the infection process (24 hpc) may reflect manipulation of host immunity by invading parasites to facilitate their differentiation/establishment in the teneral state

Studies that analysed gut responses three to six days post parasite challenge (Ellis and Evans, 1977; Gibson and Bailey, 2003) and from older adults reported more robust antimicrobial peptide and IMD pathway associated transcript expression (Hao *et al.,* 2001; Hao *et al.,* 2003; Hu and Aksoy, 2006). In a study, where temporal immune responses of older adults to *T. brucei* challenge were investigated, host responses were more strongly detected at 72 hpc relative to 48 hpc suggesting that PCF parasites may not be recognized by host epithelial responses early during the infection process in mature adults, until they bypass the PM barrier and invade the ectoperitrophic space of gut (Aksoy *et al.,* 2016). Strong immune responses, including induction of AMPs, such as Attacin and Cecropin, were also reported from older immunologically mature adults (Hu and Aksoy, 2006) and the role of these responses in parasite clearance has been functionally shown through RNAi silencing studies.

Mammalian parasite surface coat VSG proteins have been shown to modify host transcriptional responses transiently to reduce PM barrier integrity in adult flies (Aksoy *et al.,* 2016). The integrity of the PM has been shown to be an important barrier that limits parasite infections in older adults (Weiss *et al.,* 2014). In this analysis, the organ that produces PM was present in the carcass samples, and we similarly noted a reduction of PM associated peritrophin genes (Pro1, Pro2, and Pro3) during 48 hpc in the carcass samples. The immature nature of the PM in teneral adults coupled with the deliberate suppression of these responses by the BSF VSG proteins may facilitate the greater infection prevalence typically noted at the teneral state. Besides the major peritrophins, modifications with chitin metabolism was also noted, the main structure of the PM backbone. Lack of mature immune responses in the teneral state can further facility establishment of these infections in the teneral state while they would be more effectively cleared from the mature adult gut.

Follow up studies with a spectrum of pathogen and trypanosome developmental forms (BSF/ PCF) may provide insights on whether these response patterns are pathogen specific and potential role of the BSF – PCF transformations on the pattern. Additionally, the role of the

tsetse microbiome in modulating these responses also merit investigation since factors influencing establishment of the parasite in the fly midgut are not limited to tsetse-trypanosome interactions since *Wigglesworthia glossinidia* and *Sodalis glossinidius* have been established to interfere with the infection (Geiger *et al.*, 2007; Farikou *et al.*, 2010)

4.6 References

Abubakar L.U., Bulimo W.D., Mulaa F.J. and Osir E.O (2006). Molecular characterization of a tsetse fly midgut proteolytic lectin that mediates differentiation of African trypanosomes. *Insect Biochemisrty and Molecular Biology*. **36**:344-52.

Aksoy S. (2003). Control of tsetse flies and trypanosomes using molecular genetics. *Veterinary Parasitology*. **115**:125-45.

Allsopp R. (1984). Control of tsetse flies (*Diptera Glossinidae*) using insecticides a review and future prospects. *Bulletin of Entomological Research* **74**:1-23.

Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**:403-410.

Anene B.M., Onah, D.N. and Nawa, Y. (2001). Drug resistance in pathogenic African trypanosomes: what hopes for the future? *Veterinary. Parasitology.* 96: 83–100.

Baker N., de Koning H.P., Mäser P. and Horn (2013). Drug resistance in African trypanosomiasis: the melarsoprol and pentamidine story. *Trends in Parasitology*. **29**:110-8

Barrett M.P., Vincent I.M., Burchmore R.J., Kazibwe A.J. and Matovu E. (2011). Drug resistance in human African trypanosomiasis *Future Microbiology* **6:**1037-1047.

Barry J.D., Graham S.V., Fotheringham M., Graham V.S., Kobryn K. and Wymer B. (1998). VSG gene control and infectivity strategy of metacyclic stage *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **91**:93-105.

Baton L.A. and Ranford-Cartwright L.C. (2005). How do malaria ookinetes cross the mosquito midgut wall? *Trends in Parasitology*. **21**:22–28.

Boulanger N., Brun R., Ehret-Sabatier L., Kunz C. and Bulet P. (2002). Immunopeptides in the defense reactions of *Glossina morsitans* to bacterial and *Trypanosoma brucei brucei* infections. *Insect Biochemistry and Molecular Biology* **32**:369-375.

Brun R., Blum J., Chappuis F. and Burri C (2010). Human African trypanosomiasis *Lancet* **375**:148-159.

Conesa A., Götz S., García-Gómez J.M., Terol J., Talón M. and Robles M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**:3674-3676.

Ellis D.S. and Evans D.A. (1977). Passage of *Trypanosoma brucei rhodesiense* through peritrophic membrane of *Glossina morsitans morsitans*. *Nature* **267**:834-835.

Fèvre E.M., Picozzi K., Jannin J., Welburn S.C. and Maudlin I. (2006). Human African trypanosomiasis: Epidemiology and control. *Advances in Parasitology*. **61**:167-221

Franc N.C., Dimarcq J.L., Lagueux M., Hoffmann J. and Ezekowitz R.A. (1996). Croquemort, a novel Drosophila hemocyte/macrophage receptor that recognizes apoptotic cells. *Immunity* **4**:431-443.

Geiger A., Hamidou Soumana I., Tchicaya B., Rofidal V., Decourcelle M., Santoni V. and Hem S. (2015). Differential expression of midgut proteins in *Trypanosoma brucei gambiense*-stimulated vs. non-stimulated *Glossina palpalis gambiensis* flies. *Frontiers in Microbiology* **6**:444.

Geiger A., Ravel S., Mateille T., Janelle J., Patrel D., Cuny G. and Frutos R. (2007). Vector competence of *Glossina palpalis gambiensis* for *Trypanosoma brucei s.l.* and genetic diversity of the symbiont *Sodalis glossinidius*. *Molecular Biology and Evolution* **1**:102-109

Gibson W. and Bailey M. (2003). The development of *Trypanosoma brucei* within the tsetse fly midgut observed using green fluorescent trypanosomes. *Kinetoplastid Biology and Disease* **2**:1.

Giraldo-Calderón G.I., Emrich S.J., MacCallum R.M., Maslen G., Dialynas E., Topalis .P, Ho N., Gesing S., Madey G., Collins F.H. and Lawson D. (2015). VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Research.* **4**:D707-13.

Götz S., García-Gómez J.M., Terol J., Williams T.D., Nagaraj S.H., Nueda M.J., Robles M., Talón M., Dopazo J. and Conesa A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*earch **36**:3420-3435.

Haines L.R. (2013). Examining the tsetse teneral phenomenon and permissiveness to trypanosome infection. *Frontiers in Cellular and Infection Microbiology* **19:**3:84.

Haines L.R., Jackson A.M., Lehane M.J., Thomas J.M., Yamaguchi A.Y., Haddow J.D and Pearson T.W (2005). Increased expression of unusual EP repeat-containing proteins in the

midgut of the tsetse fly (*Glossina*) after bacterial challenge. *Insect Biochemistry and Molecular Biology* 35:413-423.

Haines L.R., Lehane S.M., Pearson T.W. and Lehane M.J (2010). Tsetse EP protein protects the fly midgut from trypanosome establishment. *PLoS Pathogen*.**6**:e1000793

Hao Z., Kasumba I. and Aksoy S. (2003). Proventriculus (cardia) plays a crucial role in immunity in tsetse fly (Diptera : Glossinidiae) *Insect Biochemistry and Molecular Biology* **33**:1155-1164.

Hao Z., Kasumba I., Lehane M.J., Gibson W.C., Kwon J. and Aksoy S. (2001) Tsetse immune responses and trypanosome transmission: Implications for the development of tsetse-based strategies to reduce trypanosomiasis *Proceedings in National Academy of Sciences* **98**: 12648-12653.

Hotez P.J. and Kamath A. (2009). Neglected tropical diseases in sub-saharan Africa: review of their prevalence, distribution, and disease burden. *PLoS Neglected Tropical Diseases* **3:**e412.

Hu C. and Aksoy S. (2006). Innate immune responses regulate trypanosome parasite infection of the tsetse fly *Glossina morsitans morsitans*. *Molecular Microbiology* **60:**1194-1204.

Hu Y. and Aksoy S. (2005). An antimicrobial peptide with trypanocidal activity characterized from *Glossina morsitans morsitans*. *Insect Biochemistry and Molecular Biology* **35:**105-115.

Jordan A.M. 1986. Trypanosomiasis control and African Rural Development. *Harlow and New York: Longman Group Limited*.

Kal A.J., van Zonneveld A.J., Benes V., van den Berg M., Koerkamp M.G., Albermann K., Strack N., Ruijter J.M., Richter A., Dujon B., Ansorge W. and Tabak H.F. (1999). Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources. *Molecular Biology of the Cell*. **10**:1859-1872.

Kal A.J., van Zonneveld A.J., Benes V., van den Berg M., Koerkamp M.G., Albermann K. Strack N., Ruijter J.M., Richter A., Dujon B., Ansorge W. and Tabak H.F. (1999). Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources. *Molecular Biology of the Cell* **10**:1859-1872.

Lehane M.J. (1997). Peritrophic matrix structure and function. Annual Review of Entomology 42:525-550.

MacLeod E,T., Maudlin I., Darby A.C., Welburn S.C. (2007). Antioxidant promote establishment of trypanosome infections in tsetse. *Parasitology*.**134**:827-831

Mika A., Reynolds S.L., Pickering D., McMillan D., Sriprakash K.S., Kemp D.J. and Fischer K. (2012b). Complement inhibitors from scabies mites promote streptococcal growth—a novel mechanism in infected epidermis *PLoS Neglected Tropical Diseases* **7**:e1563

Mika A., Reynolds SL., Mohlin F.C., Willis C., Swe P.M., Pickering D.A., Halilovic V., Wijeyewickrema L.C., Pike R.N., Blom A.M., Kemp D.J.and Fischer K (2012a). Novel scabies mite serpins inhibit the three pathways of the human complement system. *PLoS One*. **7** e40489

Moloo S.K. (1971). An artificial feeding technique for *Glossina*. *Parasitology*. **63**:507-512.

Moloo S.K. and Kutuza S.B. (1988). Comparative study on the infection rates of different laboratory strains of *Glossina* species by *Trypanosoma congolense*. *Medical and Veterinary Entomology* **2**:253-257.

Moloo S.K., Zweygarth E and Sabwa C.L (1994). Comparative study on the susceptibility of different laboratory strains of *Glossina* species to *Trypanosoma simiae. Medical Veterinary Entomology* **8:**225-30

Mortazavi A., Williams B.A., McCue K., Schaeffer L. and Wold B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**:621-628.

Nappi A.J., Vass E., Frey F. and Carton Y(2000). Nitric oxide involvement in *Drosophila* immunity. *Nitric Oxide* **4:**423-430.

Nayduch D. and Aksoy S. (2007). Refractoriness in tsetse flies (*Diptera: Glossinidae*) may be a matter of timing. Journal of *Medical Entomology* **44** :660-665.

Onyango R.J., Van Hoeve K. and De Raadt P. (1966). The epidemiology of *Trypanosoma rhodesiense* sleeping sickness in Alego location, Central Nyanza, Kenya. I. Evidence that cattle may act as reservoir hosts of trypanosomes infective to man. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **60**:175-182

Peacock L., Ferris V., Bailey M., Gibson W. (2012). The influence of sex and fly species on the development of trypanosomes in tsetse flies. *PLoS Neglected Tropical Diseases*. **6**:e1515

Roditi I, Schwarz H, Pearson TW, Beecroft RP, Liu MK, Richardson JP, Bühring HJ, Pleiss J, Bülow R, Williams RO, *et al.* (1989) Procyclin gene expression and the loss of the variant surface glycoprotein during differentiation of *Trypanosoma brucei*. *Journal of Cell Biology* 108:737-746.

Sambrook J., Fritsch E.F. and Maniatis T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Seed J.R and Wenck M.A. (2003). Role of the long slender to short stumpy transition in the life cycle of the African Trypanosomes. *Kinetoplastid Biology and Diseases* **2**:3.

Telleria E.L., Benoit J.B., Zhao X., Savage A.F., Regmi S., Alves e Silva T.L., O'Neill M. and Aksoy S (2014). Insights into the Trypanosome-Host Interactions Revealed through Transcriptomic Analysis of Parasitized Tsetse Fly Salivary Glands. *PLoS Neglected Tropical Diseases* **8**: e2649.

Vale G.A. (1993). Development of baits for tsetse flies (*Diptera: Glossinidae*) in Zimbabwe. *Journal of Medical Entomology*. **30**:831-842.

Vale G.A., Mutika, G. and Lovemore D.F. (1999) Insecticide-treated cattle for controlling tsetse flies (*Diptera: Glossinidae*): some questions answered, many posed. *Bulletin of Entomological Research* **89**: 569-578.

Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., De Paepe A. and Speleman F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*. **3**:RESEARCH0034.

Vassella E., Reuner B., Yutzy B. and Boshart, M. (1997). Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *Journal of. Cell Science* **110**:2661-2671.

Vlachou D. and Kafatos F.C. (2005). The complex interplay between mosquito positive and negative regulators of Plasmodium development. *Current Opinion in Microbiology* **8**:415–421.

Vlachou D., Schlegelmilch T., Christophides G.K. and Kafatos F.C. (2005). Functional genomic analysis of midgut epithelial responses in Anopheles during Plasmodium invasion. *Current Biology* **15**:1185–1195.

Walshe D.P., Lehane M.J., Haines L.R (2011). Post eclosion age predicts the prevalence of midgut trypanosome infections in *Glossina*. *PLoS One*. **6**:e26984.

Wang J., Wu Y., Yang G. and Aksoy S. (2009). Interactions between mutualist *Wigglesworthia* and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission *Proceedings in the National Academy of Sciences* **106** :12133-12138.

Weiss B.L., Savage A.F., Griffith B.C., Wu Y. and Aksoy S. (2014). The peritrophic matrix mediates differential infection outcomes in the tsetse fly gut following challenge with commensal, pathogenic, and parasitic microbes. *Journal of Immunology* **193**:773-82.

Weiss B.L., Wang J., Maltz M.A., Wu Y. and Aksoy S. (2013) Trypanosome infection establishment in the tsetse fly gut is influenced by microbiome-regulated host immune barriers. *PLoS Pathogen* 9:e1003318

Willett K.C. (1955). Sleeping sickness in East Africa and its treatment. *East African Medical Journal* **32**: 273-82.

CHAPTER FIVE

FUNCTIONAL ANNOTATION OF RESPONSIVE GENES IN *G. pallidipes* CHALLENGED WITH *T. b. brucei*

5.1 Abstract

Glossina pallidipes is the major vector of human and animal trypanosomiasis in Kenya. The fly is highly refractory to trypanosome infection when compared to other related species. Factors mediating this phenomenon are not well understood Therefore understanding molecular processes and pathways affected by trypanosome challenge can help in development a parasite transmission blocking mechanism and hence disease control. Functional annotation of transcriptional molecular responses elicited in teneral *G. pallidipes* challenged with *Trypanosoma brucei brucei* at 24 and 48 hours post parasite challenge (hpc), was performed using Blast2GO™ software. Differentially expressed transcripts were classified into different functional groups by Gene Ontology annotations (GO) and Interpro databases and enriched pathways identified using R spider analysis. The analysis revealed suppression of trancripts associated with chitin metabolism at 24hpc while at 48hpc they were related to ribosome and serine-type endopeptidase activity. Enriched pathways included those involve in induction of transmembrane transport, induction of metabolic processes and suppression of adult life span at 24hpc. Enriched immune related pathways at 48hpc include those involved in zinc ion binding, axon guidance and immunoglobulin fold pathway. These results indicate that, upon challenge the parasite encounters minimal immunological response in the fly gut at 24hpc but there is gradual maturity of the immune system at 48hpc. Suppression of chitin metabolism at 24 hpc indicated reorganization of the fly cytoskeleton so as to surmount or accommodate the parasite. Therefore early immune responses in parasite challenged flies can be targeted from 48 hpc since most of the affected pathways at 24 hpc are non immune related.

5.2. Introduction

Tsetse flies of *Glossina genus* are vectors of African trypanosomes which causes human and animal trypanosomiasis in Sub Saharan Africa. *Glossina pallidipes* is the most important vector of these diseases in Kenya. For successful transmission to occur the parasite changes its metabolism where it transform from the blood stream form (BSF) to procyclic form (PCF) (Gibson and Bailey, 2003). The parasite then establishes in the midgut and matures in the mouth

parts (proboscis or salivary gland) depending on the parasite species. The parasite undergoes a process of attrition in the midgut during the infection where a large number is cleared by day three post challenge (Van den Abbeele *et al.,* 1999). This resistance to trypanosome infection is highly pronounced in *Glossina pallidipes* than a related species *Glossina morsitans moristans* that has been intensely studied (Peacock *et al.,* 2012).

Studies have shown that the tsetse innate immune system is partly responsible for the parasite transmission resistance phenomenon. The fly innate immune system is stimulated at three days post infection (Hao *et al.,* 2001; 2003; Boulanger *et al.,* 2002) and immunodeficiency signaling pathway effectors have been shown to offer immediate defense against invading microbes in *G. m moristans* (Hu and Aksoy, 2006; Weiss *et al,* 2014). The peritrophic matrix (PM) offers first line defense in tsetse midgut against trypanosome infection (Miller, 1991). Earlier studies in *G. m. moristans* have also shown up regulation of genes involved toll signaling pathway upon trypanosome infection (Lehane *et al.,* 2003).

Transcriptional analysis of *G. m. morsitans* midgut infected with trypanosome between one and seven days post challenge, has shown an up regulation of thiol ester-containing protein families, oxidative stress genes, proteases and inhibitors, chitin binding proteins and genes involved in iron metabolism . Several down regulated genes with unknown function were identified (Lehane *et al.,* 2003).

A proteomic study on *Glossina palpalis gambiensis* midgut challenged with *Trypanosoma brucei gambiense* at three, ten and twenty days post challenge revealed induction of proteins involved in oxidoreduction (Geiger *et al.,* 2015), due to oxidative stress resulting from breakdown of heme in the blood meal (Lehane *et al.,* 2003; MacLeod *et al.,* 2007). Other up regulated proteins included those involved in nucleotide binding, catalytic activity, peptidases, ion binding and structural proteins (Geiger *et al.,* 2015). The enzyme lectizyme involved in trypanosome establishment was found to be over expressed in a similar transcriptional study (Hamidou *et al.,* 2015). Other processes that were induced at three day post challenge include: signal transduction, metabolic processes, system development, transport, response to stimuli and gene expression (Hamidou *et al.,* 2015).

The purpose of this study was to determine the functional role of responsive genes *G. pallidipes* challenged with *T. b. brucei* during the initial 24-48 hrs when BSF-PCF transformation occurs in the midgut. This duration constitute a critical event in the establishment

of the parasite when relatively small proportion of trypanosomes (~10%) will continue to multiply in number (Gibson and Bailey, 2003). This developmental stage of the parasite in the fly midgut is a major bottleneck in trypanosome survivorship, and a critical determinant of competence of the fly to transmit the parasite, and hence an optimum target for transmission blocking initiatives. The availability of the complete and annotated *G. pallidipes* genome data and its associated transcriptome (Giraldo-Calderón. *et al.*, 2015) presents a unique resource and opportunity to investigate molecule(s) and pathways responsible for the differential resistance of *G. pallidipes* to trypanosome infection.

5.3 Materials and Methods

5.3.1 Functional annotations of the differentially expressed transcripts in *G. pallidipes* challenge with *T. brucei brucei*

Functional annotations of the differentially expressed gene sets (identified in chapter 4) were performed using Blast2GO^{TM} software (Conesa *et al*., 2005; Götz *et al.*, 2008). Differentially expressed transcripts were classified into different functional groups by Gene Ontology annotations (GO) and Interpro databases The GO terms related to biological processes, molecular function and cell component were assigned to the transcripts. An E-value of 0.001 was used to perform the BLAST step (BLASTx) while mapping was carried out by default settings and the annotation step performed using an E-value of 0.001. The annotation had a cut-off of 20, GO weight of at least 5 and high scoring pair hit coverage cut-off of 30. GO enriched terms in individual differentially expressed transcripts were established against those in the entire *G. pallidipes* gene sets using fisher's exact test at $p < 0.05$.

5.3.2 Pathway enrichment analysis of transcripts in challenged and unchallenged *G. pallidipes* midguts 24 and 48hpc

Enriched pathways in the challenged and unchallenged midguts were determined through R spider network analysis (Antonov *et al.*, 2010) with *Drosophila melanogaster* genes as a proxy for the *G. pallidipes* genes as previously described (Telleria *et al*., 2014). Hubs within the interactome were classified as minor or major if they had two or more edges respectively. Immune-specific and associated genes in *D. melanogaster* were acquired from Flybase (Marygold *et al*., 2013) as previously described (Telleria *et al.*, 2014), and from *G. m. mostitans* (IGGI, 2014). These genes were used to determine immune associated transcripts among those

differentially expressed, through tBLASTx (Altschul *et al.*, 1990) homology searches. Heat map of consistently and differentially expressed transcripts at 24 and 48 hpc was developed by comparing fold changes of respective RPKM

5.4 Results

5.4.1 Functional annotations of the differentially expressed transcripts in *G. pallidipes*

Blast2GO Fishers Exact test (Conesa and Götz, 2008) revealed enrichment of suppressed transcript categories 24 or 48 hpc, in relation to the entire *G. pallidipes* gene sets. The suppressed transcripts in the 24 hpc dataset were predominantly associated with chitin metabolism, while those at 48 hpc were associated with ribosome and related serine-type endopeptidase activity (Table 6 A). The ProfCom GO and ProfCom InterPro pathway analyses (Antonov *et al*., 2008) revealed enrichment of pathways associated with induction of trans-membrane transport, or suppression of determination of adult lifespan 24 hpc (Table 6 B). Pathways associated with induction of zinc ion binding, nucleus and axon guidance and Immunoglobulin-like fold pathways were among those enriched 48 hpc (Table 6 B).

Table 6. Blast2GO, ProfCom Gene Ontology (GO), InterPro and R-spider Pathway enrichment analyses of differentially expressed genes in *G. pallidipes* guts 24 or 48 hrs post-challenged by *T. b. brucei*.

A

Database/Tool	Hpc	Differential Expression	Category	Pathway ID	Description of Pathway	FDR	P-Value	Test*	Ref**
Blast2GO	24	Up-regulated	None	-	-	-	-	-	-
		Down-regulated	Cell Component	GO:0005576	Extracellular region	6.18E-04	9.78E-07	14	110
			Molecular Function	GO:0008061	Chitin binding	7.30E-07	7.74E-10	12	37
			Biological Process	GO:0006030	Chitin metabolic process	7.30E-07	3.89E-10	13	44
	48	Up-regulated	None	-	-	-	-	-	-
		Down-regulated	Cell Component	GO:0005840	Ribosome	5.92E-04	3.12E-07	7	117
			Molecular Function	GO:0003735	Structural constituent of ribosome	5.92E-04	2.49E-07	7	113
				GO:0004252	Serine-type endopeptidase activity	2.34E-03	4.31E-06	7	176
			Biological Process	GO:0006412	Translation	9.60E-04	7.59E-07	8	200

B

Database/Tool	Hpc	Differential Expression	Pathway ID	Description of Pathway	Number of Genes	P-Value
ProfCom GO						
	24	Up-regulated	GO:0055085	Transmembrane transport	15	0.020
			GO:0006355	Regulation of transcription, DNA dependent	12	0.130
			GO:0007507	Heart development	5	0.140
		Down-regulated	GO:0008340	Determination of adult lifespan	9	0.010
			GO:0007498	Mesoderm development	7	0.050
			GO:0005509	Calcium ion binding	10	0.250
	48	Up-regulated	GO:0008270	Zinc ion binding	18	0.010
			GO:0005634	Nucleus	24	0.010
			GO:0007411	Axon guidance	7	0.020
		Down-regulated	GO:0006508	Proteolysis	5	0.280
ProfCom InterPro						
	24	Up-regulated	IPR011701	Major facilitator superfamily MFS-1	7	0.523
			IPR007110	Immunoglobulin-like	5	0.990
			IPR013783	Immunoglobulin-like fold	5	0.990
		Down-regulated	IPR008271	Serine/threonine-protein kinase, active site	5	0.990
			IPR016196	Major facilitator superfamily, general substrate transporter	7	0.990
			IPR000719	Protein kinase, catalytic domain	6	0.990
	48	Up-regulated	IPR013783	Immunoglobulin-like fold	11	0.006
			IPR003599	Immunoglobulin subtype	7	0.076
			IPR007110	Immunoglobulin-like	8	0.088
		Down-regulated	None	-	-	-

A- Enriched pathways determined through Blast2GO (Conesa and Götz, 2008), and **B**-ProfCom (Antonov *et al.*, 2008). * Differentially exprressed dataset. ** Entire *G. pallidipes* gene set. Hpc= Hours post challenge.

5.4.2 Pathway enrichment analysis of transcripts in challenged and control midguts 24 and 48hpc

The R-spider analysis (Antonov *et al.*, 2010) for potential protein-protein interactions among the differentially expressed genes revealed significant interactions between genes in pathways associated with induced transcripts 24 or 48 hpc (Figure 10). The G protein beta-subunit 13F (FBGN0001105) was a major hub in the 24 hpc interactome, with G protein α i subunit (FBGN0001104), Dynamin associated protein 160 (DAP 160), and LEA as major products. The 48 hpc interactome revealed cGMP as a major hub with cycalpha99B, dunce (encoding cAMP phosphodiesterase (FBGN0000479), cAMP-dependent protein kinase (PKAC1), adenine phosphoribosyltransferase (APRT) and Adenylosuccinate Lyase (FBGN0038467) as major products.

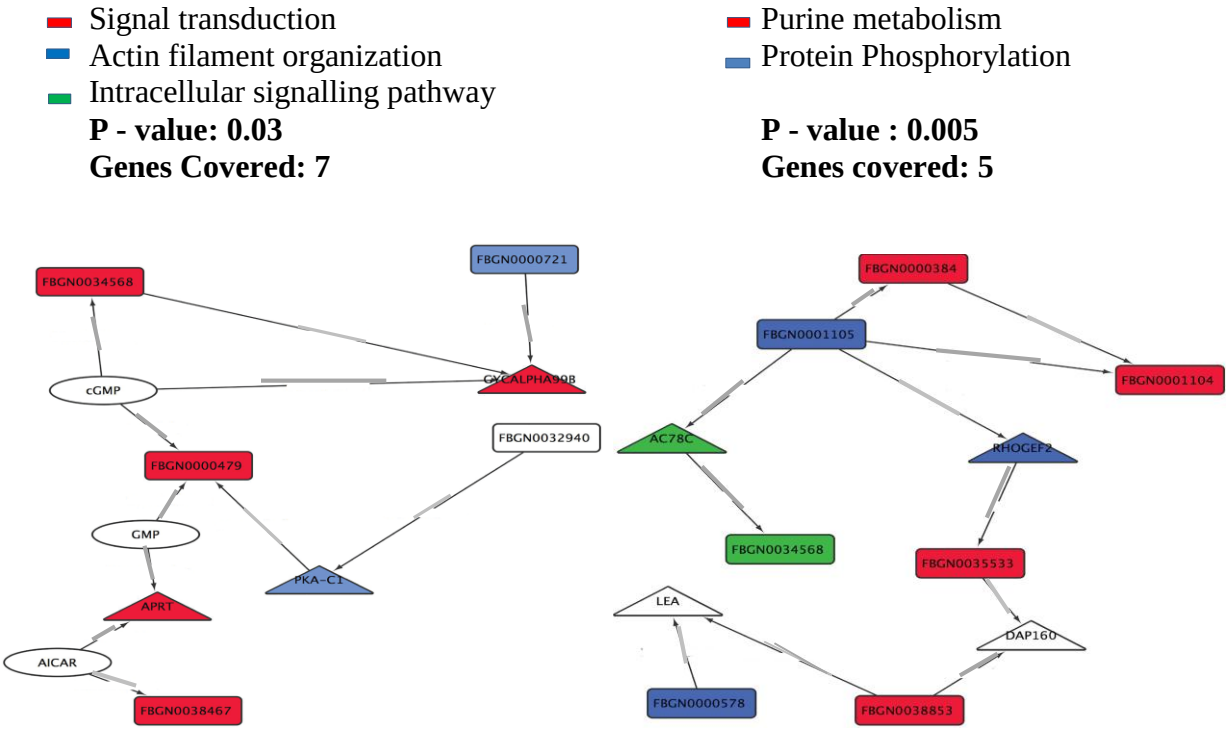


Figure 10 Enriched networks in *G. pallidipes* guts 24 or 48 hrs post-challenged by *T. b. brucei* identified by R- spider analysis (Antonov *et al.*, 2010) by allowing for one missing gene based on *Drosophila*

5.4.3 Predominant transcripts at 24 or 48 hpc

Analysis of the most DE (> 90 percentile) transcripts, and supported by at least 5000 reads in the guts or carcasses were interrogated, midgut specific transcripts induced 24 hpc were associated with lipid remodeling/lipogenesis, proteolysis, urea cycle, carnitine trafficking, collagen metabolism, apoptosis, and cell growth/differentiation were identified (Appendix 11). About 23 and 27 % of these transcripts were secreted and associated with immunity, respectively. Those induced 48 hpc were associated with embryonic, growth and development, muscle/motility, suppression of tumor, serine endopeptidase and related protesomal degradation of target protein, enhanced translation of mRNA and neuronal development. About 33 and 10% of the transcripts where secreted and they were associated with immunity respectively.

The suppressed midgut specific transcripts were associated with nervous system development, neurotransmitter transport/ cellular calcium ion hemostasis and cuticular structure 24 hpc, and ATP-dependent degradation of ubiquitinated proteins and cell proliferation/migration 48 hpc. About 37 and 32 % of these transcripts were secreted and

associated with immunity respectively 24 hpc. About 50% were secreted or associated with immunity at 48 hrs post challenges.

5.5 Discussion

Results from this study indicate that transcripts associated with metabolic processes dominated the early (24 hpc) responses, with immune associated genes increasing in expression later by 48 hpc. These findings demonstrate that upon parasite entering the gut lumen, parasites encounter minimal immunological challenge early during the infection process, which potentially permits the differentiation of the parasite from BSF to PCF forms in the gut. There was an increased immunological responses to parasite challenge at 48 hpc where transcripts for immune associated zinc ion binding (Brazão *et al.,* 2009), Estradiol 17-beta-dehydrogenase 11 (Regan *et al.,* 2013), Croquemort (Franc *et al.,* 1996), serine peptidase (Gorman *et al.,* 2000), purine metabolism (Simmonds *et al.,* 1978), axon guidance and Immunoglobulin-like fold pathways (Govind and Nehm 2004, Dong *et al.,* 2006) were significantly induced.

Temporal increase in expression of transcripts associated with microfilament re-organization and metabolism, indicate development of the cytoskeleton in the young teneral flies and digestion of the ingested blood meal. The observed induction of transcripts related to apoptosis associated keratinocyte signaling pathway (Bowen *et al* 2003) signals recognition of dead cells or parasite antigens by the host defences during the course of parasite challenge. Parasite released products have been shown to compromise the integrity of host physiology during the early course of parasite infection in adult flies.

Chitin metabolism was induced in the carcass at 24 hpc, but suppression of the same transcripts was noted at 48 hpc. The transcript induction observed at 24 hpc may be linked to insect growth and morphogenesis due to chitin requirement for structures (Merzendorfer and Zimoch, 2003) since in this study teneral tsetse with under developed exoskeleton were used. Suppression of these transcripts at 48 hpc, however, may be tied to suppression of chitin metabolism in the gut associated with impaired PM structure to facilitate trypanosome escape to the ectoperitrophic space.

5.6 References

Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**:403-410.

Antonov A.V., Schmidt T., Wang Y. and Mewes H.W. (2008). ProfCom: a web tool for profiling the complex functionality of gene groups identified from high throughput data. *Nucleic Acids Res*e*arch* **36**: W347–351

Antonov A.V., Schmidt E.E., Dietmann S., Krestyaninova M. and Hermjakob H. (2010). R spider: a network-based analysis of gene lists by combining signaling and metabolic pathways from Reactome and KEGG databases. *Nucleic Acids Res*e*arch* **38**: W78–83.

Bowen A.R., Hanks A.N., Allen S.M., Alexander A., Diedrich M.J. and Grossman D. (2003). Apoptosis regulators and responses in human melanocytic and keratinocytic cells. *The Journal of Investigative Dermatology* **120**:48-55.

Brazão V., Caetano L.C., Filipin Mdel V., Santello F.H., Toldo M.P. and do Prado J.C. Jr. (2009). *Trypanosoma cruzi*: The effects of zinc supplementation in the immune response during the course of experimental disease. *Experimental Parasitology* **121**:105-109.

Conesa A., Götz S., García-Gómez J.M., Terol J., Talón M. and Robles M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**:3674-3676.

Conesa A. and Götz S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* 2008:619832.

Dong Y., Taylor.E. and Dimopoulos G. (2006). AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. *PLoS Biology* **7:**e229.

Franc N.C., Dimarcq J.L., Lagueux M., Hoffmann J. and Ezekowitz R.A. (1996). Croquemort, a novel Drosophila hemocyte/macrophage receptor that recognizes apoptotic cells. *Immunity.* **4**:431-443.

Geiger A., Hamidou Soumana I., Tchicaya B., Rofidal V., Decourcelle M., Santoni V. and Hem S. (2015). Differential expression of midgut proteins in *Trypanosoma brucei gambiense*-stimulated vs. non-stimulated *Glossina palpalis gambiensis* flies. *Frontiers in Microbiology* **6**:444.

Gibson W. and Bailey M. (2003). The development of *Trypanosoma brucei* within the tsetse fly midgut observed using green fluorescent trypanosomes. *Kinetoplastid Biology and Disease* **2**:1.

Giraldo-Calderón G.I., Emrich S.J., MacCallum R.M., Maslen G., Dialynas E., Topalis .P, Ho N., Gesing S., Madey G., Collins F.H. and Lawson D. (2015). VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Research.* **4**:D707-13.

Gorman M.J., Andreeva O.V. and Paskewitz S.M. (2000). Sp22D: a multidomain serine protease with a putative role in insect immunity. *Gene*. **251**:9-17.

Götz S., García-Gómez J.M., Terol J., Williams T.D., Nagaraj S.H., Nueda M.J., Robles M., Talón M., Dopazo J. and Conesa A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* **36**:3420-35

Govind S. and Nehm R.H. (2004). Innate immunity in fruit flies: A textbook example of genomic recycling. *PLoS Biology* **2**: e276.

Hamidou Soumana I., Klopp C., Ravel S., Nabihoudine I., Tchicaya B., Parrinello H., Abate L., Rialle S. and Geiger A. (2015). RNA-seq de novo Assembly Reveals Differential Gene Expression in *Glossina palpalis gambiensis* Infected with *Trypanosoma brucei gambiense* vs. Non-Infected and Self-Cured Flies. Front Microbiol. 6: 1259.

Lehane M.J., Aksoy S., Gibson W., Kerhornou A., Berriman M., Hamilton J., Soares M.B., Bonaldo M.F., Lehane S. and Hall N. (2003). Adult midgut expressed sequence tags from the tsetse fly *Glossina morsitans morsitans* and expression analysis of putative immune response genes. *Genome Biology* **4**:R63.

MacLeod E,T., Maudlin I., Darby A.C., Welburn S.C. (2007). Antioxidant promote establishment of trypanosome infections in tsetse. *Parasitology*.**134**:827-831

Marygold S.J., Leyland P.C., Seal R.L., Goodman J.L., Thurmond J., Strelets V.B. and Wilson R.J. (2013). FlyBase: improvements to the bibliography. *Nucleic Acids Research* **41**: D751–757.

Merzendorfer H. and Zimoch L. (2003). Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases *Journal of Experimental Biology* **206:**4393-412

Peacock L., Ferris V., Bailey M., Gibson W. (2012). The influence of sex and fly species on the development of trypanosomes in tsetse flies. *PLoS Neglected Tropical Diseases*. **6**:e1515

Telleria E.L., Benoit J.B., Zhao X., Savage A.F., Regmi S., Alves e Silva T.L., O'Neill M. and Aksoy S (2014). Insights into the Trypanosome-Host Interactions Revealed through Transcriptomic Analysis of Parasitized Tsetse Fly Salivary Glands. *PLoS Neglected Tropical Diseases* **8**: e2649.

Van den Abbeele J., Claes Y., van Bockstaele D., Le Ray D. and Coosemans M. (1999). *Trypanosoma brucei* spp. development in the tsetse fly: characterization of the postmesocyclic stages in the foregut and proboscis. *Parasitology* **118**:469-478.

CHAPTER SIX

GENERAL DISCUSSION, CONCLUSIONS AND RECCOMENDATIONS

6.1 General discussion

The current research focused on the genetic diversity and population structure of *T. b. brucei* and *T. b. rhodesiense* in Uganda and western Kenya in and the molecular response elicited by *G. pallidipes* challenged with *T. b. brucei*. This work was carried out with the aim of providing evidence that will be useful in the development of tsetse and trypanosomiasis control policies and strategies.

Earlier studies on population structure of *T. b. brucei* and *T. b. rhodesiense* isolates from Uganda and Kenya have been carried out using isoenzyme analysis. In this study variable microsatellite locus were used and they allowed partitioning *T. brucei* isolates in three genetic clusters. Ugandan *T. b. rhodesiense* isolates were found as a mix in the first and third clusters while the second cluster contained western Kenya isolates. *T. b. rhodesiense* populations from the two different geographical sites appeared to be genetically distinct however; there was also evidence for genetic admixture. Earlier studies have shown that *T. b. rhodesiense* isolates vary, depending on their geographical foci (Gibson *et al*., 1980; Gibson *et al.,* 1985; Hide *et al.,* 1994; Gibson and Stevens 1999; Goodhead *et al.,* 2013 and Duffy *et al.,* 2013) although a great an overlap would be expected between isolates from southeast Uganda and western Kenya. This could be due to the proximity of the two sites (Baldry, 1972). The vector that transmits *T. b. rhodesiense* in southeast Uganda changed from *G. pallidipes* to *G. f. fuscipes.* This led to selection of specific genotypes and also mixing of two different trypanosome populations. Previous studies also showed that isolates from Uganda's new and old foci are genetically distinct; however, this study indicate that isolates from the new focus are as a result of expansion from the old focus and this concurs with studies carried out in Tanzania on the spread of HAT (Komba *et al.,* 1997). Findings from this study did not concur with earlier studies that had shown sub-structuring of trypanosome population on the basis host and geographical locations (Hide *et al.,* 1994; MacLeod *et al.,* 2001; Goodhead *et al*., 2013). Presence of different genotypes in both the old and new foci show occurrence of multiple lineages during disease epidemics and these agrees with what had been observed earlier. It was also noted that temporal variation does not result in genetic portioning of *T. b. brucei* and *T. b. rhodesiense* in Uganda unlike what Duffy *et*

al. (2013) observed using *T. b. rhodesiense* isolates only collected between 1970s and 1990 but we instead found geographic structuring.

The observed linkage disequilibrium at most loci demonstrated the clonal nature of *T. brucei* which agrees with what was observed by Duffy *et al.* (2013). The presence of SRA gene in isolates collected from cattle confirmed cattle as reservoirs of *T. b. rhodesiense* Therefore, continued movement of livestock from southeast Uganda to Northern Uganda where *T. b. gambiensis* is found, could fuel the fusion of *T. b. gambiensis* and *T. b. rhodesiense* related diseases. In addition, livestock trade across the Kenya-Ugandan border could result in increased disease incidences as this region is currently recording very low cases (Rutto *et al.,* 2013).

G. pallidipes has been shown to transmit both HAT and AAT in Kenya; therefore studies on molecular responses induced in the fly upon trypanosome challenge can lead to identification of molecules that can be targeted for transmission blocking. These flies have been shown in previous studies to be refractory to trypanosome infection (Peacock *et al.,* 2012). Transcriptional analysis of midgut and carcass of *G. pallidipes* challenged with *T. b. brucei* demonstrated the prominence of induced non-innate immune responses in the teneral fly after parasite challenge prior to establishment in the gut. Induction of innate immune responses in *G. m. morsitans* has been shown to occur at three to six days post challenge (Ellis and Evans, 1977; Gibson and Bailey, 2003). Immune associated responses were more prevalent at 24hpc when compared to 48hpc; this could be due to fly response to BSF or establishment of vector-parasite endemic stability in the guts that facilitate the establishment in immunity naïve young and teneral flies Suppression of the fly immune system to enable parasite establishment in the gut was observed; this was due to induction of CD109 antigen and suppression of serine protease sp24d at both 24 and 48hpc as well as suppression of transcripts involved in chitin metabolism at 24hpc, hence damaging the PM. The results of the study also showed that a temporal enrichment of immune associated genes occurs at 48hpc relative to 24hpc indicating gradual maturity of the immune system in the fly. The study also showed a global suppression of genes at 24 relative to 48hpc indicating transcriptional reorganization in order to overcome parasite challenge. There were transcripts consistently expressed 24 and 48 hpc that play a role in the establishment of the infection demonstrating changes in molecular responses in the fly to the parasite challenges.

6.2 Conclusions

This study showed that *T. b. rhodesiense* and *T. b. brucei* populations from Uganda and Kenya fall into three genetic clusters and there is continuing genetic exchange perhaps due to movement of cattle as they are reserviours for both *T. b. brucei* and *T. b. rhodesiense* and dispersal among the two strains of parasites.

G. pallidipes challenged with *T. b. brucei* induces predominant early non-innate immune responses and there is a more global suppression than induction of transcripts in the fly that help in the reorganization of metabolic and physiological processes in the tsetse to withstand parasite challenge. There is an enhanced induction of immune response transcripts 24 relative to 48 hpc, indicating progressive maturity of the fly's immune system. The under-developed immune system in teneral flies could account for their susceptibility to infection

6.3. Recommendations

This study revealed that the origin of new HAT disease foci in central Uganda was as a result of movement of the *T. b. rhodesiense* from the traditionally endemic region of south eastern Uganda to the new foci. This could be as a result of movement of cattle which are reservoirs for the parasite or movement of the vector to the new foci. Therefore vector and parasite control measures should be reinforced to prevent the disease from spreading to new foci or reintroduction of the disease in western Kenya which is currently reporting very low disease incidences.

Affected pathways such as chitin binding pathway during the 24 hpc time point can be targeted for parasite transmission blocking in tsetse. A repeat study on molecular responses of *G. pallidipes* challenged by *T. b .brucei* need to be carried out at later post challenge time points like 72hrs post challenge in order to be able to capture early time point tsetse immune responsive genes. Since at 24hpc non-innate immune responses we predominant while at 48hpc there is an indication of on-going maturity of the fly's immune system.

6.4 References

Baldry D.A.T. (1972) A history of Rhodesian sleeping sickness in the Lambwe Valley. *Bulletin of the World Health Organisation* **47**: 699-718

Duffy C.W., MacLean L., Sweeney L., Cooper A., Turner C.M., Tait A., Sternberg J., Morrison L.J. and MacLeod A. (2013). Population Genetics of *Trypanosoma brucei rhodesiense*: Clonality and Diversity within and between Foci. *PLoS Neglected Tropical Diseases* **7**.11: e2526

Ellis D.S. and Evans D.A. (1977). Passage of *Trypanosoma brucei rhodesiense* through the peritrophic membrane of *Glossina morsitans morsitans*. *Nature* 267:834-835

Gibson W.C., de C Marshall T.F. and Godfrey D.G (1980). Numerical analysis of enzyme polymorphism: a new approach to the epidemiology and taxonomy of trypanosomes of the subgenus *Trypanozoon*. *Advances in Parasitology* **18**: 175–246.

Gibson W.C. and Wellde B.T. (1985). Characterization of *Trypanozoon* stocks from the South Nyanza sleeping sickness focus in Western Kenya. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **79**: 671–676

Gibson W. and Stevens J. (1999). Genetic exchange in the *trypanosomatida*e. *Advances in* Parasitology **43**: 1–46

Gibson W. and Bailey M. (2003). The development of *Trypanosoma brucei* within the tsetse fly midgut observed using green fluorescent trypanosomes. *Kinetoplastid Biology and Disease* **2**:1

Goodhead I., Capewell P., Bailey J.W., Beament T., Chance M., Kay S., Forrester S., MacLeod A., Taylor M., Noyes H. and Hall N. (2013). Whole-genome sequencing of *Trypanosoma brucei* reveals introgression between subspecies that is associated with virulence." *MBio* pii: e00197-13.

Hide G., Tait A., Maudlin I. and Welburn S.C (1994). Epidemiological relationships of *Trypanosoma brucei* stocks from South East Uganda: Evidence for different population structures in human and non-human trypanosomes. *Parasitology* **109**: 95–111

Komba E.K., Kibona S.N., Ambwene A.K., Stevens J.R. and Gibson W.C. (1997). Genetic diversity among *Trypanosoma brucei rhodesiense* isolates from Tanzania. *Parasitology* **115:** 571–579

MacLeod A., Tait A. and Turner C.M.R (2001). The population genetics of *Trypanosoma brucei* and the origin of human infectivity. *Philosophical Transactions of the Royal Society of London* **356**: 1035-1044.

Peacock L., Ferris V., Bailey M. and Gibson W (2012). The influence of sex and fly species on the development of trypanosomes in tsetse flies. *PLoS Neglected Tropical Diseases.* **6:**e1515.

Rutto J.J., Osano O., Thuranira E.G., Kurgat R.K. and Odenyo V.A. (2013). Socio-Economic and Cultural Determinants of Human African Trypanosomiasis at the Kenya–Uganda Transboundary." *PLoS Neglected Tropical Diseases* **7**: e2186.

Walshe D.P., Lehane M.J., Haines L.R (2011). Post eclosion age predicts the prevalence of midgut trypanosome infections in *Glossina*. *PLoS One*. **6**:e26984.

6.5 APPENDICES

Appendix I. Details of the 269 *T.b.brucei* and *T.b.rhodesiense* samples used in the study. The first three columns list the sample name, and its geographic origin (Country and District). The fourth column shows the code used in this study to identify a district. The following columns identify the named subspecies for each isolate (Taxon), the presence/absence of the SRA gene (SRA), the isolate host (Host), and the year of collection (Year). The next three columns report the Q values (the probability an individual to be assigned to each of the three clusters detected by the Structure analysis). The last column report the individual assignment based on the DAPC analysis.

								Structure Clusters			
Sample ID	**Country**	**District**	**Code**	**Taxon**	**SRA**	**Host**	**Year**	**1**	**2**	**3**	**DAPC cluster**
H120	Uganda	Apac	AP	*Tbr*	+	Human	2010	0.369	0.003	0.628	3
H573	Uganda	Bugiri	BG	*Tbb*	-	Pig	2001	0.989	0.004	0.007	1
H588	Uganda	Bugiri	BG	*Tbb*	-	Pig	2001	0.990	0.005	0.005	1
H839	Uganda	Bugiri	BG	*Tbr*	+	Human	1977	0.976	0.020	0.004	1
H592	Uganda	Bugiri	BG	*Tbr*	+	Human	1990	0.021	0.005	0.974	3
H593	Uganda	Bugiri	BG	*Tbr*	+	Human	1990	0.017	0.004	0.979	3
H594	Uganda	Bugiri	BG	*Tbr*	+	Human	1990	0.017	0.003	0.981	3
H364	Uganda	Bugiri	BG	*Tbr*	+	Human	2009	0.052	0.004	0.945	3
H100	Uganda	Bukedea	BKD	*Tbr*	+	Human	2011	0.003	0.003	0.994	3
H849	Uganda	Busoga	BS	*Tbb*	-	*G.pallidipes*	1969	0.933	0.038	0.029	1
H201	Uganda	Busoga	BS	*Tbb*	-	*G.f.f*	1971	0.970	0.004	0.027	1
H291	Uganda	Busoga	BS	*Tbb*	-	*G.pallidipes*	1969	0.886	0.036	0.078	1
H851	Uganda	Busoga	BS	*Tbb*	-	Pig	1991	0.950	0.037	0.013	1
H832	Uganda	Busoga	BS	*Tbb*	-	Wildlife	1966	0.962	0.007	0.031	1
H262	Uganda	Busoga	BS	*Tbb*	-	Wildlife	1970	0.994	0.003	0.003	1
H243	Uganda	Busoga	BS	*Tbr*	+	Human	1959	0.994	0.003	0.003	1
H646	Uganda	Busoga	BS	*Tbr*	+	Human	1960	0.891	0.034	0.075	1
H303	Uganda	Busoga	BS	*Tbr*	+	Human	1961	0.876	0.017	0.107	1
H640	Uganda	Busoga	BS	*Tbr*	+	Human	1961	0.910	0.064	0.026	1
Try055	Uganda	Busoga	BS	*Tbr*	+	Human	1972	0.374	0.621	0.005	2

H837	Uganda	Busoga	BS	*Tbr*	+	Human	1976	0.180	0.127	0.693	3
Try009	Uganda	Busoga	BS	*Tbr*	+	Human	1976	0.008	0.981	0.011	2
H834	Uganda	Busoga	BS	*Tbr*	+	Human	1979	0.006	0.011	0.983	3
H841	Uganda	Busoga	BS	*Tbr*	+	Human	1991	0.005	0.003	0.992	3
H843	Uganda	Busoga	BS	*Tbr*	+	Human	1991	0.305	0.005	0.691	3
H844	Uganda	Busoga	BS	*Tbr*	+	Human	1991	0.008	0.003	0.989	3
H845	Uganda	Busoga	BS	*Tbr*	+	Human	1991	0.016	0.054	0.929	3
H846	Uganda	Busoga	BS	*Tbr*	+	Human	1993	0.009	0.013	0.979	3
H848	Uganda	Busoga	BS	*Tbr*	+	Human	1993	0.070	0.010	0.920	3
H836	Uganda	Busoga	BS	*Tbr*	+	Human		0.060	0.013	0.927	3
H853	Uganda	Busoga	BS	*Tbr*	+	Human		0.066	0.006	0.928	3
Try045	Uganda	Busoga	BS	*Tbr*	+	Tsetse fly	1963	0.012	0.985	0.003	2
H581	Uganda	Busia	BU	*Tbb*	-	Cattle	2009	0.834	0.011	0.156	1
H879	Uganda	Busia	BU	*Tbb*	-	Cattle	2009	0.976	0.004	0.020	1
H558	Uganda	Busia	BU	*Tbb*	-	Pig	2007	0.938	0.003	0.060	1
Try019	Kenya	Busia	BU	*Tbr*	+	Cattle	1987	0.002	0.995	0.003	2
Try020	Kenya	Busia	BU	*Tbr*	+	Cattle	1995	0.066	0.929	0.006	2
Try013	Kenya	Busia	BU	*Tbr*	+	Dog	2001	0.016	0.979	0.005	2
Try005	Kenya	Busia	BU	*Tbr*	+	Human	1989	0.003	0.993	0.004	2
H598	Uganda	Busia	BU	*Tbr*	+	Human	1990	0.003	0.003	0.994	3
H599	Uganda	Busia	BU	*Tbr*	+	Human	1990	0.003	0.003	0.994	3
H600	Uganda	Busia	BU	*Tbr*	+	Human	1990	0.003	0.003	0.994	3
H602	Uganda	Busia	BU	*Tbr*	+	Human	1990	0.066	0.006	0.928	3
H619	Uganda	Busia	BU	*Tbr*	+	Human	1990	0.003	0.003	0.994	3
H865	Uganda	Busia	BU	*Tbr*	+	Human	1990	0.905	0.020	0.075	3
H869	Uganda	Busia	BU	*Tbr*	+	Human	1990	0.003	0.003	0.994	3
H577	Uganda	Busia	BU	*Tbr*	+	Human	1991	0.003	0.003	0.994	3
H610	Uganda	Busia	BU	*Tbr*	+	Human	1991	0.004	0.006	0.990	3
H611	Uganda	Busia	BU	*Tbr*	+	Human	1991	0.003	0.003	0.994	3
Try006	Kenya	Busia	BU	*Tbr*	+	Human	1997	0.003	0.992	0.005	2
Try054	Kenya	Busia	BU	*Tbr*	+	Human	1997	0.005	0.984	0.011	2

Try028	Kenya	Busia	BU	*Tbr*	+	Human	1999	0.003	0.993	0.004	2
Try007	Kenya	Busia	BU	*Tbr*	+	Human	2000	0.003	0.994	0.003	2
Try031	Kenya	Busia	BU	*Tbr*	+	Human	2000	0.003	0.991	0.006	2
Try032	Kenya	Busia	BU	*Tbr*	+	Human	2000	0.003	0.992	0.005	2
Try026	Kenya	Busia	BU	*Tbr*	+	Human	2001	0.003	0.994	0.003	2
Try029	Kenya	Busia	BU	*Tbr*	+	Human	2001	0.009	0.987	0.004	2
Try008	Kenya	Busia	BU	*Tbr*	+	Human	2002	0.003	0.993	0.004	2
Try021	Kenya	Busia	BU	*Tbr*	+	Human	2002	0.002	0.995	0.003	2
Try022	Kenya	Busia	BU	*Tbr*	+	Human	2002	0.003	0.994	0.003	2
Try023	Kenya	Busia	BU	*Tbr*	+	Human	2002	0.011	0.986	0.003	2
Try024	Kenya	Busia	BU	*Tbr*	+	Human	2002	0.003	0.985	0.012	2
Try025	Kenya	Busia	BU	*Tbr*	+	Human	2002	0.884	0.088	0.028	1
Try027	Kenya	Busia	BU	*Tbr*	+	Human	2002	0.003	0.993	0.004	2
Try030	Kenya	Busia	BU	*Tbr*	+	Human	2002	0.003	0.994	0.003	2
Try015	Kenya	Central Nyanza	CN	*Tbr*	+	Cattle	1964	0.002	0.995	0.003	2
Try016	Kenya	Central Nyanza	CN	*Tbr*	+	Cattle	1967	0.003	0.994	0.003	2
Try002	Kenya	Central Nyanza	CN	*Tbr*	+	Human	1961	0.002	0.995	0.003	2
Try004	Kenya	Central Nyanza	CN	*Tbr*	+	Human	1977	0.006	0.991	0.003	2
Try052	Kenya	Central Nyanza	CN	*Tbr*	+	Human	1977	0.002	0.995	0.003	2
Try012	Kenya	Central Nyanza	CN	*Tbr*	+	Wildlife	1958	0.071	0.892	0.037	2
Try0 51	Kenya	Central Nyanza	CN	*Tbr*	+	Cattle	1967	0.025	0.968	0.008	2
Try033	Kenya	Central Nyanza	CN	*Tbr*	+	Tsetse fly	1961	0.007	0.946	0.047	2
Try0 50	Kenya	Central Nyanza	CN	*Tbr*	+	Wildlife	1958	0.006	0.985	0.009	2
H458	Uganda	Dokolo	DK	*Tbr*	+	Human	2009	0.003	0.115	0.882	3
H460	Uganda	Dokolo	DK	*Tbr*	+	Human	2009	0.009	0.009	0.982	3
H467	Uganda	Dokolo	DK	*Tbr*	+	Human	2009	0.492	0.004	0.505	3
H472	Uganda	Dokolo	DK	*Tbr*	+	Human	2009	0.010	0.010	0.980	3
H624	Uganda	Dokolo	DK	*Tbr*	+	Human	2009	0.005	0.005	0.990	3
H111	Uganda	Dokolo	DK	*Tbr*	+	Human	2010	0.784	0.026	0.190	1
H114	Uganda	Dokolo	DK	*Tbr*	+	Human	2010	0.003	0.003	0.994	3
H117	Uganda	Dokolo	DK	*Tbr*	+	Human	2010	0.844	0.023	0.133	1

H493	Uganda	Dokolo	DK	*Tbr*	+	Human	2010	0.003	0.003	0.994	3
H511	Uganda	Dokolo	DK	*Tbr*	+	Human	2010	0.088	0.005	0.907	3
H109	Uganda	Dokolo	DK	*Tbr*	+	Human	2011	0.953	0.008	0.039	1
H170	Uganda	Kaberamaido	KA	*Tbb*	-	Cattle	2011	0.957	0.024	0.019	1
RE12_058	Uganda	Kaberamaido	KA	*Tbb*	-	Cattle		0.989	0.007	0.004	1
H498	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2006	0.003	0.003	0.994	3
H356	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2009	0.006	0.004	0.990	3
H359	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2009	0.005	0.005	0.991	3
H360	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2009	0.085	0.013	0.901	3
H361	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2009	0.017	0.003	0.980	3
H362	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2009	0.003	0.003	0.994	3
H459	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2009	0.003	0.002	0.995	3
H462	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2009	0.004	0.200	0.796	3
H463	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2009	0.002	0.003	0.995	3
H464	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2009	0.004	0.004	0.992	3
H465	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2009	0.049	0.004	0.947	3
H468	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2009	0.004	0.003	0.993	3
H470	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2009	0.016	0.004	0.980	3
H474	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2009	0.333	0.334	0.333	3
H104	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.004	0.003	0.993	3
H103	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.237	0.023	0.739	3
H110	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.654	0.213	0.133	1
H112	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.949	0.039	0.012	1
H113	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.005	0.002	0.993	3
H115	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.004	0.004	0.992	3
H116	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.004	0.003	0.993	3
H118	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.003	0.003	0.995	3
H119	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.759	0.006	0.236	1
H121	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.876	0.087	0.037	1
H122	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.921	0.021	0.058	1
H123	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.007	0.005	0.988	3

H125	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.003	0.002	0.995	3
H128	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.449	0.007	0.544	3
H129	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.018	0.003	0.979	3
H130	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.004	0.003	0.993	3
H131	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.007	0.003	0.990	3
H461	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.002	0.003	0.995	3
H476	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.101	0.012	0.887	3
H478	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.003	0.002	0.996	3
H479	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.015	0.003	0.982	3
H481	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.395	0.027	0.579	3
H483	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.162	0.022	0.816	3
H485	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.165	0.752	0.084	2
H487	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.165	0.025	0.810	3
H489	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.003	0.003	0.994	3
H490	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.004	0.003	0.993	3
H491	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.002	0.002	0.996	3
H492	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.003	0.003	0.994	3
H495	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.002	0.002	0.996	3
H505	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.004	0.002	0.994	3
RE045	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.003	0.003	0.994	3
RE12_045	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2010	0.986	0.006	0.008	1
H101	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2011	0.003	0.003	0.994	3
H105	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2011	0.789	0.185	0.026	1
H106	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2011	0.003	0.003	0.994	3
H107	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2011	0.976	0.007	0.017	1
H108	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2011	0.981	0.008	0.011	1
H124	Uganda	Kaberamaido	KA	*Tbr*	+	Human	2011	0.003	0.003	0.994	3
H351	Uganda	Kaberamaido	KA	*Tbr*	+	Human		0.003	0.002	0.995	3
H353	Uganda	Kaberamaido	KA	*Tbr*	+	Human		0.003	0.003	0.994	3
H354	Uganda	Kaberamaido	KA	*Tbr*	+	Human		0.002	0.002	0.996	3
H355	Uganda	Kaberamaido	KA	*Tbr*	+	Human		0.003	0.003	0.994	3

H014	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.986	0.004	0.010	1
H015	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.990	0.004	0.006	1
H017	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.974	0.016	0.011	1
H018	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.918	0.076	0.006	1
H025	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.971	0.017	0.012	1
H027	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.473	0.520	0.008	1
H029	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.990	0.005	0.005	1
H038	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.977	0.014	0.009	1
H045	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.971	0.010	0.019	1
H054	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.965	0.004	0.031	1
H055	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.980	0.014	0.006	1
H056	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.959	0.018	0.023	1
H065	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.987	0.007	0.006	1
H070	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.992	0.003	0.005	1
H073	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.933	0.060	0.007	1
H075	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.933	0.009	0.058	1
H085	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.990	0.003	0.007	1
H145	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.979	0.021	0.007	1
H152	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.222	0.004	0.774	3
H153	Uganda	Kole	KO	*Tbb*	-	Cattle	2010	0.752	0.131	0.117	1
H019	Uganda	Kole	KO	*Tbr*	+	Cattle	2010	0.056	0.021	0.923	3
H031	Uganda	Kole	KO	*Tbr*	+	Cattle	2010	0.415	0.007	0.564	3
H034	Uganda	Kole	KO	*Tbr*	+	Cattle	2010	0.955	0.004	0.038	1
H095	Uganda	Kole	KO	*Tbr*	+	Cattle	2010	0.983	0.014	0.013	1
H151	Uganda	Kole	KO	*Tbr*	+	Cattle	2010	0.979	0.005	0.016	1
H621	Uganda	Kampala	KP	*Tbr*	+	Human	2010	0.276	0.353	0.371	1
H574	Uganda	Kayunga	KY	*Tbb*	-	Pig	2001	0.987	0.003	0.010	1
H591	Uganda	Kayunga	KY	*Tbb*	-	Pig	2001	0.980	0.004	0.016	1
H570	Uganda	Lira	LR	*Tbb*	-	Pig	2001	0.951	0.005	0.044	1
H575	Uganda	Lira	LR	*Tbb*	-	Pig	2001	0.977	0.003	0.020	1
H629	Uganda	Lira	LR	*Tbb*	-	Pig	2001	0.984	0.008	0.008	1

H500	Uganda	Lira	LR	*Tbr*	+	Human	2006	0.015	0.227	0.758	3
H521	Uganda	Lira	LR	*Tbr*	+	Human	2006	0.177	0.004	0.819	3
H522	Uganda	Lira	LR	*Tbr*	+	Human	2006	0.003	0.002	0.995	3
H614	Uganda	Lira	LR	*Tbr*	+	Human	2006	0.310	0.012	0.678	3
H616	Uganda	Lira	LR	*Tbr*	+	Human	2006	0.002	0.002	0.996	3
H633	Uganda	Lira	LR	*Tbr*	+	Human	2006	0.058	0.009	0.933	3
H878	Uganda	Lira	LR	*Tbr*	+	Human	2006	0.005	0.003	0.992	3
H569	Uganda	Mukono	MK	*Tbb*	-	Cattle	2001	0.791	0.199	0.010	1
H571	Uganda	Mukono	MK	*Tbb*	-	Pig	2001	0.989	0.004	0.007	1
H576	Uganda	Mukono	MK	*Tbb*	-	Pig	2001	0.989	0.004	0.007	1
H411	Uganda	Pallisa	PL	*Tbb*	-	Cattle	2009	0.900	0.004	0.096	1
H335	Uganda	Pallisa	PL	*Tbr*	+	Human	2008	0.003	0.003	0.994	3
H339	Uganda	Pallisa	PL	*Tbr*	+	Human	2008	0.003	0.003	0.994	3
H340	Uganda	Pallisa	PL	*Tbr*	+	Human	2008	0.003	0.003	0.994	3
H341	Uganda	Pallisa	PL	*Tbr*	+	Human	2008	0.003	0.002	0.995	3
H613	Uganda	Pallisa	PL	*Tbr*	+	Human	2008	0.004	0.005	0.991	3
H872	Uganda	Pallisa	PL	*Tbr*	+	Human	2008	0.008	0.095	0.897	3
H873	Uganda	Pallisa	PL	*Tbr*	+	Human	2008	0.004	0.005	0.991	3
H345	Uganda	Pallisa	PL	*Tbr*	+	Human	2009	0.011	0.003	0.986	3
H346	Uganda	Pallisa	PL	*Tbr*	+	Human	2009	0.017	0.010	0.972	3
H348	Uganda	Pallisa	PL	*Tbr*	+	Human	2009	0.003	0.003	0.994	3
H875	Uganda	Pallisa	PL	*Tbr*	+	Human	2009	0.004	0.005	0.991	3
RE150	Uganda	Pallisa	PL	*Tbr*	+	Human	2009	0.006	0.005	0.989	3
H595	Uganda	Pallisa	PL	*Tbr*	+	Human	2010	0.003	0.003	0.994	3
H876	Uganda	Pallisa	PL	*Tbr*	+	Human	2010	0.006	0.005	0.989	3
Try010	Uganda	Sidende	SD	*Tbr*	+	Tsetse fly	1970	0.319	0.675	0.006	2
Try017	Kenya	South Nyanza	SN	*Tbr*	+	Cattle	1970	0.005	0.992	0.003	2
Try018	Kenya	South Nyanza	SN	*Tbr*	+	Cattle	1980	0.008	0.989	0.003	2
Try003	Kenya	South Nyanza	SN	*Tbr*	+	Human	1969	0.002	0.994	0.004	2
Try014	Kenya	South Nyanza	SN	*Tbr*	+	Sheep	1970	0.077	0.912	0.011	2
Try035	Kenya	South Nyanza	SN	*Tbr*	+	Tsetse fly	1969	0.320	0.676	0.005	2

Try037	Kenya	South Nyanza	SN	*Tbr*	+	Tsetse fly	1969	0.002	0.994	0.004	2
Try034	Kenya	South Nyanza	SN	*Tbr*	+	Tsetse fly	1970	0.006	0.924	0.070	2
Try036	Kenya	South Nyanza	SN	*Tbr*	+	Tsetse fly	1981	0.019	0.978	0.003	2
Try011	Kenya	South Nyanza	SN	*Tbr*	+	Wildlife	1970	0.002	0.995	0.003	2
Try048	Kenya	South Nyanza	SN	*Tbr*	+	Wildlife	1970	0.004	0.993	0.003	2
H541	Uganda	Soroti	SR	*Tbb*	-	Cattle	2003	0.006	0.006	0.988	3
H547	Uganda	Soroti	SR	*Tbb*	-	Cattle	2003	0.938	0.003	0.059	1
H864	Uganda	Soroti	SR	*Tbb*	-	Cattle	2003	0.971	0.003	0.026	1
H446	Uganda	Soroti	SR	*Tbb*	-	Cattle	2009	0.338	0.005	0.657	3
H527	Uganda	Soroti	SR	*Tbb*	-	Cattle	Dates?	0.973	0.007	0.020	1
H528	Uganda	Soroti	SR	*Tbb*	-	Cattle		0.983	0.004	0.013	1
H529	Uganda	Soroti	SR	*Tbb*	-	Cattle		0.982	0.011	0.007	1
H531	Uganda	Soroti	SR	*Tbb*	-	Cattle		0.979	0.015	0.006	1
H533	Uganda	Soroti	SR	*Tbb*	-	Cattle		0.990	0.006	0.004	1
H540	Uganda	Soroti	SR	*Tbb*	-	Cattle		0.989	0.007	0.004	1
H543	Uganda	Soroti	SR	*Tbr*	+	Cattle	2003	0.972	0.019	0.010	1
H625	Uganda	Soroti	SR	*Tbr*	+	Cattle	2003	0.005	0.018	0.978	3
H583	Uganda	Soroti	SR	*Tbr*	+	Cattle	2008	0.002	0.002	0.996	3
RE053	Uganda	Soroti	SR	*Tbr*	+	Human	1999	0.002	0.002	0.996	3
H585	Uganda	Soroti	SR	*Tbr*	+	Human	2000	0.002	0.002	0.996	3
H589	Uganda	Soroti	SR	*Tbr*	+	Human	2001	0.002	0.002	0.996	3
H590	Uganda	Soroti	SR	*Tbr*	+	Human	2001	0.002	0.002	0.996	3
H880	Uganda	Soroti	SR	*Tbr*	+	Human	2003	0.004	0.002	0.994	3
H586	Uganda	Soroti	SR	*Tbr*	+	Human	2005	0.427	0.023	0.550	1
H365	Uganda	Soroti	SR	*Tbr*	+	Human	2009	0.003	0.002	0.995	3
H367	Uganda	Soroti	SR	*Tbr*	+	Human	2009	0.628	0.022	0.350	1
H369	Uganda	Soroti	SR	*Tbr*	+	Human	2009	0.962	0.025	0.013	1
H515	Uganda	Soroti	SR	*Tbr*	+	Human	2009	0.900	0.007	0.093	1
H517	Uganda	Soroti	SR	*Tbr*	+	Human	2009	0.174	0.011	0.815	3
H519	Uganda	Soroti	SR	*Tbr*	+	Human	2009	0.004	0.005	0.991	3
H854	Uganda	Tororo	TR	*Tbb*	-	Cattle	1988	0.972	0.011	0.017	1

H857	Uganda	Tororo	TR	*Tbb*	-	Cattle	1988	0.986	0.008	0.006	1
H858	Uganda	Tororo	TR	*Tbb*	-	Cattle	1988	0.973	0.009	0.018	1
H862	Uganda	Tororo	TR	*Tbb*	-	Cattle	1988	0.977	0.007	0.016	1
H578	Uganda	Tororo	TR	*Tbb*	-	Cattle	1991	0.399	0.003	0.598	3
H582	Uganda	Tororo	TR	*Tbb*	-	Cattle	2005	0.916	0.007	0.078	1
H285	Uganda	Tororo	TR	*Tbb*	-	Hippo	1961	0.995	0.002	0.003	1
H596	Uganda	Tororo	TR	*Tbr*	+	Human	1990	0.990	0.003	0.007	1
H601	Uganda	Tororo	TR	*Tbr*	+	Human	1990	0.028	0.003	0.969	3
H605	Uganda	Tororo	TR	*Tbr*	+	Human	1990	0.003	0.003	0.994	3
H617	Uganda	Tororo	TR	*Tbr*	+	Human	1990	0.003	0.003	0.994	3
H618	Uganda	Tororo	TR	*Tbr*	+	Human	1990	0.004	0.008	0.988	1
H620	Uganda	Tororo	TR	*Tbr*	+	Human	1990	0.004	0.005	0.991	1
H840	Uganda	Tororo	TR	*Tbr*	+	Human	1990	0.009	0.007	0.984	1
H847	Uganda	Tororo	TR	*Tbr*	+	Human	1990	0.014	0.006	0.980	3
H866	Uganda	Tororo	TR	*Tbr*	+	Human	1990	0.005	0.003	0.992	1
H868	Uganda	Tororo	TR	*Tbr*	+	Human	1990	0.035	0.011	0.954	1
H579	Uganda	Tororo	TR	*Tbr*	+	Human	1991	0.382	0.007	0.611	3
H601	Uganda	Tororo	TR	*Tbr*	+	Human	1991	0.028	0.003	0.969	3
H607	Uganda	Tororo	TR	*Tbr*	+	Human	1991	0.978	0.005	0.017	1
H612	Uganda	Tororo	TR	*Tbr*	+	Human	1991	0.010	0.050	0.940	3
H838	Uganda	Tororo	TR	*Tbr*	+	Human	1991	0.619	0.003	0.378	1
H838_1	Uganda	Tororo	TR	*Tbr*	+	Human	1991	0.011	0.006	0.983	3
H850	Uganda	Tororo	TR	*Tbr*	+	Human	1992	0.004	0.006	0.990	3
Try056	Kenya	Tororo	TR	*Tbr*	+	Human	1992	0.003	0.993	0.004	2
H855	Uganda	Tororo	TR	*Tbr*	+	Human	1988	0.005	0.007	0.988	3
H856	Uganda	Tororo	TR	*Tbr*	+	Human		0.003	0.004	0.993	3
H859	Uganda	Tororo	TR	*Tbr*	+	Human		0.004	0.004	0.992	3
H860	Uganda	Tororo	TR	*Tbr*	+	Human		0.023	0.018	0.960	3
H861	Uganda	Tororo	TR	*Tbr*	+	Human		0.004	0.005	0.991	3
Try046	Kenya	Tororo	TR	*Tbr*	+	Tsetse fly	1960	0.003	0.994	0.003	2
Try053	Kenya	Tororo	TR	*Tbr*	+	Tsetse fly	1960	0.002	0.995	0.003	2

Try001	Kenya	Teso	TS	*Tbr* $_+$	Human	2009	0.002	0.995	0.003	2

Appendix II. Information on microsatellite loci and primers used in the analyses. The first two columns report the locus name. The next two columns show the DNA sequence of the forward and reverse primers, specifying in parenthesis the type of fluorescent dye used for each one. The next two columns list the repeat motif for each locus and the range of length of the alleles in base pairs (bp). The second to the last column reports the *c*hromosomal location of each locus.

Locus	Forward Primer	Reverse Primer	Motif	Size range (bp)	Chrom. Loc.
Tryp51	[FAM]-TGACCCGTGAGAAGTGAAC	GCGCATCTACAGGCATAGAC	(ATT)	187-238	9
Tryp52	[ALEXA 532]-GCATCATTGACGTCGACCC	TAACAACCACTGGGACCGC	(GT)	201-231	11
Tryp53	[ALEXA 546]-GTACAGCCACGTGCAAACC	TGTACACAATCGGGTGGATG	(AC)	200-254	7
Tryp54	[ROX]-AGTCGGCGTGATGGTACTC	TTCAGCCCACAAACAACCG	(AAAT)	144-176	10
Tryp55	[FAM]-AATTCAACCCCAACAGCCC	CTCGTTCAATGACTTGCCCC	(GT)	208-246	5
Tryp59	[ALEXA 532]-GAGGCAATCGCAGTGTGTG	CGCACGTTTCACCATCCTC	(GT)	209-225	9
Tryp61	[ALEXA546]-ACTCGCGACAGACCATGAG	ACAGGAGAGTGTTGTGAGTG	(ATT)	179-215	11
Tryp62	[ROX]-AAGGCGACCAACTTCAACC	GTTGTCATCGGCTTGCTCC	(AC)	153-177	11
Tryp65	[ALEXA 546]-GGAGGTAAACTTGATTCGGGTG	ACGACAACAGCGACAAAGC	(ATT)	207-234	9
Tryp66	[ROX]-TCCTCGTACCTTTTCTCTCAC	ACGAAATTTAGGTGTGAAAGCTG	(ATT)	384-396	5
Tryp67	[FAM]-GTTGCTGAGGTGCAACTGG	GTCGTCAGGCACCAAAACG	(GTT)	151-178	7
TB1/8	[FAM]-AGGTTTAGTGCATGTCGGA	CCTGTTGTACGGAGGTCA	(CA)	97-117	1
TB5/2	[HEX]-CAACCGAAAGTAAGGGGAAC	TCTCGCCTTCTTTGCCC	(AT)	83-107	5
TB6/7	[HEX]-AAGCTGACAGGTGGTTGA	GAACATGCGTGCGTGTG	(AT)	104-136	6
TB9/6	[HEX]-TGATTCATTGGTTAAGACAGG	AATGATAACTGCGGATTACAC	(AC)	124-158	9
TB10/5	[FAM]-AAAGGCGATATGTTATTATTGA	ATTGGGTATACTGTCCCTCA	(TA)	79-115	10
TB11/13	[FAM]-CAAGAACTCTGCATTGAGC	ATCTGTTGGCGATGGTGA	(AT)	125-161	11

Appendix III. Results of ITS and SRA screening of animal trypanosome isolates. The first three columns list the geographic origin (Origin), the district abbreviations (Code), and the host (Host). The forth column shows the number of strains for each host (N).

The following three columns provide information on the *Trypanosoma* species infection in the samples other than *T. brucei* (*T. vivax = Tv; T.congolense*) and the occurrence of mixed *Tv* and *Tc* infections. The next two columns summarize the number of *Tbb* and *Tbr* samples, according to the SRA test. The final column reports the number of samples that did not produce PCR products, likely due to low DNA concentration and/or poor quality.

Origin	**Code**	**Host**	**N**	***Tv***	***Tc***	***Tv* & *Tc***	***Tbb***	***Tbr***	**No amplification**
Amuru	AM	Cattle	46	45	1	0	0	0	0
		Goats	2	2	0	0	0	0	0
Bugiri	BG	Pig	2	0	0	0	2	0	0
Busia	BU	Cattle	11	3	1	0	1	2	4
		Pig	1	0	0	0	1	0	0
Busoga	BS	Tsetse	1	0	0	0	0	1	0
		Dog	1	0	0	0	0	1	0
Central Nyanza	CN	Cattle	4	0	0	0	0	4	0
		Tsetse	1	0	0	0	0	1	0
		Wildlife	2	0	0	0	0	2	0
Dokolo	DK	Cattle	33	28	1	0	1	0	3
Kaberamaido	KA	Cattle	21	13	1	5	2	0	0
Kayunga	KY	Pig	2	0	0	0	2	0	0
Kole	KO	Cattle	101	51	5	2	20	5	18
Lira	LR	Cattle	3	2	1	0	0	0	0
Mukono	MK	Cattle	1	0	0	0	1	0	0
		Pig	5	0	0	0	5	0	0
Sidende	SD	Tsetse	1	0	0	0	0	1	0
South Nyanza	SN	Cattle	2	0	0	0	0	2	0
		Sheep	1	0	0	0	0	1	0
		Wildlife	2	0	0	0	0	2	0
		Tsetse	4	0	0	0	0	4	0
Soroti	SRT	Cattle	15	1	0	98 0	10	3	1
Tororo	TR	Cattle	5	0	1	0	2	0	2

	Tsetse	2	0	0	0	0	2	0
Total		**269**	**145**	**11**	**7**	**49**	**31**	**28**

Appendix IV. Linkage disequilibrium (LD) for all pairs of the seventeen microsatellites tested at 10,000 permutations in Arlequin (Excoffier *et al.,* 2005). P-values and their Standard Errors (S.E>) are reported in the last column.

Locus 1	Locus 2	P-Value	S.E
Locus_51	Locus_55	0.000	0.000
Locus_52	Locus_55	0.000	0.000
Locus_53	Locus_55	0.024	0.014
Locus_54	Locus_55	0.000	0.000
Locus_51	Locus_59	0.000	0.000
Locus_52	Locus_59	0.000	0.000
Locus_53	Locus_59	0.000	0.000
Locus_54	Locus_59	0.000	0.000
Locus_55	Locus_59	0.000	0.000
Locus_51	Locus_61	0.083	0.024
Locus_52	Locus_61	0.000	0.000
Locus_53	Locus_61	0.042	0.017
Locus_54	Locus_61	0.000	0.000
Locus_55	Locus_61	0.025	0.012
Locus_59	Locus_61	0.003	0.002
Locus_51	Locus_62	0.004	0.004
Locus_52	Locus_62	0.000	0.000
Locus_53	Locus_62	0.000	0.000
Locus_54	Locus_62	0.000	0.000
Locus_55	Locus_62	0.000	0.000
Locus_59	Locus_62	0.000	0.000
Locus_61	Locus_62	0.000	0.000
Locus_51	Locus_65	0.000	0.000
Locus_52	Locus_65	0.000	0.000
Locus_53	Locus_65	0.000	0.000
Locus_54	Locus_65	0.000	0.000

Locus_55	Locus_65	0.003	0.003
Locus_59	Locus_65	0.000	0.000
Locus_1	Locus_65	0.022	0.009
Locus_62	Locus_65	0.000	0.000
Locus_51	Locus_66	0.068	0.016
Locus_52	Locus_66	0.000	0.000
Locus_53	Locus_66	0.000	0.000
Locus_54	Locus_66	0.000	0.000
Locus_55	Locus_66	0.000	0.000
Locus_59	Locus_66	0.225	0.019
Locus_61	Locus_66	0.000	0.000
Locus_62	Locus_66	0.000	0.000
Locus_65	Locus_66	0.018	0.006
Locus_51	Locus_67	0.000	0.000
Locus_52	Locus_67	0.000	0.000
Locus_53	Locus_67	0.000	0.000
Locus_54	Locus_67	0.000	0.000
Locus_55	Locus_67	0.000	0.000
Locus_59	Locus_67	0.000	0.000
Locus_61	Locus_67	0.010	0.008
Locus_62	Locus_67	0.000	0.000
Locus_65	Locus_67	0.000	0.000
Locus_66	Locus_67	0.003	0.002
Locus_51	Tb5_2	0.000	0.000
Locus_52	Tb5_2	0.000	0.000
Locus_53	Tb5_2	0.000	0.000
Locus_54	Tb5_2	0.000	0.000
Locus_55	Tb5_2	0.000	0.000
Locus_59	Tb5_2	0.000	0.000

Locus_61	Tb5_2	0.521	0.032
Locus_62	Tb5_2	0.000	0.000
Locus_65	Tb5_2	0.000	0.000
Locus_66	Tb5_2	0.018	0.009
Locus_67	Tb5_2	0.000	0.000
Locus_51	Tb6_7	0.000	0.000
Locus_52	Tb6_7	0.000	0.000
Locus_53	Tb6_7	0.000	0.000
Locus_54	Tb6_7	0.000	0.000
Locus_55	Tb6_7	0.000	0.000
Locus_59	Tb6_7	0.000	0.000
Locus_61	Tb6_7	0.000	0.000
Locus_62	Tb6_7	0.000	0.000
Locus_65	Tb6_7	0.000	0.000
Locus_66	Tb6_7	0.000	0.000
Locus_67	Tb6_7	0.000	0.000
Tb5_2	Tb6_7	0.000	0.000
Locus_51	Tb9_6	0.000	0.000
Locus_52	Tb9_6	0.000	0.000
Locus_53	Tb9_6	0.000	0.000
Locus_54	Tb9_6	0.000	0.000
Locus_55	Tb9_6	0.000	0.000
Locus_59	Tb9_6	0.000	0.000
Locus_61	Tb9_6	0.000	0.000
Locus_62	Tb9_6	0.000	0.000
Locus_65	Tb9_6	0.000	0.000
Locus_66	Tb9_6	0.000	0.000
Locus_67	Tb9_6	0.000	0.000
Tb5_2	Tb9_6	0.000	0.000

Tb6_7	Tb9_6	0.000	0.000
Locus_51	Tb10_5	0.010	0.008
Locus_52	Tb10_5	0.000	0.000
Locus_53	Tb10_5	0.000	0.000
Locus_54	Tb10_5	0.000	0.000
Locus_55	Tb10_5	0.000	0.000
Locus_59	Tb10_5	0.000	0.000
Locus_61	Tb10_5	0.000	0.000
Locus_62	Tb10_5	0.000	0.000
Locus_65	Tb10_5	0.000	0.000
Locus_66	Tb10_5	0.000	0.000
Locus_67	Tb10_5	0.000	0.000
Tb5_2	Tb10_5	0.000	0.000
Tb6_7	Tb10_5	0.000	0.000
Tb9_6	Tb10_5	0.000	0.000
Locus_51	Tb1-8	0.000	0.000
Locus_52	Tb1-8	0.000	0.000
Locus_53	Tb1-8	0.000	0.000
Locus_54	Tb1-8	0.000	0.000
Locus_55	Tb1-8	0.000	0.000
Locus_59	Tb1-8	0.000	0.000
Locus_61	Tb1-8	0.000	0.000
Locus_62	Tb1-8	0.000	0.000
Locus_65	Tb1-8	0.000	0.000
Locus_66	Tb1-8	0.000	0.000
Locus_67	Tb1-8	0.000	0.000
Tb5_2	Tb1-8	0.000	0.000
Tb6_7	Tb1-8	0.000	0.000
Tb9_6	Tb1-8	0.000	0.000

Tb10_5	Tb1-8	0.000	0.000
Locus_51	Tb11-13	0.000	0.000
Locus_52	Tb11-13	0.000	0.000
Locus_53	Tb11-13	0.000	0.000
Locus_54	Tb11-13	0.000	0.000
Locus_55	Tb11-13	0.000	0.000
Locus_59	Tb11-13	0.000	0.000
Locus_61	Tb11-13	0.000	0.000
Locus_62	Tb11-13	0.000	0.000
Locus_65	Tb11-13	0.000	0.000
Locus_66	Tb11-13	0.000	0.000
Locus_67	Tb11-13	0.000	0.000
Tb5_2	Tb11-13	0.000	0.000
Tb6_7	Tb11-13	0.000	0.000
Tb9_6	Tb11-13	0.000	0.000
Tb10_5	Tb11-13	0.000	0.000
Tb1-8	Tb11-13	0.000	0.000

Appendix V. Average pairwise FST (Weir and Cockerham 1984) values among 19 *Trypanosoma brucei* sampling sites obtained using Arlequin (Excoffier *et al.*, 2005) and averaged across 17 loci. Asterisks denote statistically significant values (*P<0.05; **P<0.01).

	AP	BG	BU	BS	DK	KA	KP	KY	KB	KO	LR	MK	PL	SR	TR	TS	CN	SN	SD
AP	0.00																		
BG	-0.10	0.00																	
BU	0.16	0.13	0.00																
BS	-0.02	0.03	0.05**	0.00															
DK	0.16	0.05	0.09**	0.01	0.00														
KA	0.17	0.14**	0.2**	0.1**	0.03	0.00													
KP	0.38	-0.82	-0.97	-0.79	-0.15	-0.23	0.00												
KY	-0.06	0.05	0.14*	0.03	0.18	0.33*	-0.89	0.00											
KB	0.00	-0.11	0.05	-0.11	-0.04	-0.34	-0.50	-0.06	0.00										
KO	0.32	0.01	0.13**	0.01	0.23**	0.26**	-0.10	-0.02	0.19	0.00									
LR	-0.01	-0.03	0.12**	0.00	-0.01	0.02	-0.35	0.07	-0.33	0.18**	0.00								
MK	0.19	0.04	0.27**	0.16**	0.35**	0.48**	-0.50	0.06	0.19	0.25**	0.22*	0.00							
PL	0.11	0.24**	0.24**	0.15**	0.06**	0.03*	-0.50	0.37**	-0.51	0.26**	0.05	0.49**	0.00						
SR	0.02	0.08*	0.18**	0.07**	-0.03	0.01	-0.60	0.13	-0.21	0.12**	-0.04	0.32**	0.09**	0.00					
TR	-0.08	0.05**	0.05**	0.01	-0.01	0.07**	-0.78	0.08	-0.14	0.01*	0.01	0.2**	0.15**	0.06**	0.00				
TS	0.44	0.24	0.14	0.17	0.47	0.52	-0.43	0.11	0.26	0.44	0.38	0.27	0.48	0.37	0.23	0.00			
CN	0.33	0.23**	0.05*	0.16**	0.3**	0.42**	-1.07	0.2*	0.22	0.31**	0.29**	0.27	0.39**	0.31**	0.18**	-0.05	0.00		
SN	0.40	0.25**	0.08**	0.16**	0.31**	0.43**	-0.75	0.21**	0.26	0.28**	0.29**	0.31**	0.41**	0.32**	0.19**	-0.02	-0.05	0.00	
SD	0.33	-0.12	-0.06	-0.12	0.10	0.22	-0.75	-0.39	-0.25	0.08	0.10	-0.06	0.21	0.05	-0.06	-0.29	-0.26	-0.02	0.00

Appendix VI. Pairwise FST estimation among the three *Trypanosoma brucei* genetic structure/DAPC inferred clusters (1, 2 and 3) estimated using Arlequin (Excoffier *et al.*, 2005) and averaged across 17 loci. Two asterisks indicate significance at P<0.01.

Cluster	1	2	3
1	0		
2	0.16**	0	
3	0.21**	0.27**	0

Appendix VII. Primers utilized on *G. pallidipes* DNA and cDNA PCR

Putative tsetse Gene ID	**Vectorbase Accession ID**	**Forward primer**	**Reverse primer**
Tsetse β-tubulin	GPAI022614	ACGTATTCATTTCCCTTTGG	AATGGCTGTGGTGTTGGACAAC
GAPDH	GPAI033271	CTGATTTCGTTGGTGATACT	CCAAATTCGTTGTCGTACCA
Glutamine Synthase	GPAI006387	TGGTGGAAGCTCATGCCTTG	CAACGCTGGGTCCTACTTGG
Heat Shock protein 83	GPAI002368	GGAATGACCAAGCCGGATTTG	CAGCAATCAAATACGCGGAATAG
Multi Copper oxidase	GPAI025756	ACTGAGCCAATGCCACCTATACTG	GCACACTTACCGCAAGCAACTC
Pyruvate Carboxylase	GPAI003647	GAGGGTATGGGCATACGTTTG	TTGTAAGTCGCTGGCATGTGAG
Chitinase 4	GPAI022616	GCTATGTATGGGCACTCTTTTCAG	TGTCTCGCAGATTTCATTGTAACC
Transferrin 1	GPA1033230	CATCCACGCTAACAACTA	AAGTAGGTTCCACAACAG
Tep2	GPAI040205	CCAACGCCCATTACCCCTACA	GCCCCGCTGAAGGTGGTA
Serpin 4	GPAI011576	CGGCTATGGCTCGCTTAG	TGTCGCTTTTCTCGTATTGC
Serpin 6	GPAI011576	AATGAGGCTGCTGCTGCTAG	CGCCCAGTTCCCTTTGAAATG
PGRP-LB	GPAI047520	CAACAACAACCCAAAAGG	GAGTTGGTACTGCCGATGT

Appendix VIII. Putative immune responsive genes in *G. pallidipes* guts challenged with *T. b. brucei* for 24 or 48 hrs

***T. b. brucei* Challenge**	**Vectorbase Acccession No.**	**Fold Change**	**FDR Corrected p-Value**	**Treatment* - RPKM**	**Reference**- RPKM**	**Best BLAST Hit, SwissProt database**	**BLAST E - Value**	***G. m. morsitan homolog***	**Vectorbase Acccession No.**	**BLAST E - Value**	***Drosophila melanogaster* homolog**	**Lowest E-value**
24hr challenged* vs unchallenged gut**	GPAI040204	7.59	0	90.64	12.45	Myosin-viia	0	-		-	Myo61F-PB	4.75E-159
	GPAI030117	3.89	0	51.85	13.9	Leucine-rich repeat-containing protein 15	2.15E-22	Toll-7	GMOY009848	1.06E-16	Tl-PD	1.43E-16
	GPAI038505	3.6	0	278.06	80.55	ATP-dependent zinc metalloprotease yme1 homolog	0	Asrij	GMOY010861	0	-	-
	GPAI040207	3.58	0	675.83	197.08	CD109 antigen	4.27E-50	Thioester-containing protein 1	GMOY010996	0	Tep2-PF	1.53E-163
	GPAI040414	3.47	0.002	20.08	6.03	Myosin heavy non-muscle	1.20E-53	-		-	Myo61F-PA	3.01E-24
	GPAI002368	3.25	0	694.2	222.51	Heat shock protein 83	0	-		-	-	-
	GPAI039274	3.09	0.001	25.5	8.61	Tankyrase	0	Multiple ankyrin repeats single KH domain	GMOY002483	7.14E-52	Mask-PD	8.22E-34
	GPAI003307	3.06	0.001	24.68	8.4	Zinc finger protein gfi-1	9.04E-56	Zn finger homeodomain 1	GMOY006469	2.98E-10	Zfh1-PB	7.39E-11
	GPAI033707	3	0	67.86	23.57	E3 ubiquitin-protein ligase su	0	-		-	-	-

GPAI038345	2.98	0	31.09	10.88	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	3.14E-70	Oligosaccharide transferase Delta subunit	GMOY005213	0	-	-
GPAI011692	2.98	0	36.35	12.73	Guanylate cyclase 32e	0	Pole hole	GMOY000610	6.49E-09	P38b-PA	8.00E-14
GPAI021346	2.97	0.009	19.51	6.85	E3 ubiquitin-protein ligase trim9	3.18E-41	-		-	-	-
GPAI000661	2.88	0.002	25.09	9.08	ATP-dependent protease subunit	1.23E-114	-		-	-	-
GPAI017278	2.84	0	91.46	33.62	Protein rmd5 homolog a	1.51E-107	Sphinx1	GMOY010124	3.03E-12	-	-
GPAI025572	2.69	0.007	23.42	9.07	Esterase b1	2.55E-109	Immune induced molecule 3	GMOY005924	8.85E-28	-	-
GPAI044630	2.69	0.001	30.26	11.73	Teneurin-m	0	Serrate	GMOY002009	4.06E-18	-	-
GPAI000772	2.62	0	136.11	54.11	Elongation factor tu	0	Roughened	GMOY000821	1.88E-23	-	-
GPAI001769	2.51	0.003	28.87	11.97	Peptidoglycan-associated lipoprotein	7.03E-59	-		-	-	-
GPAI012722	2.48	0	40.71	17.11	Serine protease sp24d	1.33E-29	Spheroide	GMOY005310	0	Spheroide-PA	2.80E-63
GPAI023150	2.48	0	51.66	21.74	Protein phosphatase 1 regulatory inhibitor subunit 16b	5.90E-106	Multiple ankyrin repeats single KH domain	GMOY002483	8.28E-20	Mask-PD	9.74E-20
GPAI032289	2.46	0	46.2	19.61	Myeloid zinc finger 1	7.83E-51	Zn finger homeodomain 1	GMOY006469	3.67E-30	Zfh1-PB	1.94E-15
GPAI043377	2.42	0	260.27	111.94	CD109 antigen	2.23E-156	Thioester-containing protein 2	GMOY010996	0	Tep2-PB	0

GPAI039400	2.4	0.014	24.6	10.69	Trypsin	1.64E-99	Spheroide	GMOY005310	1.01E-18	Grass-PA	6.82E-25
GPAI042995	2.36	0	42.27	18.66	Ras-responsive element-binding protein 1	3.60E-19	Zn finger homeodomain 1	GMOY006469	1.22E-14	Zfh1-PA	4.12E-07
GPAI005597	2.32	0	98.19	44.16	-		-		-	Sick-PE	6.06E-125
GPAI003074	2.29	0.028	23.39	10.66	Zinc finger protein 821	4.66E-08	-		-	-	-
GPAI015300	2.26	0.027	23.94	11.02	U4 small nuclear ribonucleoprotein 27 kda protein	1.28E-24	Yantar	GMOY009365	1.09E-122	-	-
GPAI008345	2.25	0	83.08	38.44	Trypsin-1	1.34E-36	Sphinx1	GMOY010124	1.47E-21	CG6639-PA	2.49E-18
GPAI014733	2.22	0	118.16	55.54	Chymotrypsin-2	6.33E-54	Sphinx2	GMOY004422	3.05E-15	Spirit-PD	7.89E-20
GPAI048304	2.22	0.003	37.1	17.44	Cytokine receptor	2.42E-137	Domeless	GMOY007646	0	Et-PB	1.88E-18
GPAI006344	2.19	0	108.4	51.68	1-phosphatidylinositol - bisphosphate phosphodiesterase classes I and II	0	No receptor potential A	GMOY012089	0	-	-
GPAI005373	2.09	0	56.81	28.36	Bifunctional methylenetetrahydrofolate dehydrogenase mitochondrial	6.18E-141	Relish	GMOY001673	0	-	-
GPAI029001	2.09	0.037	26.42	13.21	Transcription factor hivep2	9.85E-46	Zn finger homeodomain 1	GMOY006469	3.02E-14	Zfh1-PA	8.93E-08

GPAI021330	2.09	0	74.4	37.19	Integrin beta-nu	0	Integrin betanu subunit	GMOY004265	0	-	-
GPAI048270	2.08	0.047	25.07	12.57	Homeobox protein vnd	3.78E-123	-	GMOY000032	-	Zfh1-PA	0.00649243
GPAI039813	2.02	0	136.57	70.34	Uncharacterized serpin-like protein tk1782	1.07E-31	Serpin 3	GMOY002444	-	Spn1-PB	2.01E-38
GPAI040411	2.02	0	157.78	81.51	Myosin heavy non-muscle	0	-		-	Myo61F-PB	9.27E-100
GPAI005595	2.02	0	120.14	62.11	Protein sickie	0	-		-	Sick-PJ	0
GPAI005575	-2.05	0	893.09	1911.69	Trypsin epsilon	1.20E-79	Spheroide	GMOY005310	9.33E-21	Spirit-PD	0
GPAI046533	-2.12	0	36.33	80.35	Peroxidasin	0	Dual oxidase	GMOY011845	2.70E-35	Duox-PB	0
GPAI041345	-2.12	0.021	13.54	29.95	Autophagy-related protein 16	6.14E-45	-		-	-	-
GPAI010312	-2.14	0.018	13.59	30.34	Protein fem-1 homolog b	0	Multiple ankyrin repeats single KH domain	GMOY002483	2.99E-21	Mask-PD	0
GPAI006122	-2.14	0.002	19.35	43.28	Forkhead box protein o	5.63E-79	-		-	Foxo-PC	0
GPAI019637	-2.15	0.017	13.65	30.56	Ral guanine nucleotide dissociation stimulator-like 1	7.70E-117	PDZ domain-containing guanine nucleotide exchange factor	GMOY006487	6.01E-11	-	-
GPAI001864	-2.17	0.032	11.6	26.24	Peptidoglycan-recognition protein lc	1.16E-45	Peptidoglycan recognition protein LC	GMOY006094	0	PGRP-LC-PD	0
GPAI006132	-2.17	0	40.23	91.09	Forkhead box protein o	6.45E-111	-		-	Foxo-PC	0

GPAI039275	-2.18	0.005	16.15	36.68	Regulator of gene activity	3.31E-129	-		-	-	-
GPAI001089	-2.18	0	22.7	51.58	Protein amalgam	5.80E-154	-	GMOY002876	-	Pvr-PO	0
GPAI006406	-2.2	0.001	21.08	48.39	Ubiquitin-conjugating enzyme e2-24 kda	8.08E-119	Lesswright	GMOY007017	2.69E-19	-	-
GPAI045573	-2.22	0.012	13.04	30.24	-		Fondue	GMOY010266	0	-	-
GPAI019793	-2.23	0.002	17.06	39.68	Protein decapentaplegic	1.48E-163	Decapentaplegic	GMOY005849	0	-	-
GPAI035174	-2.26	0	22.07	52.01	Ras-related protein rab-7a	3.93E-123	Rab11	GMOY009988	2.85E-29	Rab11-PA	0
GPAI012651	-2.27	0.001	18.82	44.47	Mitochondrial inner membrane protein cox18	1.41E-68	Sphinx1	GMOY010124	3.74E-25	-	-
GPAI045393	-2.31	0.001	18.26	43.92	Ring finger protein nhl-1	2.70E-82	-		-	-	-
GPAI002812	-2.31	0.01	12.21	29.4	Calcium calmodulin-dependent protein kinase type 1	1.21E-168	Par-1	GMOY002090	8.32E-31	Par-1-PS	0
GPAI027740	-2.37	0.004	13.32	32.9	Microtubule-associated protein rp eb family member 3	1.07E-102	Eb1	GMOY007672	0	-	-
GPAI045204	-2.37	0	119.21	295.02	Croquemort	1.58E-125	Croquemort	GMOY005165	8.97E-116	-	-
GPAI038915	-2.43	0	119.49	302.36	Serine protease sp24d	1.20E-58	-	GMOY010861	-	Spheroide-PA	0
GPAI013797	-2.45	0.006	11.26	28.76	Heat shock protein 67b2	3.90E-36	-		-	-	-

GPAI031847	-2.47	0.031	8.14	20.93	l-asparaginase	3.57E-160	Multiple ankyrin repeats single KH domain	GMOY002483	1.58E-14	Mask-PC	0
GPAI043307	-2.54	0	17.75	47	Tyrosine-protein kinase btk29a	0	heartless	GMOY007684	7.70E-43	Hop-PA	0
GPAI040738	-2.57	0	17.44	46.81	Harmonin	9.94E-31	PDZ domain-containing guanine nucleotide exchange factor	GMOY006487	4.70E-06	CG6509-PA	0
GPAI001250	-2.59	0.048	6.4	17.32	Protein-l-isoaspartate(d-aspartate) o-methyltransferase	1.06E-112	-		-	Pcmt-PA	0
GPAI047877	-2.63	0.001	12.28	33.67	Dual specificity tyrosine-phosphorylation-regulated kinase 2	7.23E-103	Cyclin-dependent kinase 5	GMOY010352	1.06E-08	Srpk79D-PN	0
GPAI018260	-2.64	0.009	8.62	23.74	Mitochondrial 2-oxodicarboxylate carrier	9.48E-115	Hemipterous	GMOY008135	0	-	-
GPAI040425	-2.72	0.03	6.37	18.09	Endocuticle structural glycoprotein bd-1	3.52E-22	Cuticular protein 49Ac	GMOY005324	1.07E-09	-	-
GPAI010579	-2.82	0.018	6.54	19.23	Serine protease easter	1.74E-29	-	GMOY002535	-	SPE-PA	0
GPAI026455	-2.87	0	32.39	97.03	ATP-dependent rna helicase me31b	0	Rm62	GMOY004318	2.17E-51	Rm62-PI	0
GPAI012725	-2.89	0.036	5.28	15.94	Histone h4 transcription factor	4.66E-79	Zn finger homeodomain 1	GMOY006469	2.16E-06	Zfh1-PB	0

GPAI026443	-2.9	0.038	5.19	15.69	Katanin p60 atpase-containing subunit a-like 1	0	Spastin	GMOY009501	6.62E-51	-	-
GPAI032726	-2.91	0.014	6.36	19.31	Probable atp-dependent rna helicase ddx10	0	Rm62	GMOY004318	1.55E-33	Rm62-PI	0
GPAI017996	-2.92	0	14.6	44.4	Zinc finger protein 566	4.64E-41	Zn finger homeodomain 1	GMOY006469	6.39E-22	Zfh1-PA	0
GPAI020540	-2.98	0	14.42	44.85	Proliferating cell nuclear antigen	1.59E-176	Mutagen-sensitive 209	GMOY003241	0	Mus209-PA	0
GPAI012599	-3.02	0.03	5.01	15.75	DNA replication factor cdt1	7.55E-73	Double parked	GMOY007729	0	Dup-PA	0
GPAI000460	-3.04	0.005	6.72	21.35	Major facilitator superfamily domain-containing protein 1	1.08E-166	Kismet	GMOY001860	2.76E-11	-	-
GPAI011862	-3.07	0.011	5.84	18.69	Cytochrome p450	0	-		-	Cyp6a13-PA	0
GPAI008344	-3.08	0	35.68	114.62	Trypsin eta	1.67E-41	Sphinx1	GMOY010124	1.79E-16	Spirit-PD	0
GPAI030991	-3.11	0	11.65	37.79	Protein piwi	0	Argonaute 2	GMOY004940	1.73E-12	AGO2-PE	0
GPAI027612	-3.19	0.007	5.9	19.62	Protein pellino	0	Pellino	GMOY007135	0	-	-
GPAI033230	-3.22	0	56.7	190.2	Transferrin	0	-		-	Tsf3-PA	0
GPAI034602	-3.29	0	10.43	35.81	Wee1-like protein kinase	0	Mekk1	GMOY011804	7.34E-11	Par-1-PT	0
GPAI040468	-3.41	0	21.48	76.45	Endocuticle structural glycoprotein bd-2	1.04E-12	Cuticular protein 49Ac	GMOY005324	1.77E-08	-	-
GPAI037499	-3.46	0	33.42	120.57	Cysteine proteinase cg12163	0	Secreted Wg-interacting	GMOY006991	4.80E-12	-	-

							molecule				
GPAI034724	-3.7	0	662.69	2555.15	TLD domain-containing protein 2	4.51E-59	-		-	Mtd-PI	0
GPAI046500	-3.7	0	229.25	884.95	Serine protease sp24d	8.05E-34	Spheroide	GMOY005310	8.27E-30	Spheroide-PA	0
GPAI038923	-3.84	0	9.44	37.75	Serine protease sp24d	3.32E-39	Spheroide	GMOY005310	8.88E-39	Spheroide-PA	0
GPAI029165	-4.02	0	15.65	65.66	Cytochrome b561	7.21E-28	-		-	-	-
GPAI029774	-4.42	0	147.23	678.15	Ejaculatory bulb-specific protein 3	1.40E-47	-		-	PebIII-PA	0
GPAI031817	-4.42	0	15.62	71.97	Scavenger receptor class b member 1	7.98E-84	Croquemort	GMOY005165	7.76E-38	-	-
GPAI023936	-4.55	0	10.74	50.9	Phosphatidylethanolamine-binding protein homolog	3.17E-62	-		-	Pebp1-PA	0
GPAI006396	-5.09	0	8.15	43.25	Ras-related and estrogen-regulated growth inhibitor-like protein	1.79E-29	Roughened	GMOY000821	1.19E-06	Rab11-PA	0
GPAI019955	-5.15	0	5.65	30.37	Protease inhibitor	6.80E-20	-		-	-	-
GPAI046151	-5.55	0	11.22	64.93	Zinc finger protein 865	1.08E-05	Longitudinals lacking	GMOY001502	7.05E-05	lola-PE	0

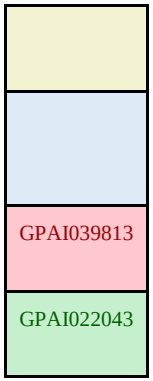
	GPAI035068	-5.93	0	5.34	33.02	Von willebrand factor type egf and pentraxin domain-containing protein 1	1.96E-22	Scavenger receptor class C, type I	GMOY002720	0	Sr-CI-PA	0
	GPAI000012	-6.1	0	20.27	128.98	Icarapin-like	9.99E-06	Virus-induced RNA 1	GMOY003759	5.55E-117	Vir-1-PH	0
	GPAI019770	-6.93	0	5.55	40.13	Defensin-a	3.55E-43	-		-	Def-PA	0
	GPAI014520	-7.17	0	5.83	43.56	Zinc finger protein 512b	3.06E-07	-		-	-	-
	GPAI018121	-7.71	0	2.64	21.21	Protein toll	0	Toll	GMOY011790	0	-	-
	GPAI011288	-9.7	0	12.63	127.71	Eukaryotic translation initiation factor 4e-binding protein 1	5.20E-18	Thor	GMOY000206	0	Thor-PA	0
	GPAI040427	-62.41	0	234.19	15240.23	Endocuticle structural glycoprotein abd-4	1.82E-36	Cuticular protein 49Ac	GMOY005324	0.00082922	-	-
	GPAI007525	-89.97	0	14.62	1371.68	Endocuticle structural protein bd-6	8.60E-25	Cuticular protein 49Ac	GMOY005324	2.12E-05	-	-
48hr challenged* vs unchallenged gut**	GPAI014246	4.24	0	64.6	16.88	Metallopeptidase 1	1.04E-106	ECSIT	GMOY007478	3.28E-06	-	-
	GPAI014735	3.58	0	209.15	64.77	Chymotrypsin-1	8.88E-62	Spheroide	GMOY005310	3.54E-19	Spirit-PD	1.25E-24

GPAI003879	3.33	0	101.12	33.66	Myc protein	2.34E-27	Immune induced molecule 4	GMOY002876	0.0021725	-	-
GPAI046220	3.29	0.01	19.56	6.59	Ring-box protein 1a	6.39E-73	Hemese	GMOY004358	8.55E-17	-	-
GPAI027452	3.26	0	29.53	10.03	Leucine-rich repeat and fibronectin type iii domain-containing protein 1	2.91E-26	Tollo	GMOY000543	4.17E-14	Toll-7-PA	2.15E-11
GPAI008007	3.13	0.001	29.23	10.34	Serine threonine-protein kinase tousled-like	0	Tousled-like kinase	GMOY005326	0	Tlk-PI	0
GPAI022043	3.13	0.006	22.47	7.95	Sh3 and multiple ankyrin repeat domains protein 3	2.01E-87	Notch	GMOY004815	4.20E-06	N-PB	1.02E-08
GPAI010960	3.07	0	54.15	19.57	AP2-associated protein kinase 1	1.68E-134	Downstream of raf1	GMOY011401	1.99E-09	Par-1-PP	2.10E-12
GPAI045098	3.06	0.05	15.78	5.71	Disintegrin and metalloproteinase domain-containing protein 10	1.59E-75	kuzbanian	GMOY006223	3.19E-70	-	-
GPAI046158	2.94	0.007	23.83	9	Longitudinals lacking isoforms f i k t	3.52E-20	longitudinals lacking	GMOY001502	7.44E-54	Lola-PI	7.04E-23
GPAI022873	2.74	0	104.82	42.42	Serine protease easter	8.35E-32	Serine protease 7	GMOY006266	1.64E-46	SPE-PA	5.05E-29

GPAI021165	2.62	0.006	28.06	11.86	Rho guanine nucleotide exchange factor 17	3.91E-118	Vav ortholog (H. sapiens)	GMOY003538	5.30E-06	-	-
GPAI029535	2.58	0.019	24.2	10.38	Tyrosine-protein phosphatase 99a	0	Myopic	GMOY011192	1.39E-06	Mop-PA	1.30E-10
GPAI016346	2.55	0.002	33.7	14.67	Suppressor of cytokine signaling 4	3.38E-76	-		-	-	-
GPAI042995	2.54	0	54.52	23.76	Ras-responsive element-binding protein 1	3.60E-19	Zn finger homeodomain 1	GMOY006469	1.22E-14	Zfh1-PA	4.12E-07
GPAI040205	2.53	0	60.24	26.35	CD109 antigen	4.94E-46	Thioester-containing protein 2	GMOY010996	0	Tep2-PD	1.98E-159
GPAI040411	2.48	0	222.37	99.39	Myosin heavy non-muscle	0	-		-	Myo61F-PB	9.27E-100
GPAI020801	2.47	0.005	32.22	14.48	Nucleoprotein tpr	0	Multiple ankyrin repeats single KH domain	GMOY002483	0.0047118	-	-
GPAI003056	2.44	0.033	23.85	10.82	Serine threonine-protein kinase tao1	7.96E-58	Licorne	GMOY001289	2.18E-18	-	-
GPAI005826	2.42	0	128.73	59	Autophagy-related protein 13 homolog	0	-		-	-	-
GPAI034601	2.32	0	60.78	29.08	Heterogeneous nuclear ribonucleoprotein 27c	0	ECSIT	GMOY007478	3.10E-12	-	-
GPAI007049	2.24	0.048	25.73	12.72	Rho-associated protein kinase 2	0	Atypical protein kinase C	GMOY012181	7.06E-45	Gprk2-PB	1.26E-33

GPAI011404	2.23	0.001	49.67	24.69	E3 ubiquitin-protein ligase trip12	0	-		-	-	-
GPAI033230	2.21	0	1304.24	653.84	Transferrin	0	-		-	Tsf3-PA	5.49E-53
GPAI037392	2.16	0.003	43.49	22.35	Chromodomain-helicase-dna-binding protein 1	0	Brahma	GMOY007000	3.03E-113	Kis-PD	2.47E-138
GPAI034740	2.16	0	56.81	29.21	Segmentation protein cap n collar	1.88E-141	Kayak	GMOY007416	1.69E-07	Kay-PB	3.98E-07
GPAI034487	2.11	0	74.04	38.92	Transcription factor hnf-4 homolog	0	-		-	Eip75B-PC	2.93E-27
GPAI003078	2.09	0.003	46.27	24.55	Gastrula zinc finger protein	1.06E-06	Zn finger homeodomain 1	GMOY006469	1.17E-05	Zfh1-PA	0.0022139
GPAI002368	2.05	0	591.92	319.8	Heat shock protein 83	0	-		-	-	-
GPAI031988	2.04	0	63.81	34.75	Ras gtpase-activating protein	0	-		-	-	-
GPAI045204	2.01	0	630.95	347.56	Protein Croquemort	1.58E-125	Croquemort	GMOY005165	8.97E-116	-	-
GPAI040425	-2.22	0.023	13.67	33.65	Endocuticle structural glycoprotein bd-1	0	Cuticular protein 49Ac	GMOY005324	1.07E-09	-	-
GPAI012728	-2.22	0	17177	42339.32	Serine protease sp24d	0	Spheroide	GMOY005310	1.85E-47	Spheroide-PA	0
GPAI046500	-2.32	0	484.51	1247.18	Serine protease sp24d	0	Spheroide	GMOY005310	8.27E-30	Spheroide-PA	0
GPAI014231	-2.37	0.013	12.54	32.93	Ubiquitin-fold modifier-conjugating enzyme 1	0	-		-	-	-
GPAI033252	-2.76	0	55.56	170.02	Lectin subunit alpha	0	Galactose-specific C-type lectin	GMOY000781	1.01E-14	Lectin-galC1-PB	0

GPAI039398	-2.82	0	1140.84	3568.85	Trypsin	0	Spheroide	GMOY005310	5.58E-21	Spirit-PD	0
GPAI008345	-3.02	0.001	10.18	34.1	Trypsin-1	0	Sphinx1	GMOY010124	1.47E-21	CG6639-PA	0



Up-regulated genes

Down-regulated genes

Toll pathway genes

Imd pathway genes

-

-

-

Appendix IX. Predominantly (90 percentile) differentially expressed genes in guts and carcasses of teneral *G. pallidipes* tsetse flies challenged with *T. b. brucei* parasites 24 or 48 hours

***T. b. brucei* Challenge**	**Vectorbase Acccession No.**	**Fold Change**	**FDR Corrected p-Value**	**Treatment* - RPKM**	**Reference**- RPKM**	**Best BLAST Hit, None-Redundant (nr) NCBI database**	**E - Value**
24hr challenged* vs unchallenged gut**	GPAI036743	9.67	0	279.75	30.17	Abhydrolase domain-containing protein 3	1.42E-96
	GPAI020062	8.71	0	840.50	100.63	MID1-interacting protein 1	7.34E-04
	GPAI006387	8.51	0	2631.58	322.57	Glutamine synthetase mitochondrial	0
	GPAI040204	7.59	0	90.64	12.45	Myosin-viia	0
	GPAI030960	5.78	0	184.57	33.32	Flagellar attachment zone protein 1	3.76E-08
	GPAI008216	5.41	0	120.85	23.29	Protein serac1	7.01E-37
	GPAI025756	5.39	0	216.61	41.89	l-ascorbate oxidase	4.72E-51
	GPAI003006	5.17	4.77489E-14	61.46	12.40	Organic cation transporter protein	2.07E-20
	GPAI016956	4.78	0	200.80	43.76	Mitochondrial glutamate carrier 1	5.66E-116
	GPAI013696	4.65	9.67144E-07	31.90	7.16	Sparc-related modular calcium-binding protein 1	4.06E-26
	GPAI039815	4.53	0	181.16	41.66	Glycogen debranching enzyme	0
	GPAI034320	4.48	2.07554E-08	40.87	9.51	Organic cation transporter protein	7.02E-99
	GPAI043406	4.43	0	2870.01	675.59	Myosin-8	0.261954
	GPAI030117	3.89	1.76062E-09	51.85	13.90	Leucine-rich repeat-containing protein 15	2.15E-22
	GPAI048071	3.88	0	216.83	58.30	Alkaline phosphatase 4	0
	GPAI040274	3.80	3.63417E-12	67.95	18.63	Carbohydrate sulfotransferase 7	0.00195991

GPAI038505	3.60	0	278.06	80.55	ATP-dependent zinc metalloprotease yme1 homolog	0
GPAI040207	3.58	0	675.83	197.08	CD109 antigen	4.27E-50
GPAI000767	3.36	0	132.84	41.23	Chaperone protein	0
GPAI029826	3.36	0	612.76	190.42	Indirect flight muscle	0
GPAI025179	3.31	0	224.77	70.76	---NA---	
GPAI043944	3.28	0	988.94	314.23	Xaa-pro dipeptidase	0
GPAI002368	3.25	0	694.20	222.51	Heat shock protein 83	0
GPAI045866	3.24	0	424.40	136.71	---NA---	
GPAI027766	3.15	0	263.31	87.20	Inorganic phosphate cotransporter	1.62E-86
GPAI039274	3.09	0.000978895	25.50	8.61	Tankyrase	0
GPAI016948	3.08	9.04146E-06	39.96	13.51	Solute carrier family 35 member f6	1.99E-34
GPAI033707	3.00	1.99132E-09	67.86	23.57	E3 ubiquitin-protein ligase su	0
GPAI022045	-4.30	8.00488E-05	5.56	24.96	---NA---	
GPAI046173	-4.34	0	52.21	236.43	Carnitine o-palmitoyltransferase liver isoform	0
GPAI029774	-4.42	0	147.23	678.15	Ejaculatory bulb-specific protein 3	1.40E-47
GPAI031817	-4.42	4.77489E-14	15.62	71.97	Scavenger receptor class b member 1	7.98E-84
GPAI023188	-4.44	1.16071E-09	10.67	49.39	Fumarylacetoacetase	0
GPAI002755	-4.44	0	51.51	238.62	Cysteine-tRNA ligase	5.06788
GPAI006888	-4.78	3.2777E-12	11.89	59.19	Neprilysin-2	1.46E-165
GPAI009340	-5.17	1.10768E-07	6.55	35.27	---NA---	
GPAI023464	-5.50	0	51.90	297.50	---NA---	
GPAI009124	-5.74	0	79.40	475.46	Guanine nucleotide-releasing factor 2	0.255612
GPAI000012	-6.10	0	20.27	128.98	Icarapin-like	9.99E-06

	GPAI011288	-9.70	0	12.63	127.71	Eukaryotic translation initiation factor 4e-binding protein 1	5.20E-18
	GPAI009911	-11.46	0	5.90	70.46	Sco-spondin	2.16E-06
	GPAI009914	-13.15	0	12.66	173.65	Sco-spondin	7.46E-05
	GPAI009286	-14.24	0	6.02	89.40	PTB domain-containing adapter protein ced-6	0
	GPAI045765	-17.43	0	11.76	213.64	Synaptic vesicle glycoprotein 2b	4.13E-33
	GPAI045763	-27.98	0	7.24	211.35	Synaptic vesicle glycoprotein 2a	2.57E-30
	GPAI040427	-62.41	0	234.19	15240.23	Endocuticle structural glycoprotein abd-4	1.82E-36
	GPAI007525	-89.97	0	14.62	1371.68	Endocuticle structural protein bd-6	8.60E-25
24hr challenged* vs unchallenged carcase**	GPAI027547	13.35	0	81.51	6.13	Protein fam192a	8.43E-29
	GPAI028027	5.82	0	62.42	10.76	Uncharacterized protein	1.67137
	GPAI043373	5.19	0	992.88	192.20	Adult cuticle protein 1	5.70E-10
	GPAI006838	4.25	0	1956.51	461.77	Cuticle protein 6	4.91E-10
	GPAI030973	3.78	0	152.58	40.57	D-3-phosphoglycerate dehydrogenase	1.01E-111
	GPAI043351	3.36	0	1376.49	410.97	Adult cuticle protein 1	1.28E-10
	GPAI013819	3.19	0	271.17	85.27	Bifunctional purine biosynthesis protein purh	0
	GPAI021719	3.16	0	213.27	67.78	---NA---	
	GPAI027343	-3.65	0	25.60	93.72	NADH dehydrogenase	5.60E-29
	GPAI025883	-4.06	0	20.59	83.94	Ras-related protein rab-2a	8.78E-139
	GPAI016233	-4.86	0	31.89	155.63	ATP-binding cassette sub-family g member 1	3.52E-96
	GPAI042034	-5.20	0	15.06	78.57	---NA---	

	GPAI011288	-7.25	0	16.36	119.03	Eukaryotic translation initiation factor 4e-binding protein 1	5.20E-18
	GPAI045765	-9.08	0	29.95	273.09	Synaptic vesicle glycoprotein 2b	4.13E-33
	GPAI011054	-11.55	0	7.33	85.02	Arylphorin subunit c223	0
48hr challenged* vs unchallenged gut**	GPAI025756	4.90	0	1069.26	241.64	l-ascorbate oxidase	4.72E-51
	GPAI014246	4.24	3.39372E-12	64.60	16.88	Endoplasmic reticulum metallopeptidase 1	1.04E-106
	GPAI004211	3.60	0.000980928	24.12	7.43	Protein split ends	1.97E-163
	GPAI021719	3.59	7.06316E-07	43.19	13.34	---NA---	
	GPAI014735	3.58	0	209.15	64.77	Chymotrypsin-1	8.88E-62
	GPAI022855	3.56	0	1348.61	419.32	C3 and pzp-like alpha-2-macroglobulin domain-containing protein 8	3.13E-04
	GPAI047844	3.47	0	99.36	31.73	---NA---	
	GPAI018204	3.47	4.3781E-05	33.65	10.76	Adenomatous polypis coli protein	7.69E-139
	GPAI002694	3.38	0	227.12	74.38	Myosin-2 heavy chain	2.00371
	GPAI003879	3.33	0	101.12	33.66	Myc protein	2.34E-27
	GPAI012449	3.30	0	122.14	41.08	---NA---	
	GPAI046220	3.29	0.009964058	19.56	6.59	Ring-box protein 1a	6.39E-73
	GPAI046445	3.20	0.048050395	15.15	5.25	---NA---	
	GPAI027974	3.14	0.001830228	26.07	9.21	Enhancer of polycomb homolog 1	6.50E-111
	GPAI022043	3.13	0.005719855	22.47	7.95	SH3 and multiple ankyrin repeat domains protein 3	2.01E-87
	GPAI010220	3.12	2.79982E-13	95.14	33.78	LA-related protein 1	0
	GPAI010960	3.07	2.83921E-07	54.15	19.57	AP2-associated protein kinase 1	1.68E-134

	GPAI039965	2.98	0	1232.73	458.95	Immunoglobulin superfamily containing leucine-rich repeat protein 2	1.93E-08
	GPAI002686	2.97	0	116.63	43.51	---NA---	
	GPAI043644	2.88	0.017239711	21.00	8.08	---NA---	
	GPAI004663	2.80	3.38158E-05	44.43	17.61	Tiggrin	0
	GPAI035562	-2.95	3.66643E-13	34.19	111.82	26s proteasome non-atpase regulatory subunit 12	4.42E-154
	GPAI021775	-3.43	0	222.31	844.12	Chondroitin proteoglycan 1	0.708089
48hrs challenged* vs unchallenged carcase**	GPAI025756	2.99	7.3218E-11	82.87	27.21	I-ascorbate oxidase	4.72E-51
	GPAI027712	1.89	0	861.73	448.13	Activity-regulated cytoskeleton-associated protein	0.00771969
	GPAI047125	-2.96	0.002307459	10.02	29.10	Trophinin	4.59E-04
	GPAI035199	-3.72	0	31.16	113.78	Defective chorion-1 fc106 isoform	4.99E-14

APPENDIX X. List of publications.

1. Echodu R., Sistrom M., Bateta R., Murilla G., Okedi L., Aksoy S., Enyioha C., Enyaru J., Opiyo E., Gibson and, Caccone A (2014). Genetic diversity and population structure of Trypanosoma brucei in Uganda: implications for the epidemiology of sleeping sickness and Nagana. *PLOS Neglected Tropical Diseases:* 9(2): e0003353
2. International *Glossina* Genome Initiative (2014). Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis *Science*. 344:380-386