The Population Genetics and Genomics of the African Salivarian Trypanosomes

Craig William Duffy

Submitted in fulfilment of the requirements for the

Degree of Doctor of Philosophy

College of Medical, Veterinary and Life Sciences

University of Glasgow

Submitted September 2011

Submitted with corrections January 2012

Abstract

The African Salivarian trypanosomes are the causative agents of both Human African Trypanosomiasis, or sleeping sickness, and Animal African Trypanosomiasis, more widely known as Nagana. Primarily spread through the tsetse fly vector both diseases are distributed across the sub-Saharan tsetse belt, afflicting some of the poorest communities in the world. Three species of trypanosomes are predominantly responsible for these two diseases. *T. brucei*, which is comprised of the three morphologically identical sub-species *T. b. brucei*, *T. b. gambiense* (further separated into two subgroups) and *T. b. rhodesiense*, with the latter two sub-species exclusively responsible for infections in humans. The animal infective species *T. congolense*, comprised of the Forest, Kilifi and Savannah subtypes, and *T. vivax* meanwhile are responsible for millions of livestock and wild animal infections across the continent, with severe downstream economic consequences.

A crucial component in understanding the diseases caused by these parasites is through understanding the diversity present in the field, as it is ultimately the combination of host, vector and parasite diversity that gives rises to the disease phenotypes observed during clinical diagnosis and treatment. In order to truly understand the role of such diversity in the field it is necessary to know how individuals within a population interact with one another, if they do at all. Mating between individuals allows for the direct interaction of genomes, allowing for the generation of new chromosomal sequences through meiotic recombination and new chromosomal pairings through bi-parental inheritance of genetic material. Identified as a non-obligatory process in *T. brucei* the importance of mating in natural trypanosome populations is both a controversial and understudied topic despite the significant role of the process in shaping the evolutionary development of these clinically important parasites.

In order to further investigate the genetic diversity and role of mating in the trypanosomes populations from The Gambia, Uganda and Malawi have been examined through the use of microsatellite markers specific to the genomes of *T. brucei*, *T. congolense* and *T. vivax*. The results presented here demonstrate drastically different levels of diversity in the respective populations and evidence for a spectrum of genetic exchange, with both highly clonal and frequently mating populations identified in this manner.

T. vivax, sampled from horses, donkeys and cattle in The Gambia would appear to most closely fit with the traditional views of clonality in trypanosomes, extensive clonal reproduction of a single genotype, significant disagreement with Hardy-Weinberg principles and the presence of significant linkage between the loci examined. These results, which closely resemble those observed for *T. b. gambiense* Group 1, suggest that genetic exchange may be absent or rare in *T. vivax*, which may

lead to the eventual divergence of independent populations as they slowly accumulate unique mutations. The apparent dominant clonality of *T. vivax* is a sharp contrast to the situation observed for *T. congolense* in The Gambia, with strong evidence for frequent mating and a high rate of inbreeding. That this evidence originated from the same sample sets used in the *T. vivax* studies presented here highlights the differences between these two species and the requirement for further work independent of the studies into *T. brucei*.

The final half of this thesis has focused upon the population genetics and genomics of *T. brucei*, the species responsible for sleeping sickness in humans. Examination of five of *T. b. rhodesiense* populations, four from Uganda and one from Malawi has demonstrated the potential for variation in the population structure within a single species. The Ugandan populations are dominated by clonality; with repeated bottlenecks reducing the genetic diversity present as the parasites has spread northwards. The Malawi population, genetically distinct from the populations of Uganda, instead appears to favour genetic exchange over clonality, with a genetically diverse population and only a limited number of repeated genotypes. This provides the first evidence of mating playing a significant role in a field population of human infective trypanosomes, introducing a significant role for meiotic recombination and chromosomal reassortment which may drastically alter the way in which these parasites respond to selective pressures and evolutionary forces. Finally, this thesis has aimed to bridge the gap between traditional low resolution studies and the developing field of genomics by examining the SNP variation present between three laboratory strains of *T. brucei*, providing the building blocks in understanding genome wide variation in trypanosomes. Utilising these data, and through sequencing of progeny generated in the process of constructing the TREU 927 genetic map, it has been possible to partially reassemble the haplotypes for the megabase chromosomes of this strain, previously selected as the *T. brucei* genome reference strain. Collected together these data provide an important resource of genomic variation for both laboratory studies and as a baseline for future investigations into the genomic diversity of field populations.

In summary this thesis has demonstrated the variable nature and versatile role of genetic exchange in the trypanosomes, bringing together data not only from the human infective sub-species of *T. brucei* but from the animal infective species *T. congolense* and *T. vivax*. Finally in looking to the future this work has begun the process of transitioning from the relatively low density microsatellite markers by examining high density SNP variation in common laboratory strains, the first step towards future adoption of these markers for the purpose of population genomics.

Dedication

For those that got me here.

Acknowledgments

I'd like to thank a number of people for getting me to this point. First and foremost my supervisor, Dr Annette MacLeod for her guidance and support in the four years of work leading up to the submission of this thesis.

I would like to thank all of the members of the Trypanosome Genetics Group for their help and friendship over the years, especially Dr Liam Morrison whose help has been invaluable in getting to grips with the trickier portions of both PCR and population analysis. I'd also like to thank Prof Bill Cushley, Prof Darren Monckton and Dr Olyn Byron of the Wellcome Trust Four Year programme for their support and for giving me the opportunity to study here at the University of Glasgow.

Thank you to my parents for getting me where I am today, for your continual support and encouragement and for putting up with me on this long journey.

Finally I'd like to thank Michelle, your support has gotten me through this and I couldn't have done it without you.

Contents

Declaration

I declare that this thesis and the results presented within it are entirely my own work, with the following exceptions:

Species identification tests of the 2007 Gambian samples were performed by Alana Black and Liam Morrison. Speciation identification tests of the 2009 Gambian samples were performed with the aid of Catherine Orr who performed approximately half of the tests. Genotyping of approximately half of the Tororo pre-1991 and Tororo 2003 populations was undertaken at five markers by Liam Morrison, Anneli Cooper and Lindsay Sweeny.

The *T. vivax* specific microsatellite markers described here were developed as part of a sample study submitted as part of an MRes qualification in 2007. All genotyping of the 2007 *T. vivax* isolates presented here were undertaken subsequent to the original study. No other part of thesis has been previously submitted for a degree at any other institution.

Supporting Publications

Duffy CW, Morrison LJ, Black A, Pinchbeck GL, Christley RM, Schoenefeld A, Tait A, Turner CM and MacLeod A. (2009). *Trypanosoma vivax* displays a clonal population structure. Int J Parasitol, 39 (13).

Abbreviations

µg	Microgram
µl	Microliter
AAT	Animal African Trypanosomiasis
apoL-1	apolipoprotein L-1
BAC	bacterial artificial chromosome
BIIT	blood incubation infectivity test
bp	base pairs
CATT	card agglutination test for trypanosomiasis
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DNA	deoxyribonucleic acid
dTTP	deoxythymidine triphosphate
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
g	Gram
GFP	green fluorescent protein
GHDT	Gambia Horse and Donkey Trust
HAT	Human African Trypanosomiasis
Hb	Haemoglobin
Hp	Haptoglobin
HpHbR	haptoglobin / haemoglobin receptor
Hpr	haptoglobin related protein
indel	insertion / deletion
ITC	International Trypanotolerance Centre
kb	Kilobase
kDNA	kinetoplast DNA
L	Litre
LAMP	loop-mediated isothermal amplification
LB	Luria-Bertani medium
Mb	Megabase
ml	Millilitre
MLG	multilocus genotype
mM	Millimolar
mm	Millimetre
PCoA	principal co-ordinate analysis
PCR	polymerase chain reaction
PCV	packed cell volume
PFGE	pulsed field gel electrophoresis
RFP	red fluorescent protein
rpm	revolutions per minute
rRNA	ribosomal RNA
SNP	single nucleotide polymorphism
SRA	serum resistance-associated
TBE	Tris-Borate-EDTA

TbgHpHbR	*T. b. gambiense* Hp/Hb receptor
TgsGP	*T. b. gambiense* specific glycoprotein
TIGR	The Institute for Genome Research
TLF-1	trypanosome lytic factor 1
TLF-2	trypanosome lytic factor 2
TP	total protein
tritrypDB	Tri-Trypanosome database
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
UV	ultraviolet
v/v	volume per volume
VSG	variable surface glycoprotein
w/v	weight per volume
WGA	whole genome amplification
WHO	World Health Organisation
WTSI	Wellcome Trust Sanger Institute

Chapter 1

Introduction

1.1 General Introduction

The Salivarian trypanosomes, single cellular protozoan parasites, are members of the order Kinetoplastida, genus *Trypanosoma*. The genus encompasses almost 500 different named species, globally dispersed and with an extensive host range in both vertebrates and invertebrates. The majority of these species are able to infect, replicate and survive in their hosts without adverse effects, and thus rarely come to our attention (Hoare, 1972; Cox, 1993). Infection by a select few does however result in noticeable disease states in the host species. African sleeping sickness and Chagas disease, because of their status as human diseases, are the most widely known of these. These two diseases are the result of infections by *Trypanosoma Trypanozoon brucei* (*T. (T.) brucei*) and *Trypanosoma Schizotrypanum cruzi* (*T. (S.) cruzi*), respectively. The species of the genus *Trypanosoma* that are infective to mammalian hosts have been classified into two groups based upon their method of transmission between hosts (Hoare, 1972). Those of the Stercoraria group, which includes *T. cruzi*, are transmitted to the mammalian host by contamination of bites or wounds with faeces from the insect vector. In contrast to the Stercoraria group, members of the Salivarian group, the focus of this work, are almost exclusively transmitted through the saliva of the insect vector following biting of, and feeding upon mammalian host species.

The primary insect vectors of the Salivarian trypanosomes are the species of the genus *Glossina*, collectively known as tsetse flies (Krafsur, 2009). These large biting flies survive on blood meals from vertebrate animals and are found throughout the majority of sub-Saharan Africa within a region commonly referred to as the tsetse belt. This region spans 37 countries and an area of approximately 8.5 million km^2, approximately 28% of the African continent (Allsopp, 2001). During transmission by tsetse the Salivarian trypanosomes undergo a series of specific developmental changes capable of establishing lasting infections within the flies, a process referred to as cyclical transmission. A secondary method of transmission, lacking these developmental changes, is possible and involves uptake of bloodstream form trypanosomes during one meal and their transfer into new hosts during subsequent feedings. This mechanical transmission does not require establishment of an infection within the insect vector and can therefore use biting flies other than the tsetse (Wells, 1972). While less frequent, mechanical transmission has allowed the expansion of some Salivarian trypanosomes beyond the ranges of the tsetse belt. Amongst the Salivarian trypanosomes only *Trypanosoma equiperdum* does not use an insect vector, instead being transmitted between equines through venereal contact (Brun *et al.*, 1998).

Traditionally the members of the *Trypanosoma* genus have been grouped and defined by the techniques available at their time of discovery. At the turn of the twentieth century, a period of mass discovery, the primary methods were morphology, pathogenicity, host range and geographical distribution. With these diverse criteria it is perhaps unsurprising that many of these species were

later re-evaluated, most notably by Hoare (1972) who amalgamated many of the species and subgenera into those still used today. The Salivarian trypanosomes are separated into four subgenera, *Duttonella, Nannomonas, Pycnomonas* and *Trypanozoon* (Table 1.1). This work is focused specifically upon three species, *T. brucei*, the only Salivarian trypanosome to possess human infective variants; *Trypanosoma Duttonella vivax,* a pathogen of livestock and the most widespread of the Salivarian trypanosomes and *Trypanosoma Nannomonas congolense*, which alongside *T. vivax* is one of the primary causative agents of animal African trypanosomiasis.

Subgenus	Species
Duttonella	*Trypanosoma vivax*
	Trypanosoma uniforme
Nannomonas	*Trypanosoma congolense*[1]
	Trypanosoma simiae
	Trypanosoma godfreyi[2]
Pycnomonas	*Trypanosoma suis*
Trypanozoon	*Trypanosoma brucei*[3]
	Trypanosoma evansi
	Trypanosoma equiperdum

Table 1.1 The recognised species of the Salivarian trypanosomes.

The currently recognised subgenera and species of the Salivarian trypanosomes based primarily upon the definitions of Hoare (1972). [1] Presently recognised to be comprised of the three sub-types Forest, Kilifi and Savannah. [2] Characterised by McNamara (1994). [3] Species complex comprised of the three sub-species *T. b. brucei*, *T. b. gambiense* and *T. b. rhodesiense* with *T. b. gambiense* further subdivided into Group 1 and 2.

1.1.1 The phylogeny of trypanosomes

While the question of species appeared to have been resolved by work of Hoare (1972), determination of the phylogenetic relationships between the various species of trypanosome required the development of molecular sequencing. The results of early rRNA sequencing suggested the genus *Trypanosoma* to be paraphyletic, with multiple branches of origin from the *Trypanosomatids*, typically separating the African and American trypanosomes from one another (Maslov *et al.*, 1996; Stevens *et al*., 1999; Hughes and Piontkivska, 2003a, 2003b; Piontkivska and Hughes, 2005). Interestingly in these studies *T. vivax* varies in its positioning, occasionally appearing to have an independent origin from the remainder of the African Salivarian trypanosomes. In contrast, later studies, utilising data from rRNA, multiple protein families and whole genome sequencing have suggested a monophyletic origin of the *Trypanosoma* genus (Hamilton *et al.*, 2004; Simpson *et al.*, 2006; Leonard *et al.*, 2011). These results, upon which the weight of evidence now falls, date the divergence of *T. brucei* and *T. cruzi*, and therefore their respective clades, to an estimated 100 million years before present (Stevens and Gibson, 1999; Stevens *et al.*, 1999). Within the Salivaria the reconstructed monophylogenies support an early separation of the lineage giving rise to *T. vivax* with a closer relationship between the *T. brucei* and *T. congolense* lineages.

The same technologies that have been used to assess the relationships between species have reopened the question of species definition as, with the examination of further field samples increasing diversity has been observed within the African trypanosomes (Malele *et al.*, 2003; Adams *et al.*, 2008). Classically only *T. brucei* has been represented by multiple sub-species, *T. b. gambiense* (of which there are two recognised subgroups), *T. b. rhodesiense* and *T. b. brucei*, with *T. b. gambiense* and *T. b. rhodesiense* originally defined by their ability to survive within human hosts and distinct geographic origins. *T. congolense* is now recognised to be comprised of at least three sub groups (Forest, Savannah and Kilifi) (Garside *et al.*, 1994) based upon isoenzyme and genetic analysis while there is evidence for similar sub-groups of *T. vivax* within a single country (Adams *et al.*, 2009). Only a single new species, *Trypanosoma Nannomonas godfreyi* has been recognised following isolation and characterisation of this trypanosome (McNamara *et al.*, 1994). While originally identified from tsetse in The Gambia, the species is now recognised to have a wide distribution across Africa (Masiga *et al.*, 1996). The development of species-independent primer sets has indicated that a substantial number of trypanosome species, or subtypes of known species, have yet to be fully described or identified (Malele *et al.*, 2003; Adams *et al.*, 2006, 2008, 2009; Adams and Hamilton, 2008; Hamilton *et al.*, 2008) and it has been suggested that it may be necessary to re-evaluate the recognised species, sub-species, clades and groups to take account of these new data and clearly define the boundaries between them (Gibson, 2003, 2007).

1.2 The life cycles of *T. brucei, T. congolense* and *T. vivax*

The life cycles of *T. brucei, T. congolense* and *T. vivax* comprise two distinct segments, development within a mammalian host and transmission between hosts by an insect vector. For each of these species the primary vector is the tsetse fly, inside which the trypanosomes undergo specific developmental stages prior to their return to a new mammalian host. During the life cycle the parasites alternate between adaptive stages, prior to a change in host or vector environments and proliferative stages, which maintain the number of parasites at any given life cycle stage (Hoare, 1972; Vickerman, 1985; Vickerman *et al.*, 1988; Gardiner, 1989; Matthews, 2005; Fenn and Matthews, 2007).

1.2.1 The *T. brucei* life cycle

The classical *T. brucei* life cycle, summarised in Figure 1.1 (Adapted from Vickerman, 1985) was originally established through analysis of morphology and progression of infections in the host and vector. Infection of the mammalian host is established by initiation of a blood meal by an infected tsetse fly, during which metacyclic form trypanosomes, mixed into the saliva of the tsetse are injected into the skin of the host. This may give rise to an inflammatory reaction known as a chancre at the site of the tsetse bite. The metacyclic trypanosomes, a non-proliferative form, are pre-adapted for survival within the mammalian host and maintain a dense surface coat comprised of a single variable surface glycoprotein (VSG) which protects against initial responses from the immune system (Turner *et al.*, 1988). On reaching the bloodstream metacyclics transform into the proliferative long slender trypomastigote form, which multiplies within the bloodstream. The metacyclic VSG coat is replaced at this stage by bloodstream specific VSGs, with the trypanosome once again expressing only a single variant at a given time. It is by varying this coat, a process known as antigenic variation, that the intra-host trypanosome population is capable of evading the responses of the adaptive immune system (Vickerman, 1969; Vickerman and Luckins, 1969; Barry and McCulloch, 2001; Morrison *et al.*, 2009a). As the infection progresses and trypanosome numbers increase within the host, a proportion of the parasite population transforms again, from the long slender forms into a non-dividing short stumpy form (Reviewed in Macgregor and Matthews, 2010). The short stumpy form of *T. brucei* functions as the link between host and vector infections. It is in the tsetse developmental stages where the majority of differences are observed between the life cycles of *T. brucei, T. congolense* and *T. vivax* (Hoare, 1972).

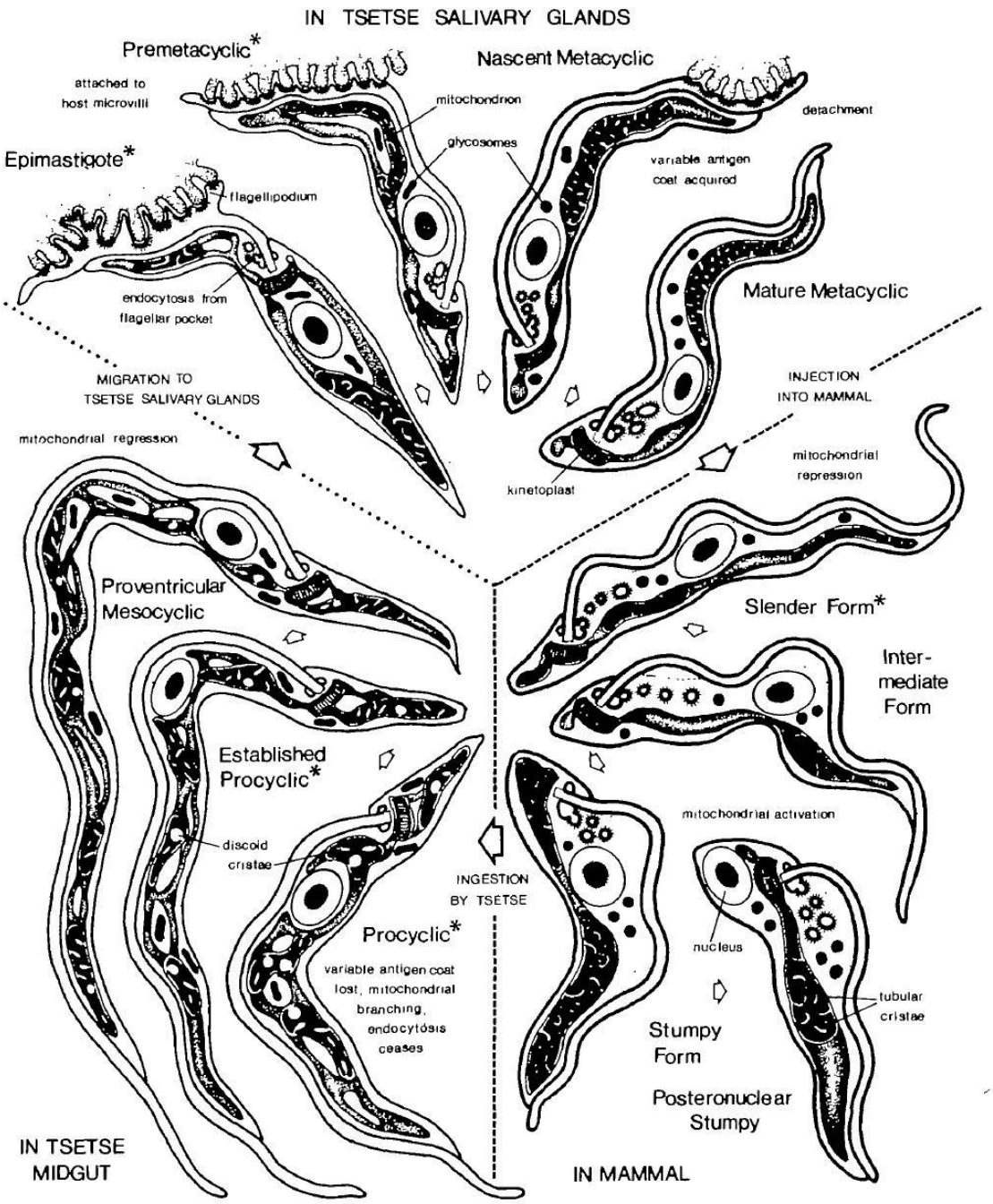


Figure 1.1 The classical *T. brucei* life cycle

Schematic diagram of the *T. brucei* developmental cycle grouped by the three major developmental locations; mammalian host, tsetse midgut and tsetse salivary glands. Developmental stages on the right half of the figure possess the VSG coat required for evasion of the mammalian immune response. * Cell proliferation occurs during these life cycle stages. Reproduced from Vickerman (1985), with more recent studies into the life cycle reviewed by Sharma *et al.* (2009).

For *T. brucei* tsetse development begins with differentiation of the trypanosomes in the gut into the procyclic form, shedding the VSG coat and replacing it with one composed of procyclins, which protects against the immune responses of the tsetse (Urwyler *et al.*, 2005). Forming an infection in the ectoperitrophic space, between the gut and the peritrophic membrane, the trypanosomes once again enter a proliferative stage before a proportion transform into the proventricular form, migrating to the salivary glands where the next form, that of the dividing epimastigotes attach to the epithelium of the salivary gland. Extensive proliferation and differentiation through a number of forms finally gives rise to the mammalian infective metacyclic form, detaching from the salivary gland in the process. These cells are then injected back into the mammalian host during future blood meals, with approximately 20 days required to reach this point after initial infection of the tsetse (Hoare, 1972; Vickerman *et al.*, 1988). Transmission through the tsetse fly involves a substantial bottleneck as trypanosomes migrate from the midgut to the salivary gland (Oberle *et al.*, 2010). In the presence of mixed infections this bottleneck can lead to significant changes in the frequency of the strains present, which may facilitate the survival of low frequency genotypes.

1.2.2 The mechanism of genetic exchange in *T. brucei*

While cyclical development represents the classically described *T. brucei* life cycle the role of mating in the life cycle should not be excluded. Early evidence of a mating cycle was presented in 1980 following isoenzyme analysis of 19 enzymes in a population of 17 isolates of *T. b. brucei* collected from Lugala, Uganda in 1969-1970 (Tait, 1980). Examination of these isolates, which contained a high proportion of all the possible allele combinations, suggested frequent and random mating of trypanosomes in the field. The small sample size of this study has however attracted criticism due to the possibility of a Type 2 error, under which agreement with Hardy-Weinberg predictions could have arisen by chance alone (Cibulskis, 1988). The existence of a mating system in trypanosomes was later confirmed by Jenni *et al.*(1986) with the first genetic cross between two strains with distinct isoenzyme and restriction fragment length polymorphism profiles. Since this first cross numerous additional crosses have been performed utilising a diverse selection of strains of *T. brucei* (Table 1.2). To date there have been no successful crosses utilising *T. b. gambiense* Group 1 reported, supporting the conclusions of field studies that suggest genetic exchange may be absent (Morrison *et al.*, 2008b; Koffi *et al.*, 2009).

While the exact events surrounding genetic exchange have yet to be elucidated the site of reproduction has been narrowed down to the salivary glands of the tsetse fly (Gibson and Whittington, 1993; Gibson and Bailey, 1994; Bingle *et al.*, 2001; Tait *et al.*, 2007). Early investigations used recombinant lines into which differing drug resistances had been conferred (Gibson and Whittington, 1993; Gibson and Bailey, 1994). Progeny which had inherited resistance

to both drugs could therefore be selected for, following isolation of trypanosomes from the differing developmental locations within the tsetse fly. These experiments indicated that the earliest stage at which double drug resistance, and therefore progeny trypanosomes, could be detected was the salivary glands with progeny generated prior to development of the metacyclic stage. Genetic exchange was therefore occurring within the salivary glands or during the migration of trypanosomes to it.

Subsequent work has expanded upon this technique, first by introduction of an inducible Green Fluorescent Protein (GFP) marker expressed only in hybrid progeny and secondly by the introduction of GFP / Red Fluorescent Protein (RFP) parental lines, with hybrid cells carrying both markers fluorescing yellow (Bingle *et al.*, 2001; Peacock *et al.*, 2007; Gibson *et al.*, 2008). In examining dissected tsetse flies infected with the fluorescent trypanosomes these studies have confirmed the salivary glands to be the site of genetic exchange, with progeny cells only detected at this site within the fly. However while a mix of the parental infections was observed in the majority of midgut infections salivary glands were observed to be highly variable, with disparity observed even between the two glands of a single fly, suggesting individual colonization of each gland by only a small number of trypanosomes. Only 37% of flies were observed to have a mixed parental infection in at least one salivary gland, limiting the potential for outcrossing and generation of progeny (Gibson *et al.*, 2008). A variation of the fluorescence technique, utilising differentially tagged populations of the same trypanosome strain has indicated that intra-strain mating of a single line is possible and does not require the presence of a second strain to trigger the mating cycle (Peacock *et al.*, 2009), in contrast to previous reports of a requirement for out crossing in order for selfing to occur (Tait *et al.*, 1996; Gibson *et al.*, 1997). Despite this Peacock *et al.*(2009) noted that, in comparison to crosses between different strains intra-strain mating appeared to produce a higher proportion of inviable progeny which could not be isolated and cloned following identification in the salivary glands. The potential therefore remains for a system allowing a strain to differentiate self from non self, capable of acting as a self limiter on the frequency of inbreeding in field populations.

Parent 1	Parent 2	Comments	Reference
STIB 247	STIB 386	First reported laboratory cross.	Jenni *et al.*(1986)
STIB 247	STIB 386	Suggested nuclear fusion to give raised DNA levels with subsequent loss to return to diploidy.	Paindavoine *et al.*(1986)
STIB 247	STIB 386	Detected increased DNA contents in hybrids compared to the parental lines.	Wells *et al.*(1987)
STIB 247	STIB 386	Suggested uniparental inheritance of the kinetoplast	Sternberg *et al.*(1988)
STIB 247	STIB 386	Observed biparental inheritance of intermediate chromosomes	Sternberg *et al.*(1989)
058	196	Suggested biparental inheritance of kDNA	Gibson (1989)
STIB 247	STIB 386	Demonstrated segregation and inheritance	Turner *et al.*(1990)
STIB 247	TREU 927/4	of alleles consistent with Mendelian	
STIB 386	TREU 927/4	inheritance.	
196	J10	Demonstrated variable chromosomal locations, and therefore inheritance of, housekeeping genes in different strains	Gibson and Garside (1991)
058	196	Analysis of trisomic and triploid progeny	Gibson *et al.*(1992)
058H	KP2N	Used drug resistant lines in order to select for double drug resistant progeny	Gibson and Whittington (1993)
STIB 247-LF	STIB 777-A	Mendelian inheritance of megabase chromosomes but not intermediate chromosomes	Schweizer *et al.*(1994)
058H	KP2N	Independent segregation of markers in drug selected progeny and high frequency of triploidy	Gibson and Bailey (1994)
STIB 831-K cl 1	STIB-831-K cl 2	Cross utilising isolates originating the same natural population in Uganda	Degen *et al.*(1995)
058H	P20 (hybrid of 058 x KP2)	First reported backcross	Gibson *et al.*(1995)
STIB 247	STIB 386	Self fertilisation of STIB 247 in presence	Tait *et al.*(1996)
STIB 247		of STIB 386	
TH2N	058H	Self fertilisation of TH2N in presence of	Gibson *et al.*(1997)
TH2N		058H	
K11 (derived from TH2)	KP2N	Used a repressed GFP reporter system with expression only in hybrids	Bingle *et al.*(2001)
STIB 247	STIB 386	Demonstrated Mendelian patterns of	MacLeod *et*
STIB 247	TREU 927	inheritance	*al.*(2005)
1738	J10	Used GFP / RFP tagged parental strains to measure dynamics of infection	Peacock *et al.*(2007)
427 variant 3	1738	Used GFP / RFP tagged parental strains, producing yellow hybrids	Peacock *et al.*(2008)
J10	1738	Used GFP / RFP tagged parental strains, producing yellow hybrids	Gibson (2008)
1738		Used GFP and RFP to detect selfing in the absence of mixed infections	Peacock *et al.*(2009)
J10			
SG3 (hybrid of 1738 x J10 cross)		Identified possible selfing of a hybrid line	
J10	1738	Identification of early meiotic cells through use of fluorescently tagged meiotic genes	Peacock *et al.*(2011)

Table 1.2 Summary of published *T. brucei* genetic crosses.

Summary of the published *T. brucei* crosses to date.

A number of genetic crosses have observed the presence of increased ploidy amongst the megabase chromosomes of the hybrid progeny (Jenni *et al.*, 1986; Paindavoine *et al.*, 1986b; Wells *et al.*, 1987; Gibson *et al.*, 1992, 1995, 1997, 2008; Gibson and Bailey, 1994; Hope *et al.*, 1999; Peacock *et al.*, 2008, 2009). This led to the suggestion that genetic exchange employed a fusion with chromosome loss model, which was first proposed by Paindavoine *et al.*(1986). This model, envisaged as a fusion event between two diploid cells followed by random chromosome loss, is similar to that proposed for *T. cruzi* based on both field and laboratory studies. In *T. cruzi* genetic exchange is proposed to be a rare fusion event, with genetic exchange between otherwise distinct clonal lineages giving rise to new ones (Gaunt *et al.*, 2003).

Two additional models invoking Mendelian inheritance have also been proposed for genetic exchange in *T. brucei*. The first requires the fusion of two diploid cells (Gibson *et al.*, 1995). In place of random chromosome loss this model proposed that following fusion of two cells the nuclei of each underwent a form of meiosis giving rise to multiple haploid nuclei within a single cell. Controlled nuclear fusion between haploid nuclei originating from each parent would then give rise to the observation of Mendelian patterns of inheritance with the remaining haploid nuclei being destroyed in the process, resulting in a single diploid cell. Under this model the observed triploid cells would arise due to additional fusion events or failure of one parent to undergo meiosis, resulting in fusion between diploid and haploid nuclei. The second Mendelian model that has been proposed for genetic exchange in trypanosomes uses a classical meiotic cycle, with generation of independent haploid cells and fusion to restore ploidy (Sternberg and Tait, 1990). Under this model the parental cells would undergo DNA duplication and homologous recombination independently of one another prior to segregation of the homologues during meiosis I, producing two daughter cells. Within each daughter cell the two sister chromatids for each chromosome are essentially identical to one another, differing only at sites which have undergone homologous recombination. A second round of cell division (meiosis II) then occurs to generate a total of four haploid gametes, two from each daughter cell. Subsequent fusion of two gametes completes the sexual cycle by restoring the 2n state and combining genetic material from either a single parent (selfing) or from two independent parents (out crossing). To date however haploid gametes have yet to be observed in the analyses of trypanosome crosses.

For both of these proposed models Mendelian inheritance is a central concept, requiring that the models follow Mendel's laws of segregation and independent assortment. The law of segregation requires that during the production of gametes (be they haploid cells or haploid nuclei) each gamete receives only a single copy of a given chromosome, thus fusion of two gametes generates a 2n progeny cell with a DNA content equal to that of the two parents and each parent contributing only half of the total genetic material. The second law, that of independent assortment, states that during the segregation of chromosomes during meiosis I the homologues of each chromosome are mixed

in a random nature, with no association between physically unlinked chromosomes. Each daughter cell therefore possesses a random assortment of chromosomal homologues, ensuring parental material is mixed during each subsequent round of sexual reproduction.

For *T. brucei* the random chromosome loss model has largely been discarded as, with an expansion in the number of progeny available for analysis it has become clear that Mendelian patterns of inheritance are observed in progeny which maintain a ploidy of 2n, inconsistent with random chromosome loss (MacLeod *et al.*, 2005a) and allowing for the generation of genetic maps for *T. b. brucei* and *T. b. gambiense* Group 2 (MacLeod *et al.*, 2005b; Cooper *et al.*, 2008). The recent work of Peacock *et al.*(2011) has provided the greatest insight into the mechanism of genetic exchange in *T. brucei* by identifying early meiotic cells, through fluorescent tagging of meiosis associated genes (Ramesh *et al.*, 2005; Schurko and Logsdon, 2008). This work demonstrated that these prophase I genes were expressed in the same order as observed in other eukaryotes and during the life cycle stage previously associated with genetic exchange in *T. brucei*. While the products of meiosis II (haploid cells or multiple haploid nuclei within a single cell) were not detected in this study it is likely that this powerful tagging technique will provide the means to observe the latter stages of genetic exchange in *T. brucei* and finally elucidate the details of this process.

For the Mendelian models employing a meiotic cycle, the observation of raised DNA contents in some hybrid lines may indicate a failure to segregate genetic material correctly, potentially due to divergence between homologous chromosomes, disrupting correct pairing at the beginning of meiosis and subsequent segregation (Gibson *et al.*, 2008). The level of divergence between homologues of a given strain may therefore explain the differing rates of aneuploidy observed in the crosses to date due to the use of different strains by different laboratories. It is also possible that incorrect segregation is merely a normal meiotic occurrence, as has been observed in humans where ~5% of pregnancies are monosomic or trisomic for at least one chromosome (Hassold *et al.*, 2007; Yanowitz, 2010). In trypanosomes the viability of many of these progeny lines would suggest that partial or full aneuploidy may be tolerated to a greater degree than in mammals, although as the products of laboratory crosses it is unclear as to whether these progeny would endure in field populations.

1.2.3 The *T. congolense* life cycle

T. congolense shares a similar developmental cycle to *T. brucei* until the migration of the proventricular form which proceeds only as far as migration to the proboscis of the insect. It is in this location that the parasite transforms into the epimastigotes, attaching to the chitinous wall and undergoing a number of rounds of proliferation. The pre-metacyclics that result from this infection then complete the life cycle by migrating to the hypopharynx where they mature into the host

infective, free swimming metacyclics (Hoare, 1972; Vickerman *et al.*, 1988). Culture conditions have now been described for the entire *T. congolense* life cycle, facilitating investigations into the species (Coustou *et al.*, 2010). Recent evidence for mating in field populations in this species and the lack of development in the tsetse salivary glands has provided the first indication that the environment present in the salivary glands is not an inherent requirement for mating in the Salivarian trypanosomes (Morrison *et al.*, 2009b). At present there have been no crosses, successful or otherwise reported for *T. congolense*.

1.2.4 The *T. vivax* life cycle

The tsetse transmitted life cycle of *T. vivax*, last extensively reviewed by Gardiner (1989) is the simplest of the trypanosomes described here. Initial mammalian infection begins as with its relatives with proliferation of bloodstream forms and build-up of the tsetse adapted 'late form' and, while typically reported as being confined to the vasculature of the mammalian host, cells have been identified within the lymph nodes, heart tissue and central nervous system, in rare cases giving rise to acute haemorrhagic infections (Magona *et al.*, 2008). Unlike *T. brucei*, where the tsetse adapted stage is represented by non-proliferating stumpy cells, these 'late' forms take on a more slender elongated morphology with a tendency to adhere to cellular material by their flagellar tips (Gardiner, 1989). This adherent nature may facilitate attachment to the inner wall of the proboscis, thereby preventing the parasite from being washed into the gut following ingestion by the tsetse.

Following ingestion by the tsetse fly the parasites differentiate into the epimastigote form, shedding their VSG coat and begin to proliferate, forming bundles of cells still attached to the wall of the proboscis. After a round of proliferation the cells detach and migrate from the inner wall of the proboscis to the hypopharynx region where they transform to the trypomastigote form and finally the metacyclic form (Hoare, 1972). The metacyclic form of *T. vivax*, covered with a metacyclic VSG coat, shares morphological similarities to the early bloodstream cells but with a reduction in the average length of the cell. At any given time a small proportion of the metacyclics will detach and become free-swimming cells, ready to be injected into a new mammalian host (Hoare, 1972; Gardiner and Wilson, 1987; Gardiner, 1989).

1.2.5 The mechanical transmission cycle

While cyclical transmission through the tsetse fly represents the predominant route of transmission for the African trypanosomes, a secondary method, mechanical transmission, is possible (Reviewed by Wells, 1972). This route, which has been demonstrated for *T. brucei*, *T. congolense* and *T. vivax*, forgoes the entirety of the insect developmental cycle, allowing for transmission of

trypanosomes by other biting flies (Roberts *et al.*, 1989; Mihok *et al.*, 1995; Sumba *et al.*, 1998; Desquesnes and Dia, 2003, 2004). Mechanical transmission is possible due to two factors, the first is that during the acquisition of a blood meal a proportion of the trypanosomes taken up have retained the proliferating bloodstream forms. Secondly, while the majority are washed into the gut and digested, a small proportion are retained within the proboscis of the fly. During subsequent blood meals these trypanosomes may be injected back into a new mammalian host, in which their existing life cycle stage allows them to establish a new infection.

Direct mechanical transmission has been demonstrated for *T. vivax* and *T. brucei* (Mihok *et al.*, 1995) and *T. congolense* (Sumba *et al.*, 1998) through use of an interrupted feeding technique, by which flies are allowed a partial feed on an infected host (typically a mouse) and then moved to an uninfected host for completion of the feed, with successful transmission establishing an infection in the second host. While the survival rate of *T. congolense* within the mouthparts of Stable flies appears to be relatively short at less than 10 minutes (Sumba *et al.*, 1998), mechanical transmission of *T. b. rhodesiense* by tsetse flies can be achieved for up to 160 minutes after feeding on infected hosts (Roberts *et al.*, 1989). Mechanical transmission has also been demonstrated through the use of fly proof enclosures to isolate herds. Through the controlled introduction of trypanosome positive animals and bloodsucking flies it is possible to detect mechanical transmission as infections are passed to uninfected animals. Experiments utilising this technique have demonstrated the ability of tabanids to mechanically transmit both *T. congolense* and *T. vivax*, with *T. vivax* successfully transmitted at a high rate (Desquesnes and Dia, 2003, 2004).

While mechanical transmission has been experimentally described for *T. brucei*, *T. congolense* and *T. vivax* the extent and significance of this transmission route is unknown for all but *T. vivax*. The geographic ranges of *T. brucei* and *T. congolense* fall exclusively within the tsetse belt of sub-Saharan Africa, strong evidence that mechanical transmission alone is insufficient to maintain populations in the absence of cyclical tsetse transmission. Epidemiological surveys of tsetse free regions within the tsetse belt typically report *T. vivax* to be responsible for >90% of observed infections when tsetse are absent, with observed incidence rates correlating with the density of biting flies in the region (Cherenet *et al.*, 2004, 2006; Sinshaw *et al.*, 2006). These results are however dependent upon the certainty of the region being tsetse free, despite being located within the tsetse belt, which is difficult to definitively prove. While mechanical transmission of *T. brucei* appears rare the species has demonstrated the ability to adapt to non-tsetse transmission, as evidenced by the existence of *T. evansi* and *T. equiperdum*. These two species, which are transmitted mechanically and venereally, are respectively believed to have originated from *T. brucei* through the partial or complete loss of the kinetoplast (Gibson, 2007; Lai *et al.*, 2008).

T. vivax appears to be well adapted to use of the mechanical transmission cycle. This is perhaps due to the nature of its cyclical life cycle, however, it is unknown whether the species possesses specific adaptations facilitating mechanical transmission. Most notable has been the accidental transfer of *T. vivax* to South and Central America, likely by the import of infected livestock originating from the west of Africa at the start of the twentieth century (Dirie *et al.*, 1993a, 1993b; Jones and Dávila, 2001; Cortez *et al.*, 2006; Rodrigues *et al.*, 2008). *T. vivax* has become entrenched within the ecosystem of South America, with transmission strongly linked to the prevalence of biting flies and with acute, often fatal, infections common in herds which have not previously encountered this parasite (Otte and Abuabara, 1991; Seidl *et al.*, 1999; Batista *et al.*, 2007, 2009; Osório *et al.*, 2008; Cuglovici *et al.*, 2010; Da Silva *et al.*, 2011). The success of *T. vivax* in South America emphasises the potential role mechanical transmission may already be playing within Africa, a role about which too little is currently known.

1.3 Genetics of the Salivarian trypanosomes

1.3.1 Organisation and Structure of the Genome

Direct examination of the ploidy and karyotype of the Salivarian trypanosomes has not been possible due to the fact that these parasites do not appear to condense their chromosomes during mitosis (Vickerman and Preston, 1970). Early understanding of the genomes of the trypanosomes therefore came from a wide range of indirect techniques and in many cases is severely limited in species other than *T. brucei*. The diploid nature of Salivarian trypanosomes was recognised based on the results of isoenzyme analysis (Tait, 1980), DNA content measurements (Borst *et al.*, 1982) and analysis of restriction site polymorphisms in housekeeping genes (Gibson *et al.*, 1985).

The development of pulsed field gel electrophoresis (PFGE) further expanded our understanding of the genomes of trypanosomes by allowing the separation of chromosomes based upon their size and migration properties within an electric field. This technique has allowed the separation of trypanosome DNA into distinct classes (Reviewed in El-Sayed *et al.*, 2000). The mini-chromosomes, which are ~50-150 kilobases (kb) in size and are composed of a combination of simple repeat motifs and non-transcribed VSG genes (Sloof *et al.*, 1983a, 1983b; Wickstead *et al.*, 2004); the intermediate chromosomes which are ~200-900 kb long and contain VSG expression sites; the 11 diploid megabase chromosomes, which carry the housekeeping genes of the cell and range in size from ~1 - 6 megabases (Mb) and the highly interlinked maxicircle DNA of the kinetoplast (Melville *et al.*, 1998).Approximately 100 mini-chromosomes are present in *T. brucei,* although the number appears to vary between strains (Van der Ploeg *et al.*, 1984; Wickstead *et al.*, 2004). These chromosomes, although transcriptionally inactive, provide an important repertoire of material for the construction of new VSG genes, allowing for continual immune evasion. While present in high numbers in *T. brucei*, it has been reported that *T. vivax* possesses only one or two copies of these chromosomes, potentially limiting antigenic variation in this species (Dickin and Gibson, 1989). These mini-chromosomes are important, as the central sequence motifs specific to each species have been used as the basis for the development of highly sensitive, species specificpolymerase chain reaction (PCR) assays (Masiga *et al.*, 1992), which have been used for diagnosis and epidemiological studies.

Eleven diploid megabase chromosomes, ranging in size from 1-6 Mb have been identified in *T. brucei* through PFGE separation with the TREU 927 genome reference strain possessing a haploid genome of approximately 35 Mb based on analysis of the megabase chromosomes (Melville *et al.*, 1998; Berriman *et al.*, 2005). The role of the megabase chromosomes is the maintenance of the housekeeping genes of the parasite, which are predominately organised into mono-directional

arrays of genes expressed as a single polycistronic unit. With the housekeeping genes located primarily within the central portion of each chromosome, the telomeric regions posses the VSG expression sites and associated genes required for continual immune evasion through antigenic variation. The telomeres of each chromosome also maintain multiple copies of the unexpressed VSG repertoire which are amongst the first to be sampled during the VSG switching.

Considerable size variation of up to four fold difference between homologous chromosomes has been observed within and between strains of *T. brucei* (Gibson and Borst, 1986; Gottesdiener *et al.*, 1990; Melville, 1997; Melville *et al.*, 1998, 1999, 2000). The majority of this size variation is believed to be the result of duplication of chromosomal segments. While many of these duplication events are located within the telomeric VSG encoding regions, increasing potential coat variation, many involve duplication of gene coding regions which has given rise to the numerous gene families observed within trypanosomes (Melville *et al.*, 1999; Tait *et al.*, 2002; Callejas *et al.*, 2006). Due to the small number of investigations into *T. congolense* and *T. vivax,* there is little information concerning chromosome variation, however analysis of sequencing data from the respective genome projects suggests a high level of synteny between the chromosomes of these species and that of *T. brucei* (A Jackson, personal communication).

1.3.2 Genome sequencing of the Trypanosomes

The advent of the genomic era has provided for substantial leaps in our understanding of the genetics of many organisms, including those of the African trypanosomes with the first publishing of the *T. brucei* reference genome in 2005, consisting of sequence for each of the 11 megabase chromosomes (Berriman *et al.*, 2005). Other projects have led to the publication of sequences for the related kinetoplastids *Leishmania majo*r (Ivens *et al.*, 2005) and *T. cruzi (*El-Sayed *et al.*, 2005), and most recently *T. b. gambiense* Group 1 (Jackson *et al.*, 2010). Sequencing of *T. congolense* and *T. vivax* is ongoing at the Wellcome Trust Sanger Institute (WTSI) with initial sequence assemblies available via the WTSI and the Tri-Trypanosome database (TriTrypDB, http://www.tritrypdb.org) (Aslett *et al.*, 2010).

The initial reference sequence used the TREU 927 *T. b. brucei* strain, which was chosen for a number of factors. The strain has been adapted for growth in laboratory culture and is readily amenable to genetic manipulation, allowing for the insertion and knockout of gene constructs. While laboratory adapted the strain has maintained the ability to differentiate through each of the life cycle stages, including transmission through the tsetse fly as it would in natural populations (Turner *et al.*, 1990).

Sequencing of the reference strain used two distinct approaches. The WTSI was responsible for sequencing of chromosomes 1, 9, 10 and 11 utilising a whole chromosome shotgun approach, while The Institute for Genome Research (TIGR) sequenced chromosomes 2 – 8 through the use of bacterial artificial chromosome (BAC) with the project initially focusing on chromosomes I (Hall *et al.*, 2003) and II (El-Sayed *et al.*, 2003). The whole chromosome shotgun approach of the WTSI separated the chromosomes initially through PFGE prior to the generation of small insert libraries and subsequent sequencing. The chromosomes were then reassembled through the joining of overlapping sequence segments to generate increasingly large contigs. Where contigs could not be joined to one another, a chromosomal walking technique was used to expand contigs until the sequences overlapped. As the two homologues of each chromosome could not be fully separated through PFGE, the sequence obtained consists of a haploid mosaic, with single nucleotide polymorphism (SNP) information discarded from the final assemblies. The second approach, used by TIGR, generated numerous BACs from the entire genome which were mapped to the individual chromosomes. The individual BACs were then sequenced and assembled prior to chromosome level assembly of the BAC contigs. As with the whole chromosome sequencing, gaps in the sequence and between contigs were filled by directed PCR and sequence walking. The BAC generated sequences again represent a mosaic haplotype of the two homologues with sequence from each BAC originating from a single one of the homologues. Due to the potential for sequence variation between homologues of any given chromosome and the mosaic assembly these techniques have yielded, it is likely that regions of unique or duplicated sequence have been lost from the final reference sequence.

The first release of this reference sequence covered 26 Mb of the megabase chromosomes containing 9068 predicted genes, of which approximately 1700 are specific to *T. brucei* (Berriman *et al.*, 2005). The majority of these species specific genes are located within the sub-telomeres and are related to antigenic variation. Further extension, finishing and analysis of the sequence has extended the total number of predicted genes to 11,425 (TriTryDB, data retrieved April 2011).

The second published genome of an African trypanosome was that of the *T. b. gambiense* Group 1 strain DAL 972, generated through whole genome shotgun sequencing of plasmid and BAC clones (Jackson *et al.*, 2010). Assembly of the sequence was aided through the use of the *T. b. brucei* reference sequence as a scaffold with which to identify contig order and orientation. Analysis of the DAL 972 and TREU 927 genomes indicated a high level of similarity in terms of gene content, order and the level of sequence identity. No coding sequences specific to *T. b. gambiense* Group 1 could be identified within either the assembly or the unassembled contigs, while 86% of the identified coding sequences were 99% identical to the *T. b. brucei* orthologues. Despite this high level of conservation, 92,794 SNPs were identified within coding regions of which 49% were non-synonymous compared to the *T. b. brucei* reference sequence. Further sub-species specific

variation was found to be located within tandem gene arrays following comparison of the *T. b. gambiense* Group 1 and *T. b. brucei* homologues.

Publication of this second genome sequence has furthered our understanding of the genomics of the African trypanosomes however much work still remains to be done. The genomes of the related *T. congolense* and *T. vivax* species have yet to be completed to the same depth as those of *T. brucei* (Berriman *et al.*, 2005; Jackson *et al.*, 2010) and remain unpublished, while genome wide SNP data are not available for even the published strains. In order to fully understand the diversity and variation of natural populations additional sequencing of isolates from across Africa will be required allowing for truly comparative genomic studies.

1.3.3 Molecular variation within trypanosomes

Beyond gross differences in the size and numbers of (non megabase) chromosomes present, trypanosomes possess extensive variation at the sequence level which can be assayed through a multitude of techniques, even if the actual sequence variation underlying the differences is not known. These markers differ in terms of their application, the polymorphisms being measured and the information that can be obtained from their use. Amongst the most widely used techniques in trypanosome studies are isoenzymes, minisatellites and microsatellites, while SNPs and whole genome sequencing are likely to become the dominant approaches in the near future.

1.3.3.1 Isoenzymes

Isoenzyme analysis relies upon the presence of amino acid variation in assayable enzymes. Sequence variation is observed through alteration of the electrophoretic properties of the proteins on starch or polyacrylamide gels and visualisation of the resulting banding patterns by staining for specific enzymatic activity. Isoenzyme analysis is limited by a number of factors. The technique requires a relatively large amount of starting material; is reliant upon sequence variation unlikely to be selectively neutral and is often difficult to interpret due to many isoenzymes representing multiple genes. Isoenzyme analysis has been widely applied to trypanosome studies in the analysis of field populations, supplying early understanding of the variation present and the first evidence for genetic exchange in trypanosomes (Kilgour *et al.*, 1975; Kilgour and Godfrey, 1977; Gibson *et al.*, 1978; Tait, 1980; Young and Godfrey, 1983). The technique was also commonly used in the characterisation of laboratory isolates and analysis of early genetic crosses (Jenni *et al.*, 1986; Sternberg *et al.*, 1988, 1989).

1.3.3.2 Micro- and minisatellites

Micro- and minisatellites are amongst the widest used markers of genetic variation, being highly abundant throughout eukaryote genomes while the PCR is both highly sensitive and specific to individual sequences. As the majority of micro- and minisatellites are located within non-coding regions of the genome, they are expected to be subject to a low level of selection in field populations, as opposed to isoenzymes where variation may be selected against if it disrupts protein function sufficiently. Micro- and minisatellite markers have been widely employed in the analysis of trypanosomes due to their abundance, ease of use, single copy nature, specificity and sensitivity (Oliveira *et al.*, 1998; Hope *et al.*, 1999; MacLeod *et al.*, 1999, 2000; Jamonneau *et al.*, 2002; Balmer *et al.*, 2006; Koffi *et al.*, 2007; Morrison *et al.*, 2009b). Their presence throughout the genome has allowed the development of genetic maps of *T. brucei*, identifying hotspots of recombination across the 11 megabase chromosomes (MacLeod *et al.*, 2005b; Cooper *et al.*, 2008).

1.3.3.3 Single Nucleotide Polymorphisms

Single nucleotide polymorphisms (SNPs) represent the most abundant form of variation within the genome and may be present between individuals or between homologous chromosomes within an individual. As SNPs primarily arise through point mutations they are distributed throughout the entirety of the genome with no initial bias towards coding or non coding regions. Codon alteration as a result of SNPs within a coding sequence can lead to selection against their presence as many of these mutations will result in a detrimental effect upon the protein. As SNPs are by definition constrained to variation of a single base, only, four different allelic states are possible, one for each nucleotide, in contrast to markers such as microsatellites where there is no theoretical limit to the number of alleles that may be observed at a single locus. In reality the majority of identified SNP positions are bi-allelic as pyrimidines (C / T) are most likely to mutate into the other pyrimidine, likewise for the purine (A / G) bases. This low informational content is, however, compensated for by the sheer number of SNPs typically present in genomes.

Advances in genotyping technologies (reviewed in Ragoussis, 2009) have allowed for the rapid expansion in the pool of available SNPs with high throughput approaches currently focused upon high density microarray and next generation sequencing platforms. At large scales, hybridisation platforms are readily available off the shelf for many organisms or can be designed at a price if not already in production. When only a small number of SNPs need to be analysed, primer specific hybridisation assays, direct sequencing and variable restriction site cleavage techniques are all possible with prior SNP data. While genome wide SNP data have yet to be widely deployed in the study of African trypanosomes, high density SNP mapping has already been used in genome wide studies of *Plasmodium* (Neafsey *et al.*, 2008; Milet *et al.*, 2010; Mu *et al.*, 2010; Orjuela-Sánchez *et al.*, 2010) providing insight into how these studies might proceed in trypanosomes.

1.3.3.4 Whole genome sequencing

While the last decade has seen substantial leaps in sequencing technologies, it is only recently that the costs and time required have reached the point where individual groups can afford to sequence isolates for specific purposes. This has largely been achieved through the development of so called second generation sequencing platforms, capable of multiple parallel sequencing events on a single chip (Reviewed in Zhou *et al.*, 2010). The most widespread of these is the Illumina Solexa platform, which uses fluorescently labelled terminator nucleotides to extend reads one base at a time over multiple cycles. At present this platform is capable of generating paired end reads in the 100 – 150 bp length generating upwards of 25 gigabases of data per full run. Barcoding during preparation allows for multiple samples to be combined into a single run, reducing the monetary cost which is offset by lower fold coverage per sample (Kozarewa and Turner, 2011).

The power of whole genome sequencing arises from the ability to examine the genome with a resolution of a single base. While SNPs form the most widely examined form of genomic variation, genome sequencing also allows for the examination of a far wider range of sequence and structural variation. Insertion / deletion (indel) events, sequence rearrangements and duplication of sequence may all be detected through genome sequencing, allowing for identification of variation both between the homologues of a single diploid individual and between individuals. At present, however, both the sequencing platforms and analysis methodologies have yet to reach their full potential. The relatively short reads employed by second generation platforms mean that unless high read depths are employed sequence assembly typically requires an existing reference to act as a scaffold, limiting the potential to detect larger rearrangements, indels and regions of unique sequence. Read length further constrains the ability to assemble repetitive or duplicated regions with accuracy, limiting the analysis possible of closely related gene families or the mini- and microsatellites previously used in genetic studies. It is hoped that future platforms will overcome these problems by providing sufficiently long reads to bridge the length of repetitive or non-unique sequences, allowing for more complete assemblies. Further to this the software for assembling and analysing the datasets generated remains under constant development with many of the approaches currently inaccessible to those without a bioinformatics background (Reviewed in Miller *et al.*, 2010; Paszkiewicz and Studholme, 2010). Due to the recent and rapid development of whole genome sequencing platforms trypanosome genomic data are currently sparse with the available sequences limited to a handful available on TriTrypDB (Aslett *et al.*, 2010), however heterozygote SNP data are presently unavailable and data from multiple strains of a single species are only available for one of the African trypanosomes, *T. brucei*.

1.4 The variable surface glycoprotein coat and antigenic variation

Central to the ability of *T. brucei*, and the other Salivarian trypanosomes, in maintaining infections within mammalian hosts is the process of antigenic variation, mediated by the VSG coat that covers the surface of metacyclic and blood stream form trypanosomes (Vickerman, 1969; Vickerman and Luckins, 1969; Barry and McCulloch, 2001; Donelson, 2003). With the mammalian stages of infection occurring primarily within the bloodstream of the host the parasite is directly exposed to both the innate and adaptive immune systems. The VSG coat, which is tightly packed over almost the entire surface of the trypanosome, protects against complement mediated lysis by the innate immune system, while antigenic variation, whereby the VSG protein is continually varied, protects against antibody mediated lysis by the adaptive immune system and correlates with new peaks of parasitaemia (Balber, 1972). The continued replacement of the coat may also aid the clearance of antibodies that successfully bind the parasite (Engstler *et al.*, 2007).

The VSG repertoire of *T. brucei* accounts for a large proportion of the genome and is therefore a major source of genetic diversity (Reviewed in Donelson, 2003; Taylor and Rudenko, 2006). VSGs have been observed in gene families found across the megabase chromosomes and within the mini-chromosomes, however they are expressed only from metacyclic or bloodstream expression sites, dependent upon the life cycle stage, with only a single VSG expressed at any one time. *T. brucei* maintains a vast repertoire of VSGs, estimated at over 1000 genes (Van der Ploeg *et al.*, 1982), however, in TREU 927, the genome reference sequence, only 7% of all VSGs are believed to encode complete and functional genes (Berriman *et al.*, 2005) with the remainder serving as a reservoir of variation, copied into expression sites in order to generate new variants. By possessing such a large number of VSGs, with only one required to be functional at any given time, trypanosomes facilitate the accumulation of mutations in the repertoire, providing novel sources of variation that may then be incorporated into functional VSG sequences. Due to this process individual strains possess vastly different repertories of VSGs (Hutchinson *et al.*, 2007), which may then be shuffled through genetic exchange, generating previously unseen combinations.

1.5 Human African Trypanosomiasis

Human African Trypanosomiasis (HAT), commonly known as sleeping sickness, is the disease state resulting from infection by *T. b. gambiense* or *T. b. rhodesiense*, acquired through the bite of an infected tsetse fly. While the risk zone encompasses the entirety of the tsetse belt, with an estimated 60 million human inhabitants, the World Health Organisation (WHO) recorded ~17000 new cases across Africa in 2004 (WHO, 2006), dropping to less than 10,000 in 2009 (Simarro *et al.*, 2011). Due to the use of active screening in detecting the majority of cases this, however, is likely to be an underestimate of the true number of infections per year (Mumba *et al.*, 2011), which has been estimated to be in the range of 50 – 70,000 individuals (WHO, 2006). Although this number is greatly reduced upon the 1995 estimate of 300,000 cases per year (WHO, 1998) the disease remains a substantial risk to public health with the potential for new outbreaks outside of established foci. Social upheaval and civil instability have historically limited access to many of the endemic countries, limiting the potential for surveillance and treatment. Regional improvements across much of Africa, coupled with the renewed support of governmental, non-governmental and pharmaceutical organisations have in recent years allowed for the substantial increase in monitoring and treatment responsible for observed reductions (Welburn *et al.*, 2009).

Although identified as a single disease, HAT comprises two forms, one chronic and traditionally restricted to Western Africa, the other acute and typified as East African, caused by *T. b. gambiense* and *T. b. rhodesiense* respectively (Figure 1.2). The picture for *T. b. gambiense* is further complicated by the existence of two subgroups, Group 1 and 2, both of which are human infective. *T. b. gambiense* infections, almost exclusively of the Group 1 subgroup are responsible for the vast majority of reported cases in humans, with 17036 (97%) of the new cases in 2004 attributed to this sub-species (WHO, 2006). These two sub-species are unique in their ability to successfully infect humans, displaying the ability to resist lysis by human serum as opposed to *T. b. brucei* which is sensitive to lysis by human serum. The innate immunity of humans to infection by *T. b. brucei* functions through the cytotoxic activity of a subset of human serum high density lipoprotein (Rifkin, 1978), trypanosome lytic factors 1 and 2 (TLF-1 and TLF-2) (Hajduk *et al.*, 1989; Tomlinson *et al.*, 1997; Raper *et al.*, 1999). Both particles function through two protein components, apolipoprotein L-1 (apoL-1) and haptoglobin (Hp) related protein (Hpr), with haemoglobin (Hb) increasing the trypanolytic activity by acting as a cofactor (Widener *et al.*, 2007). Uptake of TLF-1 by trypanosomes utilises the high affinity haptoglobin / haemoglobin receptor (HpHbR) (Vanhollebeke *et al.*, 2008), located within the flagellar pocket. Binding the Hpr/Hb complex within TLF-1 triggers trafficking of TLF-1 to the trypanosomes lysosome, where the low pH triggers activation of apoL-1, formation of pores across the membrane of the lysosome and killing of the cell. Uptake of TLF-2 in contrast may involve the use of antibodies, found on the surface of the particle, to bind to VSGs on the surface of the trypanosome with uptake occurring as

VSGs are internalised for recycling by the parasite (Vanhollebeke and Pays, 2010). As TLF-2 possesses ApoL1 it is likely that the same lytic mechanism observed for TLF-1 is likely to be used by TLF-2.

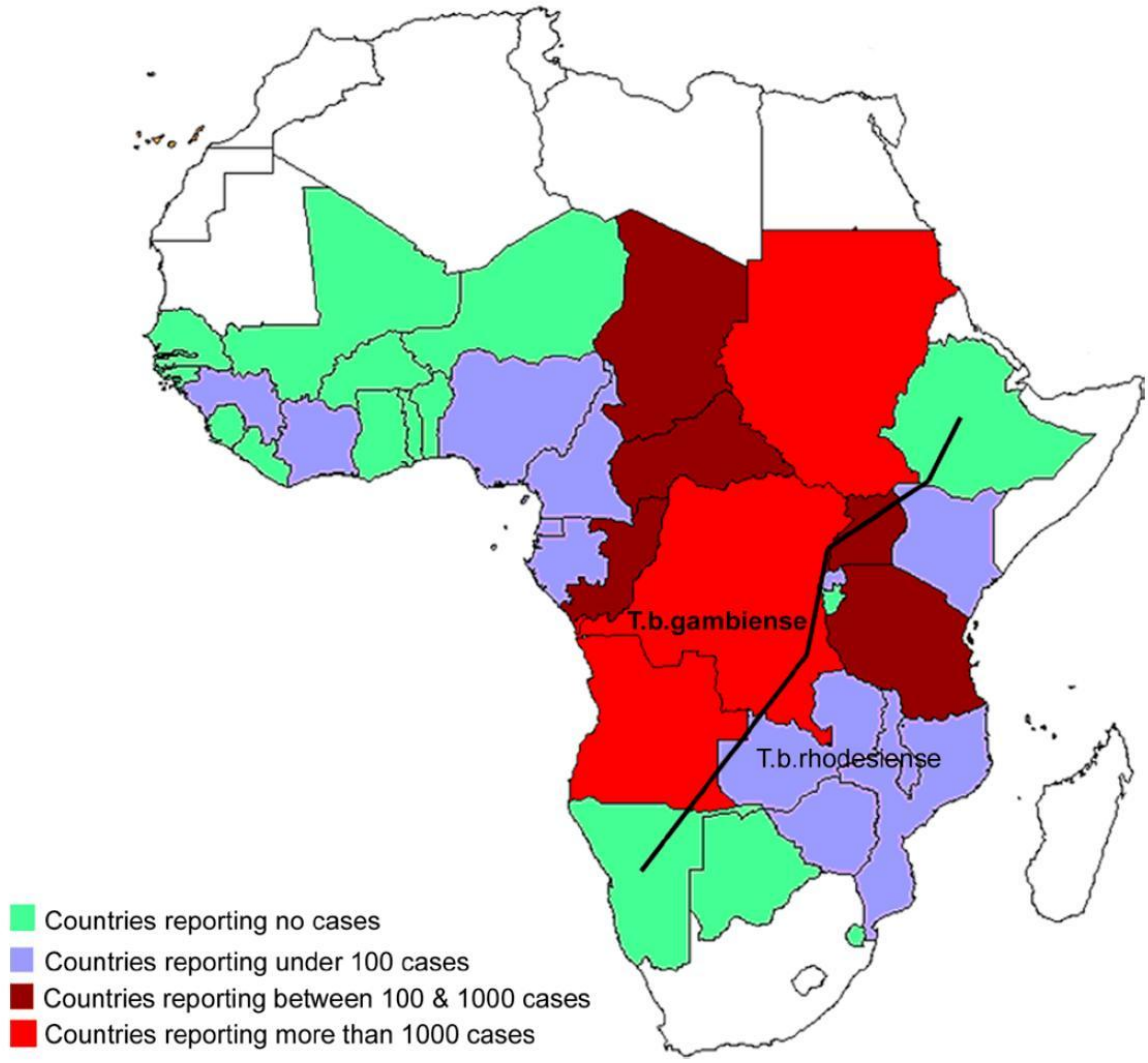


Figure 1.2 Distribution of human trypanosomiasis across Africa

Distribution and prevalence across Africa for the countries considered endemic for HAT due to their inclusion within the tsetse belt and indicating the approximate lower and upper boundaries of *T. b. gambiense* and *T. b. rhodesiense* respectively. Incidence is based upon WHO reported case number with the figure reproduced from Simarro *et al.*(2008).

1.5.1 *T. b. gambiense* Group 1

The dominant agent of HAT, *T. b. gambiense* Group 1 is the most distinct of the three *T. brucei* sub-species and clinically the most important (Fèvre *et al.*, 2008). Traditionally the sub-species has been defined by a number of distinct foci distributed across West and Central Africa; constitutive ability to infect the human host and a subsequent slow, chronic infection with progression to the symptomatic disease state typically requiring in excess of a year (Checchi *et al.*, 2008a, 2008b). Infections in the early stage are characterised by low parasitaemia, to the extent that they are sometimes undetectable by microscopic examination, instead being detectable by only the CATT or PCR based techniques (Kanmogne *et al.*, 1996a; Kaboré *et al.*, 2011; Wastling *et al.*, 2011). *T. b. gambiense* Group 1 is currently localised to a relatively small number of foci scattered over its wide geographic range. Angola, the Democratic Republic of the Congo and Sudan presently possess the greatest number of cases, with >1500 recorded in 2004. The Central African Republic, Chad, Congo, Côte d'Ivoire, Guinea and Uganda currently present intermediate levels of 50 – 1500 cases per year while Burkina Faso, Cameroon, Equatorial Guinea, Gabon and Nigeria all recorded less than 50 cases (WHO, 2006).

While morphologically indistinguishable from the other sub-species of *T. brucei*, *T. b. gambiense* Group 1 can be separated at the molecular level as a genetically distinct and relatively homogenous group (Mathieu-Daudé and Tibayrenc, 1994; Mathieu-Daudé *et al.*, 1995; Kanmogne *et al.*, 1996b; Morrison *et al.*, 2008b; Balmer *et al.*, 2011). Genetic variation within the group is limited as isolates from a given focus are highly related with the majority of variation observed between the different foci (Morrison *et al.*, 2008b). To date only a single unique sub-species specific gene, *T. b. gambiense* specific glycoprotein (*TgsGP*) has been identified in *T. b. gambiense* Group 1, encoding a truncated VSG (Berberof *et al.*, 2001). The 47 kilodalton protein localises to the flagellar pocket where it is believed to function as a sub-species specific receptor. Expression of *TgsGP* in *T. b. brucei* fails to confer resistance to human serum, suggesting either the presence of a multifactor resistance system in *T. b. gambiense* Group 1 or that *TgsGP* has a function not directly linked to human serum resistance, although it may be involved in further adaptation to human hosts. Resistance to human serum in *T. b. gambiense* Group 1 has yet to be fully elucidated, in part due to the difficulty in transfecting the sub-species. However, recent work has suggested mutations in the *T. b. gambiense* Hp/Hb receptor (TbgHpHbR) play an important role in conferring resistance by reducing the expression and function of this receptor. As a result *T. b. gambiense* Group 1 takes up less TLF-1, preventing exposure to the lytic factors contained within (Kieft *et al.*, 2010) while the constitutive reduction in TbgHpHbR expression and activity provides the fixed resistance to TLF-1 observed in the subgroup. At present the mechanism by which *T. b. gambiense* Group 1 resists lysis by TLF-2 remains unknown as this particle is not taken up by the same receptor as TLF-1.

1.5.2 *T. b. gambiense* Group 2

T. b. gambiense Group 2 represents a separate genetic grouping of West African trypanosomes capable of infecting humans, primarily isolated from the Côte d'Ivoire region and presenting a more acute disease progression than Group 1 isolates with a variable resistance to human serum (Gibson *et al.*, 1980, 1986; Borst *et al.*, 1981; Mehlitz *et al.*, 1982; Tait *et al.*, 1984; Paindavoine *et al.*, 1989; Richner *et al.*, 1989; Agbo *et al.*, 2001). Genetically the group is more closely related to *T. b. brucei* than *T. b. gambiense* Group 1 (Mathieu-Daudé and Tibayrenc, 1994; Stevens and Tibayrenc, 1996; Balmer *et al.*, 2011; Capewell, 2011), with a higher level of genetic diversity, ability to undergo genetic exchange and evidence for mating in field populations (Capewell, 2011). In contrast to Group 1, isolates from Group 2 display a variable resistance to lysis by human serum, with no consistent patterns of survival in the blood incubation infectivity test (BIIT), which measures the ability of isolates to infect laboratory animals following incubation in human blood (Rickman and Robson, 1970a, 1970b). The variable resistance of *T. b. gambiense* Group 2 implies that the subgroup possesses a different method of resisting lysis by human serum than *T. b. gambiense* Group 1. While the resistance phenotype of *T. b. gambiense* Group 2 is variable the subgroup does not possess the serum resistance associated (*SRA*) gene that has been shown to be responsible for resistance to lysis by human serum in *T. b. rhodesiense*, suggesting *T. b. gambiense* Group 2 may represent a third independent development of the human serum resistant phenotype.

1.5.3 *T. b. rhodesiense*

T. b. rhodesiense, restricted to the eastern regions of Africa is responsible for the acute form of HAT, with progression from initial infection to death typically requiring only a matter of months and with infections presenting a level of parasitaemia consistently higher than that observed for *T. b. gambiense* Group 1. As a sub-species *T. b. rhodesiense* forms a relatively homogeneous group, however strains from geographically distinct foci are typically distinguishable from one another (Gashumba *et al.*, 1994; Hide *et al.*, 1994; Komba *et al.*, 1997; MacLeod *et al.*, 2000) and genotypes may persist within an individual foci for long periods of time (Hide *et al.*, 1998). While genetically distinct from sympatric *T. b. brucei* populations *T. b. rhodesiense* has been shown to be more closely related to local *T. b. brucei* populations than other foci of *T. b. rhodesiense* indicating gene flow is occurring between the sub-species (MacLeod *et al.*, 2000, 2001c; Balmer *et al.*, 2011).

The number of reported cases of *T. b. rhodesiense* is significantly lower than that of *T. b. gambiense* and in 2004 the primary foci were located in Malawi, Uganda and the United Republic of Tanzania, each reporting 50 – 1500 cases (WHO, 2006). Sporadic cases, classed as fewer than 50 per year, are occasionally reported for Kenya, Mozambique, Rwanda, Zambia and Zimbabwe. Uganda is the only country where foci of both human infective sub-species can be found at present.

While presently distinct the potential for overlap poses significant concern for the diagnosis and treatment of HAT in this region (Picozzi *et al.*, 2005), due to the two sub-species being morphologically identical but being proscribed different courses of drug treatment. Overlapping foci of *T. b. gambiense* Group 1 and *T. b. rhodesiense* would therefore require more sophisticated screening techniques, likely employing PCR amplification of the *TgsGP* and *SRA* genes in order to distinguish the two, measures which are costly and time consuming compared to microscopic diagnosis.

While defined as a separate sub-species, *T. b. rhodesiense* may perhaps be more accurately described as a host variant of *T. b. brucei* that possesses the capability to survive in human serum. The ability to resist lysis by human serum in *T. b. rhodesiense* is a variable phenotype and not constitutively active as in *T. b. gambiense* Group 1. The resistance phenotype of *T. b. rhodesiense* is reliant upon the active expression of the *SRA* gene, an expression site associated gene encoding a truncated VSG (De Greef and Hamers, 1994; Xong *et al.*, 1998). SRA functions through the ability to bind apoL-1, the pore forming component of TLF-1 and TLF-2, preventing its activity following trafficking to the lysosome and therefore ensuring survival of the trypanosome in human blood. Experimental insertion of the *SRA* gene into human serum sensitive *T. b. brucei* lines is sufficient to confer resistance to human serum (Xong *et al.*, 1998; Oli *et al.*, 2006). The variable phenotype of *T. b. rhodesiense* can be explained by the phenomenon of VSG switching between the bloodstream expression sites, with resistance to lysis by human serum only conferred when the correct site is transcriptionally active and expressing the *SRA* gene. Due to this variable expression it is possible that samples identified as *T. b. brucei* based upon the BIIT instead represent isolates of *T. b. rhodesiense* that were not expressing *SRA* at the time of testing.

1.5.4 Symptoms and disease progression

The progression of HAT features two distinct disease stages, a haematolymphatic early stage followed by a meningo-encephalitic late stage with the stages defined by the respective localisation patterns of the trypanosomes. Early stage symptoms including nausea, headache, fever and lethargy are confined to the period of initial infection as trypanosomes migrate from the chancre and establish themselves within the host. Subsequent to the initial point of infection the early stage is largely asymptomatic with trypanosomes primarily confined to the blood and lymph of the host. Being transient and non specific the initial symptoms are easily missed or ascribed to other causes, in particular malaria which is typically of higher concern to individuals than HAT.

Parasitaemia and length of the early stage differentiates the infections of *T. b. gambiense* Group 1 and *T. b. rhodesiense*. *T. b. gambiense* Group 1 infections consist of low parasitaemia infections with occasional peaks in parasite numbers and a long early stage, typically requiring at least a year

to progress to late stage (Checchi *et al.*, 2008a, 2008b). *T. b. rhodesiense* in comparison presents consistently higher parasitaemia from the onset of infection and as such the early stage of East African HAT typically lasts no more than 6 – 12 months with rapid progression to the late stage. Virulence and thus the rate of progression in both forms of HAT are however variable with considerable differences reported between individuals and distinct foci which have been attributed to genetic factors of both the host and parasite (MacLean *et al.*, 2004, 2007; Sternberg and Maclean, 2010; Kuepfer *et al.*, 2011; Morrison, 2011).

Late stage HAT is defined by the migration of trypanosomes from the bloodstream into the brain and cerebral spinal fluid by traversal of the blood brain barrier (Kennedy, 2006a; Rodgers, 2010). As the trypanosomes reproduce within the central nervous system the resulting damage from both the trypanosome themselves and the host immune response gives rise to the neurological symptoms of the late stage, affecting cognitive and motor functions in addition to alteration of circadian rhythms. Early symptoms may include loss of concentration, anxiety, irritability and personality shifts, progressing to violent mood swings, manic episodes and hallucinations. Motor functions slowly degrade, making speech, fine manipulation and walking increasingly difficult. The disturbance of circadian rhythms from which the term 'sleeping sickness' arises are characterised by a reduction in alertness, daytime drowsiness and night time restlessness, leading to eventual coma and finally death.(Kennedy, 2006b, 2008)

While late stage symptoms increase the ease of diagnosis, traversal of trypanosomes across the blood brain barrier necessitates the use of different treatment regimes from infections identified during the early stage (Wilkinson and Kelly, 2009; Burri, 2010). The side effects of the drugs required during late stage, and existing damage from the infection itself, makes early stage diagnosis preferable. The largely asymptomatic nature of the early stage however requires the use of active screening programmes, which are costly, time consuming and difficult to implement in many regions. It is perhaps unsurprising then that the majority of cases are believed to go unreported.

1.5.5 Diagnosis of infection

Despite the many recent advances in medicine, microscopic examination of blood smears or buffy coat preparations remains the gold standard for field diagnosis, due to its 100% specificity in identifying infection. While highly specific, microscopic examination is hindered by a lack of sensitivity and the level of parasitaemia in the blood, especially in *T. b. gambiense* Group 1 where infections are typified by an initial asymptomatic, low parasitaemia stage. Diagnosis of late stage infections, where trypanosomes access the cerebral spinal fluid and migrate across the blood brain barrier, requires lumbar puncture and collection of cerebral spinal fluid samples. Late stage

diagnosis is based upon either visual identification of trypanosomes or an elevated white blood cell count (Reviewed in Chappuis *et al.*, 2005).

Although microscopy remains the gold standard, advances have led to the development of a number of kits and procedures in order to improve upon the sensitivity of microscopic techniques with a minimal loss in specificity (Reviewed in Adams and Hamilton, 2008; Wastling and Welburn, 2011). These can be loosely grouped into two categories, concentration and molecular. Concentration techniques, such as the buffy coat technique, mini-anion exchange column and microhaematocrit centrifugation function by concentrating trypanosomes into a smaller volume than the initial blood sample (Murray *et al.*, 1977; Lumsden *et al.*, 1979; Büscher *et al.*, 2009; Camara *et al.*, 2010). It is therefore possible to identify infection in low parasitaemia individuals where examination of the required volume of blood would not be feasible by traditional microscopy. A recent advance in the process of concentration is the development of a lateral displacement technique (Holm *et al.*, 2011), which aims to separate blood cells and parasites based on cell size and morphology. While still in development, it is hoped that this technique will provide a new simple to use and cheap technique for diagnosis in the future.

Molecular techniques rely upon indirect methods of diagnosis by the identification of trypanosome specific molecules (Reviewed in Wastling and Welburn, 2011). The most prominent of these in diagnosis has been the card agglutination test for trypanosomiasis (CATT) which detects the presence of antibodies raised against the VSG LiTat 1.3, which appears to be widespread amongst *T. b. gambiense* Group 1 (Magnus *et al.*, 1978). While sensitivities of between 87 and 98% and specificity of 95% have been reported (Lejon *et al.*, 2010) the test has a number of limitations. Incidences of seropositive, parasitologically negative individuals are common, this can arise through prior infections that have been treated (Truc *et al.*, 1994), exposure that did not lead to infection or low parasitaemia below the sensitivity of microscopy (Kanmogne *et al.*, 1996a; Kaboré *et al.*, 2011). Incidences of seronegative yet parasitologically positive individuals have also been reported, both prior to and following treatment for trypanosomiasis, potentially from trypanosome strains lacking LiTat 1.3 (Dukes *et al.*, 1992; Enyaru *et al.*, 1998; Lejon *et al.*, 2010).

Alternative molecular techniques for diagnosis are predominantly focused upon the amplification and detection of specific DNA fragments. The 177 bp repeat found in multiple copies of the *T. brucei* mini-chromosomes has allowed for the development of a highly sensitive, *T. brucei* specific marker (Masiga *et al.*, 1992) while additional primers are available for the amplification of *TgsGP* and *SRA*, specific to *T. b. gambiense* and *T. b. rhodesiense* respectively (De Greef and Hamers, 1994; Xong *et al.*, 1998; Berberof *et al.*, 2001; Radwanska *et al.*, 2002). While amplification based techniques possess extremely high sensitivity this is offset by the higher costs of these tests and requirement for laboratory analysis, making them impractical for screening of patients in the field

(Deborggraeve and Büscher, 2010). Loop-mediated isothermal amplification (LAMP), an amplification technique which functions at a lower temperature than traditional PCR (~60°C as opposed to a 65-95°C cycle) requires only a single fixed temperature water bath and allows for direct visual inspection of amplification products (Reviewed in Adams and Hamilton, 2008; Mori and Notomi, 2009). Species specific LAMP protocols have already been developed for *T. brucei* and its sub-species (Kuboki *et al.*, 2003; Thekisoe *et al.*, 2007b; Njiru *et al.*, 2008a, 2008b, 2011b) with the development of this technique having gone some way to bringing molecular techniques to the field by providing a quick, sensitive test utilising reagents which are stable in the higher temperatures encountered in the field (Thekisoe *et al.*, 2009; Njiru, 2011). Further development is, however, required in order to make the technique viable and cost effective.

1.5.6 Treatment

Treatment following diagnosis of HAT is currently reliant upon the chemotherapeutic action of four drugs, each of which is complicated by the disadvantages of toxicity, limited efficacy, restriction to specific sub-species or disease stage and spreading parasite resistance leading to treatment failure and relapse (Docampo and Moreno, 2003; Barrett *et al.*, 2007; Kennedy, 2008; Burri, 2010). The potential for a vaccine against HAT is negligible due to the constantly changing nature of the VSG coat on the surface of trypanosomes and ability of the parasite to disrupt vaccine induced memory B cell responses (Radwanska *et al.*, 2008; Reviewed in Magez *et al.*, 2010).

If diagnosed during the early stage, prior to invasion of the central nervous system, HAT may be treated with either suramin or pentamidine, both of which are effective against *T. b. gambiense* and *T. b. rhodesiense*. While the mechanisms by which these two drugs act are not known the specificity of each is due to selective uptake of the drugs by trypanosomes, likely by endocytosis for suramin and by carrier mediated transporters for pentamidine (Reviewed in Barrett and Gilbert, 2006). Late stage treatments for HAT differ due to the requirement for drugs capable of crossing the blood brain barrier. Two drugs, melarsoprol and eflornithine are currently available for treatment of this stage (Kennedy, 2004). Melarsoprol remains the primary drug for treatment of late stage infection by both *T. b. gambiense* and *T. b. rhodesiense* despite the highly toxic side effects that result in the death of 5% of all patients. Uptake of melarsoprol is mediated by the same transporter families responsible for uptake of pentamidine and once again the mechanism by which the drug acts remains unknown. Resistance to pentamidine and melarsoprol, acquired through down regulation of the active transporters involved in uptake (de Koning, 2001; Reviewed in Mäser *et al.*, 2003), has been observed in both laboratory and field settings with treatment failures as high as 30% reported in the field (Burri and Keiser, 2001; Pépin and Mpia, 2005; Robays *et al.*, 2008; Kazibwe *et al.*, 2009).

Eflornithine is the newest drug to market for the treatment of HAT (Van Nieuwenhove *et al.*, 1985; Van Bogaert and Haemers, 1989). It is also the only HAT drug for which a definitive mode of action is known (Reviewed in Fairlamb, 2003), functioning through inhibition of the polyamine biosynthetic enzyme ornithine decarboxylase. The drug is only active against *T. b. gambiense* Group 1 due to the slow turnover of ornithine decarboxylase in this sub-species, allowing for effective disruption of enzymatic activity. Eflornithine may be deployed in combination with nifurtimox, a drug typically used in the treatment of *T. cruzi*, in order to shorten the treatment program, lowering the associated costs without an increase in adverse reactions (Checchi *et al.*, 2007; Priotto *et al.*, 2007; Yun *et al.*, 2010). While eflornithine resistance has yet to be reported in the field it has been generated *in vitro*, linked to the loss of the amino acid transporter *TbAAT6* (Vincent *et al.*, 2010; Baker *et al.*, 2011).

1.5.7 Animal reservoirs

While a human disease, the agents of HAT are not restricted solely to utilising humans as a host, with *T. brucei* being able to infect a wide range of domestic and wild fauna, potentially providing reservoirs of genetic material. *T. b. rhodesiense*, essentially a host variant of *T. b. brucei* retains the capability to infect the same wide host range of its human serum sensitive relative (Hutchinson *et al.*, 2003). The ability of *T. b. rhodesiense* and *T. b. brucei* to interbreed with one another further increases the potential size of the genetic reservoir, allowing for the incorporation of material from human serum sensitive individuals into human serum resistant populations. Investigations which have demonstrated sympatric *T. b. brucei* and *T. b. rhodesiense* populations to be genetically distinct from one another suggests interbreeding between the sub-species may not be a frequent occurrence (Hide *et al.*, 1994; MacLeod *et al.*, 2001c). The role of an animal reservoir in *T. b. rhodesiense* has been clearly demonstrated in Uganda where the outbreak of East African HAT has been linked to the movement of infected livestock between regions (Fèvre *et al.*, 2001; Welburn *et al.*, 2001; Hutchinson *et al.*, 2003; Waiswa *et al.*, 2003; Njiru *et al.*, 2004b; Enyaru *et al.*, 2006).

The role of animal reservoirs in *T. b. gambiense* Group 1 is less clear and traditionally the disease has been described as being restricted to humans where the long asymptomatic early stage allows for the formation of an effective human reservoir. It has become apparent however that this sub-species can be maintained within sheep, pigs and goats, raising the potential for an animal reservoir (Paindavoine *et al.*, 1986a; Cordon-Obras *et al.*, 2009). These reservoirs may play a role in maintaining, or re-establishing, foci even after the treatment of all the human infections in the region.

1.5.8 Focal nature of outbreaks

Despite the potential for a wide geographical distribution across much of Africa, HAT is a highly localised and focal disease, with a tendency for long lasting, established foci despite frequent fluctuations in the number of reported cases. New foci may arise through the splintering of larger established foci during periods of contraction or through displacement of the host reservoir, as has been observed in Uganda with the establishment of the Soroti focus following import of *T. b. rhodesiense* infected livestock from the nearby Tororo region (Fèvre *et al.*, 2001). Further to host mobility is the role of vector availability and mobility (Bouyer *et al.*, 2009; Solano *et al.*, 2010). While a diverse range of tsetse species are capable of acting as trypanosome vectors the exact species present has an important role and the efficiency of transmission is likely to be dependent upon the combination of parasite strain and tsetse species. In addition species specific blood meal preferences may control the frequency with which a fly will bite a human host (Njiokou *et al.*, 2004b; Simo *et al.*, 2008). The distinct foci of human disease may therefore be controlled by the presence of a localised tsetse population with a preference towards human blood meals and the ability to efficiently transmit human infective trypanosome strains. Expansion of a focus would, under these conditions require outward geographical expansion of the tsetse population while establishment of new foci would require transfer of infected hosts or vectors to a region already meeting the requirements for a successful transmission cycle.

1.5.9 Human trypanotolerance

Trypanotolerance, the ability to naturally control the parasitaemia and symptoms of trypanosome infection has been widely reported in animal trypanosomiasis while in humans death has traditionally been reported as the inevitable outcome of infection in the absence of clinical intervention and treatment, a position that is now being questioned (Reviewed in Checchi *et al.*, 2008a; Bucheton *et al.*, 2011).

The strongest evidence for control without clearance in humans has come from Jamonneau *et al.*(2004) in a study of six asymptomatic patients from the Côte d'Ivoire who consistently refused treatment despite being found to be microscopically positive for trypanosomes during initial evaluation in 1995. Over the course of seven years the patients were assessed for trypanosomiasis by a range of measures including microscopy, CATT, inoculation of blood into the Kit for In Vitro Inoculation and immunosuppressed mice and PCR based assays. While all the individuals were positive in 1995 by serology and microscopy all but one had converted to be seronegative by 2002. None of the patients had manifested symptoms associated with progression to late stage, however, the transient and non-specific symptoms of the early stage were recorded in multiple individuals during the study. Microscopy and inoculation tests were likewise variable but dropped to being

negative in all patients by the end of the study with only the PCR based assays returning positive results through the entirety of the study, suggesting continued infection despite the negative result for the other diagnostic methods. The results of this study clearly show the potential for humans to control the parasitaemia and symptoms.

1.6 Animal African Trypanosomiasis

Animal African Trypanosomiasis (AAT), more commonly known as Nagana is found throughout the entirety of the tsetse belt of Africa and has three causative agents, *T. congolense*, *T. vivax* and to a lesser extent *T. brucei*. The disease has been reported in a wide range of host animals including, but not limited to, cattle, horses, donkeys, and sheep with the most pathogenic infections typically observed in the infection of domestic animals.

1.6.1 Symptoms

Due to the wide range of factors controlling the disease state in Nagana the symptoms of infection are generally non-specific and can be caused by a wide range of other infections. Fever, weight loss, anaemia and lethargy are the most commonly reported symptoms attributed to animal trypanosomiasis. However, due to the wide range of controlling factors, including multiple infectious agents, strain variability and host control the observed severity of these symptoms can vary considerably. Infections are typically observed to be severest within domesticated animals with many wild animals showing little or no disease symptoms. Of the three species responsible for the disease, *T. congolense* is typically reported as the most important due to its association with significantly higher levels of anaemia compared to *T. brucei* and *T. vivax* (Pinchbeck *et al.*, 2008; Dayo *et al.*, 2010), although variations in virulence have been observed when comparing different subtypes and strains of *T. congolense* (Bengaly *et al.*, 2002; Masumu *et al.*, 2006, 2009; Van den Bossche *et al.*, 2011).In cattle *T. brucei* is often reported as being responsible for the mildest of disease states, while *T. vivax* varies considerably between East and West Africa with a severer acute disease profile found reported in the west of the continent (Fasogbon *et al.*, 1990).Invasion of the central nervous system has been reported in some of the most severe infections of *T. brucei, T. congolense* and *T. vivax* leading to lesions of the central nervous system and typically death even following treatment (Masake *et al.*, 1984; Wellde *et al.*, 1989; Batista *et al.*, 2007, 2011; Galiza *et al.*, 2011). Further to this, rare but deadly cases of acute haemorrhagic syndrome as a result of *T. vivax* infections have been periodically reported in East Africa (Catley *et al.*, 2002; Magona *et al.*, 2008), in contrast to the typically low pathogenicity of *T. vivax* in the east of the continent.

1.6.2 Diagnosis

Definitive diagnosis of Nagana in the field is, like HAT, predominately through microscopic examination of blood films or buffy coat preparations, due to the ease and low cost of this technique. As anaemia is a frequent symptom of Nagana the reduced packed cell volume (PCV), a crude measurement of anaemia, may be used alongside microscopic examination as a further

method of diagnosis. While a number of possible infections may give rise to anaemia, there is a strong correlation between microscopic diagnosis of trypanosomiasis and reduced PCV values, allowing for diagnosis of low parasitaemia infections that can be easily missed by the limited sensitivity of microscopic examination (Pinchbeck *et al.*, 2008). Due to the simplicity of this diagnosis it can be quickly applied in the field even in the absence of electrical power through the employment of small human powered centrifuges (Figure 1.3). Coupled with veterinary evaluation of animals via body condition scores, these techniques can be easily deployed in the field to supplement microscopic diagnosis.

Molecular markers for the diagnosis of *T. brucei, T. congolense* and *T. vivax* are available, normally falling into the categories of antigen-based enzyme-linked immunosorbent assays (ELISA) or PCR based methods. The available ELISA based tests are typically indirect in nature, detecting the presence of host antibodies raised against infection as opposed to the trypanosomes themselves (Eisler *et al.*, 1998; Magona *et al.*, 2002). These tests are therefore unable to differentiate current infections from those that have already been cleared or treated. While some antigen specific tests have been developed the majority use crude antigen prepared from cell lysates, bypassing the need to identify specific antigens and increasing sensitivity by allowing for capture of antibodies raised against multiple targets (Magona *et al.*, 2002; Madruga *et al.*, 2006). ELISAs, while more costly than microscopy diagnosis offer a significantly increased sensitivity yet are simpler and easier to employ than the PCR based techniques, facilitating widespread use in serological surveys in both Africa and South America (Eisler *et al.*, 1998; Mahama *et al.*, 2005; Delafosse *et al.*, 2006; Cabrera *et al.*, 2009; Bossard *et al.*, 2010).

Alongside ELISA based diagnosis a number of PCR based techniques have been developed to allow for the amplification of species specific motifs (Desquesnes and Dávila, 2002), with the most sensitive of these directed against the multicopy repeats primarily found in the mini-chromosomes (Masiga *et al.*, 1992; Wickstead *et al.*, 2004). The number of copies present varies between each of the species, with *T. vivax* suggested to only have one or two mini-chromosomes compared to the hundreds of copies in *T. brucei* (Dickin and Gibson, 1989). While both highly sensitive and species specific the cost and requirement for laboratory analysis has limited their use to epidemiological surveys as opposed to diagnosis for the purpose of treatment. Alongside the *T. brucei* specific LAMP protocols employed in diagnosis of HAT additional protocols have been described for the identification of both *T. congolense* (Kuboki *et al.*, 2003; Thekisoe *et al.*, 2007a, 2007b) and *T. vivax* (Njiru *et al.*, 2011a).

Figure 1.3 Hand spinning of blood samples to assess blood PCV

Capillary tubes filled with blood from animals are manually spun to determine the PCV as a measure of anaemia. The application of relatively simple technologies such as this is of great benefit to diagnosis in the field. Personal photograph taken in The Gambia, 2009.

1.6.4 Prevalence

As the disease is endemic over much of Africa and present in a wide range of hosts prevalence varies considerably from region to region and from host to host as well as being dependent upon the sampling and identification employed. In bovines, reported total trypanosome prevalence in the range of 5 – 40% is common, with the dominant species varying from region to region (Rowlands *et al.*, 1993; Waiswa and Katunguka-Rwakishaya, 2004; Simukoko *et al.*, 2007b; Miruk *et al.*, 2008; Enwezor *et al.*, 2009; Cox *et al.*, 2010; Dayo *et al.*, 2010). Prevalence in other domestic host species likewise varies considerably from close to zero to upwards of 40% (Snow *et al.*, 1996; Kalu *et al.*, 2001; Simukoko *et al.*, 2007a; Nimpaye *et al.*, 2011). While fewer studies have focused upon trypanosome infections in wild animals, prevalences of up to 25% have been reported, again dependent upon the species (host and parasite) being examined (Njiokou *et al.*, 2004c; Anderson *et al.*, 2011)

Outside of the tsetse belt the reported incidences of *T. brucei*, *T. congolense* and *T. vivax* is severely reduced with *T. vivax* typically the dominant species, facilitated by the ease with which it is mechanically transmitted (Cherenet *et al.*, 2006; Sinshaw *et al.*, 2006). As with tsetse transmission mechanical transmission is highly dependent upon the density of biting flies (Cherenet *et al.*, 2004). In comparison to HAT which is typified as a focal disease, Nagana presents as a disperse and widespread endemic disease.

1.6.5 Treatment

Treatment and control of Nagana is primarily achieved through the use of three drugs, ethidium bromide (homidium), diminazene aceturate (berenil) and isometamidium chloride (samorin) (Kinabo, 1993), with all three active against *T. brucei*, *T. congolense* and *T. vivax*. Due to extensive use of homidium during the 1960s and 1970s widespread resistance has been reported, which can potentially provide cross resistance to samorin due to their related structures (Kinabo, 1993). The exact mechanism of these drugs is not fully understood, however it is believed that they may function through their interaction with nucleic acids although a wide range of other possible modes of action have also been proposed (Reviewed in Wainwright, 2010). It has been recently reported (Roy Chowdhury *et al.*, 2010) that homidium may act through both inhibition of minicircle replication with subsequent loss of kinetoplast DNA and by secondary inhibition of nuclear DNA replication. Due to an extremely long half life, with continual circulation of up to three months, samorin is widely applied a prophylactic agent which has potentially facilitated the emergence of drug resistance if applied at insufficient doses (Matovu *et al.*, 2001).

Berenil is an aromatic diamidine that is related to pentamidine, used in the treatment of HAT and possesses a similar structure, biochemical properties and probably mode of action (Reviewed in Peregrine and Mamman, 1993; de Koning, 2001). The drug is active against each of the species involved in Nagana and like pentamidine the drug is slow to diffuse over biological membranes, requiring instead active uptake, primarily via the P2 aminopurine permease transporter in *T. brucei* (de Koning and Jarvis, 1999; Teka *et al.*, 2011). Due to the similarity to pentamidine the frequent use of berenil in treating animals may facilitate the onset of diamidine resistance in human infective trypanosomes, with a downstream impact on the treatment of HAT cases.

1.6.6 Trypanotolerance

In humans, death from trypanosome infections has, until recently, been considered an inevitable outcome of the disease if it is not identified and treated. Trypanotolerance or the ability to control the symptoms of infection is, however, not a new concept with regard animal trypanosomiasis (Murray *et al.*, 1982; reviewed in Courtin *et al.*, 2008). The phenomenon has been most extensively studied in bovines utilising breeds which are able (trypanotolerant) and unable (trypanosusceptible) to control the symptoms and damage caused by infection. Experiments in bovines have suggested that trypanotolerance is comprised primarily of two independent traits, the ability to control parasitaemia and the ability of the animal to control anaemia (Naessens *et al.*, 2002, 2003; Naessens, 2006). Trypanotolerance, however, is not a simple phenotype, being instead under the control of multiple loci, with the end phenotypes of parasitaemia and anaemia being complex multivariable phenomenon (Hanotte *et al.*, 2003; Hill *et al.*, 2005; de Koning *et al.*, 2005; Kemp *et al.*, 2008; Rennie *et al.*, 2008; Noyes *et al.*, 2011). Amongst the genes and pathways identified in these studies are differential expression of the antimicrobial peptides and acute phase proteins of the innate immune system (Meade *et al.*, 2009); a more rapid and greater immune response in the trypanotolerant N'Dama breed compared to the trypanosusceptible Boran breed (O'Gorman *et al.*, 2009); an increased type 1 T helper cell response and proinflamatory cytokine response during the initial stages of infection in trypanotolerant N'Dama (O'Gorman *et al.*, 2006) and the presence of particular major histocompatatibility complex alleles (Gautier *et al.*, 2009; Karimuribo *et al.*, 2011).

Due to the costs and difficulties associated with bovine studies a number of studies have instead chosen to focus on the genes regulating infections in mice. These studies have identified and mapped three major trypanosome resistance loci linked to survival time following infection (Iraqi *et al.*, 2000; Goodhead *et al.*, 2010; Nganga *et al.*, 2010); identified the innate response to be a major contributor in controlling the inflammation associated with anaemia (Noyes *et al.*, 2009) and demonstrated higher expression of select chemokines, interferon receptors and complement components to be associated with susceptibility (Kierstein *et al.*, 2006) amongst others.

1.7 Genetic exchange and clonality

In order to fully understand the population genetics of the Salivarian trypanosomes upon which this work focuses it is essential to understand the role of genetic exchange in a wider context. This is of particular importance given the non-obligatory nature of genetic exchange in *T. brucei* and the differences between strict clonality, where genetic exchange is completely absent from the life cycle of an organism and effectively clonal populations, where genetic exchange may be possible but lacks a significant role in influencing a population over short time scales (Balloux *et al.*, 2003). However even infrequent genetic exchange can have substantial effects on the evolution of a species over long time frames and organisms may inhabit a spectrum with regards the frequency and role of genetic exchange, at one end strictly clonal species that employ purely mitotic reproduction, at the other species with obligatory sexual cycles.

Prior to discussing this spectrum it is first necessary to clearly define some of the terms which shall be utilised in this discussion as there is no strict consensus within the literature and many terms are therefore used differently depending upon the author or species being studied. For the purpose of this discussion the following terms shall be described as follows:

Genetic exchange. Fusion of two (parental) cells to give rise to a new (progeny) cell with inheritance of nuclear DNA from each, though not necessarily equally. While this will often involve sexual recombination as defined by the meiotic cycle this is not obligatory. The definition used here deliberately excludes horizontal gene transfer between cells.

Strictly clonal. Organisms in which genetic exchange, as defined above, is biologically impossible. All reproduction is through mitotic growth.

Clonal reproduction / expansion. Mitotic growth giving rise to genetically identical daughter cells.

Effectively clonal; clonal populations. Species and populations in which genetic exchange is possible but occurs rarely enough that it has no significant effect on the population structure in the short term. When genetic exchange does occur it gives rise to new effectively clonal lineages which then propagate through clonal expansion and may therefore have significant impact in the long term.

Epidemic population. A population in which genetic exchange is occurring at a significant rate with the effects masked by the clonal expansion of individuals in the population, as defined by Maynard Smith (1993).

Panmixia. A population in which random genetic exchange between individuals is the dominant mode of reproduction. Deviation from panmixia may occur in obligatory sexual organisms if there are barriers to random mating.

Inbreeding. Genetic exchange between two individuals that are genetically identical or closely related to one another.

Selfing. Genetic exchange between two gametes originating from either a single individual or two genetically identical individuals.

In examining a range of parasitic protozoa Tibayrenc *et al.*(1990a) outlined a number of criteria by which an effectively clonal population could be identified (Table 1.3). The techniques used in assessing these criteria was further expanded by Maynard Smith *et al.*(1993) in defining the epidemic population structure (see above) and the proposal that population structure should be assessed twice, once in the presence of all individuals and a second time with the removal of repeated genotypes that have arisen through clonal expansion. In this manner it is possible to observe the underlying shape of the population in organisms where genetic exchange and clonal reproduction coexist and discern whether evidence for genetic exchange is being masked by clonal expansion.

Criteria for segregation at individual loci	Fixed heterozygosity Absence of segregation genotypes Deviation from Hardy-Weinberg equilibrium
Criteria for recombination between loci	Multiple isolates with identical genotypes Absence of recombinant genotypes Significant linkage disequilibrium Correlation of results from multiple independent molecular markers

Table 1.3 Criteria for clonality

Criteria for clonality in parasitic protozoa as defined by Tibayrenc *et al.*(1990a).

1.7.1 A spectrum of genetic exchange

With a wide range of potential reproductive strategies available it is of no surprise that examples can be found from across the entire spectrum, with a range of consequences on the evolution of the species involved. At one extreme of the spectrum strict clonality represents the simplest form of reproduction, with progeny genetically identical to the parent and with change in the lineages arising through mutation. For diploid organisms strict clonality eventually leads to a divergence in sequence between individuals of different lineages and between the homologues of each chromosome within an individual. Known as the Meselson effect (Judson and Normark, 1996; Mark Welch and Meselson, 2000), this divergence occurs as each homologue accumulates unique mutations, eventually leading to genome wide heterozygosity. As selective pressures generally require only the retention of function for a single allele of each gene the Meselson effect may lead to loss of function, through the accumulation of deleterious mutations. This is an example of Muller's ratchet (Muller, 1964), which states that it is easier to acquire deleterious mutations than to lose them as the absence of meiotic recombination prevents the generation of new haplotypes combining sequence free from these deleterious effects. While the vast majority of mutations are deleterious, on occasion the Meselson effect will give rise to a new function in the second allele of a gene, effectively providing a new gene to the organism and once again placing the allele under selective pressures. Over time the divergence between the homologous chromosomes may lead to an effectively haploid state, as the vast majority of, or even all genes are represented by only a single functional allele within the genome.

From a population standpoint strictly clonal organisms exist as numerous genetic lineages, which, due to the lack of genetic exchange, cannot interact and therefore evolve independently of one another. This can have significant consequences on the spread of traits throughout a population as traits in two lineages will also never interact, therefore placing the traits in permanent competition with one another. Within a population, strict clonality leads to high levels of heterozygosity at each locus, significant linkage between pairs of loci and significant deviation from Hardy-Weinberg predictions. With the absence of genetic exchange strictly clonal populations may also come to be dominated by a single lineage, which can occur following bottlenecks in the population or when one lineage possesses a strong competitive advantage. In both cases this leads to an initial loss of diversity within the population, which slowly increases as the members of the dominate lineage acquire new mutations, generating multiple new lineages that once again enter into competition with one another.

Perhaps the most widely recognised ancient asexual, strictly clonal eukaryote are the Bdelloid rotifer, microscopic freshwater invertebrate which are believed to have been asexual for millions of years (Reviewed in Rice and Friberg, 2007). This has led to a gradual reduction in the identity

between homologous chromosomes and generation of divergent gene copies (Judson and Normark, 1996; Mark Welch and Meselson, 2000; Mark Welch *et al.*, 2004a, 2004b; Pouchkina-Stantcheva *et al.*, 2007). While sexual reproduction is absent in bdelloid rotifers they have retained the ability to acquire new genetic material through a form of horizontal gene transfer. In order to survive periods of drought, rotifers are capable of entering a period of suspended activity, reviving only when there is sufficient water available in the environment to rehydrate the cells (Lapinski and Tunnacliffe, 2003; Ricci *et al.*, 2003). While allowing for survival the process does result in damage to cell membranes, allowing for foreign material including DNA fragments to enter from the surrounding environment. As the cells revive they are able to incorporate this genetic material into their own (Gladyshev *et al.*, 2008; Boschetti *et al.*, 2011), facilitated by the presence of effective genome repair mechanisms (Gladyshev and Meselson, 2008). This process does not appear to be limited to material originating from other rotifers and therefore may represent a powerful driver of bdelloid evolution.

Sexual reproduction has been proposed to provide a number of evolutionary advantages over asexual reproduction despite the associated costs. In sexual species the primary cost of sexual reproduction is the requirement for two individuals in order for mating to occur (Maynard Smith, 1978). This was illustrated by Maynard Smith (1978) by comparing sexual and asexual populations where the only difference was the presence or absence of sexual reproduction and where each mating event produced a single progeny. In a mating population with each individual reproducing twice (Figure 1.4A) the second generation will be comprised of the same number of individuals as the first. The asexual population, however, requires only a single reproductive event in order to maintain the size of the population in the second generation and when each individual reproduces twice doubles the size of the next generation (Figure 1.4B).

The costs associated with sexual reproduction are further increased by the requirement to find a mate and the presence of two distinct sexes in many sexual organisms, halving the size of the population with which an individual may mate (assuming the two sexes are balanced in number). Parasites such as trypanosomes, however, overcome a portion of this cost by lacking a two sex mating system (Turner *et al.*, 1990). By lacking distinct sexes each individual is therefore capable of mating with every other member of the population, as opposed to only those of the opposite sex, increasing the probability of mating occurring. It is unclear how *T. brucei* limits the occurrence of selfing, although it is likely a self incompatibility system is present preventing genetic exchange between closely related individuals.

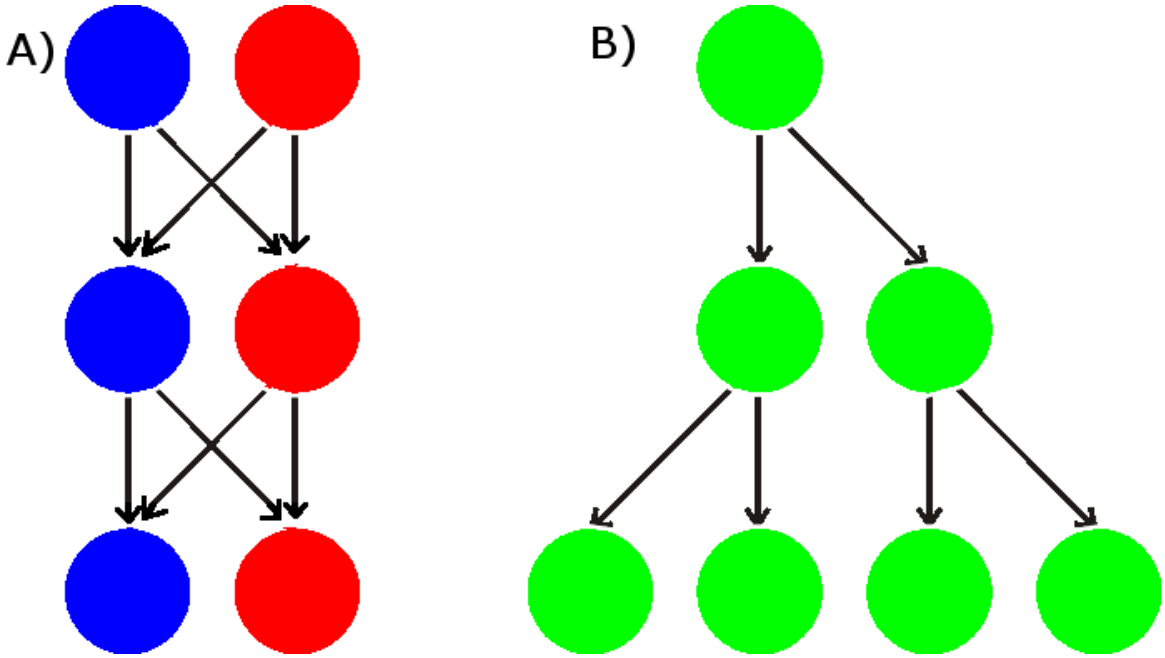


Figure 1.4 The two fold cost of sex

The two fold cost of sex as defined by Maynard Smith (1978). A) Two sexually reproducing individuals, each reproducing twice, produce two progeny in the next generation, maintaining the size of the population. B) An asexual organism, reproducing twice, produces two progeny in the next generation. The population therefore doubles in size with each generation.

Another disadvantage of mating is associated with the inheritance of genetic material. Asexual organisms pass their entire genome on to their progeny however in sexual reproduction each parent contributes only half of their genome to the progeny. Sexual organisms must therefore reproduce multiple times before there is a high probability of passing on their entire genome to the next generation.

Two of the primary advantages put forward for the evolution and maintenance of sexual reproduction are the introduction of variation and the promotion of DNA repair, which is itself associated with the introduction of further variation through meiotic recombination (Reviewed in Barton and Charlesworth, 1998; Birdsell and Wills, 2003; Hadany and Comeron, 2008; Hörandl, 2009; Neiman *et al.*, 2009; Otto, 2009).Genetic exchange may promote variation through the reassortment of chromosomes generating new combinations in progeny which may facilitate selection by occasionally generating individuals of very high or very low fitness. Such individuals

will be strongly selected for or against, leading to the maintenance of the fittest alleles and loss of the most deleterious. The majority of individuals, however, will fall within the centre of the fitness bell curve, with mildly deleterious alleles being masked by the presence of fitter alleles elsewhere, especially in diploid organisms where two alleles of each feature are typically present. By maintaining mildly deleterious alleles genetic exchange promotes diversity within populations and subsequent adaptation to shifting circumstances, as alleles that are deleterious in one situation may be beneficial in another.

DNA repair has been proposed as a further advantage of genetic exchange as the process of chromosome pairing during meiosis facilitates homologous recombination, increasing the ability of organisms to repair lesions or double stranded breaks. The mechanisms of DNA repair are also used in meiotic recombination, generating new variation by swapping material between chromosomal homologues. Without this recombination physically linked alleles will always be inherited together and therefore selection will act upon the entire chromosome, with fitter alleles forever influenced by the presence of deleterious alleles at other positions on the homologue. Meiotic recombination, in breaking these associations, allows for deleterious alleles to be selected against without affecting the fitter segments of the homologue. However, both chromosomal reassortment and meiotic recombination may break up advantageous combinations, potentially reducing the fitness of the progeny, a sharp contrast to clonal organisms where the most advantageous allele combinations are never broken up. In a truly panmictic population genetic exchange leads to agreement with Hardy-Weinberg predictions, with all allele combinations at a particular locus observed at frequencies dependent upon the allele frequency; a complete lack of linkage between loci on different chromosomes and the absence of repeated genotypes within the population. There are however many sources of deviation from this ideal, generating a spectrum of population structures.

Where genetic exchange occurs extremely rarely a species may be effectively clonal, existing as essentially independent lineages that occasionally recombine to generate a new hybrid lineage. *T. cruzi* is a prime example of an effectively clonal species, with six primary lineages currently recognised (Zingales *et al.*, 2009). A wide range of molecular markers have provided substantial evidence for clonality in the species, with a frequent observation of significant linkage disequilibrium between loci within individual populations and genetic differences more common between than within populations (Tibayrenc *et al.*, 1986, 1991; Tibayrenc and Ayala, 1987; Zhang *et al.*, 1988; Brenière *et al.*, 1991; de Luca D'oro *et al.*, 1993; Sanchez *et al.*, 1993; Oliveira *et al.*, 1998; Llewellyn *et al.*, 2009). Early evidence for genetic exchange in *T. cruzi* came from the detection of isoenzyme profiles of single genes that demonstrated the existence of homozygotes and heterozygotes within a single population, allowing the possibility that these profiles had arisen through mating of the two homozygous isotypes (Bogliolo *et al.*, 1996; Carrasco *et al.*, 1996).

Souto *et al.*(1996) further supported the existence of hybrids through the use of a dimorphism in the *T. cruzi* mini exon gene repeat that correlated with previously identified rRNA gene variation (Souto and Zingales, 1993). Of the six *T. cruzi* lineages four, TcIII – TcVI are now recognised as hybrid lineages (Reviewed in Sturm and Campbell, 2010), with TcIII and TcIV believed to have originated from fusion of the ancestral TcI and TcII lineages while TcV and TcVI arose from a backcross between isolates of the hybrid TcIII lineage with the 'parental' TcII isolate, giving rise to two widely heterozygous hybrids.

The possibility for genetic exchange in *T. cruzi* was experimentally confirmed in 2003 through the use of drug resistant isolates, allowing for the selection of double drug resistant progeny following infection of mammalian cell cultures (Gaunt *et al.*, 2003). In contrast to mating in *T. brucei* and *L. major* genetic exchange in *T. cruzi* does not appear to be Mendelian in nature and does not appear to occur during transmission through the vector. Genetic exchange in *T. cruzi* likely involves fusion between the parental lines, as evidenced by the presence of aneuoploidy in the progeny clones resulting in 1.65-1.72 times more DNA than the parental cells and the inheritance of multiple parental alleles at the loci which were examined (Gaunt *et al.*, 2003; Lewis *et al.*, 2009). It has been suggested that diploidy in progeny hybrids may be subsequently restored through recombination and chromosome loss until stability is restored. It is unclear at present how long this process takes. Alongside the apparent fusion model of genetic exchange in *T. cruzi* the apparent localisation of genetic exchange, occurring within the host as opposed to the vector suggests further differences from genetic exchange in *T. brucei* and *L. major*.

Leishmania sp., like *T. cruzi*, has long been assumed to be effectively clonal based upon early isoenzyme studies which detected significant linkage disequilibrium; deviation from Hardy Weinberg equilibrium and the existence of stable zymodemes (Evans *et al.*, 1987; Desjeux and Dedet, 1989; Tibayrenc *et al.*, 1990; Ayala, 1993; Cupolillo *et al.*, 1997; Bañuls *et al.*, 1999a). Coupled to geographic restrictions and differences in clinical progression these genetic differences led to the creation of numerous species although more recently it has been proposed that the current *L. donovani* complex should be condensed to only two distinct species, *L. infantum* and *L. donovani* (Bañuls *et al.*, 1999a; Lukes *et al.*, 2007).

While the use of microsatellite markers has supported the conclusions of effective clonality based upon excesses of heterozygotes, homogeneous populations and linkage disequilibrium (Schwenkenbecher *et al.*, 2006; Kuhls *et al.*, 2007; Al-Jawabreh *et al.*, 2008; Alam *et al.*, 2009) a number of studies have now identified the presence of putative hybrid populations where the individuals share ancestry from two neighbouring populations (Kuhls *et al.*, 2008; Seridi *et al.*, 2008; Chargui *et al.*, 2009; Rougeron *et al.*, 2009; Gelanew *et al.*, 2010). While these studies suggest a role for genetic exchange in *Leishmania sp.* they supported the dominance of clonality,

with genetic exchange being relatively rare. The identification of potential interspecies hybrids between *L. braziliensis* and *L. guyanensis* (Bañuls *et al.*, 1997, 1999b; Delgado *et al.*, 1997); *L. braziliensis* and *L. panamensis* (Belli *et al.*, 1994; Bañuls *et al.*, 1997); *L. braziliensis* and *L. peruviana* (Dujardin *et al.*, 1995; Nolder *et al.*, 2007); *L. infantum* and *L. major* (Ravel *et al.*, 2006) supports both the occurrence of genetic exchange and requirement for a streamlined taxonomy (Bañuls *et al.*, 1999a; Lukes *et al.*, 2007).

While the majority of studies support rare genetic exchange, leading to effectively clonal populations some have suggested sexual reproduction may occur more frequently based upon the detection of populations more in line with Hardy-Weinberg expectations and the presence of epidemic population structures (Bastien *et al.*, 1992; Blaineau *et al.*, 1992; Nolder *et al.*, 2007). Rougeron *et al.* (2009), Gelanew *et al.* (2010) and Rougeron *et al.* (2011), in analysing isolates of *L. braziliensis, L. major* and *L. guyanensis* identified the presence of significant homozygote excesses and high F_{IS} values, concluding that extensive and frequent inbreeding had to be occurring within the populations. Rougeron *et al.* (2009) proposed that the high levels of inbreeding may have been facilitated by the relatively low incidence of *Leishmania sp.*in the vector (Martín-Sánchez *et al.*, 2006; Rogers and Bates, 2007) and that due to the relatively limited dispersal range of the sandfly over its lifetime (Morrison *et al.*, 1993) *Leishmania sp.* may exist in micro-populations which facilitate inbreeding while being too small to separate during sample collection, giving rise to apparent clonality due to the Wahlund effect (Wahlund, 1928).

Experimental demonstration of genetic exchange in *Leishmania sp.* was only recently demonstrated through the use of two drug resistant clones of *Leishmania major*, allowing for the selection of double drug resistant progeny cells following transmission through the insect vector *Phlebotomus duboscqi* (Akopyants *et al.*, 2009). SNP genotyping of progeny indicated each to have inherited a full set of chromosomes from the two parental lines. Through comparison with lines of known ploidy (Cruz *et al.*, 1993) it was determined that 7 of the 18 progeny clones possessed raised DNA contents equivalent to a ploidy of 3n, while 4 of the 2n lines were also indicated to have raised ploidy for chromosome 31, consistent with the potential for aneuploidy that has been observed in *Leishmania sp.*(Cruz *et al.*, 1993; Ubeda *et al.*, 2008; Sterkers *et al.*, 2011). Genetic exchange in *Leishmania major* as reported by Akopyants appears to resemble that of *T. brucei*; occurring within the insect vector; biparental inheritance of chromosomes; observations of raised DNA contents for some but not all of the progeny and apparent non obligatory nature of the process.

The five human infective species of *Plasmodium* provide contrasting examples on the role of genetic exchange in parasite populations, as while the parasite uses obligatory sexual cycle both near panmictic and effectively clonal population structures have been observed. During the vast majority of the life cycle the *Plasmodium* parasites exist as haploid cells, with the development

cycle in the host culminating in the differentiation of the merozoite stage into the distinct male and female gametocytes (Reviewed in Talman *et al.*, 2004; Dixon *et al.*, 2008). Fertilisation in the vector generates the only diploid stages of the life cycle in the form of the zygote and ookinete. Following formation of the oocyst (Reviewed in Baton and Ranford-Cartwright, 2005) the progeny returns to a haploid state and undergoes multiple rounds of mitotic reproduction. While genetic exchange is an obligatory stage of the *Plasmodium* life cycle the parasite highlights that more than just the frequency of genetic exchange may determine the population structure. With both effectively clonal and panmictic populations reported in the literature the consensus for *Plasmodium* is that it is the intensity of transmission that primarily controls the population structure by determining the rate of mixed host infections and therefore the frequency of outcrossing (Tibayrenc *et al.*, 1990, 1991; Walliker, 1991; Read and Day, 1992; Paul *et al.*, 1995; Tibayrenc, 1995; Babiker and Walliker, 1997; Hastings and Wedgwood-Oppenheim, 1997; Ayala, 1998; Paul and Day, 1998; Rich *et al.*, 1998; Conway *et al.*, 1999; Anderson *et al.*, 2000a, 2000b; Urdaneta *et al.*, 2001; West *et al.*, 2001; Hartl *et al.*, 2002; Mu *et al.*, 2005; Razakandrainibe *et al.*, 2005).

In high transmission regions such as Africa and Papua New Guinea the prevalence of host infections maintains a sufficient level of mixed vector infections for out crossing to play a significant role on the population structure, shifting it towards panmixia (Conway *et al.*, 1999; Anderson *et al.*, 2000a; Anthony *et al.*, 2005; Schultz *et al.*, 2010). High transmission rates need not necessarily lead to panmixia however, in Papa New Guinea it has been proposed that an intermediate population structure was present in the early 1990s with neither selfing nor out crossing dominating (Paul *et al.*, 1995) while in Cameroon and Kenya significant levels of linkage disequilibrium and selfing has been detected even in the presence of high rates of transmission (Durand *et al.*, 2003; Razakandrainibe *et al.*, 2005; Annan *et al.*, 2007; Schultz *et al.*, 2010). When transmission intensity drops the population structure often shifts towards an effectively clonal population due to the increased rate of inbreeding (Babiker and Walliker, 1997; Anderson *et al.*, 2000a; Anthony *et al.*, 2005). With a lower frequency of transmission mixed infections in the host and therefore the mosquito vector will be rarer with a corresponding drop in the frequency of out crossing. As with selfing in high transmission regions low transmission regions do not automatically exclude significant levels of out crossing (Pumpaibool *et al.*, 2009).

1.7.2 Genetic exchange in *T. b. gambiense* Group 1

The role of genetic exchange is clearest for *T. b. gambiense* Group 1, the only *T. brucei* subtype for which genetic exchange has not been experimentally demonstrated (although the possibility of rare mating cannot be fully excluded). Early analysis of Group 1 isolates identified the presence of a highly homogeneous group (Gibson, 1986; Paindavoine *et al.*, 1986a; Richner *et al.*, 1989) with Tibayrenc describing the sub-species as "just an instance of a successful, ubiquitous human-host

clone" (Tibayrenc *et al.*, 1990). Later work comparing Group 1 populations with the use of more sensitive markers indicated that while small levels of genetic variability were present at the local level there was a higher degree of variability between geographically distant populations, consistent with a lack of gene flow between them and independent evolution of the loci (Mathieu-Daudé *et al.*, 1995; Kanmogne *et al.*, 1996b; Biteau *et al.*, 2000; Jamonneau *et al.*, 2000; Koffi *et al.*, 2007; Morrison *et al.*, 2008b). However in many of these studies the extent of homozygosity within individual populations has prevented empirical testing for clonality.

If *T. b. gambiense* Group 1 is a strictly clonal organism, lacking the ability to undergo any type of genetic exchange, then this is likely to be only a recent development in the evolution of the sub-species. The lack of an observable Meselson effect (Judson and Normark, 1996; Mark Welch and Meselson, 2000) within natural populations (Morrison *et al.*, 2008b) and the *T. b. gambiense* Group 1 reference sequence (Jackson *et al.*, 2010) suggest the sub-species has not had sufficient time to accumulate extensive heterozygosity. The low level of divergence between the *T. b. gambiense* Group 1 and *T. b. brucei*, as determined by comparison of kDNA haplotypes (Balmer *et al.*, 2011) and genome sequences (Jackson *et al.*, 2010) provides further evidence that the evolution of clonality in *T. b. gambiense* Group 1 is a recent event.

1.7.3 Genetic exchange in *T. b. brucei*, *T. b. gambiense* Group 2 and *T. b. rhodesiense*

The role of genetic exchange in *T. brucei* has caused considerable controversy over the years as while mating has been empirically demonstrated to occur under laboratory conditions (Jenni *et al.*, 1986) its frequency and significance in the field remains disputed. The issue is complicated by existence of multiple sub-species and subgroups, namely *T. b. gambiense* Group 1 (discussed above), *T. b. gambiense* Group 2, *T. b. brucei* and *T. b. rhodesiense*. Whether the respective sub-species are genetically isolated or can interbreed freely could have profound effects upon the observed population structures if incorrectly grouped or separated.

While the possibility for genetic exchange in trypanosomes was first proposed based on the observation of potential *T. b. brucei* recombinant genotypes in a field population (Tait, 1980; Gibson *et al*., 1980), a number of other isoenzyme studies found evidence for significant linkage between loci; deviation from Hardy-Weinberg equilibrium and the presence of many repeated genotypes within populations, leading to the conclusion of effective clonality in *T. brucei* (Tait *et al.*, 1985; Paindavoine *et al.*, 1989; Tibayrenc *et al.*, 1990, 1991; Mathieu-Daudé and Tibayrenc, 1994; Stevens and Tibayrenc, 1995). With the identification of *T. b. rhodesiense* to be more genetically homogenous than and genetically distinct from sympatric *T. b. brucei* populations, it has became apparent that this may have an impact upon the population structures of the respective sub-species and separate investigations into the two sub-species have suggested the presence of

epidemic and effectively clonal population structures for *T. b. brucei* and *T. b. rhodesiense* populations respectively (Hide *et al.*, 1994; Stevens and Tibayrenc, 1996; MacLeod *et al.*, 2000, 2001c; Njiokou *et al.*, 2004a; Hide and Tait, 2009).

These results of differing population structures in *T. b. brucei* and *T. b. rhodesiense* are of particular interest given the role of a single gene, *SRA* in conferring resistance to human serum, effectively defining *T. b. rhodesiense* (De Greef *et al.*, 1989; De Greef and Hamers, 1994; Xong *et al.*, 1998) and observations that while local *T. b. rhodesiense* and *T. b. brucei* populations may be genetically distinct from one another they are more closely related to one another than geographically separated populations of *T. b. rhodesiense* (MacLeod *et al.*, 2001c), suggesting at least some level of interaction between the sub-species. The observed genetic separation however indicates barriers to complete gene flow remain despite the frequent detection of *T. b. rhodesiense* in non human hosts where *T. b. brucei* can also be found (Hide *et al.*, 1998; Welburn *et al.*, 2001; Waiswa *et al.*, 2003; Njiru *et al.*, 2004b; Enyaru *et al.*, 2006). In addition to any barriers to mating between the sub-species further ones must be present in *T. b. rhodesiense* populations in order to account for the effective clonality that has been observed. The relatively homogeneous nature of *T. b. rhodesiense* within a given population may limit the rate of genetic exchange if there are mechanisms in place to minimise the frequency of inbreeding. Alternatively outbreaks of *T. b. rhodesiense* may occur due to the presence of exactly the right combination of genetic material beyond merely the presence of *SRA*, with the genetic disruption caused by mating sufficient enough to eliminate the ability to successfully infect humans.

Studies into *T. b. gambiense* group 2, identified as distinct from *T. b. gambiense* Group 1 due to their more heterogeneous population and closer relationship to *T. b. brucei* than *T. b. gambiense* Group 1 (Gibson, 1986; Hide *et al.*, 1994; Balmer *et al.*, 2011; Capewell, 2011) have suggested that the subgroup undergoes frequent mating and appears like *T. b. rhodesiense* to be a human host variant of *T. b. brucei*. It has been proposed that the ability of this subgroup to infect humans may have arisen through a rare mating event with *T. b. gambiense* Group 1 (Radwanska *et al.*, 2002), with the Group 1 signature diluted through subsequent mating with *T. b. brucei*. Alternatively Group 2 may represent a novel genesis of human serum resistance opening the potential for new human infective sub-species to arise in the future.

1.7.4 Genetic exchange in *T. congolense*

Like *T. vivax* there has been relatively little study into the population genetics of *T. congolense* and it has long been considered to be effectively clonal based upon the Tibayrenc's criteria and analysis (Tibayrenc *et al.*, 1990, 1991). This analysis was based upon the isoenzyme typing of 114 stocks of *T. congolense* from across Africa with 6 isoenzymes (Gashumba *et al.*, 1988). The initial analysis

of this study identified 71 unique zymodemes which fell into 5 distinct groups, defined by a combination of geographic and ecological origins. The use of this sample set in assessing the frequency of genetic exchange in *T. congolense* is however limited. The samples were collected from across not only a wide geographic region but had been collected over a 23 year period (1960 – 1983), both of which are likely to have introduced a Wahlund effect (Wahlund, 1928), disrupting the ability to detect the recombination events associated with mating. The conclusion of clonality based upon this sample set is therefore of limited value.

Since that initial declaration of effective clonality there has been little further research. The only publication since Tibayrenc's analyses has been a microsatellite study into *T. congolense* in The Gambia (Morrison *et al.*, 2009b). Utilising a panel of 7 polymorphic microsatellites Morrison *et al.* genotyped 89 members of a *T. congolense* population isolated from horses, donkeys and cattle in The Gambia and detected a high frequency of unique multilocus genotypes (MLGs); an excess of homozygotes at all loci; a lack of observable linkage between pairs of loci and significant deviation from Hardy-Weinberg equilibrium, consistent with the occurrence of frequent genetic exchange and a high level of inbreeding. If the conclusions from both the isoenzyme and microsatellite data are considered valid then it would appear that genetic exchange is not only possible in *T. congolense* but may vary in frequency in different populations and subgroups, as has been observed for *T. brucei*.

1.7.5 Genetic exchange in *T. vivax*

T. vivax has received exceedingly little attention with regards its population genetics due to the low parasitaemia associated with many of its natural hosts and the difficulty in isolating and culturing the species (Gardiner, 1989). There have therefore been only a handful of studies into the population genetics of the species studies (Kilgour *et al.*, 1975; Kilgour and Godfrey, 1977; Allsopp and Newton, 1985), later analysed by Tibayrenc (1991). The studies of Kilgour (1975) and Kilgour and Godfrey (1977) were the first to investigate the diversity of *T. vivax* in populations, identifying dominant zymodemes which were maintained during the three year period between the studies. This was later expanded upon by Allsopp and Newton (1985) which identified a further 9 unique zymodemes.

These studies provided the first evidence for clonality in the species due to the over representation of a small number of genotypes and significant levels of linkage disequilibrium between markers (Tibayrenc *et al.*, 1990, 1991). It should be noted however that the power of these studies were extremely limited, namely by the number of markers or number of samples used. The conclusion of clonality in *T. vivax* based upon these studies should therefore be treated with caution. Despite the advances in molecular and analytical techniques in the time since this conclusion was reached

there have been no attempts to further the investigation of genetic exchange in *T. vivax* prior to the results presented in this volume.

1.8 Aims of this study

The primary aim of the work presented here was to assess the population diversity, population structures and frequency of genetic exchange in the African Salivarian trypanosomes *T. b. rhodesiense*, *T. congolense* and *T. vivax* populations from Uganda, Malawi and The Gambia. Through the use of populations separated by both time (*T. b. rhodesiense* and *T. congolense*) and geography (*T. b. rhodesiense*) this work has examined the dynamics of genetic flux within these populations. These studies shed new light on the frequency and role of genetic exchange within these three species, which together are responsible for the clinically and economically important diseases of HAT and AAT.

The microsatellite markers used in these population studies are, however, relatively low density markers, limiting the studies which may be undertaken. With the rapidly declining cost of whole genome sequencing the use of SNP based markers is becoming increasingly viable. The final portion of this project is therefore focused upon the identification and comparative analysis of SNP based genomic diversity in three laboratory strains, providing the foundation for a future resource of trypanosome genomic variation. This SNP variation was then employed in the construction of a high density SNP map, which with the aid of the microsatellite genetic map (MacLeod *et al.*, 2005b), has allowed for an initial reconstruction of the haplotypes of the *T. brucei* reference strain.

Chapter 2

Materials and Methods

2.1 Chemicals, enzymes and reagents

Company	Reagent
BD	BD 4.0 ml vacutainers
Cambrex Bioscience	SeaKem LE agarose, NuSieve GTG agarose
GE Healthcare	GenomiPhi DNA Amplification Kit
MWG-Biotech	Custom oligonucleotides
New England Biolabs	100bp ladder
Qiagen	QIAamp DNA blood midikits
Strataclone	*Pfu* DNA polymerase, StrataClone PCR cloning kit
Thermo Scientific	Custom PCR mastermix, *Taq* DNA polymerase
Whatman	FTA Purification Reagent

2.2 Buffers and solutions

LB agar – 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar (autoclaved)

LB medium – 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl (autoclaved)

6 x electrophoresis loading buffer – 0.25% (w/v) bromophenol blue, 0.25% (w/v) orange-G, 30 (v/v) glycerol. Stored at 4°C.

TE buffer – 10mM Tris-HCl ph 8, 1 mM EDTA pH 8 (autoclaved).

5 x Tris-Borate-EDTA buffer (TBE) – 450mM Tris, 450mM Boric acid, 10mM EDTA, pH 8. Diluted 1/10 in dH_2O prior to use as gel electrophoresis buffer.

X-Gal solution – 2% (w/v) stock solution dissolved in dimethylformamide. Stored at -20°C, protected from light.

2.3 Strains, F1 progeny and field samples

2.3.1 Trypanosome strains

Five cloned trypanosome strains were used in this investigation as PCR positive controls or material for genome sequencing. DNA was extracted using DNA extraction kit (Qiagen) following manufacturers' protocols. The five stocks were TREU 927 (*T. b. brucei*), STIB 247 (*T. b. brucei*), STIB 386 (*T. b. gambiense* Group 2), ILRAD V34 (*T. vivax*) and IL3000 (*T. congolense*).

2.3.2 F1 Progeny strains

Two progeny clones, hybrid 77 and hybrid 86, derived from a genetic cross between STIB 247 / TREU 927 were used as samples in the whole genome sequencing project (Chapter 6). The generation and derivation of these progeny clones has been previously described (Turner *et al.*, 1990; MacLeod *et al.*, 2005a) while the purified DNA for high throughput genome sequencing was generated by A. Cooper as previously described (Cooper, 2009).

2.3.3 Field Samples

The origins and trypanosome species present of the field sample populations used in this investigation are described in Table 2.1.

Population	Country of origin	Sampling period	Number of samples	Sample type	Host	Trypanosome species present
Gambia 2007	The Gambia	2006-2007	323 equine 208 bovine	Blood spot on FTA card	Equines Bovines	*T. brucei* *T. congolense* *T. vivax*
Gambia 2009	The Gambia	2009	198	Blood from host	Equines	*T. brucei* *T. congolense* *T. vivax*
Tororo pre-1991*	Uganda / Kenya	1961-1990	52	DNA from cloned stabilates	Humans Bovines	*T. b. rhodesiense*
Tororo 2003	Uganda	2002-2003	30	Blood samples and blood spots on FTA cards	Humans	*T. b. rhodesiense*
Soroti 2003	Uganda	2002-2003	84	Blood spots on FTA cards	Humans	*T. b. rhodesiense*
Kaberamaido 2009	Uganda	2008-2010	86	Blood spots on FTA cards	Humans	*T. b. rhodesiense*
Malawi 2003	Malawi	2002-2003	28	Blood spots on FTA cards	Humans	*T. b. rhodesiense*

Table 2.1 Origin of field samples used in this study

* This population includes samples from the wider Tororo focus including regions in western Kenya. All except four isolates originate from the 1988-1990 period.

2.4 Molecular biology techniques

The methods used during this investigation predominantly followed standard molecular biological techniques or manufacturers' protocols unless otherwise described. The methods described here therefore give only brief descriptions of the protocols used, with additional details listed for modified or non-standard protocols.

2.4.1 Preparation of FTA card bloodspots for use as template in Whole Genome Amplification or PCRs

Blood spots on FTA filter cards were prepared for use as previously described (Morrison *et al.*, 2007). Briefly, discs were punched from blood spots using a clean 2 mm diameter Harris micropunch (Whatman), washed 3 times with 200 µl FTA purification reagent and twice with 200 µl 1 mM TE buffer, with incubation during each wash for 5 minutes as per manufacturer's protocol. Discs from the same blood spot were washed in a single tube, with a maximum of 6 discs per tube. Washed discs were air-dried for 1 hour prior to whole genome amplification or use in PCRs.

2.4.2 Whole genome amplification from FTA card bloodspots

Washed PCR filter card discs were used as direct substrate for whole genome amplification, in a final volume of 20 µl as described by the manufacturer protocol for the GenomiPhi DNA amplification kit, with 2 washed discs per reaction. Following amplification 3 independent reactions were pooled together into a single 60 µl aliquot in order to account for the possibility of allele dropout. 1 µl of the pooled reactions was used as template for subsequent PCRs.

2.4.2 Collection of whole blood samples from equines

For collection of the 2009 Gambian equine samples jugular whole blood samples were collected into 4 ml ethylene diamine tetra-acetic acid (EDTA) coated vacutainer tubes with collection performed by trained staff members of the Gambia Horse and Donkey Trust (GHDT; http://www.gambiahorseanddonkey.org.uk). The collected blood samples were used for blood measurements of packed cell volume (PCV), total protein (TP) and microscopic examination of whole blood films immediately following collection as previously described (Pinchbeck *et al.*, 2008). PCV was assessed using a 'Spindoctor™' manual centrifuge (Figure 1.3) while a solar-powered, portable microscope (Diamedica Limited, UK) was used for visual screening. Each sample was screened for a minimum of 20 fields of view or 5 minutes screening time. Total blood

cholesterol was measured utilising the BeneCheck Plus total cholesterol test following the manufacturer's instructions. The remaining blood was stored in a chilled cool bag and processed immediately upon return to the GHDT headquarters at Sambel Kunda as described in 2.4.3.

2.4.3 Extraction of DNA from blood samples

DNA extraction from whole blood samples was performed using Qiagen mini-, midi- and maxi-blood kits, following manufacturer's protocols. The Gambian 2009 equine blood samples were part processed at the GHDT headquarters prior to shipping to the UK. Briefly, 2 ml of blood, collected in EDTA vacutainers was transferred to fresh tubes prior to the addition of QIAGEN protease and lysis buffer (Kit buffer AL) as dictated by the QIAamp DNA blood midikit protocol. Samples were incubated at 70°C for 10 minutes and stored at 4°C (protocol steps 1-4) until being frozen at -20°C on return to the UK. DNA extraction in the UK, continued from protocol step 5 following thawing of the samples.

2.4.4 General PCR procedure

The high sensitivity of the PCR technique required precautions to be taken in order to prevent contamination of the equipment and reagents by genomic DNA and PCR products. Aliquots of dedicated PCR reagents and pipette tips were used and kept separate from PCR products where possible. Negative controls were included in all PCRs in order to identify the presence of contamination, when this occurred reagents and tips were replaced with fresh materials and PCRs repeated.

The standard PCR reaction mixes, unless otherwise stated were composed of: 45 mM Tris-HCl pH 8.8, 11 mM $(NH_4)_2SO_4$, 4.5 mM $MgCl_2$, 6.7 nM 2-mercaptoethanol, 4.4 µM EDTA, 113 µg ml-1 BSA, 1 mM dATP, 1 mM dGTP, 1 mM dTTP and 1mM dCTP (Custom PCR mastermix), 1 µM each oligonuclotide primer, 0.5 units *Taq* DNA polymerase and 1 µl DNA template. PCR grade H_2O was used during the preparation of reaction mixes as required. For nested PCR first round products were diluted 1/100 prior to use as template in the second round. PCRs were performed in a volume of 10 µl unless otherwise stated.

Cycling conditions for the speciation primer sets TBR, TCS and TVW were 30 cycles of:

Denature	94°C for 50 seconds
Anneal	60°C for 50 seconds
Extend	72°C for 60 seconds

Cycling conditions for the *T. brucei* and *T. vivax* microsatellite markers were 28 cycles of:

Denature	95°C for 50 seconds
Anneal	55°C for 50 seconds
Extend	65°C for 60 seconds

for both first and second round reactions.

Cycling conditions for the *T. congolense* microsatellite markers were 28 cycles of:

Denature	95°C for 50 seconds
Anneal	52°C for 50 seconds
Extend	65°C for 60 seconds

for both first and second round reactions.

2.4.5 Modifications to the general PCR protocol for use of washed FTA discs

For PCRs utilising amplification directly from FTA discs the volume of each reaction was increased to 20 µl with 2 µl of PCR grade H_2O replacing the 1 µl DNA template used in the standard 10 µl reactions.

2.4.6 Modifications to the general PCR protocol for cloning and DNA sequencing

Where amplified DNA products were required for downstream cloning and sequencing the standard PCR protocol was adapted through the addition of the proofreading polymerase *pfu* DNA polymerase, with inclusion of *pfu* in the PCR mix at a ratio of 1:9 with *Taq* polymerase.

For cloning an additional step was required in order to add the 3' adenine residue required for ligation to the Strataclone PCR cloning vector pSD-A-amp/kan. While the *Taq* polymerase catalyses the addition of this 3' overhang during PCR amplification the proofreading polymerase *pfu* does not. In order to correct this following completion of the initial PCR cycle 0.1 µl *Taq* was added to PCR reactions and subsequently incubated for 10 minutes on a robocycler preheated to 72°C. Reactions were then placed on ice prior to proceeding to the ligation step as described in section 2.4.9.

2.4.7 Gel electrophoresis and product size determination

PCR product from speciation reactions utilising the TBR, TCS and TVW primer sets were separated by gel electrophoresis on 1% (w/v) Seakem LE agarose gels, loaded alongside 100 bp DNA ladder to facilitate estimation of product size and visualised by UV using a transilluminator.

Microsatellite PCR products were separated by gel electrophoresis on 3% (w/v) Nusieve GTG agarose gels, loaded alongside 100 bp DNA ladder to facilitate estimation of product size. For accurate size determination one internal primer of each nested microsatellite primer set was labelled with a 5' FAM or HEX modification. Following dilution based upon the intensity of visible band on agarose gels products were size-separated on a capillary-based sequencer (ABI 3100 Genetic analyser; Applied Biosystems; Dundee sequencing service http://www.dnaseq.co.uk) alongside a set of ROX labelled size standards (GS400HD markers, Applied Biosystems). This allowed for the determination of DNA fragment length using the Peak Scanner v1.0 software (Applied Biosystems) with microsatellite alleles defined as individual peaks on the trace.

2.4.8 Olignonuclotides

All oligonucleotide primers used in this investigation were synthesised by MWG-Biotech based upon previously described material (Masiga *et al.*, 1992; MacLeod *et al.*, 2005b; Duffy *et al.*, 2009; Morrison *et al.*, 2009b). The sequences of all primers used are listed in Appendices 1 - 4.

2.4.9 Cloning and sequencing

2.4.9.1 Ligation of PCR product to the cloning vector

Cloning of PCR products followed the standard manufacturers protocol for the Strataclone PCR cloning kit. Products were ligated into the cloning vector pSC-A-amp/kan by mixing of 3 µl Strataclone cloning buffer, 2 µl undiluted PCR product and 1 µl Strataclone pSC-A-amp/kan cloning vector which was incubated at room temperature for five minutes before being placed on ice.

2.4.9.2 Transformation of *E. coli* competent cells

PCR product, ligated into the cloning vector was used for the transformation of *E. coli* Strataclone Competent cells following the manufacturers protocol. Cells were thawed on ice before 1 µl of the ligation mix was added and incubated on ice for 20 minutes. Following incubation the mixture was

heat shocked at 42°C for 45 seconds before being returned to ice for a further two minutes. During this time 250 µl of LB media was warmed to 37 °C before being added to the cells which were allowed to recover by incubation at 37 °C with shaking at 200 rpm for a total time of 1 hour. LB plates containing 100 µg/ml ampilcillin and prepared in advance were warmed to room temperature and 40 µl of 2% (w/v) X-Gal was spread on each to allow for blue / white colour screening of the transformed cells. At the end of the 1 hour incubation the transformed cells were spread onto the agar plates at a range of dilutions and incubated overnight at 37 °C to allow for colony growth.

2.4.9.3 Analysis of transformed cells

After overnight growth transformed cells were identified by the presence of white colonies which were picked and incubated in 5 ml LB media containing 100 µg/ml ampicillin and cultured overnight at 37 °C in a shaker set to 200 rpm. Aliquots of the overnight culture were pelleted by centrifugation at 5000 G for 5 minutes and the supernatant discarded. The QIAprep spin miniprep kit was utilised to extract the amplified plasmids from the cell pellets as described in the manufacturers protocol with purified plasmid DNA eluted into with 50 µl PCR grade H_2O which was stored at -20 °C until required. Plasmid preparations were checked for the successful integration of the desired PCR product by PCR amplification from the purified DNA and visualisation following gel electrophoresis. Sequencing of purified DNA was carried out by the Dundee sequencing service (University of Dundee) with samples prepared as according to the instructions provided by the service.

2.4.10 DNA quantification

Where required DNA concentration was determined utilising 1 µl of a sample, using the nucleic acid measurement program of a nanodrop ND-1000 spectrophotometer following calibration with appropriate solutions. Following dilution of samples the DNA concentration was re-measured and dilutions repeated if necessary.

2.4.11 High-throughput whole genome sequencing

2.4.11.1 Sample preparation

Genomic DNA from the parental strains TREU 927, STIB 247 and STIB 386, alongside genomic DNA from the two F1 progeny (hybrids 77 and 86) was prepared by A. Cooper from procyclic form cell cultures of the respective lines and sent to the Pathogen Sequencing Unit, WTSI for whole genome sequencing.

2.4.11.2 Genome Sequencing

Genomic DNA sample processing, library construction and Illumina Solexa sequencing was performed at the WTSI using the Illumina Genome Analyser System, following manufacturers and internal WTSI protocols. Each sample was run on a single lane of one Illumina Genome Analyser run generating paired end, 76 bp reads. Reads were filtered for initial quality control following Sanger protocols and made available via FTP for genome assembly and analysis.

2.5 Data Analysis

The figures produced for this investigation were produced as graphs in Microsoft Excel or exported directly from analysis software. Image manipulation (cropping, colour alteration, addition of text) was achieved using the GNU Image Manipulation Program (GIMP; http://www.gimp.org).

2.5.1 Statistical analysis of population genotyping data

MLGs were generated from the specific allele combinations across the markers utilised to study each field population, with each unique MLG assigned an ID number. Dendrograms of similarity were generated from MLG data utilising clustering calculator (http://www2.biology.ualberta.ca/jbrzusto/cluster.php). Genetic distances for the dendrograms were determined using Jaccard's similarity co-efficient with the unweighted pair group method with arithmetic mean (UPGMA) algorithm utilised in tree construction and visualised with Treedyn (Chevenet *et al.*, 2006). Hardy-Weinberg equilibrium, linkage disequilibrium between pairs of loci and F_{IS} were calculated using the Genetic Distance Analysis program (http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php). GenAlEx (Peakall and Smouse, 2006) was used for analysis of allele frequencies; pairwise F_{ST}; unbiased genetic distance (D) determination; and principal co-ordinate analysis (PCoA). The program MLGism (Stenberg *et al.*, 2003) was used to determine the probability of replicated genotypes occurring due to sexual recombination. Psex values were calculated for each replicated MLG and compared with values generated by 10^6 simulated populations. Multiple variable linear regression models were constructed in Minitab 15 while population sub-structuring was assessed with STRUCTURE (Pritchard *et al.*, 2000) and InStruct (Gao *et al.*, 2007).

2.5.2 Analysis of whole genome sequencing data

2.5.2.1 Assembly of genome sequences

High-throughput whole genome sequence data were provided by the WTSI as paired read sequences collected in the FASTQ format following internal quality checking. Reads were trimmed from the 3' end until the average base quality for the read was of 20 or higher. Reads were then aligned to the TREU 927 *T. brucei* reference genome (version 5) (Berriman *et al.*, 2005) using BWA v0.5.9-r16 (Li and Durbin, 2009) utilising the following commands:

```
bwa index 927-v5.fa
bwa aln -t 8 -q 20 927-v5.fa in_1.fq > out_1.sai
bwa aln -t 8 -q 20 927-v5.fa in_2.fq > out_2.sai
bwa sampe -a 1000 927-v5.fa in_1.sai in_2.sai in_1.fq in_2.fq > out.sam
```

2.5.2.2 SNP calling

Putative SNPs within the assembled genomes were called using the SAMtools software (Li *et al.*, 2009) by evaluation of the relative frequency of alleles at each position following quality and positional filtering. At each position bases are filtered for individual base quality and mapping quality, average base and mapping quality of surrounding bases on the read and proximity to other SNPs or indels. Reads which fail to meet the quality threshold for any of these conditions are excluded from SNP calling. Following quality assessment the consensus is called based upon the observed frequencies of the remaining bases and further filtered to exclude positions with an insufficient read depth and where the confidence of the SNP is too low. Homozygous SNPs indicate a single base, in disagreement with the reference sequence at the position in question while heterozygous SNPs are called when two bases are present, each at a sufficient frequency. Complex SNPs where 3 or more alleles are indicated to be present are filtered out due to the diploid nature of trypanosomes. For the TREU 927 sequence a minimum read depth of 20 aligned, high quality bases were required during SNP filtering while the lower average fold coverage for STIB 247, STIB 386, hybrid 77 and hybrid 86 necessitated the use of a minimum coverage of 10 aligned, high quality bases. A cap of 3 x the average chromosomal coverage was employed during SNP calling for all assemblies.

2.5.2.3 Development of custom scripts for analysis

Custom scripts for the extraction and analysis of genome sequence and SNP data were manually designed and written for use with Perl 5.10. The functionality of scripts was tested utilising subsets of the available data with results manually verified. The haplotype reconstruction scripts are described in Chapter 6.2.3.

Chapter 3

The population genetics of *Trypanosoma vivax* in The Gambia, 2006 – 2007

3.1 Introduction

Nagana, as described in Chapter 1, imposes a considerable burden to both animal welfare and the economies of developing countries reliant on susceptible livestock. The Gambia, a small sub-Saharan country in the west of Africa is located within the northern confines of the tsetse belt and is one of the many countries affected by animal trypanosomiasis. Running along the downstream portion of the Gambia River the seasons alternate between a long dry October to June period interspaced with a short, intense rainy season during July, August and September. Tsetse flies capable of transmitting the Salivarian trypanosomes are prevalent throughout the country during the rainy season while being constrained to river and water habitats during the dry season. The country has a population estimated at 1.6 million, of whom ~80% are involved in the agricultural sector and ~60% live at or below the national poverty line (WHO, 2009). Trypanosomiasis therefore has the potential to have a significant impact upon the livelihood of the majority of the population.

Animal usage within The Gambia is primarily split between herd animals such as cattle, goats and sheep and draft animals, typically oxen, horses and donkeys which are used for transportation and farm labour. While herd animals may be kept for meat they are valued as an investment in and of themselves, with manure and milk production of herds highly important (Jaitner *et al.*, 2003). The 1993 livestock census (DLS-ITC, 1993) recorded 278,538 cattle spread over 5,030 herds, almost exclusively of the *Bos taurus* N'Dama breed. This breed of cattle is widely reported to be trypanotolerant, able to control both parasitaemia and the symptoms of trypanosomiasis during infection (Naessens, 2006). Alongside cattle the census recorded 68,721 draught animals in the country. Equines, predominantly donkeys, accounted for 75% of these while work oxen / cows comprised the remainder of the animals. While the cattle of the country may be trypanotolerant, equines are reported to be highly susceptible to the effects of trypanosomiasis (Snow *et al.*, 1996). By the agricultural census of 2001 / 2002 the number of cattle had increased to 323167 animals (DOP-DOSA, 2002). Equine numbers were not recorded in this survey.

Within The Gambia there have been a number of studies investigating the prevalence of animal trypanosomiasis utilising traditional microscopy techniques. Prevalence in cattle and small ruminants has been shown to vary considerably by breed and site; Leperre and Claxton (1994) observed significant differences between zebu and N'Dama (6.2% and 1.9% respectively) while Snow et al (1996) reported prevalences in small ruminants, horses, donkeys and cattle ranging from 0% to ~10% when comparing seven sites within the country. These results are mirrored in a number of equine studies, reporting prevalence ranges typically in the 0-10% range (Mattioli *et al.*, 1994; Snow *et al.*, 1996) although prevalences as high as 45% have been reported (Faye *et al.*, 2001). The two most recent studies (Dhollander *et al.*, 2006; Pinchbeck *et al.*, 2008), while

observing higher incidences of animal trypanosomiasis have relied upon samples collected at veterinary clinics and cannot therefore be regarded as true estimates of prevalence. While the majority of studies have relied upon microscopy, with or without the use of a buffy coat to concentrate the parasites, only a few Gambian studies have used PCR based techniques (Pereira de Almeida *et al.*, 1998; Faye *et al.*, 2001; Pinchbeck *et al.*, 2008). Faye *et al.*, using species specific primer sets for *T. brucei*, *T. congolense* and *T. vivax* reported seven times more cases detected by PCR compared to traditional buffy coat microscopy. In addition to the increased sensitivity, PCR based diagnosis facilitates easy identification of the individual species present, as opposed to microscopic diagnosis which is reliant upon the ability to distinguish the three species by morphological differences.

The work by Pinchbeck *et al.*(2008), in collaboration with the GHDT reported a total trypanosome prevalence of 91% in horses and donkeys based on species specific PCR, as opposed to 18% by microscopy. The GHDT, based in Sambel Kunda in the Central River District of The Gambia (Figure 3.1), is a small non-governmental organisation with a focus on animal welfare training of local residents and the provision of treatment clinics at local markets. This focus on clinics may explain the unusually high prevalence as samples were obtained from animals brought to the clinics by owners, introducing a bias towards infected animals. These two PCR based studies indicated *T. congolense* and *T. vivax* to be the most prevalent species with only a small proportion of cases attributed to *T. brucei.* While Faye *et al.* reported *T. congolense* to be the most prevalent species the study of Pinchbeck *et al.* identified *T. vivax* present in 210 of the 241 samples examined. Due to the non-random sampling employed in the latter study it is not possible to determine whether this is due to regional variation in the prevalence of the respective species or an increased likelihood of visible illness in animals infected with *T. vivax* leading to an increased representation at treatment clinics.

While a number of studies have examined the prevalence of *T. vivax* in The Gambia and the wider contexts of Africa and South America (as described in Chapter 1) the potential for genetic exchange in the species has not been extensively analysed as in *T. brucei*. This is in part due to the low parasitaemia observed in natural infections and the difficulty in adapting the species to growth in culture. Genetic exchange in *T. brucei*, while experimentally demonstrated has proven to be a controversial subject with regards its role in field populations despite the potential for mating to have a major role on the dynamics of populations and evolution of the parasite. For *T. vivax* only a handful of studies have attempted to examine the subject of genetic exchange, however these studies were limited by the number of available isoenzyme markers or small sample sizes (Kilgour *et al.*, 1975; Kilgour and Godfrey, 1977; Allsopp and Newton, 1985; Tibayrenc *et al.*, 1991). Analysis of these data provided evidence for clonality in *T. vivax* however there has been no further

study of genetic exchange, despite the availability of new types markers or our increased understanding of genetic exchange in *T. brucei*.

Figure 3.1 Geography of The Gambia

Country profile for The Gambia showing major population centres.

★Approximate location of Sambel Kunda, base of the GHDT and focal point for local clinics at which samples were collected. Public domain map sourced from Wikimedia commons (http://commons.wikimedia.org/), October 2010.

The sample set employed by Pinchbeck *et al.*(2008), with a high prevalence of *T. vivax* and bolstered by a further 290 sympatric samples therefore provided an ideal opportunity to once again investigate the role of mating in this neglected species. Further to this, a parallel study of *T. congolense*, utilising the same sample set employed here, demonstrated evidence for sub-structuring and frequent genetic exchange, the first to do so for this species (Morrison *et al.*, 2009b). Mating appears therefore to be a feature of at least two of the species comprising the Salivarian trypanosomes. Examining the potential for genetic exchange in *T. vivax* will be essential in determining the role and extent of mating not only within the species but the genus as a whole. While previous population studies in *T. vivax* used isoenzyme markers the work of the WTSI in making available the draft genome sequence of the species made the design, evaluation and application of *T. vivax* specific microsatellite markers viable. Microsatellites, as discussed in Chapter 1 provide a number of benefits over isoenzymes. The amplification step of PCR allows for a far higher sensitivity, the markers exhibit a far higher level of polymorphism and are typically neutral with regards selective pressure as they most often located within non-coding regions of the genome. Microsatellite markers have also been widely employed in the study of genetic exchange in *T. brucei* and more recently in *T. congolense*, examining the genetic diversity of the species and frequency of mating (Biteau *et al.*, 2000; MacLeod *et al.*, 2001a; Jamonneau *et al.*, 2002; Morrison *et al.*, 2008b, 2009b)

The results of this chapter describe an analysis of the prevalence and population genetics of *T. vivax* in The Gambia with the aim of examining the genetic diversity present and detecting evidence suggesting the potential for mating in this species. The microsatellite markers described here were developed as part of a sample study submitted as part of an MRes qualification in 2007. Primer sets were designed around microsatellites present within the Y486 *T. vivax* genome reference sequence and screen for polymorphisms using the Kenyan ILRAD v34 isolate and a subset of the Gambian 2007 samples identified as *T. vivax* positive. Identification of species present in the samples was performed by A. Hamilton and Dr L. Morrison via species specific PCR prior to the analysis described here.

3.2 Results

3.2.1 Identification of meiosis-associated genes

In order for genetic exchange to occur in *T. vivax* the species must possess the required machinery, an absence of which would imply total clonality. In contrast the presence of these genes would imply the potential for mating but would not in and of themselves be sufficient to demonstrate that mating occurs. As the exact nature of trypanosome mating is not understood a group of eight meiotic genes that have been shown to be conserved amongst across many eukaryote lineages (Schurko and Logsdon, 2008) were selected for analysis. The eight genes chosen for this analysis are *spo11, dmc1, mnd1, msh4, msh5, hop1, hop2* and *rec8* have all been demonstrated to have meiotic specific functions, of which only *spo11* appears to be completely essential across all eukaryotes.

With the aid of this meiotic toolkit Morrison *et al.*(2009b) detected the presence of all eight genes within the genomes of *T. brucei* and *T. congolense* (Table 3.1). While the expression, function and meiotic role of these genes have yet to be fully characterised in *T. brucei*, the presence of mating in this species implies that, if essential in trypanosomes they are functional. The evidence for mating in *T. congolense* would likewise suggest retention of function in this species despite the high levels of divergence between *T. brucei* and *T. congolense* (Table 3.2). Utilising these orthologous sequences from *T. brucei* and *T. congolense* it has been possible to identify these eight genes within the draft genome of *T. vivax* generated by the WTSI (Table 3.1) and generate amino acid alignments for each (Table 3.2 and Appendix 5). For *mnd1* and *msh5* in *T. vivax* the annotated genes lack start codons, although it is possible that these genes use start codons 400 and 232 bp upstream, respectively. For the purpose of the alignment only the annotated sequences of both genes were included.

Alignment of the genes for each species (Appendix 5) indicates that while the sequences have diverged there are regions where a high proportion of the codons have been conserved. Insertion / deletion events can be observed throughout the alignments, with the largest typically clustering towards the start and end of sequences. The level of identity ranges from 40% to 88% (*T. congolense* / *T. vivax* pairwise identity for *Hop2* and *T. brucei* / *T. vivax* pairwise identity for *dmc1* respectively), with the majority falling within a window of 45 – 60% (Table 3.2). Despite this divergence recognisable superfamily motifs are conserved within the sequences of all three species supporting a common ancestry. While the high level of divergence of *T. vivax* from *T. brucei* and *T. congolense* could suggest a loss of function similar values of divergence are observed between *T. brucei* and *T. congolense*, two species in which genetic exchange is either known or believed to

occur. It is possible therefore that the potential for genetic exchange has likewise been retained in *T. vivax*. As for *T. brucei* and *T. congolense* further study will be required in order to determine the function of each of these genes and whether they possess a meiotic role in trypanosomes. Recent work by Peacock *et al.* (2011) has demonstrated the expression of *dmc1*, *hop1* and *mnd1* by a subset of trypanosomes in the salivary glands of a tsetse following infection with *T. brucei*. The genes were expressed in an order consistent with the meiotic cycle of other eukaryotes, indicating a continued meiotic role for these genes in *T. brucei*.

Gene	*T. brucei* accession number	*T. vivax* accession number	*T. congolense* accession number
spo11	Tb927.5.3760	TvY486_0503200	TcIL3000.0.05590
dmc1	Tb09.211.1210	TvY486_0904120	TcIL3000.11.8740
mnd1	Tb11.02.3380	TvY486_1105890	TcIL3000.0.24610 TcIL3000.11.6040
msh4	Tb927.10.1270	TvY486_1001250	TcIL3000.10.1060
msh5	Tb927.3.4280	TvY486_0303550	congo822g05.q1k_1
hop1	Tb927.10.5490	TvY486_1005520	TcIL3000.10.4620 congo1147c06.q1k_0
hop2	Tb927.2.5190	TvY486_0007970	TcIL3000.2.1290
rec8	Tb927.7.6900	TvY486_0706620	TcIL3000.7.5560

Table 3.1 Meiotic gene accession numbers

Accession numbers for the eight conserved meiotic genes putatively identified in *T. brucei*, *T. congolense* and *T. vivax*. For *mnd1* two *T. congolense* sequences, TcIL3000.0.24610 and TcIL3000.11.6040, with sequence identity of 98% to one another are present within the genome sequence. *Hop1* in *T. congolense* has been only partially assembled within the genome assembly, with the identifiers for each half listed here.

	spo11	*dmc1*	*mnd1*	*msh4*	*msh5*	*hop1*	*hop2*	*rec8*
T. brucei – T. vivax	46	88	44	48	51	54	50	44
T. brucei – T. congolense	51	49	51 / 51[a]	55	63	56[b]	52	47
T. vivax – T. congolense	47	49	44 / 43[a]	46	52	51[b]	40	48

Table 3.2 Pairwise species identity of meiotic genes

Pairwise sequence identity (%) for the eight core meiotic genes identified by Schurko and Logsdon (2008) following alignment of *T. brucei*, *T. vivax* and *T. congolense* amino acid sequences. [a] Two sequences, TcIL3000.0.24610 and TcIL3000.11.6040, with identity of 98% to one another were identified within the *T. congolense* genome assembly, pairwise identities are therefore to each of the respective sequences. [b] Identity to the combined sequences of TcIL3000.10.4620 and congo1147c06.q1k_0.

3.2.2 Prevalence of *T. vivax*, *T. congolense* and *T. brucei* in The Gambia

In order to assess the prevalence of *T. vivax* in The Gambia a total of 531 equine and bovine samples were collected between March 2006 and January 2007. Equine samples from horses and donkeys brought to clinics run by the GHDT were collected as blood samples on FTA cards. Clinics were operated within the local region surrounding the base of the GHDT in Sambel Kunda in the Central River District (Figure 3.1) although animals brought to clinic often came from further afield. A total of 323 equine samples were collected at three time points, March 2006, August 2006 and January 2007 with prevalence based on species specific PCR (Masiga *et al.*, 1992) (Table 3.3). The total trypanosome prevalence and species prevalence values of the March and August samples, determined by microscopy and species specific PCR respectively have been previously described by Pinchbeck *et al.*(2008), with the slightly different values reported here representing corrections to the published data. Of the *T. congolense* subgroups only *T. congolense* Savannah was identified within the population; therefore for the remainder of this chapter *T. congolense* refers to *T. congolense* Savannah. *T. vivax* prevalence in equines was 85% in March 2006, 91% in August 2006 and 73% in January 2007. Mixed species infections of *T. vivax* with *T. brucei*, *T. congolense* or both were common, and occurred in 46% of *T. vivax*-infected animals.

In addition to the equine samples, sympatric bovine blood samples were collected on FTA cards through a random sampling of herds maintained by the International Trypanotolerance Centre (ITC), with the data summarised in Table 3.4. Of the 193 randomly sampled cattle, 21 (11%) were identified as positive for *T. vivax* of which only 5 were also found to be infected with *T. brucei* or *T. brucei* and *T. congolense*. No mixed infections of *T. vivax* and *T. congolense* were observed amongst these animals. The difference in bovine prevalence compared with that observed in the equine sample set is likely due to the random sampling employed and thus may be more indicative of the true prevalence of *T. vivax* in the region. However other factors such as local variation or differing host susceptibilities cannot be ruled out. An additional set of 15 cattle samples collected in 2006, which were identified as being positive by microscopy prior to collection of the FTA blood spots were also included in the later genetic analysis of *T. vivax*.

Species present	March 2006 (n = 154)		August 2006 (n = 87)		January 2007 (n = 82)		Horse (n = 251)		Donkey (n = 72)		Overall prevalence (n = 323)	
	N	P%	N	P%	N	P%	N	P%	N	P%	N	P%
T. vivax only	71	46	40	46	34	41	125	50	20	28	145	45
T. vivax + *T. congolense* mixed infection	30	19	19	22	21	26	52	20	18	25	70	22
T. vivax + *T. brucei* mixed infection	23	15	10	11	3	4	25	10	11	15	36	11
T. vivax + *T. congolense* + *T. brucei* mixed infection	7	5	10	11	2	2	14	6	5	7	19	6
Total *T. vivax*	131	85	79	91	60	73	216	86	54	75	270	84

n, number of samples; N, number of trypanosome positive samples; P%, percentage of positive samples

Table 3.3 *T. vivax* prevalence in equines

Prevalence in equines of *Trypanosoma vivax* infections as determined by species specific PCR. All samples were collection from animals brought to clinics run by the Gambian Horse and Donkey Trust, The Gambia.

Species present	August 2006 (n = 15)		January 2007 (n = 193)	
	N	P%	N	P%
T. vivax only	7	Na	16	8
T. vivax + *T. congolense* mixed infection	1	Na	0	0
T. vivax + *T. brucei* mixed infection	5	Na	4	2
T. vivax + *T. congolense* + *T. brucei* mixed infection	0	Na	1	7
Total *T. vivax*	13	Na	21	11

n, number of samples; N, number of trypanosome-positive samples; P%, percentage of positive samples; Na, not applicable, as only microscopically positive samples were examined.

Table 3.4 *T. vivax* prevalence in bovines

Prevalence in bovines of *Trypanosoma vivax* infections as determined by species specific PCR. Samples were collected by random sampling of herds owned by the ITC, The Gambia.

3.2.4 Genetic analysis of *T. vivax* in The Gambia

The primary topic of this thesis is the investigation of genetic diversity in the African trypanosomes focusing here upon *T. vivax* and addressing that of *T. congolense* and *T. brucei* in subsequent chapters. In order to achieve this genotyping of each of the 304 *T. vivax* positive samples was attempted using a panel of eight single-locus microsatellite markers developed as part of an earlier study. These eight markers, TV3, TV4, TV6, TV14, TV17, TV24, TV31 and TV49 were designed around di- and trinucleotide motif microsatellites (Appendix 2) with each microsatellite located on a separate contig from the WTSI sequencing project. While genotyping was attempted for each of the *T. vivax* positive samples the majority of samples failed to amplify for any single-locus markers (Figure 3.2), thus full MLGs (positive allele identification for all eight markers) were successfully obtained for only 31 samples (Appendix 6), 10% of the total *T. vivax* infections identified. This low amplification rate with the single copy microsatellite markers is suggestive of a population where the majority of infections are present at parasitaemias below the threshold for detection by single-locus PCR but within that of the more sensitive multicopy species specific marker.

While two alleles were identified for both TV3 and TV4 virtually all of the 31 fully genotyped samples were homozygous for the same allele, which was present at a frequency above 0.8, rendering them uninformative in the genetic analysis. These markers were therefore discarded from the analysis. The remaining six markers had between two and four alleles per locus (Figure 3.3). Of the 31 samples for which full MLGs were obtained no mixed *T. vivax* genotype infections were identified, as defined by the presence of three or more alleles for any single marker. The lower sensitivity of the single locus markers and small number of fully genotyped samples however makes the identification of mixed infections unlikely if they are present. The large differences in allele sizes, up to 63 bp, and lack of intermediate alleles in the population suggest the microsatellites may be following a non-stepwise mutation model; that the sample size is too small to detect a full range of allele sizes; or that intermediate alleles have been lost from the population.

From the 31 fully genotyped samples nine unique MLGs were identified, four of which were present in multiple samples (Figure 3.4 and Appendix 6). MLG 8 was detected 15 times, constituting almost half of the genotyped samples. The control sample, *T. vivax* ILRAD V34 (MLG 1), originating from Kenya, had unique alleles at three of the six loci (Appendix 6). In order to visualise the relationships between the MLGs a UPGMA dendrogram of similarity was constructed utilising Jaccard's coefficient of similarity to group the isolates based upon their identity to one another (Figure 3.4). While high bootstrap support was identified for node A the separation of these samples did not correlate with available sampling data.

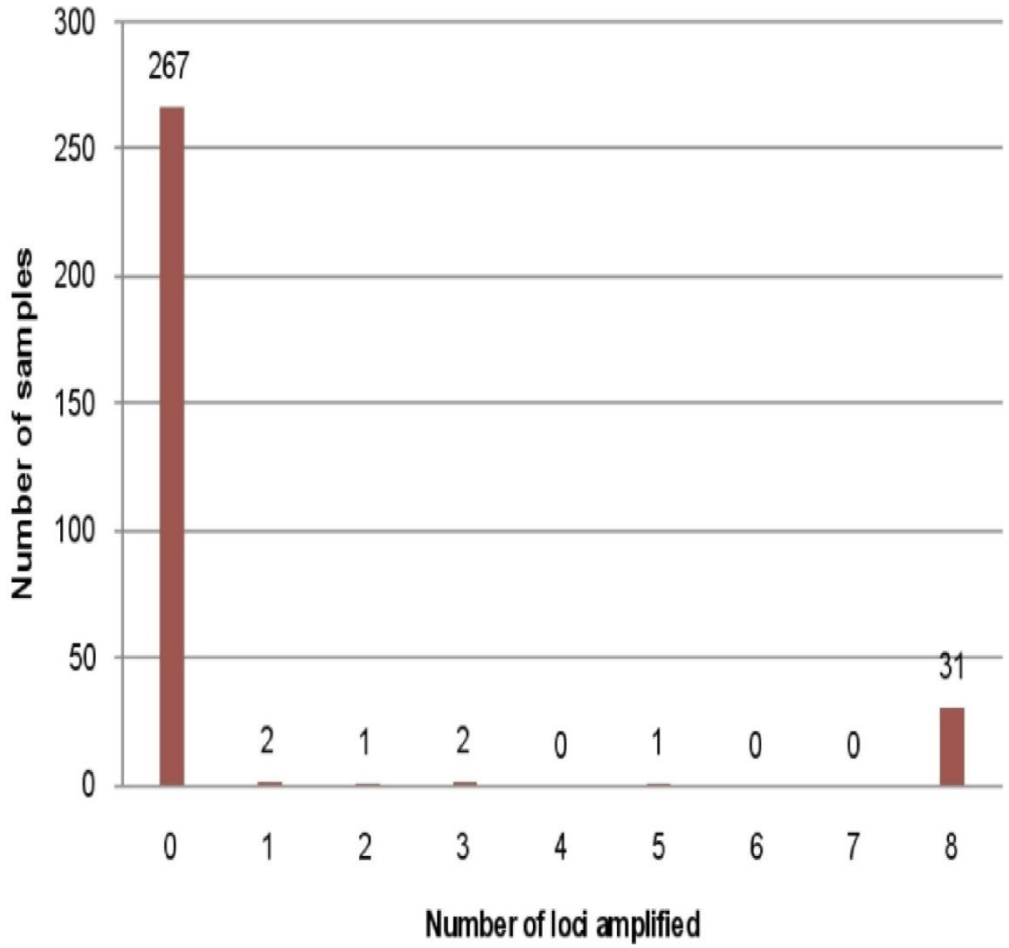


Figure 3.2 Genotyping success rate

Number of loci amplified from each *T. vivax* positive sample. Despite the use of nested PCR protocols the majority of samples identified as *T. vivax* positive failed to amplify for any of the eight microsatellite markers used in the study, which may be suggestive of a low parasitaemia population below the sensitivity of single copy markers.

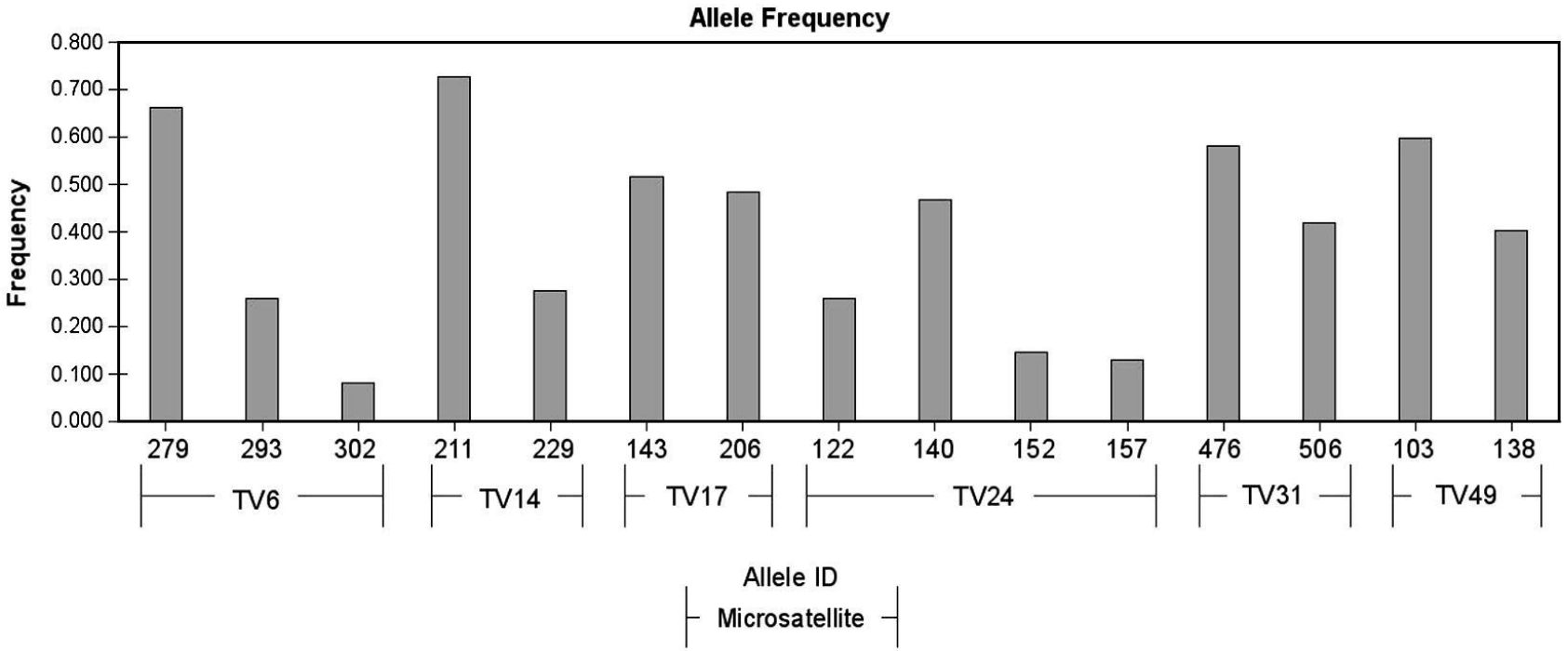


Figure 3.3 Allele frequency distribution

Distribution of allele frequencies for the 31 fully genotyped samples collected from horses, donkeys and cattle in The Gambia using six *Trypanosoma vivax* specific microsatellite markers. Alleles are group by locus and listed as their bp size.

To further our understanding of the *T. vivax* population structure, allele frequencies were examined for deviation from Hardy–Weinberg equilibrium and for the presence of linkage disequilibrium between the six loci. These tests indicate whether the observed allele combinations between loci or the genotype frequencies resemble those expected in a randomly mating population. Significant disagreements ($P < 0.05$) with Hardy–Weinberg equilibrium predictions are present for five of the six loci (TV6, TV24, TV17, TV31 and TV49) (Table 3.5). This disagreement with Hardy–Weinberg predictions is due to the global heterozygote excess observed, with TV24 and TV17 at, or close to, heterozygote fixation. No homozygotes were identified for marker TV24 despite the presence of four alleles at this locus. The population wide heterozygote excess is reflected in the F_{IS} values, ranging from -0.38 to -0.94 across the six markers (Table 3.5). Negative F_{IS} values such as those observed here are therefore the result of higher than expected heterozygosities across each of the loci. While a randomly mating population would be expected to have values close to zero, these large negative values are indicative of partial or total clonality within the population.

To test for evidence of genome re-assortment and recombination, linkage disequilibrium between alleles at all pairwise combinations of loci were examined. After accounting for significant deviation from Hardy-Weinberg equilibrium significant levels of linkage disequilibrium ($P < 0.02$ and below) (Table 3.6) were observed between all pairwise combinations, in line with expectations for a population resulting from clonal expansion. Correction for deviation from Hardy-Weinberg equilibrium, achieved by preserving the correlation between alleles at different loci is required to prevent the artificial introduction of significant linkage. As a final test to examine the role of mating in the population, samples were analysed with MLGsim, which simulates randomly mating populations from known allele frequencies in order to predict MLG frequencies. It is then possible to determine whether observed MLG frequencies are higher than expected, indicating a lack of random mating. MLGsim identified three of the four repeated MLGs (MLGs 4, 8 and 9) as having frequencies significantly ($P < 0.01$) higher than expected in a randomly mating population. MLG 10, present twice in the sample set, did not differ significantly ($P > 0.05$) in frequency from that expected in a randomly mating population.

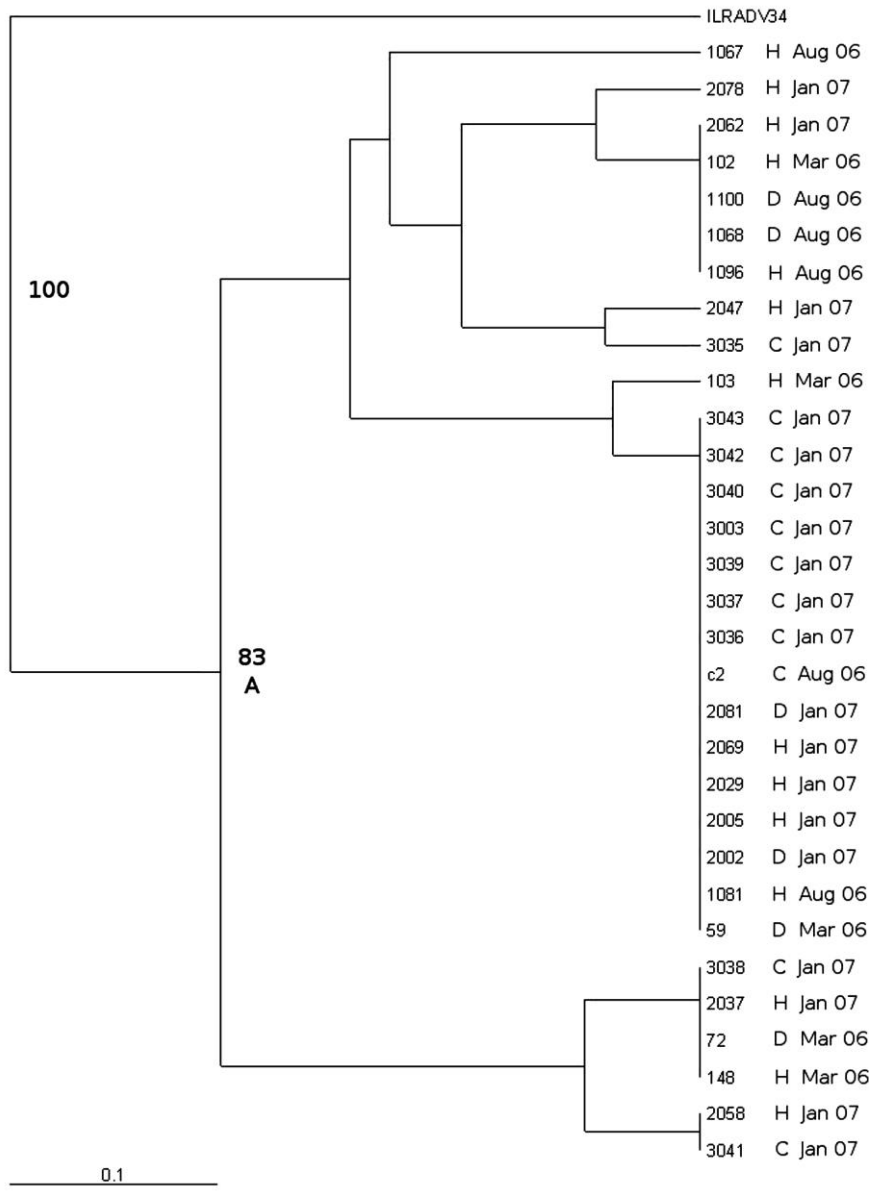


Figure 3.4 UPGMA dendrogram of similarity for the genotyped *T. vivax* samples

UPGMA dendrogram of similarity for *Trypanosoma vivax* multilocus genotypes (MLGs) for the 31 fully genotyped samples collected from horses, donkeys and cattle in The Gambia. The dendrogram was generated by Treeview from Clustering Calculator. Bootstrap values generated from 100 reiterations are displayed for the major nodes. Scale bar represents dissimilarity.

Locus	Na	Ho	He	F_{IS}	P
TV6	3	0.68	0.49	-0.38	0.04
TV14	2	0.55	0.40	-0.39	0.07
TV17	2	0.97	0.50	-0.94	<0.001
TV24	4	1.00	0.68	-0.48	<0.001
TV31	2	0.77	0.49	-0.59	<0.001
TV49	2	0.81	0.48	-0.68	<0.001

Na, number of alleles; Ho, observed heterozygosity; He, expected heterozygosity; F_{IS}, inbreeding coefficient; P, probability of agreement with Hardy-Weinberg equilibrium (P > 0.05 = agreement with Hardy-Weinberg equilibrium).

Table 3.5 *T. vivax* population statistics

Polymorphism, heterozygosity and agreement with Hardy–Weinberg equilibrium based on allele frequencies observed in 31 *Trypanosoma vivax* positive samples collected from horses, donkeys and cattle in The Gambia following genotyping at six *T. vivax* specific microsatellite markers.

Locus 1	Locus 2	P
TV6	TV14	0.02
TV6	TV17	<0.001
TV6	TV24	0.02
TV6	TV31	0.01
TV6	TV49	0.01
TV14	TV17	<0.001
TV14	TV24	<0.001
TV14	TV31	<0.001
TV14	TV49	<0.001
TV17	TV24	<0.001
TV17	TV31	<0.001
TV17	TV49	<0.001
TV24	TV31	<0.001
TV24	TV49	<0.001
TV31	TV49	<0.001

P, probability of linkage equilibrium between loci (*P*> 0.05 = no significant linkage between loci).

Table 3.6 Pairwise linkage disequilibrium

Linkage disequilibrium between all pairwise combinations of six *Trypanosoma vivax* specific microsatellite markers in 31 *T. vivax* positive samples collected from horses, donkeys and cattle in The Gambia. Significant linkage was observed in all pairwise combinations following correction for deviation from Hardy-Weinberg equilibrium.

3.3 Discussion

The primary aim of this study was to examine the prevalence and potential for genetic exchange in the African trypanosome *T. vivax* through analysis of genomic and field data. As discussed in Chapter 1, despite the extensive distribution and considerable impact of Nagana caused in part by *T. vivax* the study of this species remains neglected, partially due to the difficulties in adapting isolates for laboratory use (Gardiner, 1989). These data, collected originally by Pinchbeck *et al.*(2008), and expanded upon here indicates a high prevalence of *T. vivax* in equines that have been brought to clinic for treatment. That diagnosis of trypanosomiasis required PCR amplification techniques for the majority of animals suggests that while low parasitaemia infections are frequent they continue to place a burden upon animal welfare. Due to the sampling of equines brought to clinics it is impossible to determine the incidence rates and true prevalence of trypanosomiasis amongst equines in The Gambia from these data. The random sampling of 193 cattle from herds maintained by the ITC however indicated a far lower prevalence of trypanosomiasis, particularly of *T. vivax*. The different sampling methods employed in collecting these samples prevent a direct comparison of prevalence between equine and bovine hosts. It is therefore not possible to assess the relative susceptibility of equines and bovines to infection, which will require further studies employing random sampling.

Following examination of prevalence, evidence for genetic exchange in *T. vivax* was examined by identifying the core genes required for meiosis and examining the *T. vivax* population for signs of mating. *T. brucei* is the only member of the Salivarian trypanosomes for which mating has been unequivocally demonstrated and displays a classical Mendelian pattern of inheritance (Jenni *et al.*, 1986; MacLeod *et al.*, 2005a). More recently evidence for mating has emerged from a field study of *T. congolense*, employing the same sample set and techniques used here (Morrison *et al.*, 2009b), however mating has yet to be experimentally demonstrated in this species.

The first step in our analysis, identification of eight meiosis associated genes (El-Sayed *et al.*, 2005; Ramesh *et al.*, 2005; Schurko and Logsdon, 2008) indicated their presence within the *T. vivax* genome, with conservation of protein features. While the sequences were highly divergent between *T. brucei* and *T. vivax* similar levels of divergence were observed between the respective *T. brucei*/*T. congolense* alignments. If these eight genes have retained their function within *T. brucei* and *T. congolense* a similar retention of function, and therefore potential for genetic exchange, cannot be ruled out in *T. vivax*. The presence of these genes within the three lineages and the early divergence of the *T. vivax* line (Stevens *et al.*, 1999) indicates an ancestral nature to the meiotic machinery present and an early role for genetic exchange within the Salivarian trypanosomes. However, even in *T. brucei* our understanding of the expression, functionality and in particular the meiotic roles of these genes is limited, with only the recent work of Peacock *et*

al.(2011) providing insight into the matter. While the exact mechanism and requirements remain elusive, mating in *T. brucei* is known to occur within the salivary glands of the tsetse fly vector (Gibson and Whittington, 1993; Gibson and Bailey, 1994; Bingle *et al.*, 2001; Tait *et al.*, 2007; Gibson *et al*., 2008). However in *T. vivax* cyclical development in the tsetse is simpler in nature, being restricted to the proboscis of the fly, possibly accounting for the ease with which this species can be mechanically transmitted. It could therefore be argued that *T. vivax* is incapable of meiosis due to the lack of development within the salivary glands. The evidence for genetic exchange in *T. congolense* which, like *T. vivax,* does not colonise the salivary glands, coupled to the ancestral origin of the meiotic genes instead suggests that salivary gland meiosis may be a *T. brucei* specific adaptation of a pre-existing life cycle stage.

Evidence for genetic exchange in *T. vivax* is severely limited due to a lack of previous field data. The studies of Kilgour *et al.*(1975) and Kilgour and Godfrey (1977) addressed the diversity of *T. vivax* using samples collected from a single geographic location and time point, which were later analysed by Tibayrenc *et al.*(1991). Sampling directly from Zebu cattle in Nigeria these studies identified 61 and 31 *T. vivax* positive animals and with the use of two isoenzyme markers demonstrated the presence of three main *T. vivax* isotypes, which were still present as the dominant isotypes within the population during the second study three years later. While these studies were limited by the use of only two isoenzymes to assay diversity, together with the nine isolates from Allsopp and Newton (1985) they provided the first evidence for clonality in the species based on the criteria outlined by Tibayrenc *et al.*(1991), namely an over-representation of a small number of genotypes, an absence of recombinant genotypes and significant levels of linkage disequilibrium between markers. In order to further explore the genetic diversity and potential for genetic exchange in *T. vivax* this study has likewise used samples collected from a localised geographic area and collected over the time period of one year. The analysis suggests a high prevalence of *T. vivax* in The Gambia, characterised by low parasitaemic infections and the clonal expansion of particular genotypes.

The key features of clonality (Tibayrenc *et al.*, 1990; Maynard Smith *et al.*, 1993; Halkett *et al.*, 2005; de Meeûs *et al.*, 2006) namely excess heterozygosity at all loci, significant disagreement with Hardy–Weinberg predictions, significant levels of linkage disequilibrium between all loci combinations, an absence of recombinant genotypes and a limited number of unique genotypes, are all present within the population examined. These results show clear similarities with the analysis by Tibayrenc *et al.*(1991), supporting the idea of clonality in *T. vivax* and suggesting that it may be the primary, if not only, mode of reproduction.

The main restriction on analysing the current population set has been the non-amplification of the single-locus microsatellite markers for most *T. vivax* positive samples. This is suggestive of a

population where the majority of infections are characterised by a low parasitaemia, with occasional increase to levels examinable by the PCR techniques employed here. Parasitaemia levels are known to vary considerably depending on parasite strain, species, breed of host, geographic origin and presence of other infections (Fasogbon *et al.*, 1990; Batista *et al.*, 2007; Magona *et al.*, 2008; Pinchbeck *et al.*, 2008) while identification of infections has conventionally been by microscopy following preparation of a buffy coat, estimated by Faye *et al.*(2001) to be seven times less sensitive than the PCR technique employed here. These results, of a high prevalence of low parasitaemia infections, reinforce these observations and suggest that *T. vivax* may be more endemic than previously believed. There remains the possibility that low parasitaemia infections might act as a reservoir of genetic diversity with observed outbreaks of disease arising from a limited number of virulent clones best adapted to the local environment. It may be, therefore, necessary to examine the parasites from animals with subpatent levels of *T. vivax* parasitaemia in order to uncover the true dynamics of the population structure. Given the already high sensitivity of the WGA and PCR techniques employed in the current study, methods to concentrate trypanosome DNA by extraction from larger volumes of infected blood will be required in order to characterise these currently elusive populations.

Although there are suggestions from our data that mating may be possible, the predominance of clonal expansion is a feature that *T. vivax* shares, among the Salivarian trypanosomes, with the agents of Human African Trypanosomiasis, *T. b. rhodesiense* (MacLeod *et al.*, 2000) and *T. b. gambiense* Group 1 (Koffi *et al.*, 2007, 2009; Morrison *et al.*, 2008a). While all of the available evidence supports clonality in *T. b. gambiense* Group 1, it is possible that mating does occur in *T. b. rhodesiense*, as previous studies into this sub-species have centred around foci of human infection, where virulent strains may dominate. Other examples of trypanosomes that are likely to propagate clonally include *T. evansi* and *T. equiperdum*, although in these cases the predicted clonality is due to their modes of transmission, mechanical and venereal, respectively. These trypanosome species are essentially variants of *T. brucei* (Gibson, 2007; Lai *et al.*, 2008), that have lost the ability to undergo cyclical transmission. They are therefore likely to expand solely by asexual replication with the eventual loss of the potential for mating. As described previously, endemic field populations of those trypanosomes not under similar constraints (*T. b. brucei* and *T. congolense*) have been shown to undergo frequent mating. The *T. vivax* population in the present study is sampled from 'typical' host species (i.e. there are no known specific genetic mutations in *T. vivax* that confer survival in cattle and equines), and are sampled from a region where the predominant vector is that in which cyclical development occurs. The strong evidence for linkage disequilibrium in the *T. vivax* samples is interpreted as reflecting a clonal population and contrasts sharply with the evidence for mating in sympatric *T. congolense* subpopulations, where there is far less deviation from linkage equilibrium (Morrison *et al.*, 2009b). The discovery of evidence for mating in *T. congolense* but not in *T. vivax*, from the same host samples suggests that if meiotic

reproduction is possible in *T. vivax* the conditions required differ significantly from those of *T. congolense*.

Despite the evidence from this study that genetic exchange may be absent in *T. vivax* the possibility remains that this is a feature of this localised population as mating in *T. brucei* is known to be a non-obligatory stage of the life cycle. Detailed analysis of further distinct foci in addition to laboratory studies will therefore be required in order to fully address the potential for mating in this species.

Chapter 4

Return to The Gambia:

Trypanosomiasis in equines, 2009

4.1 Introduction

Chapter 3 introduced the issue of trypanosomiasis in the West African country of The Gambia and investigated the population genetics of *T. vivax* in the equines and bovines in the country. The study suggested that mating was absent, or occurring at undetectable levels, in the local *T. vivax* population, with a single MLG accounting for approximately half of the genotyped population. The investigation however was hampered by the inability to genotype ~90% of the 304 *T. vivax* positive samples with the majority of the samples failing for each of the eight microsatellite markers employed in the study. This may have arisen due to the low parasitaemia associated with *T. vivax* infections, an issue which has been noted previously, alongside difficulties in generating laboratory isolates by growth in rodents or adaptation to culture (Gardiner, 1989).

Concurrent to the investigation of *T. vivax* a second study, utilising the same sample set, investigated the role and extent of mating in the *T. congolense*, the other major cause of animal trypanosomiasis (Morrison *et al.*, 2009b). This study identified 133 samples that were positive for the Savannah clade of this species and was able to successfully genotype 84 isolates with seven microsatellite markers developed specifically for the study. Genetic analysis of the population indicated a high level of genetic diversity with 80 unique MLGs present, an overall excess of homozygotes across the seven loci and evidence for sub-structuring within the population. While analysis of the population as a whole did not support a panmictic population with frequent, random out crossing between individuals these data are inconsistent with that of a clonal population structure. These data, with an excess of homozygotes across all seven loci and high number of unique MLGs, were most consistent with a mating population with a high frequency of inbreeding and selfing.

While these two studies shed new light on the role of genetic exchange in *T. congolense* and *T. vivax* they were limited by the use of FTA cards, the sampling of animals across three time points and use of animals brought to clinics. FTA cards, while providing a simple and effective method for the collection of blood samples in the field are limited by the volume of blood collected. Each sample spot collects only 200 µl of blood, limiting the number of trypanosomes collected from each host while the ability to detect and genotype infections is then further limited as only a small proportion of each spot is utilised for PCR amplification of DNA. It was hypothesised that this sampling limitation, coupled to the low parasitaemia associated with *T. vivax* infections may have contributed to the inability to genotype the majority of *T. vivax* positive samples. The second limitation of the original two studies was the use of samples collected over three time points, with the first (March 2006) and last (January 2007) separated by the rainy season during which the tsetse prevalence is highest. While genetic analysis did not detect significant differences in the isolates from each time point it remains possible that the grouping of these three time points distorted the

analysis of *T. congolense* and *T. vivax*. The final limitation of the initial studies, sampling of equines brought to clinics run by the GHDT, may have introduced bias by limiting the sampling of animals to those identified as 'ill' by the owners. The genotypes identified in the equine samples may therefore be associated with only pathogenic isolates and may not be representative of the population as a whole.

At the start of 2009 a second follow up study was therefore designed in collaboration with researchers from the University of Liverpool and the GHDT in order to address the FTA card sampling and sampling period limitations. This study consisted of two months of fieldwork assisting the staff of the GHDT in the diagnosis of trypanosomiasis at field clinics run by the GHDT in the area surrounding their base at Sambel Kunda (Figure 3.1). During this time a new population of 198 samples was acquired from the equines brought to the clinics for treatment. Where the first study had employed FTA cards for the collection of blood samples, this second study collected 2 ml of whole blood from each animal. Whole blood was collected in the hope that the collection of a greater volume and concentration of trypanosome material during DNA purification would allow for the genotyping of even low parasitaemia *T. vivax* infections. Further, the collection of a large number of samples over a much smaller time period limits the potential for the introduction of Wahlund effects which may have unintentionally distorted the results of the first study.

The 2009 study had three primary aims. To conduct a survey of AAT in The Gambia in 2009, to reassess the high frequency of *T. vivax* infections observed in 2009 using techniques capable of sampling from a greater volume of blood and to examine the population structures of *T. congolense* and *T. vivax* at a new time point, allowing for comparison with the studies of these species in 2007 (Morrison *et al.*, 2009b, and Chapter 3).

4.2 Results

4.2.1 Analysis of field diagnosis data

During the two month period of fieldwork blood samples were collected from a total of 198 animals (78 horses, 39 donkeys, one mule and 80 for whom species data were not recorded) during examination at clinics run by the GHDT (Appendix 7).Field diagnosis of AAT by the GHDT consists of two components, microscopic examination of wet blood smears and determination of packed cell volume as a measure of anaemia. Anaemia, resulting in a reduced PCV, is one of the primary symptoms of trypanosome infections and as such is utilised by the GHDT as a diagnostic tool for trypanosomiasis infection with animals found to have a PCV of 20% or lower considered to be infected with trypanosomes. As dehydration can have a significant effect upon the PCV total blood protein (TP) levels was recorded in order to detect dehydrated animals where the PCV may have been artificially elevated above 20%.

Of the 198 animals from which blood was collected 28 were identified as trypanosome positive by microscopy during screening at the clinics, a prevalence of 14%. There was a significant association between microscopic infection status and packed cell volume (PCV) (2 tailed unpaired t-test, $p = <0.001$), supporting the use of PCV as a diagnostic tool. As lipid autotrophs trypanosomes rely upon the lipoproteins present within the host bloodstream (Green *et al.*, 2003) and it has previously been demonstrated that infection in bovines may induce a drop in blood cholesterol levels (Traoré-Leroux *et al.*, 1987), however no association was observed between microscopic infection status and total cholesterol level in the equines examined here ($p = 0.48$). In addition no significant association was observed between microscopic infection status and total serum protein (TP) ($p = 0.90$), a measure of hydration recorded in order to assess the reliability of PCV measurements.

4.2.2 PCR identification of infection

In order to determine the prevalence of *T. brucei*, *T. congolense* and *T. vivax* in the animals presented to the GHDT clinics 2ml of whole blood was collected in order to allow for DNA extraction, in comparison to the 2006 – 2007 studies where blood was collected on FTA cards. The use of a larger volume and concentration of trypanosome material during DNA purification was employed in order to increase the ability to identify and genotype even low parasitaemia infection. DNA was extracted from each blood sample and examined for infection through the use of species specific primer sets (Masiga *et al.*, 1992) to test for the presence of respective species. Of the 198 animals from which samples were collected 183 (92%) were found to be positive for at least one of

the three species. This exceedingly high prevalence is predominantly due to infections with *T. vivax* which were observed in 173 (87%) of the animals. *T. brucei* was observed in 84 (42%) of samples while *T. congolense* Savannah was found to be present in 70 (35%) of samples. The *T. congolense* Forest subgroup was not detected within the sample population; therefore, for the remainder of this chapter *T. congolense* refers to *T. congolense* Savannah subgroup.

Of the 28 samples identified as trypanosome positive by microscopy all bar two were also identified as positive for at least one species by PCR. The two isolates negative for all three species (X010 and ECG004) were sampled from horses, at Brikamabah and Kerrtamim respectively. *T. congolense* and *T. vivax* were observed at high frequency (23/28 and 25/28 animals respectively) while *T. brucei* was observed in 12 of the 28 microscopically positive animals. While the prevalence of *T. brucei* and *T. vivax* in the microscopically positive samples is close to that observed when examining all samples (Table 4.1) the prevalence of *T. congolense* is over double that observed in the full sample set (82% for microscopically positive; 35% for all samples). This may indicate that *T. congolense* infections are associated with high parasitaemia infections, increasing the likelihood of diagnosis by microscopy.

Mixed species infections, with at least two of the three species present were frequent and observed in 121 samples, of which 23 were positive for all three species investigated (Table 4.1 and Figure 4.1). In order to examine if there were interactions between the three trypanosome species the frequency of observed and expected mixed infections were compared based on the total observed prevalence of each individual species. Pairwise χ^2 analysis between the three species indicated no significant deviation from the expected frequencies of single / mixed infections by chance alone (*T. brucei* – *T. congolense*, $p = 0.42$; *T. brucei* – *T. vivax*, $p = 0.12$; *T. congolense* – *T. vivax*, $p = 0.41$).However log-linear analysis of the three species together indicated significant deviation from the expected values ($p = 0.05$) due to the higher than expected frequency of samples negative for all three species (15 observed, 9.3 expected). These 15 samples were collected over the duration of the study from eight different sampling locations, making temporal or geographical factors an unlikely cause for this phenomenon.

Species present	Donkeys (n = 39)		Horses (n = 78)		Species unknown (n = 80)		Total prevalence (n = 198)*	
	N	P %	N	P %	N	P %	N	P %
T. brucei only	1	3	1	1	1	1	3	2
T. congolense only	1	3	0	0	2	3	3	2
T. vivax only	9	23	23	29	24	30	56	28
T. brucei and *T. congolense* mixed infections	1	3	2	3	1	1	4	2
T. brucei and *T. vivax* mixed infections	15	38	25	32	14	18	54	27
T. congolense and *T. vivax* mixed infections	6	15	12	15	22	28	40	20
T. brucei + *T. congolense* + *T. vivax* mixed infections	5	13	9	12	8	10	23*	12
Total *T. brucei*	22	56	37	47	24	30	84*	42
Total *T. congolense*	13	33	23	29	33	41	70*	35
Total *T. vivax*	35	90	69	88	68	85	173*	87
Total (all species)	38	97	72	92	72	90	183*	92

Table 4.1 Trypanosome prevalence as determined by species specific PCR

n, number of samples; N, number of trypanosome positive samples; P %, percentage of positive samples; *, includes the single mule sample, which was positive for all three species. There were no significant differences in the number of mixed species infections when compared to the expected number based on the observed prevalence of each individual species.

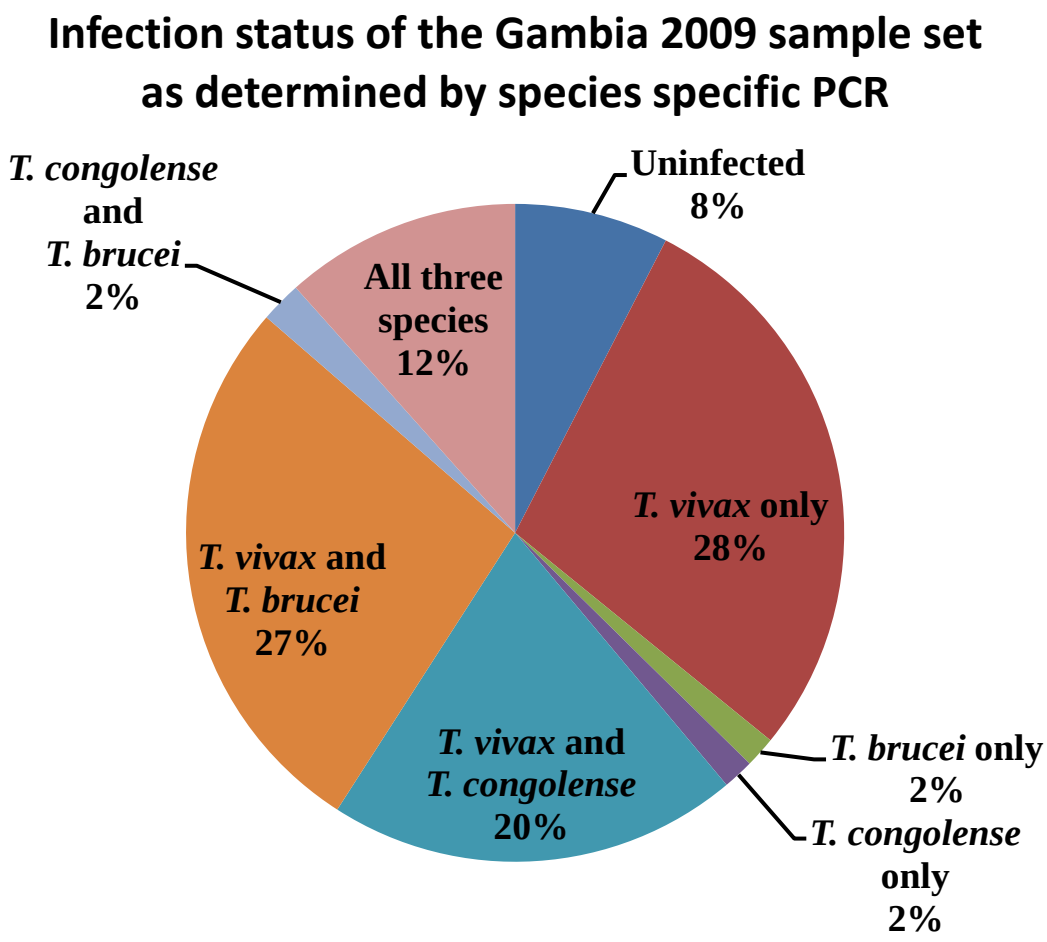


Figure 4.1 Infection status of The Gambia 2009 sample set as determined by species specific PCR

Infection status of the 198 samples following species specific PCR, highlighting the low frequency of uninfected samples (15 animals, 8%) and high frequency of *T. vivax* (173 animals, 87%).

While infection as determined by microscopy was associated with a reduced PCV the increased sensitivity of the PCR technique identified a substantially higher prevalence of AAT (microscopy 14%; PCR 92%) and therefore the associations between infection status with PCV, TP and blood cholesterol were re-examined using the higher PCR determined prevalences. There remained a significant association between PCV and PCR infection status (positive / negative for any species) ($p = <0.001$) indicating a reduced PCV was associated with infection in general and not just high parasitaemia infections detectable by microscopy. As with microscopic infection status there was no significant association between PCR infection status and TP ($p = 0.28$) or blood cholesterol ($p = 0.70$). Due to the relatively small number of control samples negative for all three species ($n = 15$) these results must be treated with caution.

Tests to assess whether PCV, TP or cholesterol were significantly influenced by the presence of the individual species indicated a significant reduction in PCV associated with infection by *T. congolense* or *T. vivax* (Table 4.2) and a significant reduction in cholesterol associated with the presence of *T. congolense*. These tests however do not account for the effect of mixed infections. In their earlier study Pinchbeck *et al.*(2008) found *T. congolense* to have the greatest individual effect on PCV, however, for the sample collection presented here 63 of the 70 samples infected with *T. congolense* were observed to also be infected with *T. vivax*. It may be therefore that the reduction in PCV observed for *T. vivax* is due to, or at least enhanced by the overlap with *T. congolense*. The small number of samples positive for only *T. brucei* or *T. congolense* ($n = 3$ in both cases) prevents independent tests free of the influence of mixed infections. The higher frequency of *T. vivax* only infections ($n = 56$) allowed for testing of the association between infection with this species and PCV. However, once again the relatively small number of completely uninfected samples means the results must be treated with caution. With the exclusion of mixed infections, there was no significant change in PCV, TP or cholesterol level associated with the presence of *T. vivax*.

Anaemia, measured as a reduced PCV, is a major symptom of AAT and its control is one of the primary components of trypanotolerance in bovines (Reviewed in Naessens, 2006; Stijlemans *et al.*, 2010). In order to assess the contribution of individual factors affecting PCV a multiple variable linear regression model was constructed. The first model assessed only three variables, whether the animals was infected by *T. brucei*, *T. congolense*, or *T. vivax*. Only infection by *T. congolense* was found to have a significant influence upon the variance of PCV (Table 4.3A). The adjusted R^2 value of the linear regression model (6.4%) indicates however that only a small proportion of the observed variation in PCV could be explained by infection with these three trypanosomes and that other factors control the majority of the variation observed.

Infection status	Association with		
	PCV (%)	TP	Cholesterol
Microscopically positive	**0.00**	0.90	0.48
PCR positive	**0.00**	0.28	0.70
T. brucei positive	0.52	0.37	0.56
T. congolense positive	**0.00**	0.31	**0.02**
T. vivax positive	**0.02**	0.71	0.72
T. vivax positive (no mixed infections)	0.29	0.38	0.78

Table 4.2 Probability of association between infection parameters

Probability of association between infection status and PCV, TP or blood cholesterol levels. The association between infection status and PCV, TP or cholesterol level was assessed through 2 tailed unpaired t-tests. The number of samples positive for only *T. brucei* or *T. congolense* prevented investigation of these species independent of the effect of mixed infections. Significant associations ($p < 0.05$) are highlighted in bold.

A

Variable	Coefficient	+/- S.E.	P
Constant	34.98	1.76	
T. brucei	-0.73	1.21	0.55
T. congolense	-4.05	1.26	<0.001
T. vivax	-3.49	1.85	0.06

B

Variable	Coefficient	+/- S.E.	P
Constant	40.95	6.21	
T. brucei	-0.34	1.25	0.79
T. congolense	-1.99	1.46	0.17
T. vivax	-4.87	1.99	0.02
Infection status (microscopy)	-5.66	1.81	<0.001
TP	-1.17	0.66	0.08
Cholesterol	1.06	0.67	0.11

Table 4.3A-B Multiple variable linear regression of infection variables

Multiple variable linear regression analysis to determine the significant variables affecting PCV. A, utilising the results of PCR based infection status only; B, a full model comprised of PCR infection status and any other factor found to have a $p < 0.25$ during initial univariable linear regression screening.

In order to further assess the contributors to PCV a second multiple variable linear regression model was constructed following the parameters employed by Pinchbeck *et al.* (2008). The individual impact of all variables on PCV was screened through an initial univariable regression analysis and those loosely associated with a change in PCV ($p < 0.25$ in initial screen) were included in the multiple variable model. PCR infection status was included in the final model without initial screening.

The results of the second multiple variable analysis (Table 4.3B) indicated only two variables to have a significant effect upon PCV, a positive microscopic infection status and infection with *T. vivax*, with microscopic infection status having the biggest effect upon PCV. The adjusted R^2 value of this second model (15.1%) indicates that while the inclusion of additional data have increased the fit of the model the majority of the observed variation in PCV remains unexplained. The results of these two models are in sharp contrast to those reported by Pinchbeck *et al.*(2008) in their analysis of samples from The Gambia in 2006 which were collected from equines under similar conditions. The study of Pinchbeck *et al.*(2008) found a wide range of variables to have a significant effect upon the PCV of animals, including species and gender of the host. However in the present study neither host species nor gender were included due to no association with PCV ($p = 0.90$ and 0.47 respectively) during initial univariable screening. The most striking differences in the results from the respective studies are for *T. brucei* and *T. congolense*, which were both found to have a significant effect on PCV by Pinchbeck *et al.*(2008) but do not in the present study when all variables are included.

4.2.3 Confirmation of the observed *T. vivax* prevalence

During the original analysis of trypanosome prevalence in The Gambia by Pinchbeck *et al.*(2008) *T. vivax* was observed in 87% of the equines, far higher than the 18% and 31% reported for *T. brucei* and *T. congolense* respectively, with the *T. vivax* prevalence declining slightly to 84% with the inclusion of the January 2007 samples as described in Chapter 3. Further to this the attempt to genotype the *T. vivax* samples utilising a panel of microsatellite markers (Duffy *et al.*, 2009) as described in Chapter 3 failed to amplify any of the markers for 90% of the samples. In order to confirm that the *T. vivax* positives in the 2009 sample set were truly *T. vivax* the band amplified by the *T. vivax* species specific primers was extracted and sequenced. BLASTing the 142 bp sequence (Appendix 8) against the WTSI *T. vivax* contigs identified numerous high identity (>90% match) sequences, as expected given the repetitive nature of the sequence that the TVW primers target (Masiga *et al.*, 1992). A subsequent BLAST search of the equine reference genome (NCBI TaxID 9796) did not identify any significant hits to the amplified sequence, indicating that the high prevalence of *T. vivax* identified by species specific PCR is not due to cross reaction with host DNA.

4.2.4 Genetic Analysis of *T. vivax*

One of the primary aims of the work in Chapter 3 was to examine the genetic diversity and frequency of mating in *T. vivax* through the use of microsatellite markers. The 2007 population was observed to be highly clonal in nature, with a single genotype dominating the population and no evidence for genetic exchange, in line with the results of earlier isoenzyme based studies (Kilgour *et al.*, 1975; Kilgour and Godfrey, 1977; Tibayrenc *et al.*, 1991). However, as described in Chapter 3 the majority of samples failed to amplify with any of the markers. This second survey therefore provided an opportunity to revisit this problem using DNA prepared from 2 ml of whole blood, which may be expected to overcome any sensitivity issues associated with blood spots on FTA cards.

Of the 173 *T. vivax* positive samples in the 2009 sample set only 11 were successfully genotyped for four or more of the microsatellite markers, a success rate of only 6% (Appendix 9). This low number of genotyped samples precludes statistical analysis of the population as the tests would lack the power to give reliable results. Grouping of the samples with those from the 2007 population described in Chapter 3 allowed for the construction of a combined dendrogram of similarity (Figure 4.2), however, this was unable to separate the 2007 and 2009 populations. While the ILRAD V34 isolate was included as an out-group, the dendrogram generated by Clustering Calculator rooted itself to X054, an isolate from the 2009 population, indicating that ILRAD V34 was more closely related to the population than expected, despite its Kenyan origins. Although the 2007 and 2009 populations could not be separated from one another, no genotypes were shared between them. While this may in part be due to the smaller number of isolates genotyped from 2009, the 2007 population was dominated by a single genotype, representing 15 of the 31 isolates. The complete absence of this genotype in 2009 suggests there may have been substantial shifts in the population in the two years between sampling, however this is impossible to confirm given the size of the populations available at this time.

The inability to genotype the majority of the *T. vivax* samples has now been observed following sampling with both FTA cards and DNA purified from 2 ml of whole blood. Following the analysis of the 2007 sample population it was considered possible that the lack of microsatellite amplification was due to low parasitaemia, a phenomenon that has been previously observed in *T. vivax* infections. If this is the case then the results here would suggest that even use of purified DNA from a larger volume of blood is insufficient to counteract this issue. It is also possible that the PCR diagnostic results are false positives, detecting DNA due to past exposure to *T. vivax* as opposed to current infection with *T. vivax*.

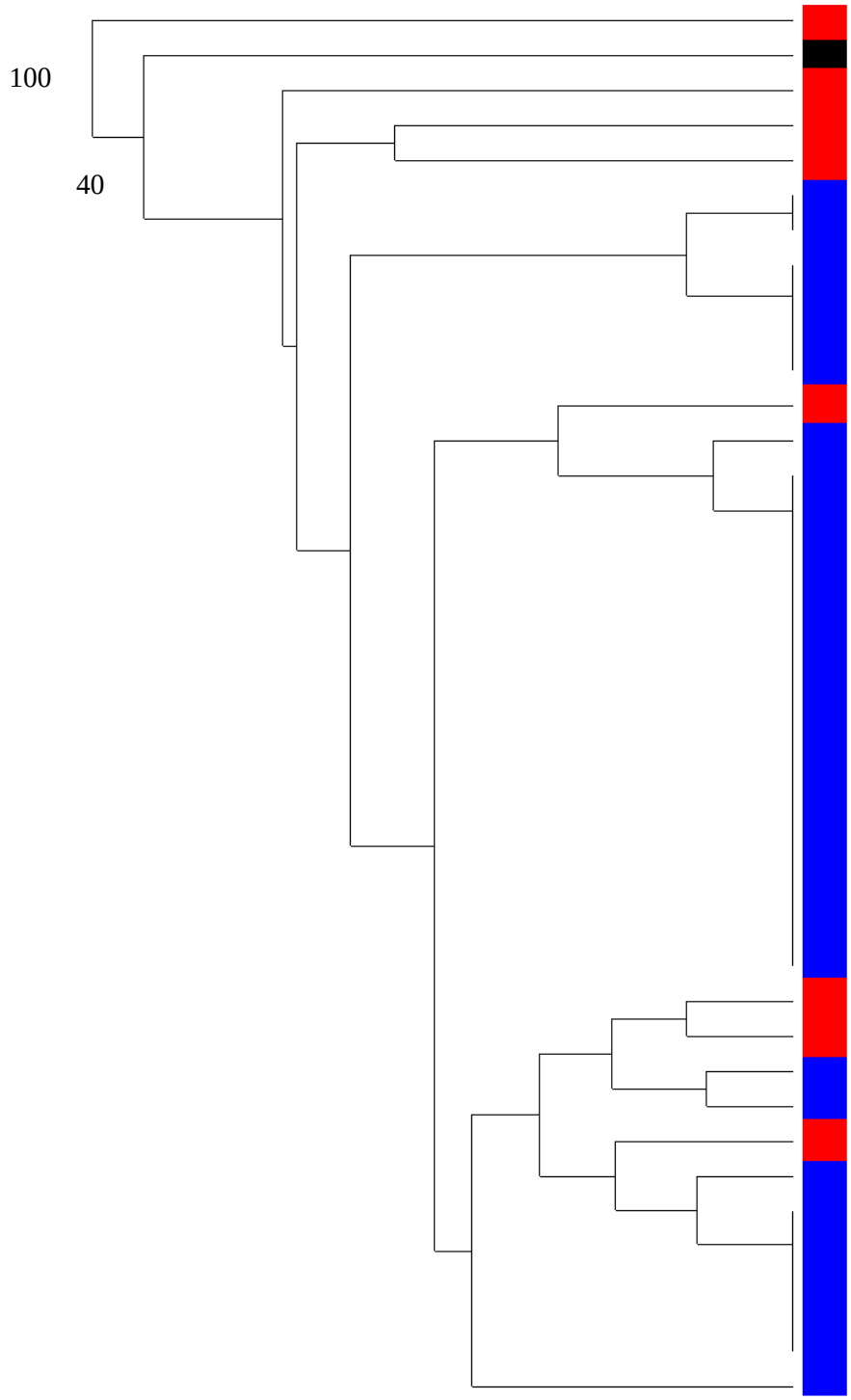


Figure 4.2 UPGMA dendrogram of similarity combining the 2007 and 2009 *T. vivax* populations

UPGMA dendrogram of similarity combining the genotyped *T. vivax* samples from 2007 and 2009. While the isolate ILRAD V34 (black) was included as an out-group the dendrogram rooted to the 2009 Gambian X054 isolate. The two (2007 and 2009) populations could not be resolved from one another based on bootstrap support (provided for the main nodes only). Blue = 2007, Red = 2009.

4.2.5 Genetic Analysis of *T. congolense*

In addition to *T. vivax*, the 2009 sample set provided the opportunity to revisit the *T. congolense* population of The Gambia as a follow-up study to that of Morrison *et al.*(2009b). That study had found a population with a high level of genetic diversity, an excess of homozygosity across seven microsatellite markers, evidence for frequent mating and existence of cryptic sub-populations. These conclusions however were based upon analysis of *T. congolense* in The Gambia collected over the period of a single year and therefore may not be representative of the species in general. The 2009 study therefore aimed to reassess the role of genetic exchange in *T. congolense* and determine whether the population structure and genetic diversity observed in 2007 was stable or fluctuated over time.

A total of 70 of the 198 samples tested positive for *T. congolense* during the initial species specific screening and of these 52 were genotyped with a minimum of four of the seven previously described microsatellite markers TCM1-7 (markers described in Morrison *et al.*, 2009b) (Appendix 10). Eleven of the samples were classified as mixed infections based upon the presence of three alleles at one or more loci. Four of these 11 samples were identified as mixed infections at two or more loci. This mixed infection prevalence of 15.7% is comparable to the 20.2% rate observed in the 2007 sample set (Morrison *et al.*, 2009b). While these samples have been classified as mixed infections it is theoretically possible that these observations instead indicate the presence of triploidy, previously observed in some *T. brucei* hybrid progeny following genetic exchange (Paindavoine *et al.*, 1986b; Wells *et al.*, 1987; Gibson *et al.*, 1992, 1995; Gibson and Bailey, 1994; Hope *et al.*, 1999; Peacock *et al.*, 2009). Samples with more than two alleles at any locus have been excluded from the remainder of the analysis. Twenty six of the remaining 41 samples were genotyped at all seven loci with each isolate possessing a unique MLG.

Between four and eight alleles were observed for the seven loci with an excess of homozygotes observed at all loci compared to that expected in a randomly mating population (Table 4.4), resulting in a population average inbreeding coefficient (F_{IS}) of 0.45. Given the high frequency of homozygotes within the population it is possible that a number of the mixed infections may represent infections with more than two distinct strains present within the host.

The number of alleles observed within the population is reduced compared to that of 2007, with many of the alleles that were observed at low frequency (< 0.1) in 2007 absent in 2009. While this could indicate an alteration in allele frequency in the two years between sampling the smaller size of the 2009 population will have also reduced the chance of these low frequency alleles being observed. Only a single new allele was observed in 2009 compared to 2007, present as a homozygous pair in only a single sample (X070).

In order to assess the genetic diversity and frequency of genetic exchange in the *T. congolense* 2009 population the genotyping data for each sample were collected together to generate MLGs for every sample. All of the 26 fully genotyped isolates possessed unique MLGs, emphasising the diversity within the population. From these 26 samples a UPGMA dendrogram of similarity was constructed (Figure 4.3), rooted to the East African *T. congolense* reference strain IL3000. The dendrogram grouped all of the isolates into a single group, with no support for the presence of genetically differentiation within the population. With multiple alleles at each locus, positive inbreeding coefficients and lack of repeated genotypes the 2009 population closely mirrors that observed by Morrison *et al.*(2009b) in the 2007 population, consistent once again with the occurrence of frequent genetic exchange in *T. congolense*. In order to assess this further, deviation from Hardy-Weinberg equilibrium and the presence of significant linkage disequilibrium was assessed within first the population as a whole and for the fully genotyped MLGs only.

For the entire population, including samples with missing data, loci TCM 1-6 were all observed to differ significantly from Hardy-Weinberg equilibrium, with observed probabilities of $p < 0.001$ for all six loci (Table 4.4). The seventh locus, TCM 7 was in agreement with Hardy-Weinberg expectations. Pairwise loci combinations were assessed for the presence of linkage disequilibrium after accounting for the deviation from Hardy-Weinberg predictions by the preservation of allele pairs at loci with significant deviation. Significant linkage was observed between only three allele pairs, TCM 1 – TCM 7, TCM 3 – TCM 7, TCM 5 – TCM 7 and TCM 6 – TCM7 (Table 4.5). In order to account for the potential effect of clonality upon a population it was proposed by Maynard Smith (Maynard Smith *et al.*, 1993) that estimations of Hardy-Weinberg equilibrium and linkage should be assessed twice, once with the whole population and again with the removal of repeated or partial genotypes. Analysis of only unique *T. congolense* MLGs resulted in a second locus, TCM 6 being in agreement with Hardy-Weinberg expectations in addition to TCM 7, which was in agreement with expectations when assessing all isolates and unique MLGs only (Table 4.4). Significant linkage disequilibrium was observed between only two of the twenty one loci pairs, TCM 1 – TCM 3 and TCM 6 – TCM 7 (Table 4.5) following removal of the repeated genotypes.

Locus	n*	A	He	Ho	F_{IS}	P
TCM 1	38 / 26	7 / 7	0.74 / 0.72	0.44 / 0.54	0.40 / 0.27	**0.00** / **0.02**
TCM 2	35 / 26	7 / 7	0.73 / 0.73	0.11 / 0.15	0.85 / 0.79	**0.00** / **0.00**
TCM 3	38 / 26	8 / 7	0.82 / 0.81	0.53 / 0.50	0.36 / 0.39	**0.00** / **0.00**
TCM 4	36 / 26	4 / 4	0.52 / 0.55	0.11 / 0.12	0.79 / 0.80	**0.00** / **0.00**
TCM 5	39 / 26	8 / 6	0.75 / 0.72	0.38 / 0.31	0.49 / 0.58	**0.00** / **0.00**
TCM 6	38 / 26	7 / 7	0.76 / 0.68	0.58 / 0.62	0.24 / 0.10	**0.00** / 0.10
TCM 7	40 / 26	7 / 7	0.76 / 0.76	0.63 / 0.69	0.17 / 0.09	0.43 / 0.85

Table 4.4 *T. congolense* 2009 descriptive statistics and agreement with Hardy-Weinberg expectations

Descriptive statistics of the 2009 *T. congolense* population. n*, sample numbers; A, observed number of alleles; He and Ho, observed and expected heterozygosities; F_{IS}, inbreeding coefficient; P, probability of agreement with Hardy-Weinberg equilibrium. Values are for all samples / unique MLGs only. Results in bold indicate the presence of significant deviation from Hardy-Weinberg expectations.

Locus 1	Locus 2	P
TCM 1	TCM 2	0.55 / 0.72
TCM 1	TCM 3	0.09 / **0.02**
TCM 1	TCM 4	0.13 / 0.23
TCM 1	TCM 5	0.84 / 0.87
TCM 1	TCM 6	0.19 / 0.15
TCM 1	TCM 7	**0.00** / 0.18
TCM 2	TCM 3	0.17 / 0.49
TCM 2	TCM 4	0.79 / 0.26
TCM 2	TCM 5	0.54 / 0.28
TCM 2	TCM 6	0.42 / 0.17
TCM 2	TCM 7	0.26 / 0.57
TCM 3	TCM 4	0.29 / 0.70
TCM 3	TCM 5	0.15 / 0.49
TCM 3	TCM 6	0.37 / 0.34
TCM 3	TCM 7	**0.00** / 0.08
TCM 4	TCM 5	0.18 / 0.08
TCM 4	TCM 6	0.98 / 0.61
TCM 4	TCM 7	0.08 / 0.21
TCM 5	TCM 6	0.71 / 0.39
TCM 5	TCM 7	**0.01** / 0.12
TCM 6	TCM 7	**0.01** / **0.05**

Table 4.5 Probability of pairwise linkage disequilibrium between *T.* congolense loci

Linkage disequilibrium between all pairwise combinations for the seven microsatellite markers TCM 1 – 7 for the entire population / unique MLGs only (after preserving genotypes of those markers that were not in Hardy-Weinberg equilibrium). Results in bold indicate the presence of significant linkage disequilibrium.

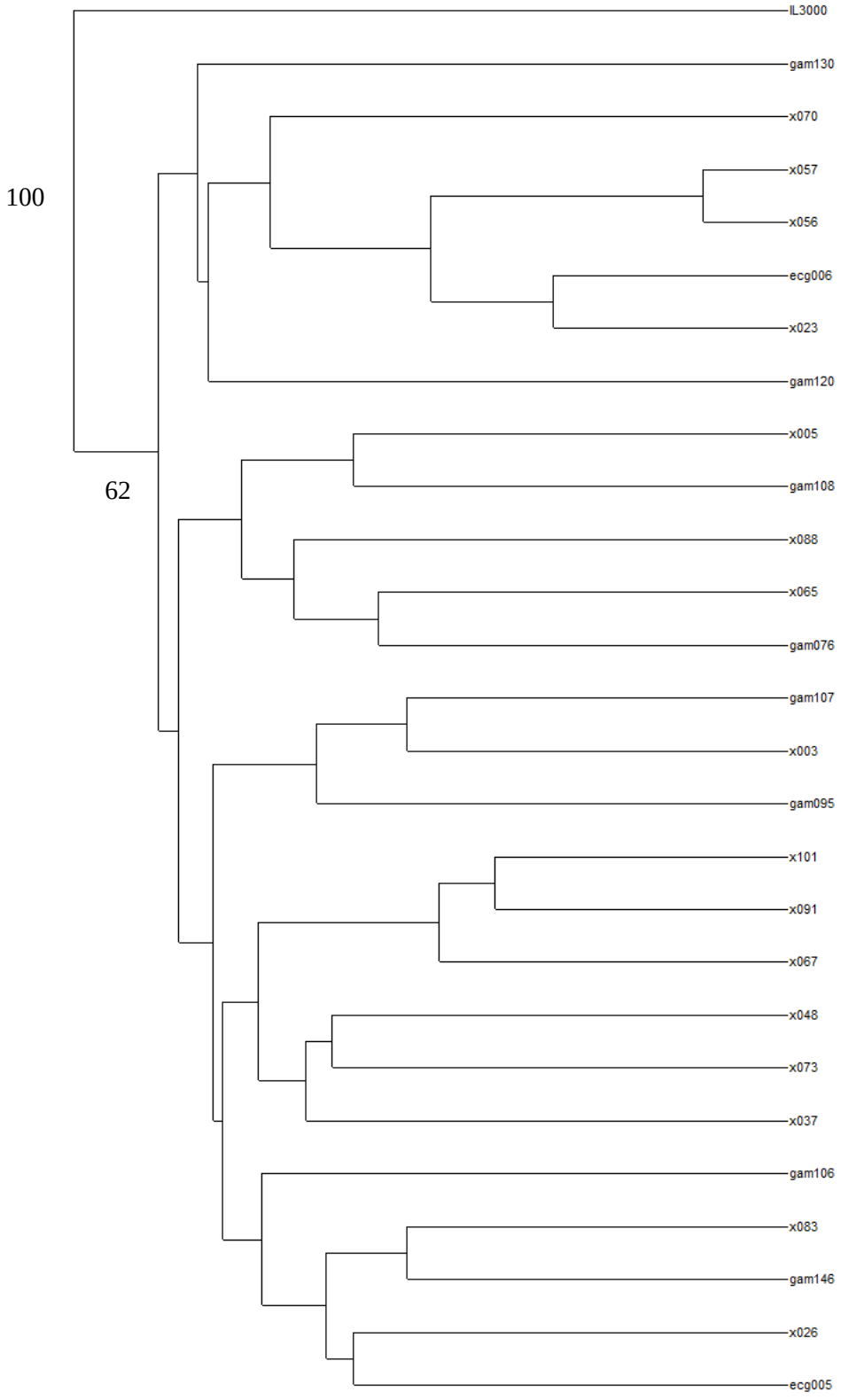


Figure 4.3 UPGMA dendrogram of similarity for the 2009 *T. congolense* population

UPGMA dendrogram of similarity constructed from the 26 fully genotyped *T. congolense* samples from 2009, rooted against the *T. congolense* isolate IL3000. The dendrogram indicated the population to form a single large group based upon bootstrap support at the major nodes (Bootstrap values for the main nodes only).

The observation of significant deviation from Hardy-Weinberg equilibrium, lack of repeated MLGs and lack of linkage disequilibrium between the majority of pairwise loci combinations does not easily fit with the idea of clonality. While this may represent the true population structure there are a number of factors which could explain the observed lack of panmixia, notably null alleles, selfing and introduction of the Wahlund effect through the grouping of independent populations (Wahlund, 1928). The observed excess of homozygotes observed across all loci could have arisen from the presence of null alleles that cannot be amplified due to polymorphisms in the primer binding sites. In order to assess this, the predicted frequency of null alleles and number of predicted homozygote nulls was calculated for each locus (Table 4.6) (Hedrick, 2005). The calculation of null allele frequency assumes that the deviation from Hardy-Weinberg expectations is due entirely to the presence of null alleles, giving rise to an excess of homozygotes and therefore calculates the frequency of null alleles required for agreement with Hardy-Weinberg principles. From the predicted null allele frequency the number of expected homozygote nulls was determined and for loci TCM 2 and TCM 5 the number of predicted homozygote nulls was very close to the observed number of homozygote nulls. It is therefore possible that null alleles are present within the population, resulting in the observed heterozygote deficit at these two loci. These calculations however preclude the possibility that other factors have contributed to the observed deviation from Hardy-Weinberg principles.

Selfing is one of the possible other explanations for the presence of the homozygote excess across all loci. While selfing in trypanosomes was initially only observed in the presence of cross fertilisation (Tait *et al.*, 1996; Gibson *et al.*, 1997; Hope *et al.*, 1999),selfing during single strain transmission has now been demonstrated to occur in a laboratory cross of *T. brucei* (Peacock *et al.*, 2009). In order to assess the role of selfing the predicted selfing rate (de Meeûs *et al.*, 2007) was calculated for each of the seven loci, with the assumption of no null alleles being present. Selfing rates ranged from 0.29 – 0.92 (Table 4.6), indicating that high levels of selfing would be required in order to explain the observed frequency of homozygotes. This could indicate that genetic exchange is a frequent occurrence in *T. congolense* with frequent single strain tsetse transmission leading to a high level of selfing. The high predicted frequency of selfing and lack of repeated genotypes in this population does not necessarily exclude genetic exchange being non-obligatory in *T. congolense*, but it would suggest that the conditions in The Gambia facilitate a high enough mating frequency to prevent a single clone from dominating the population.

Locus	p	Ne	No	s
TCM 1	0.17	1.22	3	0.57
TCM 2	0.36	5.27	6	0.92
TCM 3	0.6	1.04	3	0.53
TCM 4	0.27	2.98	5	0.88
TCM 5	0.21	1.83	2	0.66
TCM 6	0.10	0.43	3	0.39
TCM 7	0.07	0.22	1	0.29

Table 4.6 Predicted null allele frequencies, homozygote nulls and predicted selfing rates for the *T. congolense* 2009 population.

p, predicted frequency of null alleles; Ne, predicted number of homozygote nulls; No, observed number of homozygote nulls; s, predicted selfing rate. Due to the required inclusion of 'nulls' these data are presented for all samples, inclusive of those partially genotyped.

The final factor investigated here is the possibility that the 2009 *T. congolense* population is comprised of a number of separate populations, inadvertently grouped together during the initial analysis. Grouping populations in this manner can lead to deviation from Hardy-Weinberg equilibrium if the alleles present in each sub-population differ sufficiently in frequency. This distortion, known as the Wahlund effect (Wahlund, 1928), was identified by Morrison *et al.*(2009b) in their first analysis of *T. congolense* identifying four putative sub-populations in The Gambia and a similar analysis has therefore been undertaken for the 2009 population, utilising the STRUCTURE program (Pritchard *et al.*, 2000).Utilising an admixture model to allow for the possibility of gene flow between sub-populations, the optimal number of sub-populations for the dataset was determined using the delta K method of Evanno *et al.*(2005). The population structure was simulated in the presence of a known number of sub-populations (K values) of 1 – 10, with the delta K value calculated for each K value. The optimal number of sub-populations may then be identified by plotting delta K versus K and identifying the peak point of the plot (Figure 4.4). Analysis of the *T. congolense* 2009 population inferred the presence of 4 sub-populations, the same number as identified by Morrison *et al.*(2009b) for the 2007 population. Of these four putative sub-populations the largest contained only 15 samples (Figure 4.5) and due to this small sample size it was not appropriate to statistically analyse the sub-populations for agreement with Hardy-Weinberg equilibrium or the presence of linkage disequilibrium between pairs of loci.

The STRUCTURE program functions by searching for the optimal number of sub-populations that minimises the deviation from Hardy-Weinberg in each sub-population. In the presence of clonality the program is likely to separate genetically distinct populations which must then be individually assessed for agreement with Hardy-Weinberg expectations (Balmer *et al.*, 2011). The non-obligatory mating cycle observed in *T. brucei*, which may be shared with *T. congolense* therefore raises the possibility that the putative sub-populations were identified due to being genetically distinct from one another as opposed to the presence of agreement with Hardy-Weinberg expectations due to frequent genetic exchange. However, the small sizes of the putative sub-populations prevent assessment of this possibility.

During comparison of the population structuring in 2007 and 2009 it became apparent that the original conclusion of four sub-populations in the 2007 dataset was incorrect. This was due to an incorrect calculation of the delta K values due to the failure to use modal values as directed by Evanno *et al.*(2005), resulting in an inverted graph (Supplementary Figure 1 of Morrison *et al.*(2009b)). Correction of the delta K values indicated the true optimal K value to be 1 or 2 (as the delta K technique cannot distinguish between the two). While this affects the subsequent identification of a sub-population in agreement with Hardy-Weinberg principles the earlier conclusions of frequent genetic exchange and inbreeding is not affected and is supported by the results presented here.

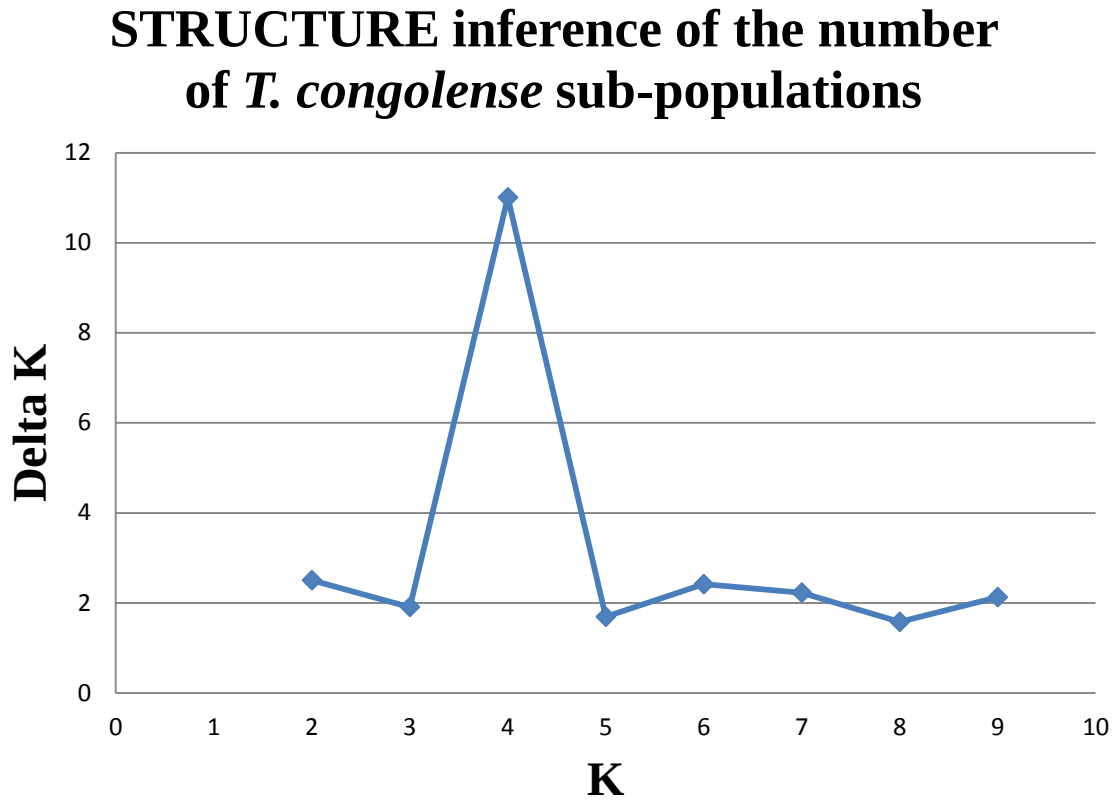


Figure 4.4 STRUCTURE inference of the number of *T. congolense* sub-populations.

The modal value of delta K indicates the most likely number of sub-populations within the sample population, indicated to be 4. Delta K was calculated as described by Evanno *et al.*(2005).

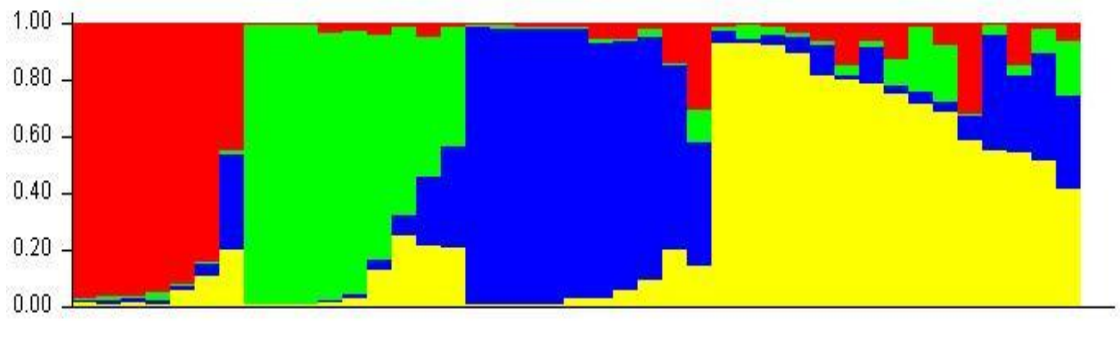


Figure 4.5 Example predicted population structure for *T. congolense* in 2009

Population structure was inferred following calculation of K = 4 as the optimal number of sub-populations.

A second limitation of the STRUCTURE program is that in its determination of population structure it does not allow for the possibility of selfing, as previously demonstrated to occur in *T. brucei* (Tait *et al.*, 1996; Gibson *et al.*, 1997; Hope *et al.*, 1999; Peacock *et al.*, 2009). Analysis of the *T. congolense* 2009 population as a whole indicated an excess of homozygotes, with positive inbreeding coefficients across all seven loci, indicating that selfing may be a frequent occurrence. A variant of the STRUCTURE program, InStruct (Gao *et al.*, 2007) allows for the possibility of selfing when searching for the optimal number of sub-populations and unlike STRUCTURE does not assume sub-populations to be in Hardy-Weinberg equilibrium. Population structures were simulated for K values of 1 – 10 with an admixture model that allowed for selfing to be occurring in the population. Delta K values (Evanno *et al.*, 2005) were calculated as for the STRUCTURE analysis and plotted versus K (Figure 4.6). The plot suggests the presence of either no sub-structuring or 2 – 3 sub-populations, as the delta K method is unable to assess the likelihood of a population lacking sub-structuring, with example predicted population structures presented in Figure 4.7. Inbreeding coefficients for K = 1 – 3 range from 0.36 – 0.52 (K = 3, sub-populations 1 and 3 respectively, Table 4.7) indicating that selfing is predicted to occur frequently in the putative sub-populations. The support for sub-structuring as detected by STRUCTURE and InStruct is indicated by the delta K values of the optimal population structures. The respective delta K values of both analyses are below 20, suggesting only low support for the presence of sub-structuring within the population.

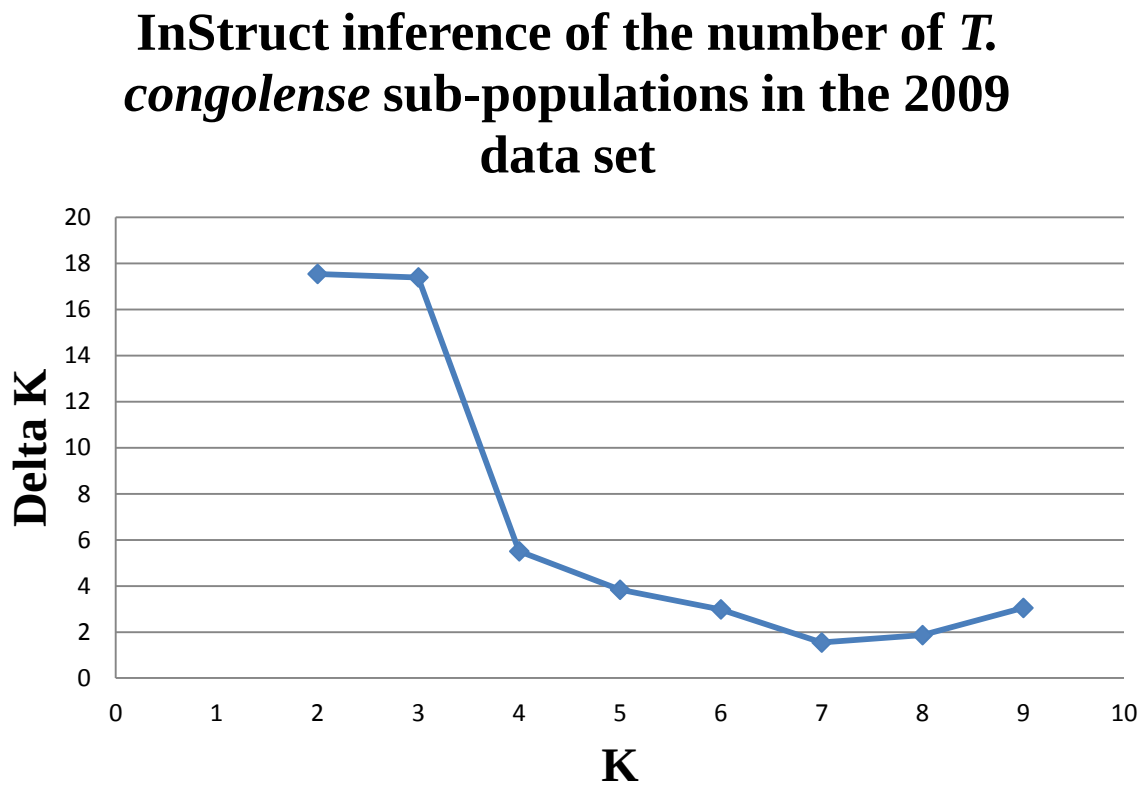
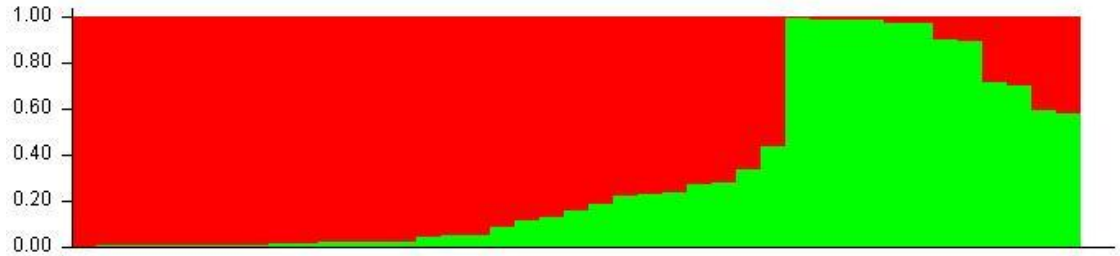


Figure 4.6 InStruct inference of the number of *T. congolense* sub-populations

InStruct inference of the number of *T. congolense* sub-populations if selfing is allowed. The greatest value of delta K indicates the most likely number of sub-populations (K) within the sample population, which is indicated to be 2 or 3. The method cannot however assess the likelihood of no sub-structuring and therefore K = 1 remains a possibility. Delta K was calculated as described by Evanno *et al.*(2005).

A



B

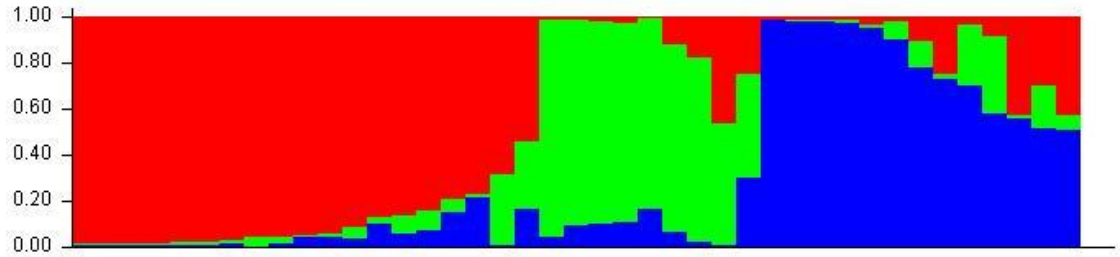


Figure 4.7A-B Example predicted population structures for *T. congolense* in 2009

The population structure was calculated utilising InStruct, which calculated an optimal K value of 2 or 3. A: K = 2, B: K = 3.

	K = 1	K = 2	K = 3
Sub-population 1	0.46	0.41	0.36
Sub-population 2	-	0.45	0.39
Sub-population 3	-	-	0.52

Table 4.7 *T. congolense* sub-population inbreeding coefficients

Average posterior distribution of inbreeding coefficients (F_{IS}) for the 2009 *T. congolense* population as calculated by InStruct for K values of 1 – 3, averaged from 20 independent runs.

4.2.6 Genetic relationship between the 2007 and 2009 *T. congolense* populations

Populations are by their very nature dynamic, with the continual introduction of new genotypes through genetic exchange, immigration of individuals from neighbouring populations and mutational drift giving rise to novel alleles. Selective pressures may further shape the population by favouring the fittest combinations and alleles that are present, contributing to changes in the frequencies of individual alleles. Even in clonal populations different genotypes can come to dominate a population over time and individual lineages can slowly drift apart, as has been observed for *T. b. gambiense* Group 1 when comparing distinct foci of this parasite (Morrison *et al.*, 2008b). With the availability of two *T. congolense* populations, sampled from the same geographical region of The Gambia two years apart it was possible to assess the relationship between them and thus gain insight into how quickly the population is changing over time. A dendrogram of similarity including both the 2007 and 2009 populations (Figure 4.8) was unable to resolve the two populations from one another or detect the presence of subgroups within the combined population. The dendrogram identified only two isolates from 2009 (X003 and X005) possessing MLGs which were present in the 2007 population (* on dendrogram), supporting the earlier analysis of genetic exchange frequently disrupting existing genotypes.

A total of 15 alleles were observed to be private to the 2007 population compared to only a single private allele unique to the 2009 population. All private alleles, regardless of the population were observed at only low frequencies of 0.05 or below and the smaller size of the 2009 population, comprising 52 isolates compared to the 84 genotyped from 2007, may have prevented observation of low frequency alleles from 2007 in isolates from 2009. This low frequency of private alleles and similar frequency of shared alleles results in virtually no observable differentiation between the populations (Nei's unbiased genetic distance = 0.01, F_{ST} = 0.01). Principal co-ordinate analysis, a second method of examining the relationships between individuals in the two populations (Figure 4.9) was unable to resolve the two from one another, as expected given their close genetic relationship. The low level of differentiation between the two populations suggests that while genetic exchange may be frequent, the diversity present within the population is being maintained, with mutation and immigration of individuals from neighbouring populations having an insignificant effect over the two year period. Further sampling over greater time periods will be required to assess the role of genetic drift, mutation and immigrationon *T. congolense* populations in The Gambia and the neighbouring regions.

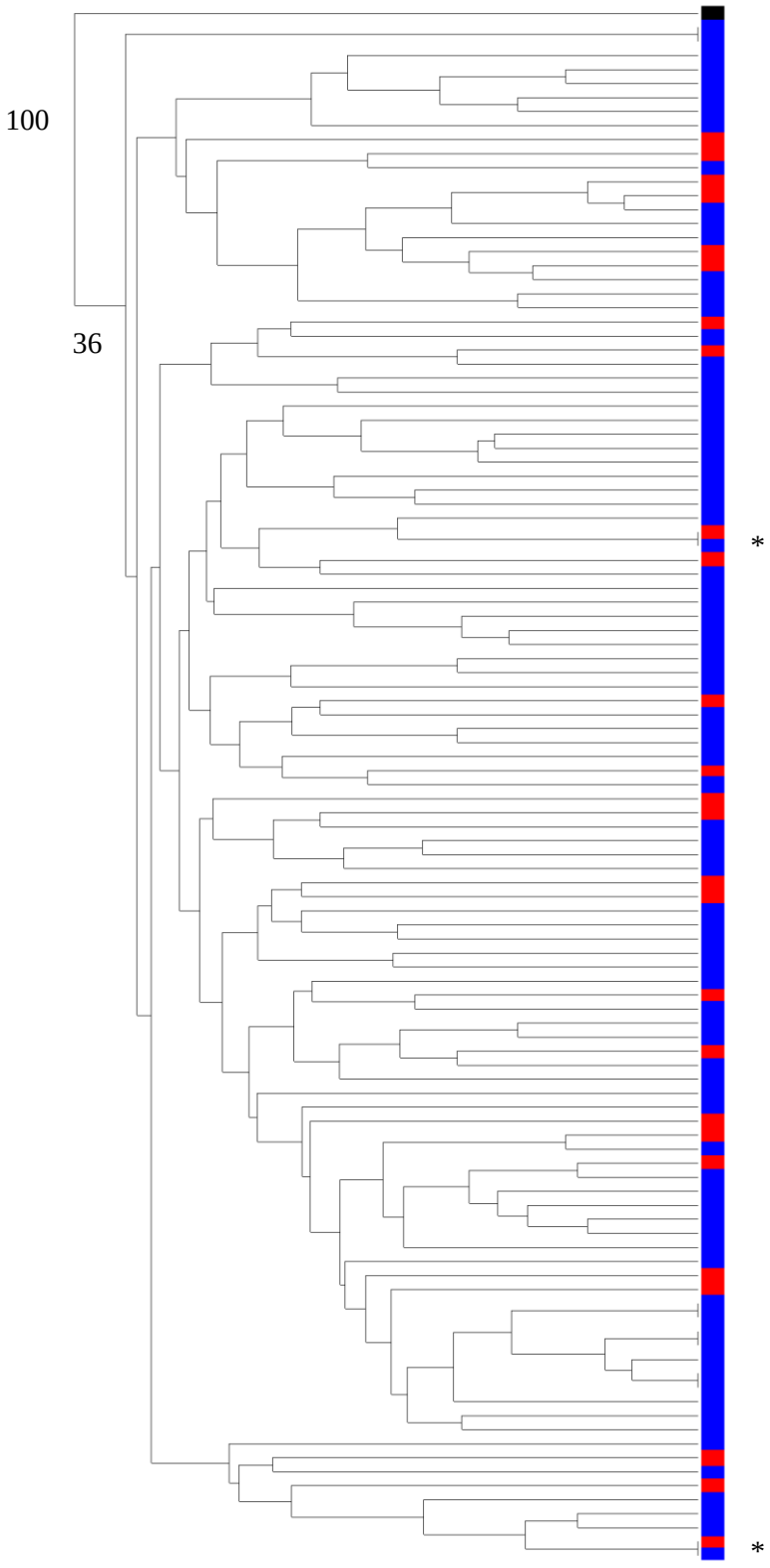


Figure 4.8 UPGMA dendrogram of similarity employing the 2007 and 2009 *T. congolense* populations

UPGMA dendrogram of similarity including both the 2007 and 2009 *T. congolense* populations. No bootstrap support was detected for the separation of the two populations. Blue = 2007, Red = 2009, * = MLGs present in both 2007 and 2009. Bootstrap values are given for the major nodes only.

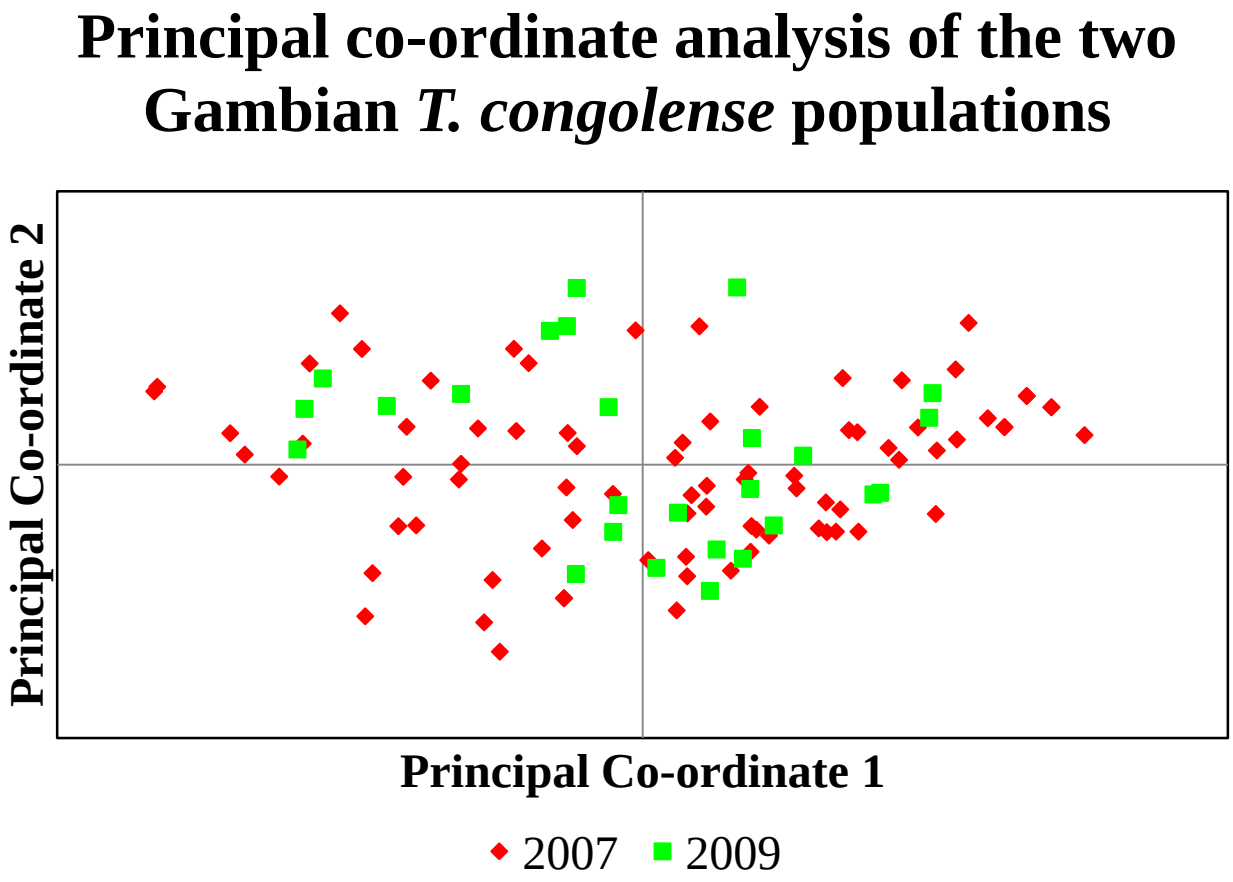


Figure 4.9 PCoA of the Gambian *T. congolense* populations from 2007 and 2009.

The analysis was unable to resolve the two populations from one another due to their close relationship. Principal co-ordinate 1 accounts for 31% of the observed variation while principal co-ordinate 2 accounts for 19% of the observed variation.

4.3 Discussion

The work described in this chapter is a direct follow up of the studies of Pinchbeck *et al.*(2008), Morrison *et al.*(2009b) and the initial *T. vivax* study described in Chapter 3. To achieve this however required the collection of new samples and I was lucky enough to be offered the opportunity to spend two months working directly with the GHDT at their base in Sambel Kunda, The Gambia, and at the clinics they run at nearby villages and markets. Up until the fieldwork my understanding of the diseases caused by trypanosomes was purely academic, gleaned from the literature alone, which cannot substitute for firsthand experience. This is especially true given the focus of this thesis primarily upon the genetics of trypanosome field populations. Working with the GHDT therefore provided the opportunity to experience firsthand the impact of animal African trypanosomiasis, the value of which cannot be conveyed in the results section of this Chapter. Most striking of the lessons learned from the fieldwork is the requirement for simple and effective diagnosis technologies that are cheap, quick and portable. Diagnosis of infection at the clinics was reliant on such methods, primarily a hand spinner for determination of PCV (Figure 1.3) and a solar powered microscope, which was typically operated from the backseat of the GHDTs pickup truck (Figure 4.10). While the highly sensitive PCR based methods employed throughout this body of work are effective for laboratory based work, much remains to be done if they are to be employed in the field.

The aims of this chapter in revisiting The Gambia were threefold, reassess the high frequency of observed *T. vivax* infections in 2007, conduct a survey of AAT in The Gambia and examine the population structures of *T. congolense* and *T. vivax* over time. There was little difference in the observed 2007 and 2009 equine prevalences for both microscopy and PCR, with the majority of the PCR positive results attributed once again to *T. vivax*. As with the original study however the use of animals brought to clinics prevents the estimation of the true prevalence of equine trypanosomiasis in The Gambia. Once again the most striking result from the species specific PCRs is the extremely high prevalence of *T. vivax*, observed in all but 15 of the samples, with the PCR results confirmed through sequencing of the amplified band. This high prevalence could possibly be explained by a number of factors. The first is simply that the equine prevalence of *T. vivax* in The Gambia is considerably higher than that of *T. congolense* and *T. brucei*, however this in sharp contrast to the results of Faye *et al.*(2001) which found *T. congolense* to be the most prevalent species infecting equines in The Gambia. Alternatively the use of equines brought to treatment clinics may have selected for animals more likely to be infected by *T. vivax* as opposed to *T. brucei* or *T. congolense*. It is not clear however as to whether infection with *T. vivax* results in the increased likelihood of an animal being brought to clinic or if existing ill health in an animal facilitates infection with *T. vivax*.

Figure 4.10 Field diagnosis setup

Diagnostic setup at the GHDT clinics, operating from the backseat of the pickup truck highlights the requirement for simple and mobile technologies when in the field.

The linear regression analysis, in indicating an association between microscopically positive infections and anaemia is in agreement with previously observed results, however the multivariable analysis found a number of differences when compared to the results of Pinchbeck *et al.*(2008), most notably that neither *T. brucei* nor *T. congolense* were significantly associated with a reduced PCV when all variables were considered. It is possible however that the low frequency of *T. brucei* or *T. congolense* only infections, coupled to the low frequency of uninfected samples has skewed the results and a considerably higher number of uninfected samples should be included in future analyses if possible.

The second half of the results in this chapter focused upon the population genetics of the *T. vivax* and *T. congolense* species. As observed in Chapter 3 the attempted genotyping of *T. vivax* yielded results for only a small proportion of the population as identified by the TVW primer set. It had been hoped that the increased material available from the collection of whole blood samples, as opposed to FTA blood spots would increase the proportion of samples which could be genotyped,

however this is clearly not the case. Sequencing of material from samples which could not be genotyped with the microsatellite markers confirmed the presence of DNA originating from *T. vivax*. There are a number of possible explanations for the failure to genotype these samples. The first is that these samples were collected from animals with low parasitaemia *T. vivax* infections, below that required for amplification with the single copy microsatellite markers. In contrast the *T. vivax* species specific TVW primer set targets a multicopy sequence which may have provided a sufficient increase in sensitivity for identification of these low parasitaemia infections. If this is the case then a method of concentrating sufficient parasites from the blood will be required in future to further examine the population diversity and structure of *T. vivax*.

An alternative possibility is that these samples are indicative of the presence of genetic variation between the West African samples examined here and Y486, the *T. vivax* genome reference strain from Nigeria, which was utilised in the design of the microsatellite markers. Geographic variation separating East and West African isolates of *T. vivax* has been previously demonstrated (Fasogbon *et al.,* 1990, Cortez *et al.,* 2006, Rodrigues *et al.,* 2008) and the failure to amplify many of the samples may indicate the presence of primer site variation giving rise to null alleles. That some of the samples did amplify may alternatively indicate the presence of multiple *T. vivax* subgroups, as has been observed in *T. congolense*, with each subgroup possessing a distinct complement of microsatellites. Subgroup associated microsatellite diversity has been previously observed in *T. congolense* (Morrison *et al.*, 2009b) while the use of fluorescent fragment length barcoding (Hamilton *et al.*, 2008) and sequencing of glycosomal glyceraldehyde phosphate dehydrogenase identified two *T. vivax* groups distinct in Tanzania distinct from West African isolates (Adams *et al.*, 2009).

Finally it is possible that these PCR positive results arise not from established infections with *T. vivax* but from the presence of DNA following exposure to the parasite. In cattle and goats it has been demonstrated that DNA remains in circulation for ~48 hours after the death of the parasite (de Almeida *et al.*, 1997; Desquesnes, 1997) while in humans cerebral spinal fluid has been observed to remain PCR positive for *T. b. gambiense* Group 1 for up to 2 years in approximately 20% of patients following successful treatment (Deborggraeve *et al.*, 2011), pointing to the possible persistence of a small number of parasites within the host. The multi-copy nature of the species specific sequences may extend the timeframe in which this material can be detected in the host compared to the single copy microsatellites used here for genotyping. If DNA clearance is as rapid in equines as in goats and cattle then it would suggest that equines in The Gambia are frequently exposed to *T. vivax*, with the majority of challenges failing to establish ongoing infections in the host. If PCR based techniques are to become more widely implemented in the diagnosis of infection then it may be necessary to determine ways to distinguish previous challenge from current infection, potentially through the use of less sensitive markers such as single copy microsatellites

that require parasite DNA concentrations unlikely to be observed in the absence of established infections.

The genotyping of *T. congolense*, in contrast to *T. vivax*, yielded a highly diverse population with a high proportion of mixed genotype infections in line with the results first observed by Morrison *et al.*(2009b). The population, with an excess of homozygotes, high allele diversity and lack of repeated MLGs does not readily fit with the idea of clonality that has been previously suggested for trypanosomes. The results from studies of *T. b. gambiense* Group 1, which is believed to propagate exclusively by clonal expansion all indicate the dominance of a small number of alleles, an excess of heterozygotes (with little variation in the heterozygote pairs present) and the presence of multiple individuals with the same MLG (Tibayrenc *et al.*, 1990; Mathieu-Daudé *et al.*, 1995; Morrison *et al.*, 2008b; Koffi *et al.*, 2009). This same pattern has now been observed for *T. vivax*, as discussed in Chapter 3 and therefore the possibility of strict or effective clonality in *T. congolense* appears highly unlikely due to the lack of the defining features of clonality.

The results presented here are most consistent with a population in which genetic exchange is a frequent occurrence, with a high frequency of inbreeding and selfing giving rise to the observed excess of homozygotes at all loci, matching the observations of Morrison *et al.*(2009b) in the analysis of the 2007 population. The high frequency of unique MLGs in 2007 (80 / 84 isolates) and 2009 (26 / 26 isolates) however would suggest that in addition to inbreeding, out crossing is occurring at a sufficient frequency to generate the observed MLG variation, contributing to the low level of linkage disequilibrium observed in the 2009 population. Outcrossing is initially dependent upon the frequency of mixed strain infections in the tsetse, which is in turn influenced by the frequency of mixed strain infections in the host. The 2007 and 2009 studies identified mixed strain infections in 20.2% and 15.7% of the host samples. In contrast no mixed infections were observed during the genotyping of *T. vivax* from 2007 or 2009 while previous studies into *T. brucei* have detected mixed infections 8.8% when examining isolates from across Africa (Balmer and Caccone, 2008) and 18% following examination of cattle from the Busoga focus (MacLeod *et al.*, 2000). In tsetse however mixed genotype *T. brucei* infections have been observed at higher frequencies of 36% - 47% (MacLeod *et al.*, 1999; Simo *et al.*, 2011), raising the possibility for significant rates of out crossing. Analysis of the 2009 *T. congolense* population with the STRUCTURE and InStruct programs indicated that sub-structuring may be present, however, the proposed sub-populations could not be associated to host, date of sampling or location of sampling. As the putative sub-populations could not be associated with any of the available sampling data, it is possible that they could reflect an association between parasite and vector as at least two species of tsetse are known to be present in The Gambia (Faye *et al.*, 2001). Other possibilities such as local environmental variation or barriers to interaction between strains remain possible. The InStruct program, which allows for the presence of selfing and deviation from Hardy-Weinberg expectations, supported the

whole population analysis in requiring a high frequency of selfing in order to explain the observed excess of homozygotes.

While selfing during single strain transmission through the tsetse fly has been demonstrated in *T. brucei* (Peacock *et al.*, 2009) genetic exchange is known to be a non-obligatory event and the frequency with which it occurs must therefore be under the control of genetic and / or environmental factors. It is plausible that for *T. congolense* in The Gambia the combination of genetics and environment are conducive to a high mating frequency with little or no clonal propagation. It is even possible that *T. congolense* is switched permanently into the 'on' position with mating a requirement for progression through the tsetse stages of the life cycle. In such a situation where genetic exchange is occurring at a high frequency single strain transmission through the tsetse would generate the observed excess of homozygotes through selfing, as is observed in many foci of *Plasmodium falciparum* where transmission intensity and population diversity control the rate of out crossing (Conway *et al.*, 1999, 2007; Anderson *et al.*, 2000a; Machado *et al.*, 2004; Anthony *et al.*, 2005; Razakandrainibe *et al.*, 2005; Annan *et al.*, 2007; Mzilahowa *et al.*, 2007). Outcrossing in *T. congolense*, facilitated by the observed mixed host infection rate of ~15-20% would subsequentially give rise to the high level of unique MLGs observed.

The frequency of mixed infections in tsetse may in part be controlled by their dispersal range and habitat size, which shrinks considerably during the dry season. Such fragmentation may give rise to micro-populations, with diversity limited due by the relatively small number of hosts and vectors. Immigration of parasites from neighbouring populations, by host movement or expansion of tsetse habitat during the rainy season, would allow gene flow and occasional out crossing. Such micro-geographic sub-structuring has been proposed to occur in *Leishmania braziliensis* (Rougeron *et al.*, 2009) in order to explain the observation of substantial genetic diversity and heterozygote deficits observed in populations from Peru and Bolivia. Further sampling will be required in order to assess this possibility, with hosts sampled in the villages at which they normally reside and with sympatric sampling from the local vector populations. As equines are frequently used as a mode of transport, the locations frequently visited by these animals may shed further light on the dispersal of trypanosomes in The Gambia.

Much work remains however in order to gain a true understanding of the genetics of *T. congolense*, the most important being confirmation that genetic exchange is occurring in this species through laboratory crosses under controlled conditions. The development of *in vitro* culture conditions for all life cycle stages of the species (Coustou *et al.*, 2010) may facilitate such an investigation, bypassing the requirement to infect and screen large numbers of tsetse under laboratory conditions while the adoption of the fluorescent tagging techniques developed for *T. brucei* (Bingle *et al.*,

2001; Gibson *et al.*, 2008; Peacock *et al.*, 2011) and *T. cruzi* (Pires *et al.*, 2008) will further aid the investigation of genetic exchange in *T. congolense*. From a perspective of genetic exchange in the wild further studies will be required as the evidence for genetic exchange at present originates from only a single country while the parasite is spread over much of sub-Saharan Africa. It will therefore be necessary to expand this analysis further afield in order to truly assess the role of mating in the species.

Chapter 5

The Population Genetics of *T. b. rhodesiense* in Uganda and Malawi, 1961 - 2010

5.1 Introduction

While *T. b. gambiense* Group 1 accounts for the vast majority of the reported HAT cases, East African trypanosomiasis resulting from *T. b. rhodesiense* infections present a second front on which the disease must be combated. With an acute disease profile and distinct response to chemotherapy the sub-species generates a different set of challenges to *T. b. gambiense*, which is further complicated by the role of animal reservoirs in the spread and maintenance of disease foci. As discussed in Chapter 1, while *T. b. rhodesiense* has classically been afforded sub-species status, within the *T. brucei* species complex, field isolates of *T. b. rhodesiense* are more closely related to the local *T. b. brucei* populations than *T. b. rhodesiense* isolates from different regions of the continent (MacLeod *et al.*, 2001c). Traditionally the ability of isolates from animal reservoirs to survive in human serum has been utilised as a proxy for human infectivity, distinguishing *T. b. gambiense* and *T. b. rhodesiense* from *T. b. brucei*. The discovery that expression of a single gene, *SRA*, confers resistance to lysis by human serum in *T. b. rhodesiense* (De Greef and Hamers, 1994; Xong *et al.*, 1998) has provided a new technique with which to accurately distinguish *T. b. rhodesiense* from sympatric *T. b. brucei* (Welburn *et al.*, 2001). Due to being defined by the presence of a single gene *T. b. rhodesiense* may be more accurately described as a host variant of *T. b. brucei* as opposed to a genetically distinct sub-species.

Of the current foci of *T. b. rhodesiense* those spread across the districts of Uganda are amongst the most extensively studied. This is in part due to Uganda being the only country presently afflicted by both *T. b. gambiense* Group 1 and *T. b. rhodesiense*, with only a relatively short distance between foci of the two sub-species (Picozzi *et al.*, 2005). Sleeping sickness was first recorded in the country at the start of the twentieth century and was one of the earliest reported outbreaks of HAT. The epidemic was focused along the shores of Lake Victoria and connected rivers and is estimated to have killed more than 200,000 people between 1900 and 1920 (Berrang-Ford *et al.*, 2006). While this outbreak has historically been attributed to *T. b. gambiense* Group 1, more recent re-analysis of case records has suggested that *T. b. rhodesiense* was at least partly, if not wholly responsible based upon the rapid progression recorded in many cases (Fèvre *et al.*, 2004).

Smaller outbreaks were recorded through the 1930s and 1940s followed by only a small number of cases until the 1970s. During this time the majority of reported cases were centred on the Busoga focus, bordering Kenya in the south east of the country. A new epidemic began towards the end of the 1970s, with increased social upheaval and the resulting reduction in screening and tsetse control measures, spreading in a northward direction as the land vacated by humans was left uncultivated, encouraging the re-colonisation by tsetse flies, primarily of the *G. f. fuscipes* species. The outbreak reached a peak in the late 1980s before the number of cases began to fall, partially in response to the return of tsetse control measures. While the number of cases in the epidemic had already

peaked by the end of the 1980s the focus continued to spread, enveloping the Tororo district. Cases were first recorded in Soroti district in 1998 (Fèvre *et al.*, 2001) and most recently in Kumi, Kaberamaido and Lira districts between 2004 and 2005 (Fèvre *et al.*, 2005; Enyaru *et al.*, 2006). Isolated reports of HAT in Masini district, bordering Lake Albert in the mid-west of Uganda investigated by Enyaru *et al* (1999) suggested the potential presence of both *T. b. rhodesiense* and *T. b. gambiense* Group 1 in this region based on isoenzyme characterisation of isolates and the identification of CATT positive individuals. There have, however, been no further reports from this region and it is therefore unclear as to whether *T. b. rhodesiense* is still present within the vector and host populations of this region. At present 14 of the districts in Uganda are reported to be afflicted by *T. b. rhodesiense* despite the present use of control measures in much of the country (Kotlyar, 2010).

The spread of the disease into the Soroti district, with the first cases being reported in December 1998, is of particular interest as it has been possible to track this outbreak from the outset through a combination of molecular and epidemiological surveys. During the political and social upheaval of the 1980s migration of human populations out of the district resulted in the return of natural habitat conditions to large areas of land previously used for agriculture, with a resurgence in tsetse numbers (Hutchinson *et al.*, 2003). The return of stability, and therefore people, to the region in the 1990s provided ideal conditions for the spread of HAT into the district. Firstly the return of wild habitat into agricultural land increased the level of contact between humans and tsetse, facilitating the transmission of *T. b. rhodesiense* to human hosts. The second major factor in the outbreak of HAT in the Soroti district was the importation of livestock to establish or enlarge herds in the district. Fèvre *et al.*(Fèvre *et al.*, 2001) identified Brookes Corner market in Soroti district as the centre of the outbreak based on early case records, and determined that in the lead up to the outbreak (1995 – 1998) 54% of the cattle traded at the market had been imported from *T. b. rhodesiense* endemic regions. Despite screening and a requirement for treatment of cattle before they are brought to market, trypanosomiasis in domestic animals has not been effectively brought under control (Waiswa *et al.*, 2003) and the movement of livestock continues to play a significant role in the spread of HAT in Uganda (Batchelor *et al.*, 2009; Wardrop *et al.*, 2010).

Animal reservoirs have long been recognised to play an important role in East African HAT. Prior to the identification of the *SRA* gene the only definitive way to classify animal isolates was through human serum resistance assays. These tests measure the ability of isolates to survive following exposure to human serum, with survival measured through the ability to successfully infect laboratory animals or through direct *in vitro* observation of cell death via the BIIT (Rickman and Robson, 1970a, 1970b; Targett and Wilson, 1973). More modern tests utilise fluorescent dyes to track the disruption of cellular membranes by human serum, allowing for estimation of rate of cell death following exposure (Turner *et al.*, 2004). Waiswa *et al.*(2003) examined a range of domestic

animals in three districts of Uganda; Kamuli, Mukono and Tororo. Across these three districts 41-46% of cattle were found to be infected with *T. brucei*, while 83-85% of pigs in Kamuli and Mukono districts were infected. No pigs were found to be infected with *T. brucei* in Tororo. Of the 185 samples investigated for human serum resistance 30% in both cattle and pigs were classified as *T. b. rhodesiense* based on this test, placing total *T. b. rhodesiense* prevalence at 13% and 25% respectively. The Busia district of Kenya, which neighbours the Tororo district and forms part of the Tororo focus also continues to maintain an animal reservoir of human infective trypanosomes, although at a lower prevalence of only 1% (von Wissmann *et al.*, 2011).

The discovery of *SRA*, located within a VSG expression site, suggests that the reported prevalence of *T. b. rhodesiense* from studies relying upon human serum resistance may be an underestimation of the true levels, as isolates carrying but not expressing the gene will have been classified as *T. b. brucei* as opposed to *T. b. rhodesiense*. Studies utilising the presence of *SRA* as a means to classify isolates have suggested the prevalence of *T. b. rhodesiense* to be 11-18% in randomly sampled cattle (Welburn *et al.*, 2001; Enyaru *et al.*, 2006). Enyaru *et al.*(2006), however, when analysing human serum resistant samples predominantly isolated from humans showed that *SRA* failed to amplify in 20% of the samples despite the use of two independent primer sets. While this may indicate the existence of *SRA* negative, *T. b. rhodesiense* like human infective isolates, the *SRA* gene has been shown to be widespread amongst isolates of *T. b. rhodesiense* from across Africa (Gibson *et al.*, 2002; Njiru *et al.*, 2004a). Enyaru *et al.*(2006) suggested that amplification failure may have occurred due to low parasitaemia in these samples or the presence of divergent *SRA* sequences. To date only two variants of *SRA* have been reported, differing from one another at three bases, with geographical partition of the types (Gibson *et al.*, 2002; MacLean *et al.*, 2004; Balmer *et al.*, 2011). At present Tanzania is the only country where the two types have been observed to coexist (Balmer *et al.*, 2011). Failure with the two independent primer sets used by Enyaru *et al.*(2006) would therefore require a higher level of diversity than has been previously observed in the gene, reducing the likelihood of this explanation being correct.

Due to the timing of the most recent epidemic in Uganda, arising at the same time that the techniques for isolation and characterisation of trypanosomes became widely implemented a large number of collections have originated from within the Ugandan foci, collected primarily from humans and cattle. Isoenzymes, RFLP, mini- and microsatellites have all been used in combination with human serum resistance tests in order to determine the relationships between isolates. Isolates of *T. b. rhodesiense* have been shown to be genetically distinct from those found in other parts of East Africa, supporting geographical sub-structuring and limited gene flow across large regions (MacLeod *et al.*, 2000, 2001b; Balmer *et al.*, 2011). A study utilising isolates of *T. brucei* from Kiboko in Kenya found a small proportion to be genetically similar to *T. b. rhodesiense* from Tororo, 500km away. Most striking was that at the time of sampling no cases of HAT had ever

been reported in Kiboko (Hide *et al.*, 2000).Within focus studies have produced a range of results with regards the level of differentiation between isolates from different districts. Hide *et al.*(1998), utilising RFLP found that isolates of *T. b. rhodesiense* from the neighbouring Busia (Kenya) and Tororo (Uganda) districts were genetically similar despite the districts being endemic and epidemic respectively and with observable differences in disease progression. In contrast studies by Enyaru *et al.*(1993),using isoenzyme markers, and MacLean *et al.*(2007), using microsatellite markers, comparing Busoga / Tororo and Tororo / Soroti respectively found isolates from the districts to be genetically distinct from one another. This would suggest that there is only limited gene flow between these foci, despite the relatively small distance between them and high likelihood that the Soroti focus is an offshoot of that in Tororo (Fèvre *et al.*, 2001; Welburn *et al.*, 2001).

As many samples from Uganda have been isolated from animals in addition to human patients it has been possible to examine the relationship between *T. b. brucei* and *T. b. rhodesiense* isolated from the same region. Intriguingly, and despite the known difference of only a single gene between the two sub-species many of these studies have found *T. b. rhodesiense* to be both genetically distinct from, and less genetically diverse than sympatric isolates of *T. b. brucei* (Hide *et al.*, 1994, 1998; MacLeod *et al.*, 2000, 2001b; Agbo *et al.*, 2003). These results would appear to suggest that *T. b. rhodesiense* may be genetically isolated from *T. b. brucei* even when found within non-human hosts that may be infected by either of the sub-species. Balmer *et al.*(2011) however, in studying isolates from all *T. brucei* subgroups, demonstrated that *T. b. rhodesiense* were often more related to *T. b. brucei* isolates than other isolates of *T. b. rhodesiense*. Together these studies may indicate that while differences can exist between sympatric populations of *T. b. brucei* and *T. b. rhodesiense* the differences are negligible when considering the continent as a whole.

While the genetic diversity and epidemiology of HAT has been extensively studied in Uganda the role of genetic exchange in natural populations of *T. brucei* remains controversial, as discussed in Chapter 1. This is in part due to the presence of the multiple sub-species which can distort the analysis of populations if grouped or separated incorrectly, as well as the non-obligatory nature of genetic exchange in *T. brucei*. The analysis of samples from Uganda has in part contributed to this controversy. Stevens and Welburn (1993), utilising 10 isoenzyme markers examined 44 isolates of *T. brucei* from Tororo district and concluded that there was a lack of regular genetic exchange occurring after grouping human serum sensitive and resistant isolates together. Agbo *et al.*(2003) in contrast, concluded that there was sufficient heterogeneity to exclude clonality, again grouping isolates that were sensitive and resistant to lysis by human serum. Based upon the genetic differentiation seen between human serum sensitive and resistant isolates a number of studies have analysed these grouped separately in order to avoid the potential bias introduced by mixing the distinct populations (Hide *et al.*, 1994; MacLeod *et al.*, 2000). These studies examined the role of population structure in the different sub-species and the potential role of clonal expansion of

individual genotypes in masking underlying genetic exchange within a population, giving rise to the epidemic population structure as defined by Maynard-Smith (1993). The results of these studies suggests that while genetic exchange may be a frequent occurrence amongst *T. b. brucei* the populations of *T. b. rhodesiense* appear to be either clonal (Hide *et al.*, 1994; MacLeod *et al.*, 2000) or epidemic and dominated by the clonal expansion of a small number of genotypes (Stevens and Tibayrenc, 1996).

In contrast to Uganda the HAT foci of Malawi, approximately 1700 km to the south, has been the subject of relatively little study. HAT in Malawi is associated with a milder and more chronic disease profile than that of Uganda (MacLean *et al.*, 2004, 2010) with distinct cytokine profiles in the two countries (MacLean *et al.*, 2004, 2007, 2010) and the observation of anaemia in a high proportion of patients in Malawi (Chisi *et al.*, 2004). The majority of cases within Malawi are associated with close proximity to game reserves and it is likely that a high proportion of cases go unreported in the country (Gondwe *et al.*, 2009; Madanitsa *et al.*, 2009).

In order to further explore the epidemiology, diversity and population structure of *T. b. rhodesiense* seven microsatellite markers were employed in order to analyse parasite samples from four different foci of disease, three in Uganda and one from Malawi. The samples from Uganda were collected from Tororo, Soroti and Kaberamaido districts and have been analysed as four separate populations, Tororo pre-1991, Tororo 2003, Soroti 2003 and Kaberamaido 2009 based upon their geography and date of origin. The population from Malawi was collected from Nkhotakota General Hospital between 2002 and 2003. These samples have therefore allowed for the comparative geographical analysis of the three populations collected in 2003 in addition to allowing us to track the genetic changes occurring in Uganda as the epidemic spread from Tororo district into Soroti and then Kaberamaido districts.

5.2 Results

Whereas Chapters 3 and 4 focused upon the population genetics of trypanosomes responsible for AAT this chapter aims to examine the population diversity and genetics of the human infective *T. b. rhodesiense* sub-species through the use of five populations. Four of these populations are from Uganda and originate from three geographically distinct foci, Tororo, Soroti and Kaberamaido. The Tororo pre-1991 population, comprising 52 samples, was collected from humans and cattle in the wider Tororo focus, including regions of western Kenya, between 1961 and 1990. All except four of the samples in this population were collected between 1988 and 1990. The samples collected from cattle had been previously identified as *T. b. rhodesiense* based upon their ability to resist lysis by human serum. During the 1990s a resurgence of HAT in Uganda led to Tororo seeding an offshoot focus in the Soroti district, linked to the movement of infected cattle. Thirty and 84 samples were collected in 2003 from human patients in Tororo and Soroti respectively. The final Ugandan population is composed of 86 samples, collected from human patients in 2009 and 2010, from the Kaberamaido district. Kaberamaido, alongside Kumi and Lira districts are amongst the most recent to have become afflicted by HAT. The shared history of these four Ugandan populations thus comprise a unique case study, allowing for examination of the progress of HAT as it spread across the country from Tororo into first Soroti and then Kaberamaido. The final population from Malawi consists of 28 samples from the Nkhotakota district focus, collected from human patients in 2003. This focus is endemic, ongoing, and in contrast to the relatively severe and acute disease observed in Uganda, is characterised by slower progression to the late stage (MacLean *et al.*, 2004, 2007, 2010).

Comparative analysis of these five populations allows for the examination of the role of both space and time in shaping the population genetics of *T. b. rhodesiense*, achieved here through the use of polymorphic, single locus microsatellite markers. These seven markers have been previously described and map to six of the megabase chromosomes of the TREU 927 *T. brucei* genome reference strain. Genotyping of the 280 samples generated complete MLGs for 214 isolates with 78 unique MLGs across the five populations. Only three samples, K237, UgE90 and liri017, were identified as possessing mixed parasite genotypes based upon the presence of three microsatellite alleles for at least one of the seven investigated loci. These samples were excluded from subsequent analysis. Partial genotypes, consisting of a minimum of four successfully genotyped loci were determined for 63 of the isolates (Appendix 11). Data from partially genotyped isolates were retained for use in single locus analysis (tests for agreement with Hardy-Weinberg expectations, per locus statistics) and pairwise locus analysis (tests for linkage disequilibrium). These samples were excluded in tests of unique MLGs or where the complete MLG was required (F_{ST} and genetic distance; dendrograms of similarity; PCoA). For genotype data see Appendix 11.

5.2.1 Geographic sub-structuring of *T. b. rhodesiense*

In order to determine if the populations were sub-structured due to geographical isolation the three populations collected during the 2002-03 time period were compared. The Tororo pre-1991 and Kaberamaido 2009 populations were excluded from the analysis in order to avoid potential sub-structuring arising from temporal variation between populations. Private alleles were determined by comparing the populations and identifying alleles unique to either a single population (Tororo / Soroti / Malawi) or country (Uganda / Malawi). There was a higher number of private alleles in the Malawi population (nine) compared to five in Tororo and three in Soroti. Of the private alleles in Malawi five were present at frequencies above 0.1 within the population, whereas only 1 private allele was above this frequency in Tororo and none were observed above this level in Soroti. Grouping Tororo and Soroti together indicated a total of 16 alleles that were unique to Uganda as a whole. Nei's unbiased genetic distance (D) and pairwise population F_{ST} assess the differences between populations based upon differences in the allele frequencies of the respective populations. Values for both range from 0 to 1, with lower values indicating increasing similarity between the two populations. By these two measures the Ugandan populations (Tororo 2003 and Soroti 2003) are closely related (Table 5.1). The Malawi population shows substantial genetic differentiation from both Ugandan populations by both of these measures (Table 5.1).

	Tororo 2003	Soroti 2003	Malawi 2003
Tororo 2003	-	0.109	0.226
Soroti 2003	0.129	-	0.266
Malawi 2003	0.669	0.680	-

Table 5.1 F_{ST} and Nei's genetic distance between Uganda and Malawi populations

Pairwise values of Wright's fixation index (F_{ST}; above diagonal) and Nei's genetic distance (*D*; below diagonal) between subpopulations of *T. b. rhodesiense* from Uganda and Malawi sampled in 2002-2003.

In order to examine the relationship between each isolate from 2003 a UPGMA dendrogram of similarity (Figure 5.1) was constructed using Jaccard's similarity and rooted against MALa1, an isolate of *T. b. gambiense* Group 1 from a focus in the Democratic Republic of the Congo (Morrison *et al.*, 2008b). The tree further supports the significant separation of the Malawi population from those in Uganda (94% bootstrap support),however the Ugandan populations could not be significantly resolved from one another (Fig. 5.2), as expected given the close genetic and historical relationships. The dendrogram highlights the different levels of clonal expansion in each population, with Malawi characterised by only two repeated MLGs representing only four isolates while in Soroti only 16 unique MLGs were observed, one of which was represented 50 times. Plotting of Principal Co-ordinates 1 and 2 following PCoA of the MLG dataset (Figure. 5.2) further highlights the separation of the Malawi population. Principal coordinate 1, accounting for 73% of the variation observed, appears to primarily separate the populations based on country of origin, while coordinate 2, which accounts for 11% of the variation, partially separates the two Ugandan populations in addition to highlighting the diversity within the Malawi focus. This method of analysis indicates that while the Tororo and Soroti foci were, by the measures tested here, closely related in 2003 the two populations could be genetically distinguished from one another. The close relationship however results in genetic overlap between the two populations, with a number of Ugandan samples that could not be allocated to their population of origin based upon genotype alone. All of these data combine to demonstrate that there is significant genetic differentiation between the Malawian and Ugandan *T. b. rhodesiense* isolates, indicating the presence of sub-structuring between spatially separated populations.

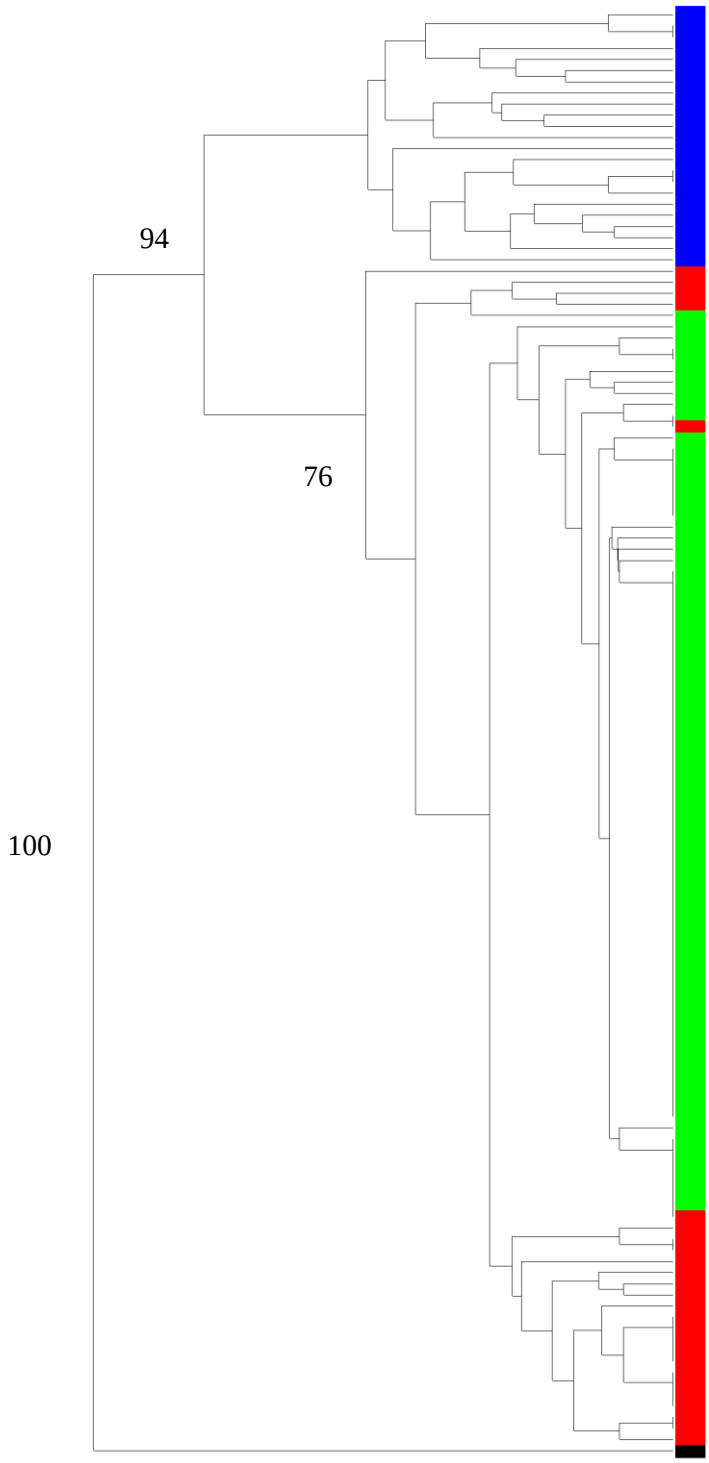


Figure 5.1 UPGMA dendrogram of similarity *T. b. rhodesiense* isolates from 2003

UPGMA dendrogram of similarity of isolates from 2003 constructed using Jaccard's similarity and rooted against MALa1, an isolate of *T. b. gambiense* Group 1 from a focus in the Democratic Republic of the Congo (Black in figure). Bootstrap values are given for the main nodes only. The Malawi population could be separated with high bootstrap support (94%) from those in Uganda while within Uganda the two populations could not be resolved.

Populations: Malawi = blue, Soroti = green, Tororo = red, MALa1 = black.

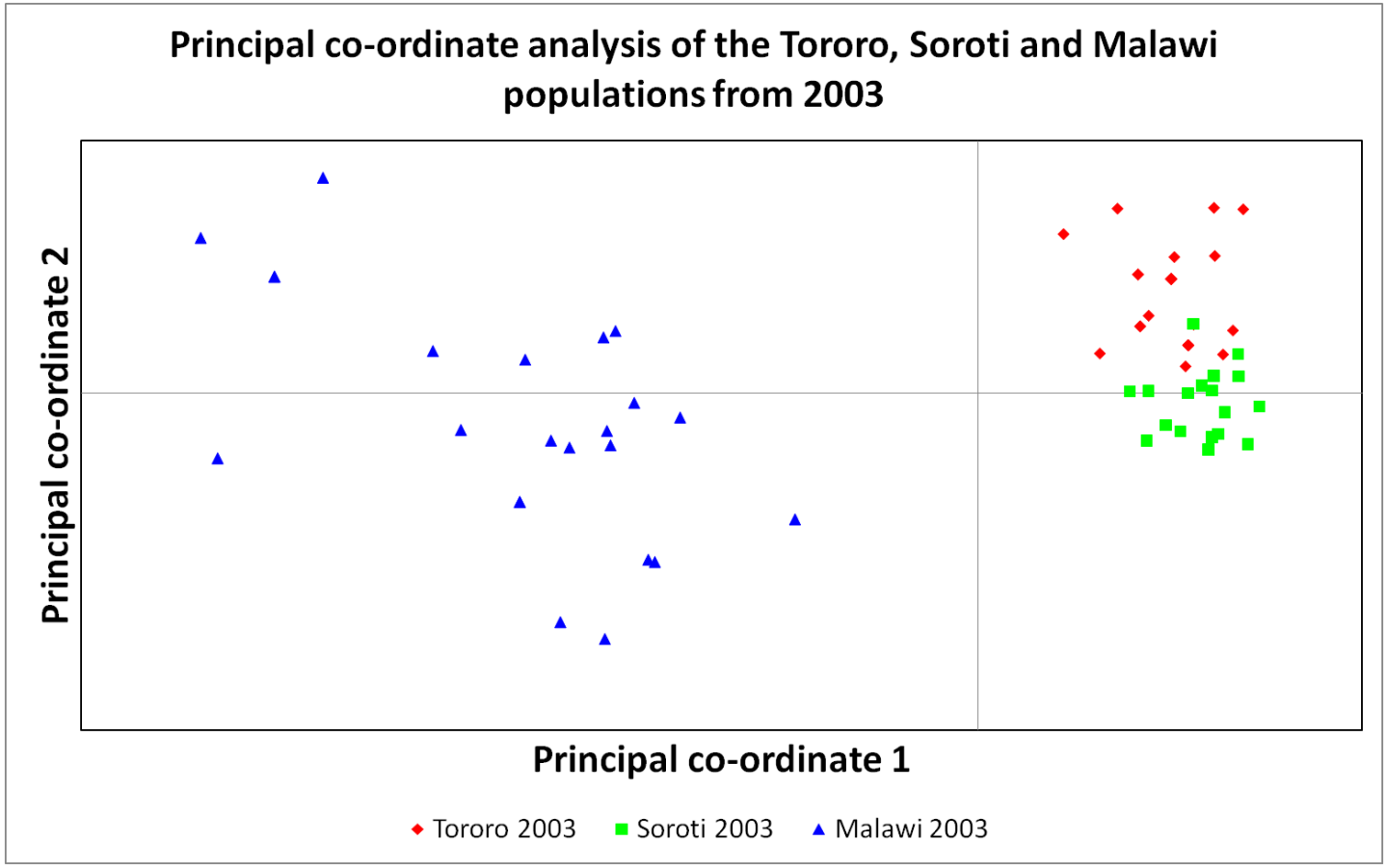


Figure 5.2 PCoA of *T. b. rhodesiense* isolates from 2003

PCoA of isolates collected in 2003. Principal co-ordinate 1 explains 73% of the variation observed and separates the Malawi population from those in Uganda. Principal co-ordinate 2 accounts for 11% of the total variation, partially separating the two Ugandan populations, in addition to highlighting the diversity within Malawi.

5.2.2 Genetic exchange in *T. b. rhodesiense*

5.2.2.1 The Tororo focus between 1961 and 1990 shows no evidence of mating

The Tororo focus lies at the heart of the *T. b. rhodesiense* outbreaks that have occurred in South East Uganda, and is an ongoing cause of HAT in the region. The sample collection from this population, comprising 52 genotyped samples from cattle and humans were collected during passive sampling of HAT and Nagana cases between 1961 and 1990. All except four samples were collected between 1988 and 1990. Two samples were identified as containing multiple infections and removed from the analysis, while 43 of the remaining 50 samples were fully genotyped with the seven loci utilised in this study. The seven partially genotyped samples were successfully genotyped at a minimum of four of the seven loci. While no single MLG dominates the population the products of clonal expansion are clearly visible, with 32 isolates, 74% of those fully genotyped, represented by just eight different MLGs. Of the four samples isolated prior to 1988 only one, UTAR 3, possessed a unique MLG which was not observed during the 1988 – 1990 period.

Multiple alleles were observed for six of the seven loci, however marker Ch4/M12C12, while polyallelic in other populations, was observed to be monoallelic in this population, rendering it uninformative and it was therefore excluded from subsequent analysis in this population. Of the remaining six loci two were observed to possess fewer heterozygotes than expected, two more than expected and two close to expectations (Table 5.2A), resulting in a slight heterozygote deficit on average. In order to examine the extent of genetic exchange in the Tororo focus, genotype frequency was examined for each locus and compared to that expected in a randomly mating population in agreement with Hardy-Weinberg expectations. Of the six polyallelic loci examined all except one was observed to display significant disagreement with Hardy-Weinberg predictions (Table 5.3). Ch3/5L5, the one locus in agreement with Hardy-Weinberg predictions, had low polymorphism, with only two alleles present at frequencies of 0.9 and 0.1 respectively. For this marker the low level of polymorphism, extremes of allele frequencies and population size limits the meaningful use of Hardy-Weinberg equilibrium as a measure of genetic exchange in the population as significant deviation is difficult to detect under these circumstances, an issue that was encountered repeatedly in the analysis of these populations. To investigate if there was evidence for recombination between loci, the alleles at the different loci were examined for agreement with linkage equilibrium in all pairwise combinations. At loci where significant deviation from Hardy-Weinberg equilibrium was observed, allele combinations were preserved, effectively treating the two alleles present in an individual as a single allele and allowing for tests of linkage disequilibrium that are not affected by deviation from Hardy-Weinberg expectations. Significant linkage disequilibrium was observed at 12 of the 15 combinations (Table 5.4A). These results agree

with the deviation from Hardy-Weinberg equilibrium in indicating a significant departure from panmixia.

While initial analysis of the population would suggest a lack of mating, the non-obligatory nature of mating in *T. brucei* allows for the possibility of an epidemic population structure, with the underlying mating masked by clonal expansion of one or more clones that then dominate the population (Maynard Smith *et al.*, 1993). The analysis was therefore repeated by removing replicate genotypes from the population, leaving 19 unique MLGs. Removal of the repeated genotypes reduced the observed heterozygote deficits and excesses with subsequent shift towards zero for the population average F_{IS} (Table 5.2A). After removal of the repeated genotypes four of the six polymorphic loci were in agreement with Hardy-Weinberg expectations (Table 5.3). However of these four loci two, Ch3/5L5 and Ch1/18, were dominated by a single allele at frequencies of 0.84 and 0.92 respectively while significant linkage disequilibrium was observed for 8 of the 15 loci combinations. Due to the presence of only 19 unique MLGs, and the dominance of single alleles at two of the six loci, care must be taken when interpreting these results due to the increased risk of a Type 2 error, which would mask deviation from Hardy-Weinberg expectations or the presence of significant linkage disequilibrium due to insufficient power during the analysis.

Taken together the evidence from both the entire population and unique genotype both suggest a lack of genetic exchange within the Tororo foci at this point in time, which is consistent with previous studies of the population utilising different types of genetic markers.

Table 5.2A-E Per population allele statistics

A, Tororo pre-1991

Locus	n*	A	He	Ho	F_{IS}
Ch3/5L5	50 / 19	2 / 2	0.18 / 0.31	0.16 / 0.37	0.12 / -0.20
Ch4/M12C12	49 / 19	1 / 1	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00
Ch2/PLC	48 / 19	5 / 5	0.42 / 0.58	0.35 / 0.47	0.15 / 0.18
Ch5/JS2	46 / 19	4 / 3	0.54 / 0.52	0.28 / 0.32	0.47 / 0.40
Ch1/18	50 / 19	5 / 3	0.13 / 0.20	0.12 / 0.21	0.11 / -0.07
Ch9/4	47 / 19	4 / 4	0.58 / 0.62	0.96 / 0.95	-0.66 / -0.54
Ch3/IJ15/1	50 / 19	7 / 7	0.60 / 0.68	0.94 / 0.95	-0.57 / -0.40
Population Average	49 / 19	4.00 / 3.57	0.35 / 0.42	0.40 / 0.47	-0.15 / -0.12

B, Tororo 2003

Locus	n*	A	He	Ho	F_{IS}
Ch3/5L5	29 / 17	2 / 2	0.49 / 0.47	0.83 / 0.71	-0.70 / -0.52
Ch4/M12C12	29 / 17	4 / 3	0.55 / 0.53	0.93 / 0.88	-0.71 / -0.69
Ch2/PLC	29 / 17	3 / 3	0.16 / 0.27	0.17 / 0.29	-0.06 / -0.11
Ch5/JS2	29 / 17	3 / 3	0.43 / 0.41	0.59 / 0.53	-0.37 / -0.29
Ch1/18	29 / 17	2 / 2	0.46 / 0.45	0.69 / 0.65	-0.51 / -0.45
Ch9/4	28 / 17	5 / 4	0.59 / 0.61	1.00 / 1.00	-0.72 / -0.66
Ch3/IJ15/1	26 / 17	4 / 4	0.54 / 0.57	0.77 / 0.65	-0.42 / -0.15
Population Average	28 / 17	3.29 / 3.00	0.46 / 0.47	0.71 / 0.67	-0.56 / -0.44

C, Soroti 2003

Locus	n*	A	He	Ho	F_{IS}
Ch3/5L5	84 / 16	2 / 2	0.50 / 0.52	1.00 / 1.00	-1.00 / -1.00
Ch4/M12C12	84 / 16	3 / 3	0.14 / 0.23	0.15 / 0.25	-0.07 / -0.08
Ch2/PLC	83 / 16	4 / 4	0.52 / 0.58	0.90 / 0.88	-0.75 / -0.54
Ch5/JS2	83 / 16	4 / 4	0.06 / 0.24	0.06 / 0.25	-0.02 / -0.06
Ch1/18	82 / 16	2 / 2	0.50 / 0.51	0.98 / 0.88	-0.95 / -0.76
Ch9/4	84 / 16	5 / 5	0.52 / 0.59	0.98 / 0.88	-0.89 / -0.50
Ch3/IJ15/1	84 / 16	2 / 2	0.01 / 0.06	0.01 / 0.06	0.00 / 0.00
Population Average	83 / 16	3.14 / 3.14	0.32 / 0.39	0.58 / 0.60	-0.81 / -0.56

D†, Kaberamaido 2009

Locus	n*	A	He	Ho	F_{IS}
Ch3/5L5	58	2	0.03	0.03	-0.01
Ch4/M12C12	83	1	0.00	0.00	0.00
Ch2/PLC	85	2	0.23	0.00	1.00
Ch5/JS2	85	1	0.00	0.00	0.00
Ch1/18	86	2	0.50	1.00	-1.00
Ch9/4	54	2	0.50	1.00	-1.00
Ch3/IJ15/1	86	2	0.08	0.08	-0.04
Population Average	77	1.71	0.16	0.30	-0.86

E, Malawi 2003

Locus	n*	A	He	Ho	F_{IS}
Ch3/5L5	28 / 21	2 / 2	0.36 / 0.37	0.46 / 0.48	-0.29 / -0.29
Ch4/M12C12	28 / 21	2 / 2	0.47 / 0.46	0.71 / 0.67	-0.54 / -0.48
Ch2/PLC	28 / 21	4 / 4	0.20 / 0.22	0.21 / 0.24	-0.06 / -0.06
Ch5/JS2	28 / 21	3 / 3	0.49 / 0.45	0.18 / 0.14	0.64 / 0.69
Ch1/18	28 / 21	3 / 3	0.20 / 0.22	0.21 / 0.24	-0.08 / -0.09
Ch9/4	24 / 21	3 / 3	0.53 / 0.53	0.42 / 0.43	0.22 / 0.19
Ch3/IJ15/1	26 / 21	4 / 4	0.70 / 0.68	0.58 / 0.57	0.17 / 0.17
Population Average	27 / 21	3.00 / 3.00	0.42 / 0.42	0.40 / 0.39	0.06 / 0.06

Table 5.2A-E Per population allele statistics

Samples numbers, observed allele numbers, observed and expected heterozygosities and inbreeding coefficient for 'all samples/unique MLGs' respectively at each locus and population average per population. Populations: A = Tororo pre-1991, B = Tororo 2003, C = Soroti 2003, D = Kaberamaido 2009, E = Malawi 2003.

† Only four fully genotyped and unique MLGs were observed within the Kaberamaido 2009 population, data are therefore shown for 'all samples' only.

*n = Mean sample number across all loci after accounting for missing data, A = mean allele number per locus, He = Expected heterozygosity, Ho = Observed heterozygosity, F_{IS} = Inbreeding coefficient.

	Tororo pre-1991	Tororo 2003	Soroti 2003	Kaberamaido 2003*	Malawi 2003
Ch3/5L5	0.40 / 1.00	**0.00** / 0.06	**0.00** / **0.00**	1.00	0.29 / 0.29
Ch4/M12C12	1.00 / 1.00	**0.00** / **0.00**	1.00 / 1.00	1.00	**0.00** / 0.05
Ch2/PLC	**0.00** / 0.07	1.00 / 1.00	**0.00** / **0.03**	**0.00**	1.00 / 1.00
Ch5/JS2	**0.00/** 0.10	0.08 / 0.64	1.00 / 1.00	1.00	**0.00** / **0.00**
Ch1/18	**0.03** / 1.00	**0.01** / 0.11	**0.00** / **0.01**	**0.00**	1.00 / 1.00
Ch9/4	**0.00** / **0.01**	**0.00** / **0.00**	**0.00** / **0.04**	**0.00**	**0.02** / **0.04**
Ch3/IJ15/1	**0.00** / **0.01**	**0.00** / **0.04**	1.00 / 1.00	1.00	0.06 / 0.15

Table 5.3 Probability of agreement with Hardy-Weinberg expectations

Probability of agreement with Hardy Weinberg predictions (Data are shown for 'all samples/unique MLGs, respectively). $P<0.05$ = Significant disagreement, shown in bold. * Only four fully genotyped and unique MLGs were observed within the Kaberamaido 2009 population, data are therefore shown for 'all samples' only.

5.2.2.2 The Tororo focus in 2003 shows no evidence for mating

The Tororo 2003 population, collected 13 years after the initial sampling period of the focus provided a unique opportunity to revisit the outbreak and assess whether genetic exchange between trypanosomes had become a frequent occurrence or whether clonality still dominated the population structure. The Tororo 2003 population consists of 30 samples collected from the region between 2002 and 2003, 27 of which were genotyped at all 7 loci. One mixed infection (isolate liri017) was identified in the population and removed from the analysis. An excess of heterozygosity was observed at six of the seven loci (F_{IS} = -0.56, Table 5.2B) and four of the seven loci were represented with fewer alleles than in the pre-1991 population (Table 5.2B). Eleven alleles present in the pre-1991 population were absent in 2003, although only one of these had been present at a frequency of above 0.1 while six new alleles, absent in the pre-1991 population, were present in 2003 with two at frequencies of above 0.1. As with the Tororo pre-1991 population clonal expansion of genotypes is visible, with four MLGs representing 13 of the 27 fully genotyped isolates.

Significant deviation from Hardy-Weinberg equilibrium was detected at five of the loci (Table 5.3). Ch2/PLC and Ch5/JS2 were found to be in agreement with Hardy-Weinberg expectations. However, both of these loci were observed to be dominated by a high frequency of a single allele. While three alleles were observed at Ch2/PLC one was observed at a frequency of 0.9, dominating the population by its presence in each sample and preventing meaningful analysis of Hardy-Weinberg equilibrium as previously discussed. Ch5/JS2 was also found to possess three alleles with the dominant allele present at a frequency of 0.71 and the second allele at a frequency of 0.28. While in agreement with Hardy-Weinberg predictions the dominant allele was observed to be present in each individual and was the only allele found in homozygotes. The absence of homozygotes for the remaining alleles may arise from the low frequency with which they are expected to occur in a population of this size. To examine the evidence for recombination, linkage disequilibrium between alleles at all pairwise combinations was examined, revealing that only 1 of the 21 loci combinations were in linkage disequilibrium (Table 5.4B) after preserving allele combinations at loci with significant disagreement from Hardy-Weinberg predictions. While this would superficially appear to be evidence for recombination within the population, these results arise from the relatively low level of variation across each of the loci, with the majority dominated by a single homo- or heterozygote genotype. This limited genetic variation hindered statistical analysis. There remained an excess of heterozygotes following removal of duplicate genotypes (F_{IS} = -0.44) while two additional loci, Ch1/18 and Ch3/5L5 show agreement with Hardy-Weinberg predictions (Table 5.3). This may, however, be a Type 2 error as after removal of the repeated MLGs only 17 individuals remain in the population. No change is observed on the number of loci

combinations in linkage disequilibrium upon removal of the replicate MLGs (Table 5.4B). These data are consistent with little or no sexual recombination in this population.

5.2.2.3 The population of the Soroti focus in 2003 is homogeneous, with extensive clonal expansion of a single genotype

The Soroti focus is a relatively new focus of HAT, in contrast to foci in Tororo and Malawi. Human cases of trypanosomiasis in the Soroti district were first reported in December of 1998 and the focus is likely to be an offshoot of the Tororo epidemic (Fèvre *et al.*, 2001). While 84 samples were genotyped for the population the majority of these samples represent replicate MLGs with only 16 unique MLGs identified in the population. A single genotype (MLG 49), dominates the population, sampled 50 times in total with the extent of this clonal expansion clearly visible in the dendrogram plots (Figures 5.1 and 5.3).

This population is dominated by high levels of both homozygosity and heterozygosity. Four of the loci (Ch3/5L5, Ch2/PLC, Ch1/18 and Ch9/4) are represented almost exclusively by a single heterozygote pair of alleles, resulting in highly negative inbreeding coefficients (Table 5.2C) and significant disagreement with Hardy-Weinberg predictions (Table 5.3). While multiple alleles were observed at the remaining three loci (Ch4/M12C12, Ch5/JS2 and Ch3/IJ15/1) (Table 5.2C) all possess one allele present at a frequency of over 0.9 and are observed to be in Hardy-Weinberg equilibrium. The low level of variation observed within this population, with each locus predominantly represented by a single homozygote or heterozygote genotype hampered meaningful analysis of linkage equilibrium (Table 5.4C). Removal of repeated genotypes had little effect on the deviation from Hardy-Weinberg predictions, in addition to substantially reducing the power of this analysis to only 16 MLGs. Overall the Soroti population, with little variation between individuals and a high number of repeated genotypes indicates no, or an undetectable level, of mating occurring.

5.2.2.4 The Kaberamaido focus in 2009 is homogeneous and dominated by a single genotype

The Kaberamaido population, like Soroti in 2003, represents a recent expansion of the HAT foci in Uganda. The sample collection consists of 86 samples, of which only 42 were fully genotyped. Missing data at two loci, Ch3/5L5 and Ch9/4 are responsible for the majority of the partially genotyped samples. The allele frequencies observed in Kaberamaido closely resemble the patterns observed in Soroti in 2003, with excesses of both homozygosity and heterozygosity. Ch4/M12C12 and Ch5/JS2 are completely monoallelic within this population while Ch3/5L5, Ch2/PLC and Ch3/IJ15/1 are all dominated by a single allele at a frequency above 0.9, resulting in inbreeding coefficients of close to zero (Table 5.2D). While these loci display agreement with Hardy-

Weinberg predictions, the high frequency of a single allele in each again renders this test susceptible to a Type 2 error (Table 5.3). The remaining two loci, Ch1/18 and Ch9/4 are both entirely represented by a single heterozygote pair giving rise to inbreeding coefficients of -1.00 (Table 5.2D) and significant disagreement with Hardy-Weinberg predictions (Table 5.3). As with the Soroti population the allele frequencies observed prevent meaningful analysis of linkage equilibrium between the loci (Table 5.4D). Removal of repeated and partially genotyped samples leaves only four unique MLGs, preventing further analysis. Of these MLG 78 dominates the population, representing 38 of the 42 fully genotyped samples (Figure 5.3).

While 86 samples were genotyped within this population there was a high frequency of missing data at two loci, with 28 samples missing data at Ch3/5L5 and 32 missing data at Ch9/4. In contrast only five samples are missing data from any of the other loci. These missing data could be explained by either failure to amplify known alleles at these loci or by the presence of null alleles within the population. The standard statistical methods for the estimation of null allele frequency, such as those of Dempster *et al.*(1977), Chakraborty *et al.*(1992) and Brookfield (1996) rely upon the assumption of panmixia within the population and that deviation from Hardy-Weinberg equilibrium at individual loci is only due to the presence of the undetected null alleles. The data from the five unaffected loci suggests however that this population is highly clonal with little or no mating occurring, preventing estimation of null allele frequency by these methods.

5.2.2.5 The population structure of the Malawi focus in 2003 indicates the presence of frequent mating

The Malawi sample collection, genetically distinct from those in Uganda (Table 5.1), comprises 28 individual samples, with 23 fully genotyped for all seven markers. Twenty-one of the 23 MLGs observed are unique within the population and none of the MLGs are shared with any of the Ugandan populations. Between two and four alleles were observed at each locus, with a close to expected number of heterozygotes when examining the population as a whole (Table 5.2E). Two loci, Ch4/M12C12 and Ch5/JS2, deviate considerably with a large heterozygote excess and deficit respectively. Examination of the markers for deviation from Hardy-Weinberg expectations revealed three loci (Ch4/M12C12, Ch5/JS2 and Ch9/4) that deviate significantly from predictions (Table 5.3). Disagreement at Ch4/M12C12 and Ch5/JS2 results from heterozygote and homozygote excesses, respectively. For marker Ch9/4 the disagreement arises from the presence of a single individual (sample NKK/T/026) homozygous for a rare allele, which was observed only in this sample. Loci Ch2/PLC and Ch1/18 are dominated by single alleles within the population, thus accounting for the complete agreement at these loci as previously described. After accounting for deviation from Hardy-Weinberg equilibrium by preserving genotypes at these loci only 2 out of 21 loci combinations were observed to show significant evidence of linkage disequilibrium (Table

5.4E). While there was a total of only two repeated genotypes within the Malawi population their removal resulted in Ch4/M12/C12 moving to agreement with Hardy-Weinberg predictions, while significant linkage disequilibrium was observed between only a single pair of loci, Ch3/IJ15/1 – Ch9/4.

Despite being the smallest of the five populations, the Malawi 2003 population is represented by the greatest number of unique MLGs with clonal expansion of two of these MLGs representing a total of only four isolates. The population is therefore clearly different from those in Uganda where clonal expansion has come to dominate the populations examined in this Chapter. The population is further differentiated by its agreement with Hardy-Weinberg expectations (Table 5.3) and low level of linkage disequilibrium (Table 5.4E). Coupled to the unique number of genotypes this evidence is consistent with the presence of frequent mating within the *T. b. rhodesiense* population in Malawi. Due to the relatively low sample size of 23 it is not possible to robustly conclude the presence of panmixia within the population and further studies will be required in order to confirm these results and determine why the *T. b. rhodesiense* populations of Uganda and Malawi differ so much.

Table 5.4A-E Per population linkage disequilibrium between pairwise loci

A†, Tororo pre-1991

	Ch3/5L5	Ch4/M12C12	Ch2/PLC	Ch5/JS2	Ch1/18	Ch9/4	Ch3/IJ15/1
Ch3/5L5	-	0.41 / 1.00	**0.00** / 0.06	0.26 / 0.10	**0.00** / 0.08	**0.00** / 0.12	**0.00** / 0.29
Ch4/M12C12	-	-	1.00 / 0.07	1.00 / 0.09	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00
Ch2/PLC	-	-	-	**0.03** / **0.03**	**0.01** / **0.02**	**0.03** / **0.03**	0.09 / 0.08
Ch5/JS2	-	-	-	-	0.19 / 0.13	**0.00** / **0.02**	**0.00** / **0.02**
Ch1/18	-	-	-	-	-	**0.00** / **0.00**	**0.00** / **0.00**
Ch9/4	-	-	-	-	-	-	**0.00** / **0.01**
Ch3/IJ15/1	-	-	-	-	-	-	-

B, Tororo 2003

	Ch3/5L5	Ch4/M12C12	Ch2/PLC	Ch5/JS2	Ch1/18	Ch9/4	Ch3/IJ15/1
Ch3/5L5	-	1.00 / 0.24	0.40 / 0.49	0.13 / 0.52	0.63 / 0.31	0.20 / 0.16	1.00 / 0.65
Ch4/M12C12	-	-	1.00 / 1.00	0.50 / 0.91	0.10 / 0.22	0.22 / 1.00	0.48 / 0.55
Ch2/PLC	-	-	-	0.53 / 0.92	0.28 / 0.33	1.00 / 0.81	0.58 / 0.85
Ch5/JS2	-	-	-	-	0.24 / 0.65	0.10 / 0.25	0.48 / 0.82
Ch1/18	-	-	-	-	-	0.61 / 0.37	0.87 / 0.67
Ch9/4	-	-	-	-	-	-	**0.01** / **0.04**
Ch3/IJ15/1	-	-	-	-	-	-	-

C, Soroti 2003

	Ch3/5L5	Ch4/M12C12	Ch2/PLC	Ch5/JS2	Ch1/18	Ch9/4	Ch3/IJ15/1
Ch3/5L5	-	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00
Ch4/M12C12	-	-	0.37 / 0.60	0.77 / 1.00	1.00 / 1.00	0.51 / 0.75	1.00 / 1.00
Ch2/PLC	-	-	-	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00
Ch5/JS2	-	-	-	-	0.11 / 0.39	**0.04** / 0.41	1.00 / 1.00
Ch1/18	-	-	-	-	-	0.07 / 0.35	1.00 / 1.00
Ch9/4	-	-	-	-	-	-	1.00 / 1.00
Ch3/IJ15/1	-	-	-	-	-	-	-

D*, Kaberamaido 2009

	Ch3/5L5	Ch4/M12C12	Ch2/PLC	Ch5/JS2	Ch1/18	Ch9/4	Ch3/IJ15/1
Ch3/5L5	-	1.00	1.00	1.00	1.00	1.00	1.00
Ch4/M12C12	-	-	1.00	1.00	1.00	1.00	1.00
Ch2/PLC	-	-	-	1.00	1.00	1.00	1.00
Ch5/JS2	-	-	-	-	1.00	1.00	1.00
Ch1/18	-	-	-	-	-	1.00	1.00
Ch9/4	-	-	-	-	-	-	1.00
Ch3/IJ15/1	-	-	-	-	-	-	-

E, Malawi 2003

	Ch3/5L5	Ch4/M12C12	Ch2/PLC	Ch5/JS2	Ch1/18	Ch9/4	Ch3/IJ15/1
Ch3/5L5	-	0.11 / 0.81	0.61 / 0.43	0.64 / 0.52	0.55 / 0.69	0.87 / 0.57	0.25 / 0.42
Ch4/M12C12	-	-	0.17 / 0.59	0.15 / 0.87	0.06 / 0.43	0.10 / 0.17	0.50 / 0.57
Ch2/PLC	-	-	-	0.79 / 0.89	0.70 / 0.70	0.29 / 0.50	0.18 / 0.24
Ch5/JS2	-	-	-	-	0.46 / 0.75	0.28 / 0.57	**0.02** / 0.06
Ch1/18	-	-	-	-	-	0.39 / 0.61	0.48 / 0.71
Ch9/4	-	-	-	-	-	-	**0.01** / **0.01**
Ch3/IJ15/1	-	-	-	-	-	-	-

Table 5.4A-E Per population linkage disequilibrium between pairwise loci

Linkage disequilibrium between loci pairs for the five populations for 'all samples/unique MLGs', respectively. Allele combinations were preserved for loci showing significant disagreement with Hardy-Weinberg predictions. $P < 0.05$ = Significant linkage disequilibrium, indicated in bold. A = Tororo pre-1991, B = Tororo 2003, C = Soroti 2003, D = Kaberamaido 2009, E = Malawi 2003. †Note that in this population locus Ch4/M12C12 is fixed for a single allele, accounting for the LD of 1.00. * Only four unique MLGs were observed within the Kaberamaido 2009 population, data are therefore shown for 'all samples' only. In addition loci Ch4/M12C12 and Ch5/JS2 were fixed for a single allele in this population accounting for the LD of 1.00.

5.2.3 Genetic variation of *T. b. rhodesiense* isolates between 1961 and 2010

Analysis of the 2003 populations indicated that whilst the Ugandan and Malawi populations were genetically and geographically distinct a low level of differentiation separated the historically related populations of Tororo and Soroti. As discussed in the introduction for this chapter, HAT in Uganda has spread since the 1970s from its historical roots in the Busoga focus outwards to envelop Tororo, Soroti and most recently Kaberamaido districts. Addition of two further study populations, Tororo pre-1991 and Kaberamaido 2009 provides respectively, an early endemic population and late, epidemic population at what is now one edge of the geographic range of *T. b. rhodesiense* in Uganda. With these populations it has been possible to examine the changing dynamics of *T. b. rhodesiense* in the country as it has spread out from Tororo district over the course of 22 years. Clonal expansion of genotypes is a common feature of the four Ugandan populations. Eight repeated MLGs were present in Tororo between 1961 - 1990 and four in Tororo in 2003, however, no single MLG was observed to dominate the respective populations. In the Tororo pre-1991 population the most frequently observed MLG represented seven isolates while in the Tororo 2003 population the most frequently observed MLG represented only four isolates. This is in sharp contrast to the Soroti 2003 and Kaberamaido 2009 populations which were both dominated by a single, high frequency MLG (MLGs 49 and 78 respectively). Only two MLGs, 29 and 31, were observed in multiple populations, representing isolates of both Tororo 2003 and Soroti 2003.

Nei's unbiased genetic distance (D) and pairwise population F_{ST} were used to comparatively assess the genetic relationship between the four populations (Table 5.5). The Tororo pre-1991 focus is likely to represent the ancestral population of the other three based upon the known spread of HAT through Uganda. The Tororo 2003 population presents a direct continuation of the foci, separated only by time. Nei's unbiased genetic distance and pairwise F_{ST} between these two populations indicated a moderate level of differentiation (Tororo pre-1991 – Tororo 2003 D = 0.411, F_{ST} = 0.201). Soroti, which is likely to have originated as an offshoot of the Tororo focus, shares a similar level of differentiation to the Tororo pre-1991 population (Tororo pre-1991 – Soroti 2003 D = 0.345, F_{ST} = 0.203) as the Tororo 2003. The similar relationships to the ancestral population is consistent with the close relationship of the Tororo and Soroti populations in 2003 (Tororo 2003 – Soroti 2003, D = 0.129, F_{ST} = 0.109).

	Tororo pre-1991	Tororo 2003	Soroti 2003	Kaberamaido 2009
Tororo pre-1991	-	0.201	0.203	0.381
Tororo 2003	0.411	-	0.109	0.159
Soroti 2003	0.345	0.129	-	0.090
Kaberamaido 2009	0.570	0.159	0.077	-

Table 5.5 F_{ST} and Nei's genetic distance between Ugandan populations

Pairwise values of Wright's fixation index (F_{ST}; above diagonal) and Nei's genetic distance (*D*; below diagonal) between subpopulations of Ugandan *T. b. rhodesiense* isolates as defined by focus and date of sample collection.

Tororo pre-1991 and Kaberamaido 2009 are the most distinct of the Ugandan populations based upon their origins and this is reflected in the high level of genetic differentiation between them (Tororo pre-1991 – Kaberamaido 2009 D = 0.570, F_{ST} = 0.381). The genetic distance between these two populations are comparable to that observed between the 2003 populations from Uganda with that of Malawi (Table 5.1), highlighting the extent to which a population within a single country may drift over time. The Kaberamaido 2009 population is much more closely related to the two populations from 2003, especially Soroti 2003, which is separated from Kaberamaido by both the shortest time period and smallest geographical distance. This would support the Kaberamaido outbreak being the most recent front of the epidemic that began back in Tororo. However, as contemporary samples are unavailable from Soroti it is not possible to formally prove that Kaberamaido represents a continuation of the existing Soroti focus or a new and separate focus with limited gene flow to and from Soroti. Further sampling will be required in order to elucidate the nature and development of the Kaberamaido outbreak.

In order to further examine the relationships between the four populations a UPGMA dendrogram of similarity was constructed using the MLGs and PCoA was undertaken. The dendrogram, rooted to the *T. b. gambiense* Group 1 isolate MALa1 (Figure 5.3) separated the isolates into two groups with 100% bootstrap support. The first (Group A) encompasses the Kaberamaido 2009 population,

all bar one isolate from the Soroti 2003 population, 22 isolates from Tororo 2003 and five isolates from Tororo pre-1991. The second (Group B) consists of 42 individuals, 38 from the Tororo pre-1991 population, three from the Tororo 2003 population and one from the Soroti 2003 population. Plotting of Principal Coordinates 1 and 2 indicated that 74% of the observed variation could be explained by the principal co-ordinate 1 (Figure 5.4). Tororo pre-1991 accounted for the majority of the diversity observed along by this principal co-ordinate and was the most visibly diverse and separate population. The differentiation identified in the dendrogram was closely correlated with separation along principal co-ordinate 1 with all bar one member of Group B falling to the right hand side of the plot. Principal co-ordinate 2 accounted for 10% of the observed variation, this time with the two Tororo populations responsible for the majority. Soroti 2003 formed a visible cluster close, but predominantly separate from Tororo 2003 while the Group A isolates from Tororo pre-1991 fall into the intersection between the Tororo 2003 and Soroti 2003 populations. Three of the four Kaberamaido 2009 MLGs are also positioned close to this intersection. The identification of these two groups, which are predominantly defined by temporal origin, suggests that there was a substantial shift in the *T. b. rhodesiense* populations of Uganda between the pre-1991 period and 2003. The PCoA analysis indicates that this is predominantly due to the loss of diversity that was present in the pre-1991 population.

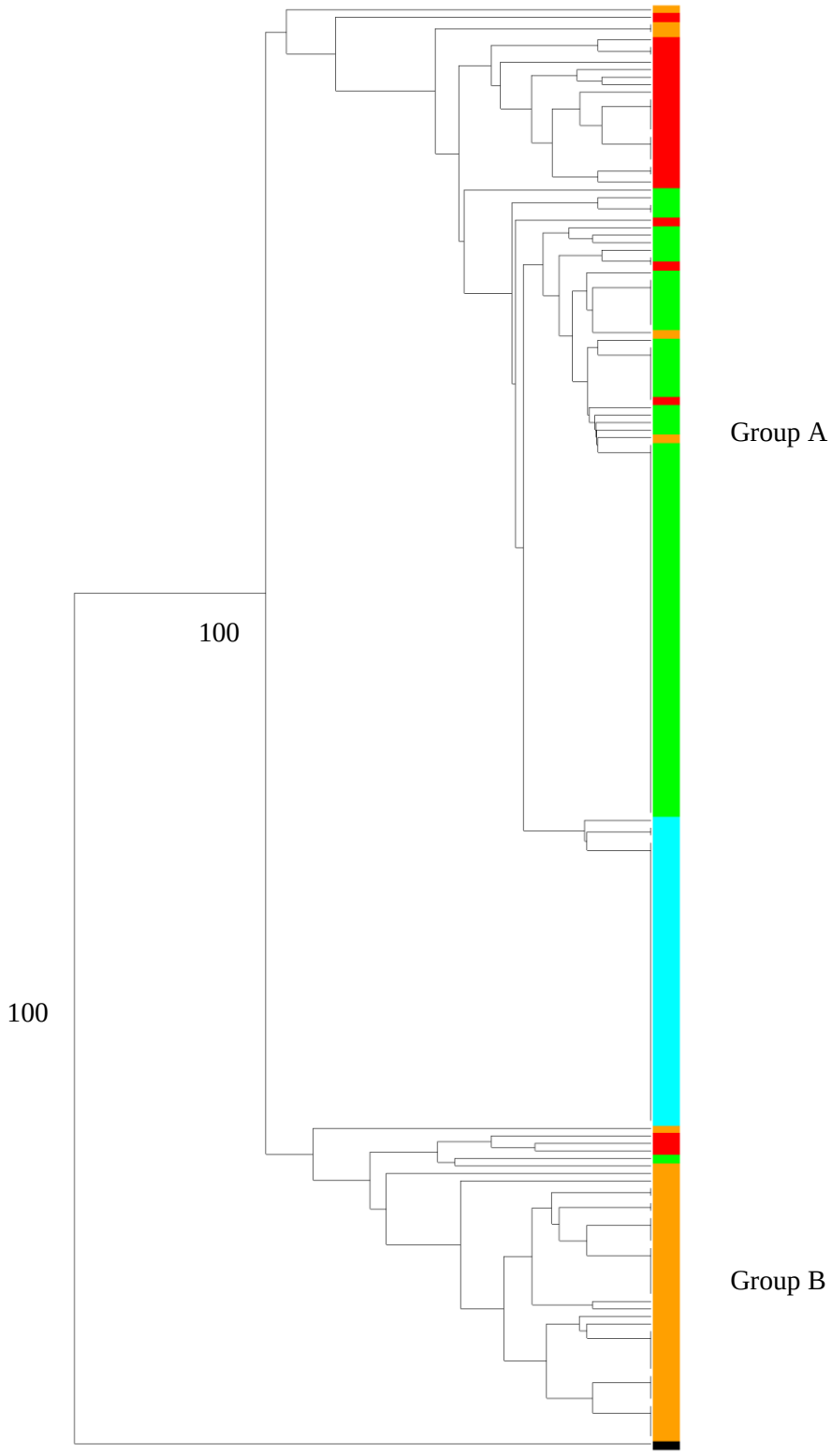


Figure 5.3 UPGMA dendrogram of similarity for the four Ugandan populations

UPGMA dendrogram of similarity for the four Ugandan populations, rooted against the *T. b. gambiense* Group 1 isolate MALa1 isolated from a focus in the Democratic Republic of Congo (black in figure). Significant bootstrap support (100%) was detected for two groups (A and B). Populations: Tororo pre-1991 = orange, Tororo 2003 = red, Soroti 2003 = green, Kaberamaido 2009 = light blue.

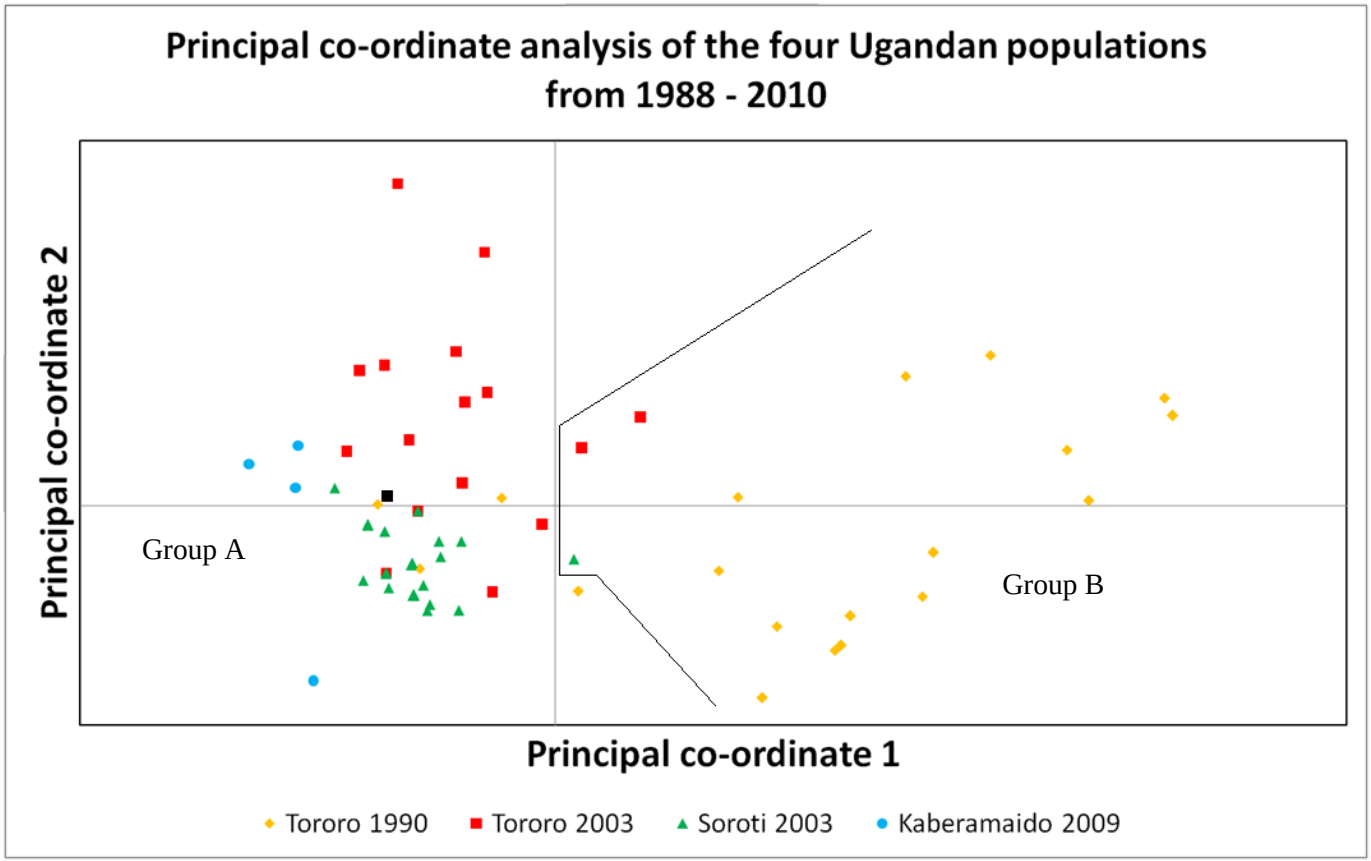


Figure 5.4 PCoA of the four Ugandan populations

PCoA of the four Ugandan populations assessing the relationships between them as the outbreak has developed over time and geography. Principal co-ordinate 1 explains 74% of the observed variation, primarily present within the Tororo pre-1991 population. Principal co-ordinate 2 accounts for 10% of the total observed variation, with the majority coming from the two Tororo populations. The separation of the two groups identified by the dendrogram is indicated, with only Liri016 (highlighted in black) not grouping as in the dendrogram.

5.3 Discussion

The focus of this chapter, like that of chapters 3 and 4 has been to examine the role and extent of mating in one of the African Salivarian trypanosomes, *T. b. rhodesiense* in addition to examining the spatial and temporal relationships that exist between populations of this sub-species. The results presented here support the existence of both genetic exchange and extensive clonality within independent populations of this sub-species, emphasising the role the non-obligatory mating cycle can play in determining population structure in *T. brucei*. This is in contrast to previous studies that have described *T. b. rhodesiense* as being a genetically homogenous member of the *T. brucei* species, much like the more extensively studied *T. b. gambiense* Group 1. This evidence for genetic exchange occurring in the field therefore represents a fundamental advance in our knowledge of the role that mating plays in shaping the evolutionary potential of this organism.

The definition of species in organisms such as trypanosomes is a complex issue, defined historically by morphological similarities while being divided by geographic localisation, host range and variations in disease profile. Molecular biology and genetics has only added to the complexity by further highlighting the similarities and differences within species. In the case of *T. brucei* species complex this has given rise to the classical description of three sub-species. The homogenous *T. b. gambiense* Group 1 is the most genetically distinct of the sub-species, with suggestions that it should be returned to its previous position as *T. gambiense* (Gibson, 2003, 2007).*T. b. brucei*, with high levels of genetic diversity appears in many respects as the archetype of the species, lacking only the ability to infect humans, while over the last decade or so the emerging hypothesis with regard to *T. b. rhodesiense* has positioned the sub-species as a host range variant of *T. b. brucei* (MacLeod *et al.*, 2001c; Gibson, 2002; Gibson *et al.*, 2002), defined only by the presence and role of the *SRA* gene (Xong *et al.*, 1998). This position as a simple variant of *T. b. brucei* however neglects that previous studies have observed that *T. b. rhodesiense* can exist as essentially stable clones over time, albeit with differences between geographically distinct populations (Hide *et al.*, 1994; MacLeod *et al.*, 2000) and *T. b. rhodesiense* populations appear distinct and separate from the sympatric, heterogeneous and frequently mating populations of *T. b. brucei* (MacLeod *et al.*, 2001a). While these differences between the sub-species have been identified, isolates of *T. b. rhodesiense* are more closely related to local *T. b. brucei* than *T. b. rhodesiense*, suggesting genetic exchange between the sub-species still occurs (MacLeod *et al.*, 2001c; Balmer *et al.*, 2011).

While the Ugandan results support this highly homogenous concept, those from Malawi demonstrate that the frequency of mating may vary, with these contrasting results highlighting the role that mating plays in shaping outbreaks and over longer periods the influence that mating may have on the evolution of the species. This means, significantly, that *T. b. rhodesiense* cannot be

classified as a genetically homogeneous human-infective variant of *T. b. brucei*, but that there are genetically *and* biologically distinct populations. Within this context, one obvious question that arises from the results is whether the observed frequent genetic exchange in Malawian samples is due to mating between *T. b. rhodesiense* and *T. b. brucei*, or between *T. b. rhodesiense* samples in a separate endemic cycle. This is a question that cannot be answered using the current sample set, and requires population analysis to be carried out between sympatric *T. b. rhodesiense* and *T. b. brucei* populations in that focus.

The Ugandan foci fit with the previously established idea of a *T. b. rhodesiense* existing as a relatively stable lineage. While variation was observed between the foci the process was gradual, with populations closely linked by time and geography sharing the closest genetic relationships. The Ugandan populations were sampled over a 22 year period, it is highly likely therefore that both mutation and population drift over time played a role in the gradual differentiation observed within the country, while the establishment of new foci likely involved severe bottleneck effects, reducing population diversity and increasing the level of genetic differentiation between the sampled populations. While these processes explain the relationships between the foci they are unable to explain the clonality observed within each. *T. b. brucei* is known to undergo frequent mating with an epidemic population structure (MacLeod *et al.*, 2000) and apart from humans shares many common hosts with *T. b. rhodesiense*, in particular livestock (Welburn *et al.*, 2001), which imply the presence of selective pressures or genetic barriers that prevent interbreeding of the sub-species. As the majority of the samples were isolated from humans the selective pressure of passage through humans may be sufficiently high that the offspring of mating events between the two sub-species are either outcompeted by the *T. b. rhodesiense* genotypes that were observed or are unable to infect humans due to the loss of the particular allele combinations required for successful human infections. Each of these scenarios would require, however, the presence of multiple genetic factors required for successful human infections. In Malawi the existence of multiple loci contributing to human infectivity could potentially be reconciled with extensive genetic exchange if the 'susceptible' alleles which reduce fitness in human hosts are present at only a low frequency within the population. If this were the case then the majority of progeny genotypes would retain sufficient fitness and survive within human hosts.

These hypotheses do not however prevent the occurrence of selfing, which does not appear to be occurring at observable levels in these populations based on the observed excesses of heterozygosity. Selfing has been observed to occur under laboratory conditions (Peacock *et al.*, 2009), however the authors noted that it resulted in an increased level of inviable progeny compared to a cross between two unrelated strains. The lack of observed selfing in Uganda may therefore suggest a level of self incompatibility, producing a sufficient level of inviable progeny that the parental strain remains dominant. Alternatively it is possible that the lack of mating

indicates an inability to mate. Loss of the ability to undergo genetic exchange has almost certainly occurred before in the evolutionary history of *T. brucei*, giving rise to *T. b. gambiense* Group 1 (Morrison *et al.*, 2008b; Koffi *et al.*, 2009). *T. evansi* and *T. equiperdum*, which are likely to have originated from *T. brucei* following partial or complete loss of the kinetoplast (Lai *et al.*, 2008), are also unlikely to be able to undergo genetic exchange as they have lost the ability to be transmitted by the tsetse fly. While genetic exchange has been demonstrated to be possible for isolates of *T. brucei* originating from Uganda (Degen *et al.*, 1995) this study used isolates derived from tsetse flies in the country and did not determine the sub-species present. It is therefore possible that mating is present within the local *T. b. brucei* population but not the *T. b. rhodesiense* population. The possibility that *T. b. rhodesiense* in Uganda lacks the ability to undergo sexual recombination would also provide an explanation over and above geographic separation for the genetic differences observed between the *T. b. rhodesiense* populations of Malawi and Uganda, as one would predict relatively rapid genetic differentiation through the Southern *T. b. rhodesiense* population incorporating diversity from sympatric *T. b. brucei* parasites, whereas the Ugandan *T. b. rhodesiense* populations would only accumulate diversity through background mutation.

These findings add to the emerging picture from studies of *T. vivax* and *T. congolense* discussed in the previous chapters, and previous studies of *T. brucei*, in emphasising that genetic exchange plays differing roles in the many trypanosome species, with a complex mating cycle where both genetic exchange and clonality can drive the evolution of populations. Much however remains to be understood, mating has yet to be directly observed although the early meiotic cells have now been identified (Peacock *et al.*, 2011) and the triggers controlling whether genetic exchange occurs or not remain to be elucidated.

Chapter 6

Genomics

6.1 Introduction

Throughout the history of genetics there has been a constant progression in the development of the markers used, from phenotype to indirect genetic markers such as isoenzymes and more recently direct genetic markers such as microsatellites and SNPs. While each of these markers have their own unique strengths and weaknesses one of the common limitations is in their association with only a small proportion of the total genomic variation. The rapid development of genome sequencing through the 1990s and start of this century (Reviewed in Metzker, 2005, 2010; Ansorge, 2009) has therefore revolutionised the field of genetics by allowing for the majority of base variants to be directly examined and analysed with respect to the whole genome.

The genome of trypanosomes is comprised of four main sections (El-Sayed *et al.*, 2000) of varying ploidy: the 11 megabase chromosomes on which the housekeeping genes are located (Melville *et al.*, 1998); the kinetoplast; the intermediate chromosomes and the mini chromosomes, with the latter two acting as a reservoir of VSG sequences. The megabase chromosomes, which are the focus of this chapter and the only chromosomes represented by the *T. brucei* genome reference sequence (Berriman *et al.*, 2005), are diploid in the majority of strains although instances of aneuploidy have been observed following laboratory crosses (Jenni *et al.*, 1986; Paindavoine *et al.*, 1986a; Wells *et al.*, 1987; Gibson *et al.*, 1992, 1995, 1997, 2008; Gibson and Bailey, 1994; Hope *et al.*, 1999; Peacock *et al.*, 2008, 2009). The most common forms of variation are SNPs and insertion / deletion (indel) of sequence (which may result in copy number variation) and sequence rearrangements.

With the increasing availability of whole genome sequencing it is now viable to survey multiple genomes and uncover the extent of SNP variation. However knowledge of SNP positions and alleles is of limited value without understanding their relationship to other SNPs in the region, which collected together form the haplotype. This is increasingly important as with an increase in SNP number in any given gene there are an increasing number of possible alleles (Table 6.1). The combination of variation in this way may have substantial effects upon how the two alleles of a sequence act if they have functional roles and variation within the coding sequences of genes provides the most visible examples of this. The presence of SNPs within a coding sequence may result in alterations to the amino acids encoded and therefore alter the structure of any protein encoded. Such a change may vastly affect the function of one allele if it changes important residues or may have more subtle effects as even some synonymous SNPs, which do not alter the amino acid sequence, can have functional roles (Komar, 2007; Hunt *et al.*, 2009). An indel event within a coding sequence will typically have a much larger effect, with the insertion or loss of bases potentially giving rise to frame shifts in downstream codons, with the two alleles subsequently encoding for vastly different proteins. While it is often feasible to determine the exact sequences of

both alleles for individual genes, through amplification, isolation and Sanger sequencing, this approach is infeasible when considering entire genomes due to the cost and time required.

Neither heterozygosity, nor its effects are confined solely to the coding regions of the genome. *Cis* acting variation within transcriptional and translational regulatory regions can have significant effects on gene expression and in humans such variation has been associated with a range of diseases (Reviewed in Epstein, 2009). As the majority of gene regulation in trypanosomes occurs post transcriptionally (Reviewed in Clayton and Shapira, 2007; Haile and Papadopoulou, 2007) due to the organisation of most genes into polycistronic expression units (Imboden *et al.*, 1987; Johnson *et al.*, 1987; Berriman *et al.*, 2005; Reviewed in Martínez-Calvillo *et al.*, 2010) *cis* acting variation will primarily function through 5' and 3' untranslated regions, with variation fine tuning the expression of specific alleles, as has been observed in humans (Kalus *et al.*, 2009; Sun *et al.*, 2011) by affecting mRNA stability (Liang *et al.*, 2003; Clayton and Shapira, 2007; Haile and Papadopoulou, 2007). Variation located within the promoter regions of the polycistronic units may, however, still play a significant role by disrupting or enhancing transcription, affecting each of the gene alleles located within the unit.

Number of variants	Number of possible alleles pairs
1	2
2	4
5	32
10	1024

Table 6.1 Allele pairs with increasing SNPs

The number of possible allele pairs for a sequence with an increasing number of heterozygous SNPs.

Haplotype variation plays a further role in the inheritance of material during reproduction following Mendel's laws of inheritance, as each parent will only contribute a single haplotype of each chromosome to any progeny. Selective pressures therefore act on combinations of haplotypes, favouring individuals with beneficial combinations and selecting against those where the particular combinations are ill suited. Ultimately selection functions on individual haplotypes, as particularly beneficial haplotypes may raise the overall fitness of the whole organism by compensating for the effects of deleterious mutations on the second homologue. The effect of selective pressures, while acting on the combination of haplotypes present in an individual, are further fine tuned through the actions of meiotic recombination. Meiotic recombination is the primary force giving rise to new haplotypes by swapping segments of homologous chromosomes during genetic exchange, generating the homologues which are subsequently inherited. The strongest linkage between genes on each homologue is therefore confined to blocks, broken up by hotspots of meiotic recombination (Rana *et al.*, 2004; Nishant and Rao, 2006).

In generating new haplotypes meiotic recombination serves two primary roles, bringing together variation and separating variation on existing haplotypes (Barton and Charlesworth, 1998; Hadany and Comeron, 2008; Hörandl, 2009). Bringing variation together onto a single homologue allows for new *cis* acting effects, such as the formation of new alleles, utilising sequence from two different lineages that would otherwise only interact through *trans*acting effects. Further to this, by physically linking variation, meiotic recombination increases the chances that the specific combination of variants will be inherited together. This is of particular importance when the new combination raises the fitness of the haplotype as selective pressures will act to maintain the presence of this combination within the population. The second role of meiotic recombination, separating variation, facilitates the continuation of beneficial mutations by sundering the physical linkage to deleterious mutations that lower the overall fitness of the haplotype. These processes of mixing through Mendelian inheritance and meiotic recombination may be exploited for the purposes of phylogenomics (Siepel, 2009; Reviewed in Boussau and Daubin, 2010); genome wide association studies, allowing for association of genomic variation with phenotypes of interest (Iles, 2008) and identification of selective sweeps, whereby regions of the genome under selection are identified through a reduction in overall allele diversity (Wootton *et al.*, 2002; Nair *et al.*, 2003, 2007).

Through the availability of whole genome sequencing it is now possible to uncover this genetic variation, however the differing sequencing platforms each function in different ways and therefore require differing approaches to the assembly of genome sequences and any attempts to resolve the haplotypes. The first genome reference sequence for the Salivarian trypanosomes used the TREU 927 *T. b. brucei* strain (Berriman *et al.*, 2005), which was chosen for a number of reasons. The strain has been adapted for growth in laboratory culture and is readily amenable to genetic

manipulation, allowing for the insertion and knockout of gene constructs. However, despite being laboratory adapted the strain has maintained the ability to differentiate through each of the life cycle stages, including transmission through the tsetse fly and completion of the non obligatory sexual cycle (Turner *et al.*, 1990). Other projects have led to the publication of sequences for the related kinetoplastids *Leishmania majo*r (Ivens *et al.*, 2005) and *T. cruzi* (El-Sayed *et al.*, 2005), and most recently *T. b. gambiense* Group 1 (Jackson *et al.*, 2010). Sequencing of *T. congolense* and *T. vivax* is ongoing at the WTSI with initial sequence assemblies available via the WTSI and the TriTrypDB (Aslett *et al.*, 2010).

The TREU 927 reference sequence (Berriman *et al.*, 2005) was constructed as a collaboration between the WTSI and TIGR with the WTSI sequencing chromosomes 1, 9, 10 and 11 through a whole chromosome shotgun approach (Hall *et al.*, 2003) and TIGR employing a BAC approach (El-Sayed *et al.*, 2003) for the sequencing of chromosomes 2 – 8. Both approaches used traditional Sanger sequencing following generation of the respective libraries. The whole genome shotgun approach of the WTSI utilised PFGE to isolate the megabase chromosomes, which were subsequently then digested and cloned into vectors for sequencing. These sequences were then reassembled through the identification of overlapping sequences, generating contigs of increasing sizes. For chromosomes 10 and 11 these assemblies represent a mosaic of both homologues while for chromosomes 1 and 9 the differing sizes of the homologous pairs allowed for enrichment for a particular homologue, thus reducing the number of heterozygous sites observed. However, as the second homologue could not be fully purified the sequences of these chromosomes remain partial mosaics of both. The approach of TIGR employed in the sequencing of chromosomes 2 – 8 used the construction of 85 – 170 kb BAC clones which were individually sequenced using a BAC walking approach and mapped to their chromosomes of origin with contigs formed through overlap between BACs. As with the WTSI approach this approach generated a mosaic sequence constructed from both homologues. The released TREU 927 genome reference sequence collected together the 11 megabase chromosomes and stripped away any information regarding heterozygous positions, generating a mosaic, pseudo haploid sequence.

With the rapid development of sequencing technologies it is now far simpler, cheaper and faster to generate genomic sequence data through massively multi-parallel, short read technologies such as the Illumina Solexa platform. The Solexa platform functions through the use of a sequencing by synthesis, reversible dye terminator method (Reviewed in Shendure and Ji, 2008; Ansorge, 2009; Pettersson *et al.*, 2009; Metzker, 2010), allowing for the extension and visualisation of sequence a single base at a time, with throughput achieved by the ability to generate millions of distinct read clusters upon a single chip. Sequence may be assembled in two manners, *de novo* or through alignment of reads to an existing reference sequence (Reviewed in Flicek and Birney, 2009; Li and Homer, 2010). While alignment to a reference sequence prevents the assembly of novel sequences

the use of *de novo* assembly, which is capable of assembling previously undetected sequences, is hampered by the requirement for a far higher average read depth, increased computational power and relatively short read length of the second generation platforms. Reference based assembly is preferred for most eukaryotic organisms as the size of the genomes makes deep sequencing uneconomical in the majority of cases. Second generation platforms provide a relatively quick method of identifying variation within the genome and as such have been widely employed in the identification of genomic variation. However, despite the advances in technology and reduction in cost, whole genome sequencing has yet to be widely deployed in the analysis of the Salivarian trypanosomes and no genome wide SNP data are available at present.

One of the limitations of the second generation sequencing platforms is the relatively short read length, which prevents effective resolution of the haplotypes present. In order to resolve two haplotypes from one another it is necessary to physically link the variation together with individual reads. Theoretically this is possible through the use of paired read technology, as the reads of each pair are sequenced from the same piece of DNA, which has originated from a single chromosomal homologue. Where haplotype specific variation repeatedly aligns to both reads of a pair it is possible to 'walk' along the haplotype, linking the variation together (Figure 6.1A). This is however only possible when both reads of a pair align to haplotype specific variation, which occurs only rarely due to the average distance between these sites (Figure 6.1B). While it is possible to overcome this through additional sequencing using multiple insert length, increasing the chance of observing reads aligned to the required positions, the depth of coverage required makes it prohibitively expensive. Due to these limitations it is therefore necessary to employ an alternative method in order to resolve the haplotypes from second generation sequencing data.

With the inability to reconstruct haplotypes through direct sequencing, alternative methods have been developed, primarily focused on the use of population data in order to assign phases to alleles. At the core of these techniques is the role of physical linkage, which asserts that alleles on the same chromosomal homologue will be associated together at a higher frequency than those on different homologues and that this association is distance dependent, dropping off due to the increased likelihood of meiotic recombination as the distance between two sites increases. Algorithms for the determination of phase data are typically broken into two categories, population based and lineage based (Reviewed in Niu, 2004). Perhaps the most ambitious project of this kind is Human HapMap project (International HapMap Consortium, 2003, 2005), which aims to determine the common haplotypes present in four initial sample populations from Nigeria, Japan, China and the USA. Through determination of these haplotypes the project aims to be able to calculate putative haplotype blocks, regions of exceptionally low recombination, and identify potential 'tag' SNPs that are representative of informative variation and can therefore be used to describe wider haplotype blocks and be utilised in association studies (Stram, 2004; Xu *et al.*, 2007; Peiffer and

Gunderson, 2009). The HapMap has provided data for a diverse range of projects, including but not limited to the identification of genetic variants associated with a range of diseases (Manolio *et al.*, 2008; Reviewed in Musunuru and Kathiresan, 2008; Manolio and Collins, 2009).

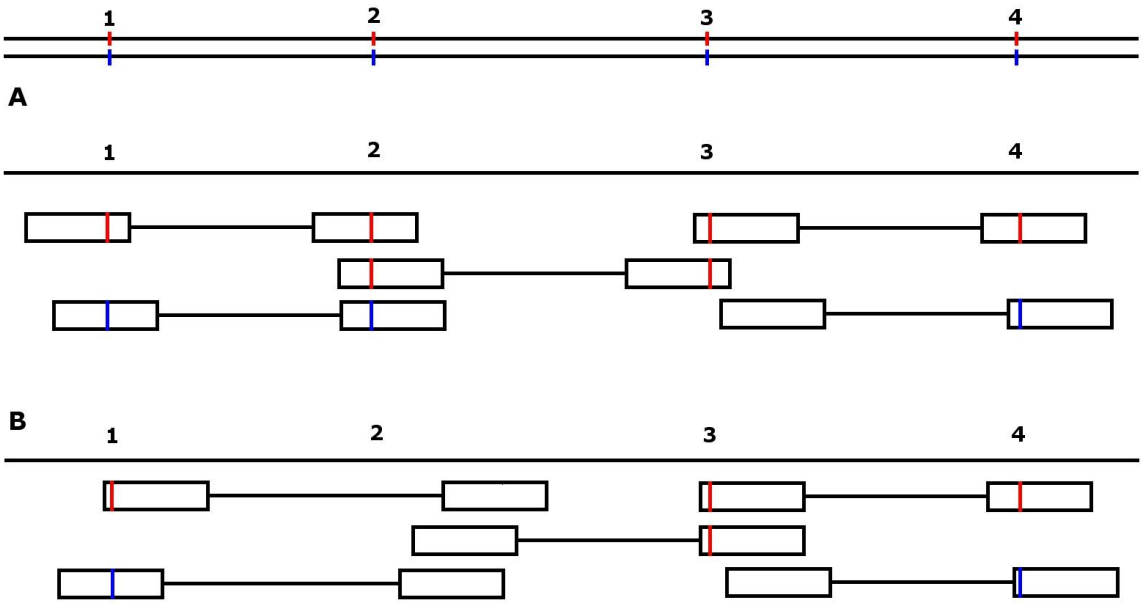


Figure 6.1 Haplotype resolution through read walking

Paired reads aligned to a reference sequence identify positions of haplotype specific variants (1 – 4), coloured red and blue for alleles from the two true haplotypes as indicated at the top of the figure. A) Where both reads in a pair align to heterozygous positions it is possible to link the alleles together and walk between sites of variation in order to reconstruct the haplotypes. B) In the majority of cases however gaps emerge when only one read of a pair aligns to a heterozygous position. The heterozygous position at position 2 may therefore not be used to link the alleles at positions 1 and 3, preventing reconstruction of the full haplotype. Haplotype inference is therefore limited to smaller blocks, such as between sites 3 and 4 in the figure.

At the level of a single individual, haplotype determination may be achieved through the generation of a genetic map following a genetic cross with another individual. With sufficient numbers of progeny it is possible to determine linkage between the haplotype specific alleles, detect the effect of recombination and work back to the original parental haplotypes. This approach has been previously used with trypanosomes in the generation of linkage maps of *T. b. brucei* TREU 927 (MacLeod *et al.*, 2005b) and *T. b. gambiense* Group 2 STIB 386 (Cooper *et al.*, 2008). The TREU 927 genetic map employed 39 progeny clones generated from a cross between TREU 927 and STIB 247, with each progeny genotyped with 182 micro- and minisatellite markers, spaced relatively evenly across the 11 megabase chromosomes. STIB 247 was selected for use as the second parent strain as the line was identified to be highly homogeneous, simplifying the process of determining which allele TREU 927 had contributed to each progeny. This homozygosity however prevents generation of a comparable STIB 247 map. Generation of the map identified 11 major and 5 minor linkage groups, which correspond to the 11 megabase chromosomes. Through comparison with the genome reference sequence an average map distance of 15.6 kb per centiMorgan was calculated, however the occurrence of recombination hot and cold spots introduces considerable variation with mapping distances of as low as 1.58 kb per centiMorgan and as high as 95.64 kb per centiMorgan present.

The later genetic map of STIB 386, once again utilising progeny from a cross with STIB 247, employed 119 markers in the analysis of 38 hybrid progeny lines. This map identified 12 linkage groups, with only chromosome 10 separated into two unique groups and an average mapping distance of 24.4 kb per centiMorgan. Comparison of the two maps by means of 47 shared markers indicated a conservation of marker order, further supported by consistency in synteny between STIB 386 and the TREU 927 reference sequence. These maps, in utilising approximately 40 progeny each have allowed for determination of the micro- / minisatellite haplotypes for TREU 927 and STIB 386 respectively. While a similar approach could be employed in generating a SNP haplotype map for these two lines the cost of sequencing approximately 40 samples for each map prevents such a project at present. However the existence of these maps, which have identified the location of recombination events, allows for an alternative approach that requires sequencing of a far smaller group of samples. This approach forms the basis of the second half of this chapter.

The results of this chapter are focused on two distinct aspects. The first half of the results concerns the assembly of three strains of *T. brucei* following Illumina Solexa sequencing. These strains are the genome reference strain *T. b. brucei* TREU 927 (Berriman *et al.*, 2005), *T. b. brucei* STIB 247 and *T. b. gambiense* Group 2 STIB 386. Each of the three strains is capable of completing the entire trypanosome life cycle, including transmission through tsetse flies and the sexual cycle. Further they have been extensively used in earlier genetic studies, including generation of genetic linkage maps (MacLeod *et al.*, 2005b; Cooper *et al.*, 2008). Presented here are the results of SNP

discovery in the three strains and comparative analysis, providing insight into the variation present within trypanosomes and an initial resource for future studies. The second half of the results detail an attempt to resolve the haplotypes of the *T. b. brucei* genome reference strain TREU 927 by mapping the inheritance patterns of SNPs in progeny from a genetic cross between TREU 927 and STIB 247 (MacLeod *et al.*, 2005b).

6.2 Results

6.2.1 Assembly of short read sequence from the parental strains TREU 927, STIB 247 and STIB 386

In order to assess genomic variation within *T. brucei* the three strains, TREU 927, STIB 247 and STIB 386 were chosen for sequencing at the WTSI utilising the Illumina Solexa sequencing platform. These strains were chosen for sequencing as each is capable of completing the entirety of the *T. brucei* life cycle, including undergoing genetic exchange and were used in construction of the TREU 927 and STIB 386 genetic maps (MacLeod *et al.*, 2005b; Cooper *et al.*, 2008). Seventy six base pair long, paired end reads were generated for each strain, providing 40,424,405, 11,824,123 and 13,038,342 read pairs for TREU 927, STIB 247 and STIB 386 respectively, equating to a pre-assembly average fold genome coverage of approximately 235, 67 and 76 reads.

Reads were assembled to the TREU 927 megabase chromosome reference sequence utilising BWA (Li and Durbin, 2009) with reads trimmed and filtered for read quality prior to alignment in order to maximise the alignment accuracy. Following alignment, reads were further filtered by alignment quality, removing reads that aligned with low quality. This second filtering step also serves to exclude reads that align to multiple positions within the reference sequence, a requirement for accurate SNP detection. Following filtering and assembly and the second round of filtering the average fold coverage of the assemblies were 101 (TREU 927), 27 (STIB 247) and 32 (STIB 386) reads, with reads of the respective assemblies covering 95%, 84% and 85% of the total reference sequence. There was substantial variation in read depths within each chromosome with regions of low coverage clustered towards the start and end of chromosomes (Figure 6.2). This was particularly notable in the STIB 247 and STIB 386 alignments, suggesting the presence of regions unique to TREU 927, such as at the end of chromosome 9 which contains an extended VSG array of unique genes and pseudogenes (Berriman *et al.*, 2005).

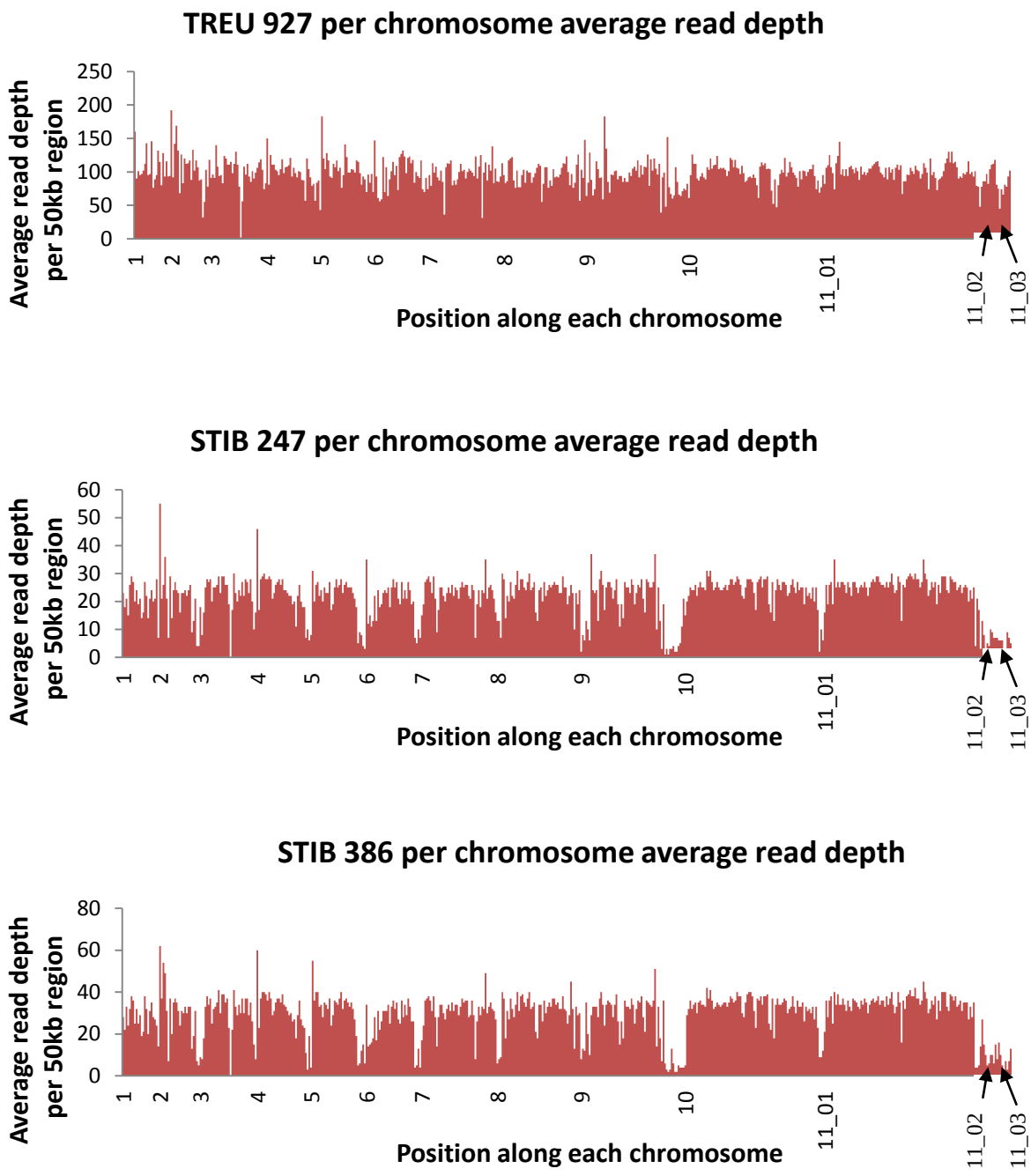


Figure 6.2 Per chromosome average read depths

Average read depths for TREU 927, STIB 247 and STIB 386 plotted for the entire genome with the average read depth calculated per 50 kb. Chromosome labels mark the start of the respective chromosomes.

6.2.2 SNP Analysis

SNPs were called from the three parental assemblies through the use of the SAMtools software package utilising the pileup SNP calling algorithm (Li *et al.*, 2009). Potential SNPs were filtered for read mapping quality (minimum required Phred quality of 30); presence of potential indels; presence of nearby SNPs and read depth at the base in question. For STIB 247 and STIB 386 a minimum read depth of 10 bases was used for SNP calling in order to compensate for the lower average read depth in these lines. The higher average read depth of the TREU 927 assembly allowed for the use of a minimum read depth of 20. For all assemblies the maximum read depth was constrained to three times the chromosomal average in order to prevent calling at positions of unusually high read depth. These regions, which are concentrated towards the telomeres, often represent repetitive regions of which only a single copy is present within the reference sequence.

Each SNP was classified into one of three types; homozygous, heterozygous type 1 or heterozygous type 2 (Table 6.2). Homozygous SNPs represent a homozygous position within the assembly that is in disagreement with the reference sequence and account for 85% and 76% of the SNPs observed in STIB 247 and STIB 386 respectively. A total of 134 homozygous SNPs were identified in the TREU 927 assembly. As this assembly used the same strain as a reference sequence the presence of these SNPs have a number of possible explanations. The first is sequencing error during generation of either the original reference sequence or the Illumina Solexa sequencing presented here. Alternatively this base may be heterozygous with identification of only a single allele that happens to be the one not represented in the reference. Finally it is possible that these bases represent real differences between the TREU 927 samples used in the two sequencing projects, reflecting mutations that have arisen during growth following the generation of the two isolates used for the respective sequencing projects.

Chromosome	TREU 927			STIB 247			STIB 386		
	Hom	Het 1	Het 2	Hom	Het 1	Het 2	Hom	Het 1	Het 2
1	21	1599	0	3783	1236	6	4528	1851	10
2	2	1480	0	3722	1598	11	4088	3067	12
3	0	2750	0	6804	1675	13	7859	2964	23
4	0	2352	0	6593	676	8	6989	3005	19
5	3	2641	0	7525	1626	14	7958	3841	26
6	2	2859	0	6410	818	8	7022	2892	16
7	1	2870	0	10001	1496	8	12386	1485	7
8	2	3469	1	10956	776	7	13251	2915	16
9	32	5959	0	12294	3005	28	12925	6726	41
10	27	5571	0	20049	1778	23	23668	3025	22
11_01*	38	5859	1	22937	4781	27	27299	6978	36
11_02*	2	755	0	754	552	14	659	1198	19
11_03*	4	264	0	631	310	7	672	569	8
Total	134	38428	2	112459	20327	174	129304	40516	255

Table 6.2 Distribution of SNPs across the megabase chromosomes

SNPs were called against the TREU 927 reference sequence (Berriman *et al.*, 2005) and filtered for quality. Hom = Homozygous positions that differ from the reference sequence; Het 1 = heterozygous positions where one of the alleles matches that present in the reference sequence; Het 2 = heterozygous positions where both SNP alleles are absent from the reference sequence. * Chromosomes 11_01, 11_02 and 11_03 respectively represent the major and two minor contigs of chromosome 11, which have yet to be fully assembled into a single sequence.

A total of 38,428, 20,327 and 40,516 heterozygous SNPs were identified in the respective TREU 927, STIB 247 and STIB 386 assemblies and were subdivided into two groups, type 1 and type 2. Heterozygous type 1 SNPs were defined as possessing a single allele in common with the reference sequence while both alleles of heterozygous type 2 SNPs differed from the reference sequence. In all strains >99% of the heterozygous positions shared a single allele with the reference sequence. Heterozygous SNPs in TREU 927 and STIB 386 were typically spread over the length of each chromosome, with increased SNP densities common towards either end of each chromosome. The plots for chromosome 10 are presented in Figure 6.3 A and C and for all other chromosomes in Appendix 12. In contrast to this, the heterozygous SNPs of STIB 247 are typically clustered into a smaller number of discrete regions, surrounded by regions of homozygosity. The plot for STIB 247 chromosome 10 is presented in Figure 6.3 B and for all other chromosomes in Appendix 12. The lower number of heterozygous positions in STIB 247 line correlates with the observations of extensive homozygosity based on micro- and minisatellite data (MacLeod *et al.*, 2005b). Pairwise comparisons of the heterozygous SNPs identified only a small proportion to be shared between strains (TREU 927 / STIB 247, 1296 shared; TREU 927 / STIB 386, 1939; STIB 247, STIB 386, 2763).

Homozygous SNPs in both STIB 247 and STIB 386 were distributed across the entirety of the megabase chromosomes (Figure 6.4 and Appendix 13). Regions of low homozygous SNP densities were typically shared between the two strains and correspond with regions of low read depth in the assemblies. The most notable difference in this pattern occurs on a region of chromosome 8, 580-720 kb from the start of the chromosome with a visible reduction in homozygous SNP density in STIB 247 that is not replicated in STIB 386. While homozygous SNPs in STIB 247 and STIB 386 are represented by base variants not found in the reference sequence, it is possible that the position corresponds to a heterozygous position in TREU 927, with the non reference sequence allele matching that observed in STIB 247 or STIB 386. In total 12% of the STIB 247 and 10% of the STIB 386 homozygous SNPs were found to correspond to the non reference alleles of heterozygous type 1 SNPs in TREU 927.

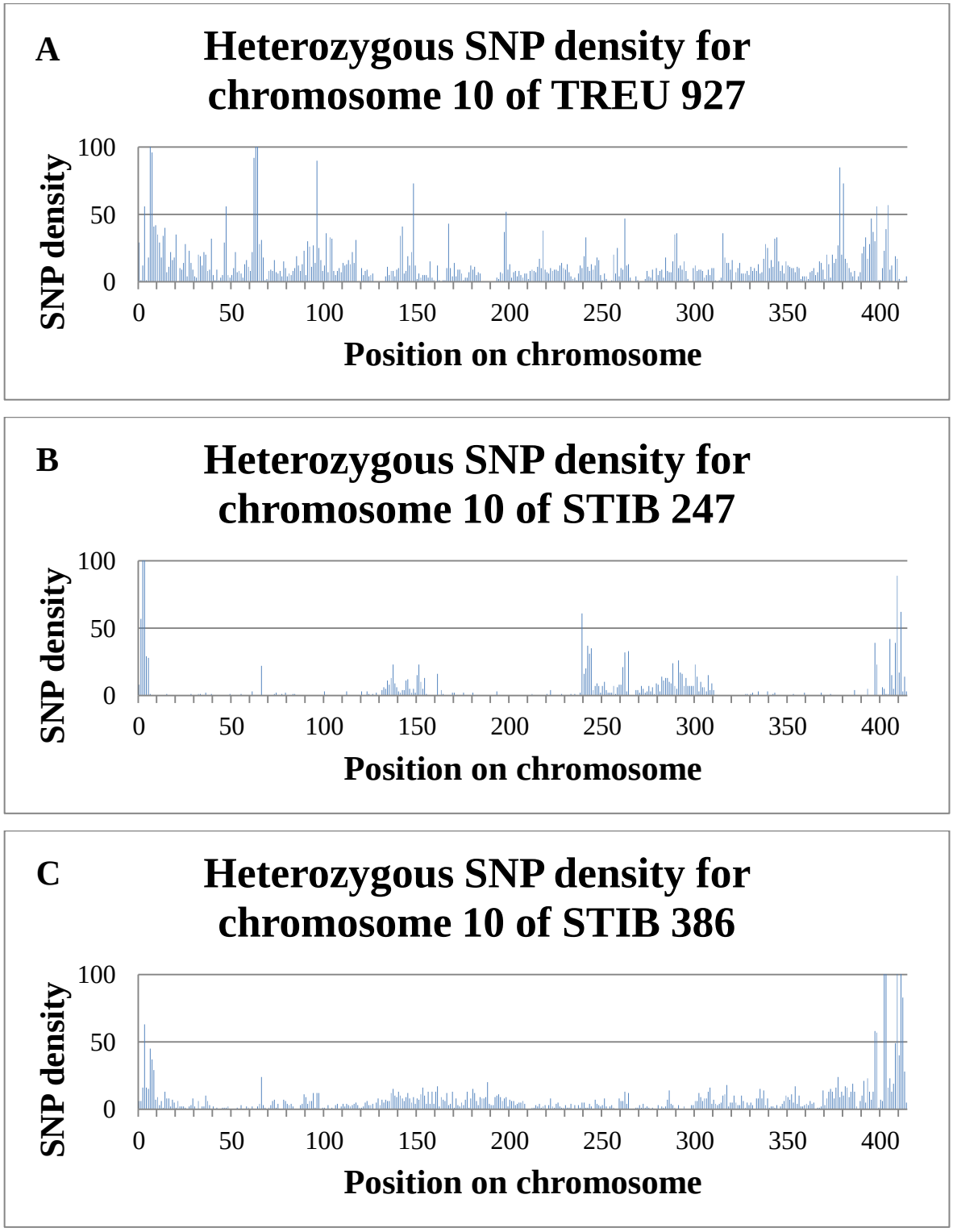


Figure 6.3 Heterozygous SNP density plots for chromosome 10

SNP density was calculated as the number of heterozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb. The Y axis of each has been limited to a maximum SNP density of 100 in order to facilitate comparison between the strains, the unrestricted plots can be found in Appendix 12. A) TREU 927, B) STIB 247, C) STIB 386.

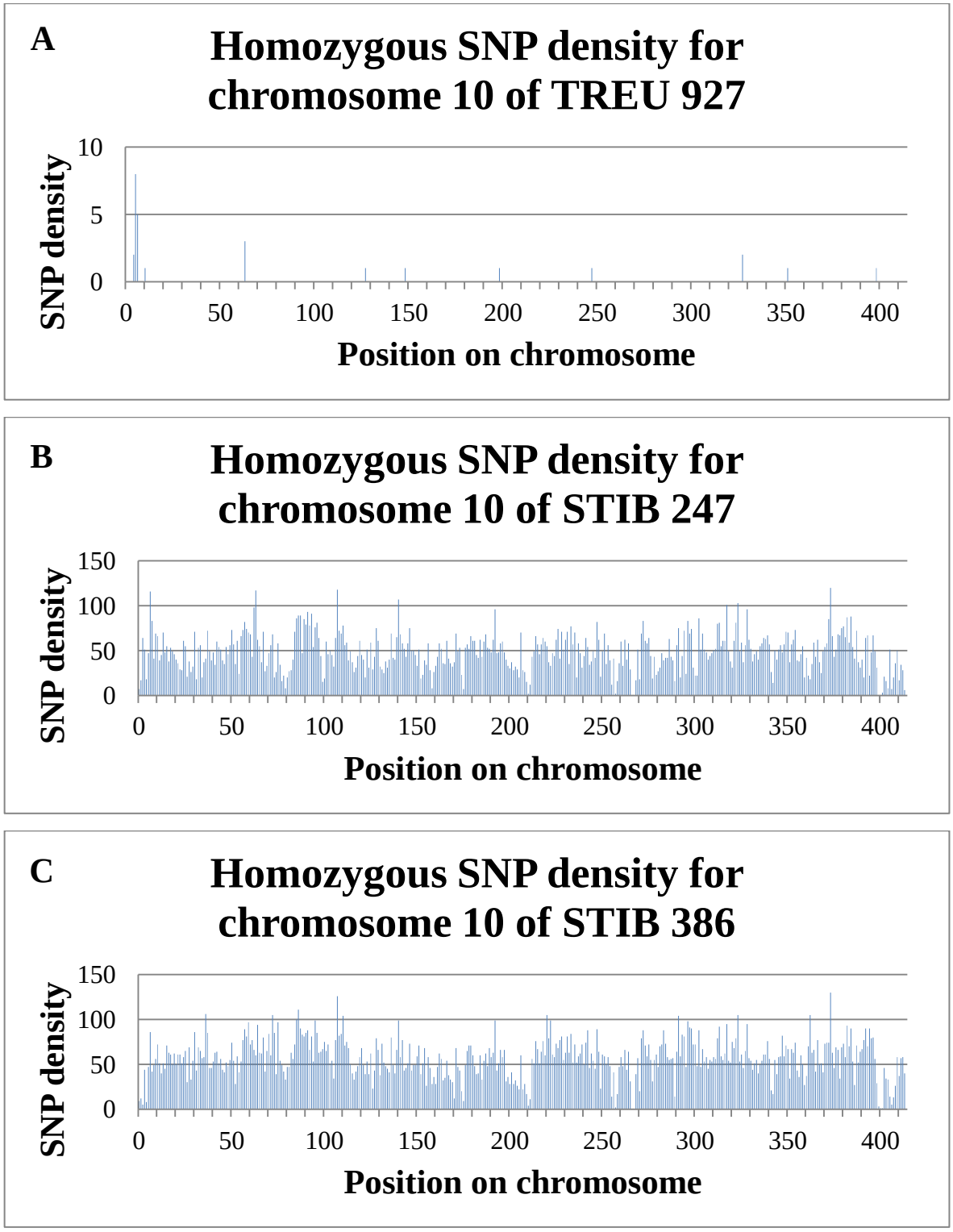


Figure 6.4A-C Homozygous SNP density plots for chromosome 10

SNP density was calculated as the number of homozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb. A) TREU 927, B) STIB 247, C) STIB 386.

One of the primary ways in which SNPs can directly affect an organism is by being located within the coding region of a gene, altering the codon sequence and potentially the amino acids the gene encodes for. Homozygous SNPs between individuals may lead to the production of unique, strain specific changes in amino acid sequences while heterozygous SNPs in diploid species allow for the introduction of protein variance within a single individual. In order to determine the extent of protein variation within these three strains the positions of homozygous and heterozygous SNPs were examined in order to identify those lying within coding regions and determine the impact upon the coding sequence.

For TREU 927 16797 heterozygous SNPs (44% of the total heterozygous SNPs) were located within coding regions, 49% of which gave rise to amino acid changes (Table 6.3). These non-synonymous codon changes gave rise to 78 new stop codons prior to the end of the annotated coding regions, which may drastically alter the function or length of the protein coded for by the gene. In total 3350 of the TREU 927 genes were determined to give rise to two unique coding alleles. As expected from the previously identified number of heterozygous SNPs STIB 386 possesses a comparable number of heterozygous genes (3479) while STIB 247 possesses less than half this number (1389). Analysis of homozygous SNPs from STIB 247 and STIB 386 indicated that 49% fell within known coding regions (Table 6.4). Of those within coding regions 43% (STIB 247) and 42% (STIB 386) led to codon changes, affecting 57% (STIB 247) and 58% (STIB 386) of the 11425 annotated genes present in the genome reference sequence.

A (TREU 927)

Chromosome	Number of heterozygous SNPs within coding sequences (% of total heterozygous SNPs)	Number of heterozygous SNPs giving rise to non synonymous codon changes (% of heterozygous SNPs in genes) [1]	Number of new stop codons	Number of genes encoding heterozygous proteins[2]
1	708 (44)	391 (55)	9	175
2	640 (43)	317 (50)	7	127
3	1143 (42)	532 (47)	3	213
4	994 (42)	479 (48)	3	190
5	1123 (43)	549 (49)	4	183
6	1184 (41)	573 (48)	8	233
7	1296 (45)	630 (49)	2	259
8	1546 (45)	737 (48)	9	294
9	2646 (44)	1463 (55)	21	504
10	2518 (45)	1182 (47)	4	539
11_01	2566 (44)	1187 (46)	4	573
11_02	295 (39)	196 (66)	7	40
11_03	138 (52)	96 (70)	2	20
Total	16797 (44)	8332 (50)	78	3350

B (STIB 247)

Chromosome	Number of heterozygous SNPs within coding sequences (% of total heterozygous SNPs)	Number of heterozygous SNPs giving rise to non synonymous codon changes (% of heterozygous SNPs in genes) [1]	Number of new stop codons	Number of genes encoding heterozygous proteins[2]
1	479 (39)	278 (58)	10	68
2	562 (35)	347 (62)	12	64
3	668 (40)	334 (50)	3	112
4	241 (35)	149 (62)	7	37
5	643 (39)	344 (53)	3	113
6	323 (39)	203 (63)	14	46
7	624 (41)	322 (52)	1	125
8	336 (43)	194 (58)	4	51
9	1307 (43)	748 (57)	29	236
10	761 (42)	381 (50)	5	152
11_01	2142 (45)	975 (46)	12	323
11_02	224 (40)	136 (61)	9	40
11_03	113 (36)	86 (76)	2	22
Total	8423 (41)	4497 (53)	111	1389

Table 6.3A-C Effect of heterozygous alleles upon coding sequences

TREU 927 (A), STIB 247 (B) and STIB 386 (C) with respect to the annotated coding sequences from the original TREU 927 pseudo haploid reference sequence. [1] Excluding those that give rise to new stop codons. [2] Total number affected irrespective of the number of codon changes within a gene.

C (STIB 386)

Chromosome	Number of heterozygous SNPs within coding sequences (% of total heterozygous SNPs)	Number of heterozygous SNPs giving rise to non synonymous codon changes (% of heterozygous SNPs in genes) [1]	Number of new stop codons	Number of genes encoding heterozygous proteins[2]
1	731 (39)	429 (59)	7	135
2	1175 (38)	648 (55)	10	179
3	1249 (42)	618 (49)	12	227
4	1340 (44)	676 (50)	9	265
5	1804 (47)	1003 (56)	20	300
6	1208 (42)	644 (53)	18	219
7	651 (44)	354 (54)	4	167
8	1409 (48)	691 (49)	9	300
9	2903 (43)	1664 (57)	53	543
10	1372 (45)	750 (55)	9	357
11_01	3197 (46)	1698 (53)	25	697
11_02	525 (43)	364 (69)	20	57
11_03	236 (41)	155 (66)	4	33
Total	17800 (43)	9694 (54)	200	3479

Table 6.3A-C Effect of heterozygous alleles upon coding sequences

TREU 927 (A), STIB 247 (B) and STIB 386 (C) with respect to the annotated coding sequences from the original TREU 927 pseudo haploid reference sequence.[1] Excluding those that give rise to new stop codons. [2] Total number affected irrespective of the number of codon changes within a gene.

A (STIB 247)

Chromosome	Number of homozygous SNPs within coding sequences (% of total homozygous SNPs)	Number of homozygous SNPs giving rise to non synonymous codon changes (% of homozygous SNPs in genes)[1]	Number of new stop codons	Number of affected genes[2]
1	2037 (54)	980 (48)	10	318
2	1720 (46)	749 (44)	6	196
3	3392 (50)	1449 (43)	5	370
4	3036 (46)	1210 (40)	6	363
5	3424 (46)	1555 (45)	4	358
6	3082 (48)	1279 (41)	7	363
7	4788 (48)	2038 (43)	9	632
8	5093 (46)	2151 (42)	8	681
9	6275 (51)	2895 (46)	30	856
10	10086 (50)	4172 (41)	13	1097
11_01	11159 (49)	4461 (40)	12	1215
11_02	338 (49)	235 (70)	9	48
11_03	262 (42)	163 (62)	3	37
Total	54692 (49)	23337 (43)	122	6534

B (STIB 386)

Chromosome	Number of homozygous SNPs within coding sequences (% of total homozygouss SNPs)	Number of homozygous SNPs giving rise to non synonymous codon changes (% of homozygous SNPs in genes)[1]	Number of new stop codons	Number of affected genes[2]
1	2449 (54)	1149 (47)	21	331
2	1922 (47)	803 (42)	6	212
3	3972 (51)	1669 (42)	7	388
4	3106 (44)	1193 (38)	3	367
5	3594 (45)	1559 (43)	7	368
6	3432 (49)	1383 (40)	10	353
7	5957 (48)	2521 (42)	7	544
8	6313 (48)	2607 (41)	11	571
9	6768 (52)	3111 (46)	36	885
10	12027 (51)	4950 (41)	17	1174
11_01	13441 (49)	5220 (39)	19	1290
11_02	285 (43)	203 (71)	5	55
11_03	336 (50)	239 (71)	6	45
Total	63602 (49)	26607 (42)	155	6583

Table 6.4A-B Effect of homozygous alleles upon coding sequences

STIB 247 (A) and STIB 386 (B) with respect to the annotated coding sequences from the original TREU 927 pseudo haploid reference sequence. [1] Excluding those that give rise to new stop codons. [2] Total number affected irrespective of the number of codon changes within a gene.

6.2.3 Reconstruction of the TREU 927 haplotypes

While Illumina Solexa sequencing has allowed us to detect the presence of heterozygous variation within the TREU 927, STIB 247 and STIB 386 lines, the available data do not directly allow for the determination of the haplotypes. As discussed in the introduction to this chapter the haplotypes could have been detected through read walking, however due to the quantity of sequencing which would have been required this was not attempted. The genetic map of TREU 927 however provides a framework for haplotype reconstruction by providing progeny lines generated following a cross between TREU 927 and STIB 247 and by identifying the regions where meiotic recombination has occurred in these progeny lines (MacLeod *et al.*, 2005b). By sequencing two of these progeny lines, hybrids 77 and 86, it is possible to reconstruct the TREU 927 haplotypes through the principals of Mendelian inheritance, utilising the genetic map to then correct for known sites of recombination.

This process is possible as the progeny lines have inherited only a single homologue of each megabase chromosome from TREU 927, in addition to one homologue of each chromosome from STIB 247. By comparing the alleles present at any given base in the three lines (TREU 927, STIB 247 and one of the progeny lines) it is possible to determine, in the majority of cases, which allele has been inherited from TREU 927 (Table 6.5). For example, a heterozygous SNP in TREU 927 possesses the two alleles G and T on the respective homologues. In the progeny line this position has been called as being homozygous for a G and therefore must have inherited the G homologue from TREU 927. By the same regard if the progeny line had been called as being heterozygous for C / G then it must be the G allele that has been inherited from TREU 927, with the C inherited from STIB 247. We know this as the TREU 927 position does not possess the C allele and therefore could not have contributed this allele to the progeny. By repeating this process for each TREU 927 heterozygous allele it is possible to obtain, for each chromosome, two haplotypes for TREU 927. However this reconstruction does not take into account possible meiotic recombination events between homologues. The original genetic map of TREU 927, generated through the use of microsatellite markers, calculated the location of the major meiotic recombination events for each megabase chromosome of each progeny line. It is therefore possible to correct for these events in the SNP haplotypes derived here, providing the original TREU 927 haplotypes.

TREU 927 SNP	STIB 247 base call	Progeny base call	TREU 927 Haplotype 1	TREU 927 Haplotype 2
G/T	*	G	T	G
G/T	*	T	G	T
A/C	*	A	C	A
A/C	*	C	A	C
A/G	*	A	G	A
A/G	*	G	A	G
C/T	*	C	T	C
C/T	*	T	C	T
C/G	*	C	G	C
C/G	*	G	C	G
A/T	*	A	T	A
A/T	*	T	A	T
A/C	*	A/T or A/G	C	A
A/C	*	C/G or C/T	A	C
A/G	*	A/C or A/T	G	A
A/G	*	C/G or G/T	A	G
A/T	*	A/C or A/G	T	A
A/T	*	G/T or C/T	A	T
C/G	*	A/C or C/T	G	C
C/G	*	G/T or A/G	C	G
C/T	*	A/C or C/G	T	C
C/T	*	G/T or A/T	C	T
G/T	*	C/G or A/G	T	G
G/T	*	A/T or C/T	G	T
G/T	T or A/T or C/T	G/T	T	G
G/T	G or A/G or C/G	G/T	G	T
A/C	A or A/G or A/T	A/C	A	C
A/C	C or C/G or C/T	A/C	C	A
A/G	A or A/C or A/T	A/G	A	G
A/G	G or C/G or G/T	A/G	G	A
A/T	A or A/C or A/G	A/T	A	T
A/T	T or C/T or G/T	A/T	T	A
C/G	C or A/C or C/T	C/G	C	G
C/G	G or A/G or G/T	C/G	G	C
C/T	C or A/C or C/G	C/T	C	T
C/T	T or A/T or G/T	C/T	T	C

Table 6.5 Allele inheritance patterns

The table summarises the resolution mechanic used to determine which allele had been inherited by the progeny line from TREU 927 and therefore reconstruct the haplotypes of TREU 927.

* = The base call here in STIB 247 is not required in order to determine inheritance.

TREU 927 haplotype 1 = The TREU 927 haplotype that has *not* been inherited by the progeny.

TREU 927 haplotype 2 = The TREU 927 haplotype which *has* been inherited by the progeny line.

These haplotypes are presented prior to correcting for meiotic recombination.

Resolution of the haplotypes required the development of two custom perl scripts. The first, haplo5-get was designed to take the TREU 927 heterozygous SNPs and determine the base call and read coverage at this position for the respective STIB 247 and hybrid progeny lines, outputting the data into a single collated file. This file served as the input for the second script, haplo5-rebuild, which then attempted to determine the inherited haplotypes by comparing the data as described in Table 6.5.

During reconstruction, positions were filtered based on minimum coverage, with a minimum coverage of 20 bases required in the progeny line and STIB 247 sequence assemblies, with the STIB 247 coverage only checked if the data from this parent were required for resolution. Where coverage was below this threshold, the positions were filtered out and left unresolved in the TREU 927 assembly. This filtering was necessary due to the way in which SNPs are called by the SAMtools software (Li *et al.*, 2009). SNP calling employs a range of filtering techniques including quality of bases aligned to the position; the read depth of high quality bases aligned to the position; average surrounding read quality and the proximity to other nearby SNPs. If a position fails one of these quality tests then it is excluded from SNP calling and the position is assumed to be identical to the reference call even if it is actually different. However positions which are excluded from SNP calling are not marked on the output file, therefore when a position is in agreement with the reference it is not possible to tell whether this is because it actually agrees with the reference or because it has been excluded from SNP calling. While it will be possible to identify these positions by filtering for the same criteria employed by the SAMtools software, this has not, at present, been implemented. For the time being a simpler filtering system, requiring a minimum coverage of 20 bases at progeny and STIB 247 non SNP positions, has been employed in the scripts developed here.

6.2.3.1 Script Validation

In order to validate the haplotype resolution script, three test pseudogenome sequences were constructed to represent the respective parental and progeny lines. The pseudogenomes were designed to include homozygous, heterozygous and unknown bases positioned such that all possible combinations were present. Low coverage calls and positions where the observed progeny base was not possible based on principles of Mendelian inheritance were included to test whether the resolution script identified these positions and marked the base as unresolved. Examination of the resolved haplotypes indicated that each position from the test sequences had been correctly resolved, demonstrating that the scripts functioned as required.

6.2.3.2 Initial resolution of the TREU 927 haplotypes

Assembly of the two progeny lines provided average genome wide read depths of 41 bases and 42 bases for hybrid 77 and hybrid 86 respectively. SNPs were called as previously described with a minimum read depth of 10 bases required for inclusion. Of the 38428 heterozygous SNPs in TREU 927 haplotype data could be inferred for 77% using the hybrid 77 data and 79% for the hybrid 86 data (Table 6.6). In both cases the majority of unresolved SNPs arose due to low coverage in the progeny or STIB 247 assemblies, with these positions therefore excluded from resolution during filtering.

These initial resolutions however represent the haplotypes inherited by the progeny lines and are therefore affected by meiotic recombination, which must be accounted for in order to generate the true TREU 927 haplotypes. By analysing the large number of progeny generated during creation of the genetic map (MacLeod *et al.*, 2005a, 2005b) it was possible to assign microsatellite alleles to the individual haplotypes of TREU 927 and therefore identify the presence of recombination events on each of the megabase chromosomes which will allow for correction of the SNP haplotypes generated here.

Chromosome	Total number of TREU 927 heterozygous SNPs	Number of TREU 927 heterozygous SNPs resolved utilising hybrid 77 (% of total TREU 927 heterozygous SNPs)	Number of TREU 927 heterozygous SNPs resolved utilising hybrid 86 (% of total TREU 927 heterozygous SNPs)
1	1599	1255 (78)	1129 (71)
2	1480	1081 (73)	991 (67)
3	2750	2035 (74)	2358 (86)
4	2352	1885 (80)	1710 (73)
5	2641	2154 (82)	2079 (79)
6	2859	2294 (80)	2158 (75)
7	2870	2330 (81)	2393 (83)
8	3470	2619 (75)	2688 (77)
9	5959	4295 (75)	5015 (84)
10	5571	4588 (82)	4731 (85)
11_01	5860	4821 (82)	4940 (84)
11_02	755	234 (31)	136 (18)
11_03	264	55 (21)	46 (17)
Total	38430	29646 (77)	30374 (79)

Table 6.6 Initial per chromosome haplotype resolution

6.2.3.3 Validation of SNP haplotypes

The haplotypes resolved by the scripts assume the use of perfect input data, however, as discussed previously when SNPs cannot be called the reference base call for that position is assumed to be correct. While the haplotype scripts attempt to filter this by identification of positions with a read coverage of lower than 20 not all positions will be identified in this manner. In addition the use of a relatively low coverage of 10 bases for SNP calling of the STIB 247 and progeny assemblies increases the probability of errors, introducing both false positives and false negatives. These two sources of error will naturally impact upon the haplotype resolution, leading to incorrect resolutions.

In order to detect these errors, and determine the frequency with which they occur, it is necessary to compare the resolved haplotypes from two progeny strains, hence the use of both hybrid 77 and hybrid 86 in this analysis. In order to prevent meiotic recombination from complicating this analysis only three chromosomes, 2, 3 and 8 were examined. These three chromosomes were selected as the TREU 927 genetic map (MacLeod *et al.*, 2005b) had identified an absence of recombination events in both hybrids 77 and 86. For the purpose of this analysis the two TREU 927 haplotypes shall be identified as H1 and H2, with the pattern of inheritance identified by the genetic map listed as chromosome – haplotype in each case, for hybrid 77 being 2-H1, 3-H2, 8-H1 and for hybrid 86 being 2-H2, 3-H1, 8-H1, indicating that the two hybrids had inherited different homologues of chromosomes 2 and 3 but the same homologue for chromosome 8. In order to assess the level of agreement between the two hybrids the reconstructed haplotypes for each chromosome were compared, with H1 and H2 from hybrid 77 compared with H1 and H2 from hybrid 86. Comparing only positions resolved in both hybrids indicated the reconstructions agreed with one another at 70%, 80% and 97% of positions for chromosomes 2, 3 and 8 respectively. Positions where the reconstructions differed from one another were spread over the length of the three chromosomes with no clustering that would indicate the presence of previously undetected recombination events.

It is possible that the visible disagreements have arisen due to mutation in the progeny lines, or from micro-recombination during meiosis, converting a small region of H1 to the H2 equivalent. However, the most likely source of this disagreement falls to the method by which SNPs calling was handled during this analysis. This explanation is the most likely as it readily explains the increased level of agreement seen between the resolved haplotypes of chromosome 8 compared to that observed for chromosomes 2 and 3. As previously discussed the SNP calling algorithms of SAMtools (Li *et al.*, 2009) do not indicate which positions have failed any of the SNP calling criteria and therefore where this occurs the original reference base call is assumed to be correct. It is the failure to filter these bases that can subsequently give rise to the disagreements observed in

chromosomes 2 and 3 but not in chromosome 8. For chromosome 8 both hybrids have inherited the 8-H1 homologue from TREU 927 and therefore where both progeny lines are marked as being identical to the reference the parental homologues resolve in the same manner (Figure 6.5A). For chromosomes 2 and 3 however the two progeny lines have inherited different homologues from TREU 927 (Figure 6.5B). At positions where both progeny lines agree with the reference sequence the H1 and H2 homologues of chromosomes 2 and 3 therefore resolve differently and generate visible disagreements.

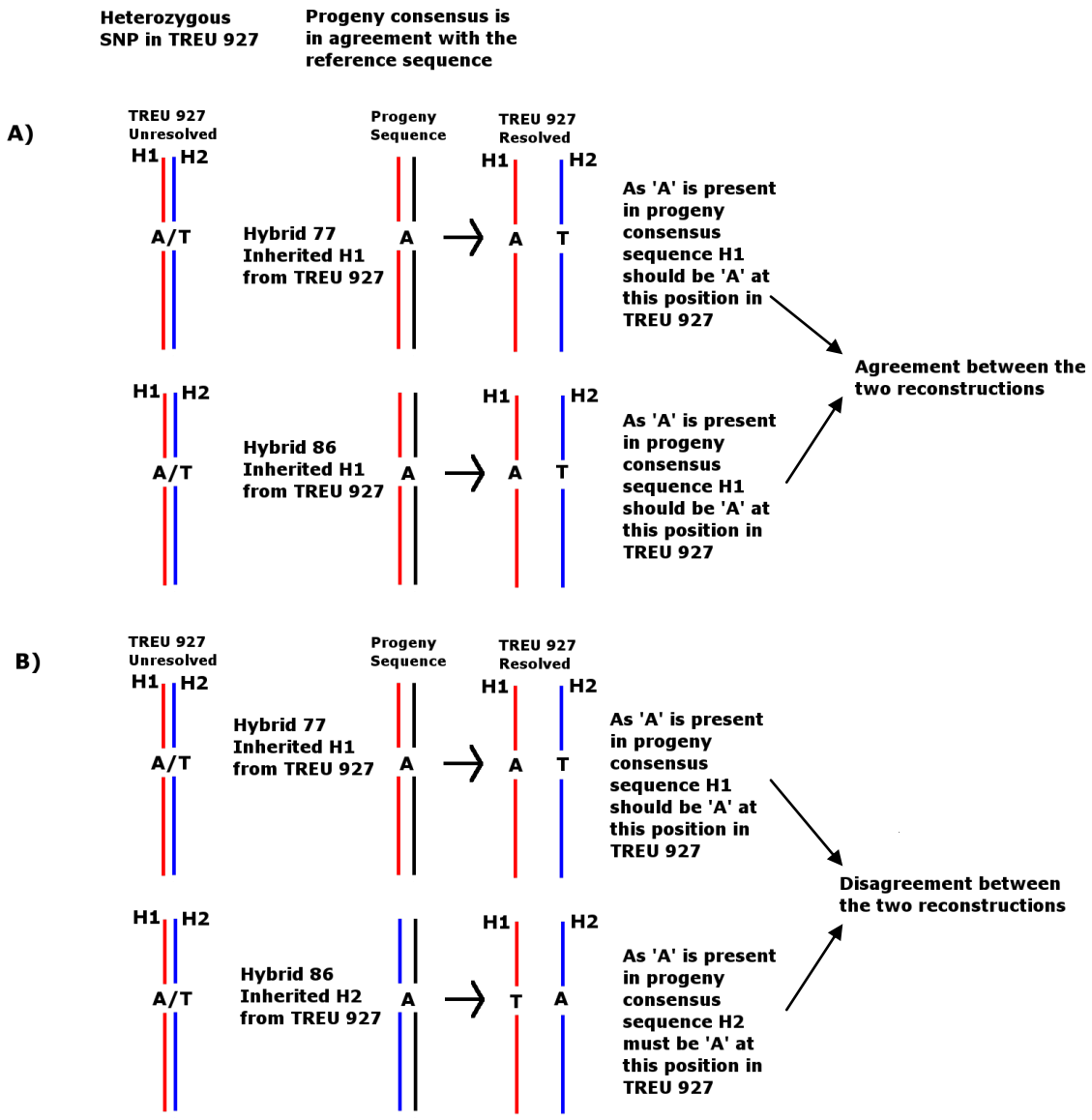


Figure 6.5 Resolution giving rise to incorrect haplotype resolution

Homologues are indicated by colour: red, H1 from TREU 927; blue, H2 from TREU 927; black, homologue inherited from STIB 247. A) Both progeny lines inherited the same homologue, as observed for chromosome 8. B) The progeny lines inherit different homologues, as occurred for chromosomes 2 and 3, giving rise to visible disagreements.

6.3 Discussion

The focus of the majority of this thesis has been the population genetics of the Salivarian trypanosomes, employing microsatellites in the investigation of field populations. The work presented in this chapter, of whole genome sequencing assembly and analysis, lays the groundwork for future trypanosome population genomics studies by providing a collection of SNP variation for three laboratory strains, including both the *T. brucei* genome reference strain TREU 927 (Berriman *et al.*, 2005) and a *T. b. gambiense* Group 2 isolate. These three strains have each been used in earlier genetic studies, including generation of two genetic maps (MacLeod *et al.*, 2005b; Cooper *et al.*, 2008). While SNP variation within laboratory strains will not reflect that present in field populations being aware of the genetic differences presents insight into the variation between these three strains, which will be of use in future laboratory studies.

Through assembly of the TREU 927, STIB 247 and STIB 386 lines it is apparent that the vast majority of inter strain variation is represented not by heterozygosity but homozygous positions comprised of alleles unique to that single line. Given the diverse origins of the three lines the number of differences observed between them is to be expected and the independent origins of these three lines is further emphasised by the low number of heterozygous positions which are shared between strains. These differences clearly demonstrate the need for local sampling during the design of future population studies if technologies such as SNP chips, which are dependent upon datasets of known variation, are to be employed in genotyping. In order to generate a database of truly representative variation it will be necessary to sequence isolates from across Africa, including a diverse range of hosts and ecologies in addition to further laboratory isolates. The inclusion of data from both field and laboratory sources will form an important step in future population studies, by allowing important variation from one to be linked or compared to variation in the other.

While the results of this study has focused upon SNP discovery and analysis this is just a starting step as numerous other forms of variation within and between individuals exists. Genomic rearrangements, indels, loss of heterozygosity, repetitive regions and copy number variation each present additional sources of variation that may have significant impact on the phenotypes of individuals. It will be important to further the analysis presented here by examining each of these in order to further determine the differences between the three strains.

The second half of this chapter has focused upon the reconstruction of the TREU 927 haplotypes through the use of sequencing data and the previously developed TREU 927 genetic map (MacLeod *et al.*, 2005b). With the availability of two sequenced progeny, in addition to the parental

lines it was necessary to employ an inference by inheritance method in order to calculate the TREU 927 haplotypes as inherited by the progeny lines. Subsequent use of the genetic map, which had identified the position of crossover events in each of the progeny lines, allowed for reconstruction of the original TREU 927 haplotypes.

Utilising the data from the two progeny lines it was possible to determine the haplotype associations for 77% (hybrid 77) and 79% (hybrid 86) of heterozygous positions identified in TREU 927. However comparison of chromosomes 2, 3 and 8 indicated that up to 30% of the resolved positions disagreed between the two reconstructions, likely due to the inclusion of low quality positions, which had been filtered during SNP calling but had not been filtered by the haplotype resolution scripts. With further development of the haplotype resolution script it should be possible to identify these positions and filter them properly, thereby increasing the accuracy of the resolved haplotypes. However this increased accuracy will be achieved by filtering out positions and therefore will reduce the total number of positions resolved by the scripts. In order to increase the proportion of the TREU 927 heterozygous positions which can be resolved it will be necessary to undertake further sequencing, primarily of the progeny lines but also of the parental STIB 247 line.

Additional sequencing will have two benefits. The first is that with additional sequencing an increased number of positions will meet the criteria for SNP calling and haplotype resolution, thereby increasing the total number of heterozygous positions resolved. In order to make full use of this increased read depth it will be necessary to further develop the haplotype resolution scripts to utilise the same filtering techniques employed by the SAMtools SNP calling algorithms (Li *et al.*, 2009). The second benefit to additional sequencing arises from the increased accuracy available during SNP calling. While there is no set read depth required for SNP calling a depth of 20 reads is typically considered a minimum for accurate calling. Due to the relatively low coverage of both the progeny lines and STIB 247, SNP calling of these assemblies used a minimum read depth of 10 bases and therefore the accuracy of SNP calling is lower than ideally desired. Additional sequencing will therefore increase the accuracy of SNP calling and subsequently increase the accuracy of haplotype resolution.

Beyond additional sequencing of the strains already described here an alternative method by which the SNP haplotype map could be improved is the sequencing of further progeny lines. Sequencing of further progeny lines has the advantage of resolving positions that could not be resolved with the existing data while increasing the accuracy of the resolution by allowing for generation of a consensus haplotype sequence. This approach has also the benefit of allowing for the identification of additional recombination events that lie between the microsatellite markers employed in the original map.

The results presented in this chapter represent an initial but important step in shifting from a low density approach employing a small number of microsatellite markers. While not the only source of genomic diversity, SNPs are perhaps the one which is most widely utilised outside of genomics studies simply due to their role in altering the expression and function of genes. Provision of genome wide SNP coverage allows for easy and quick access to this diversity, as opposed to having to sequence every gene of interest in order to uncover variation. As additional laboratory strains are sequenced, collections of genomic variation will allow for a greater understanding into the differences that give rise to the distinct and often different phenotypes observed for the most widely used laboratory strains. The future sequencing of field samples will further extend our understanding of diversity in the trypanosomes, and the variation identified here will play an important role in linking field and laboratory studies, a necessary step in truly understanding these parasites.

Chapter 7

Final Discussion

The population genetics of trypanosomes has historically been a controversial issue in the literature, originating primarily with the conclusion that effective clonality was the dominant state for the populations of a wide range of parasitic protozoa and eukaryotic microorganisms (Tibayrenc *et al.*, 1990, 1991), despite the earlier experimental demonstration of genetic exchange occurring during tsetse transmission of *T. b. brucei* (Jenni *et al.*, 1986). The study of population genetics of *T. brucei* (the best studied of the Salivarian trypanosomes) has been complicated further by a range of factors: the existence of multiple sub-species and subgroups for *T. brucei* with the potential for genetic isolation or interaction (Hide *et al.*, 1994; MacLeod *et al.*, 2001c); the differing roles of human and animal host reservoirs (Paindavoine *et al.*, 1986a; Welburn *et al.*, 2001; Njiru *et al.*, 2004b; Enyaru *et al.*, 2006; Cordon-Obras *et al.*, 2009) and the observed non obligatory nature of genetic exchange in *T. brucei*.

The role of genetic exchange in the non human trypanosomes *T. congolense* and *T. vivax* has until recently been uncontested simply due to the relative lack of interest in these pathogens and until recently the conclusions of clonality were based upon the analysis of data originating from the 1970s and 1980s (Kilgour *et al.*, 1975; Kilgour and Godfrey, 1977; Allsopp and Newton, 1985; Gashumba *et al.*, 1988) based on either a small number of available isoenzyme markers or the use of samples collected from a wide geographic range and over long periods of time. This can lead to misleading conclusions as to the role of genetic exchange within the populations, primarily due to a loss of power in statistical tests and the introduction of the Wahlund effect due to the mixing of subpopulations. The lack of investigations into the animal infective trypanosomes is surprising given the economic impact of AAT; their distribution across the entirety of the tsetse belt and spread of *T. vivax* to South America, with millions of livestock infected over the two continents. The diminished economic return from infected livestock has a significant downstream impact on the individuals reliant upon these animals, many of whom are resident amongst the poorest communities on the planet. A greater understanding of animal trypanosomiasis will therefore benefit not only the infected animals but aid in raising the prospects of the owners and wider communities.

The primary aim of this thesis has been to investigate the role of genetic exchange in the Salivarian trypanosomes with the specific inclusion of the animal infective *T. congolense* and *T. vivax* species which remain sorely understudied. In order to achieve this, populations from The Gambia, Uganda and Malawi were examined through the use of single copy, microsatellite markers specific to each species. Through these markers the diversity of the respective populations, the relationships of individuals within and between them has been assessed, allowing for inferences on the role of genetic exchange within them. The animal infective species of *T. vivax* (Chapters 3 and 4) and *T. congolense* (Chapter 4) in The Gambia were investigated utilising samples from horses, donkeys and cattle collected in 2007 and 2009, allowing for assessment of the relationships between the

species and role of genetic exchange within each. Meanwhile five populations of *T. b. rhodesiense* (Chapter 5) from Uganda and Malawi have been examined, allowing for a unique comparison of populations separated by geography (Uganda – Malawi) and also within a single country as the disease has spread over time (Uganda 1961 – 2010). The role of technological advancement however means that new techniques are constantly becoming available; as such the final portion of this work (Chapter 6) focuses on the assembly and analysis of whole genome sequence data of three laboratory strains and reconstruction of the megabase chromosome haplotypes for TREU 927, the *T. brucei* genome reference strain. This discussion will draw together the results presented across these four results chapters in order to discuss the broader picture of mating in the Salivarian trypanosomes and the future direction of population studies into these parasites.

At the core of genetic exchange is the mating system, which had remained until recently a black box. While genetic exchange has yet to be directly observed, details of the process are being filled in using a number of different approaches. Fluorescent tagging techniques (Bingle *et al.*, 2001; Gibson *et al.*, 2006, 2008; Peacock *et al.*, 2007) have allowed for the direct observation of hybrids in the salivary glands of tsetse flies, providing greater detail into the timings, clustering and frequency of mating as well providing evidence for selfing during single strain transmissions (Peacock *et al.*, 2009). Through bioinformatics it has also been demonstrated that the potential for meiosis is an ancestral trait amongst the kinetoplastids, with conserved meiotic genes having been previously identified in *T. brucei*, *L. major* and *T. cruzi* (El-Sayed *et al.*, 2005), *T. congolense* (Morrison *et al.*, 2009b) and now *T. vivax* (Chapter 3).

The expression of three of these meiosis associated genes, *mnd1*, *dmc1* and *hop1* has been recently demonstrated through fluorescent tagging in *T. brucei* (Peacock *et al.*, 2011), providing the greatest insight to date into the molecular events surrounding meiosis in trypanosomes. This work demonstrated that meiosis is initiated prior to the fusion of cells, with expression of the three tagged genes observed to occur in the same order as other eukaryotes. The majority of cells did not fuse with another during expression of these genes, however, a small proportion appeared to have fused during, or prior to, the expression of *hop1*. It is possible that these early fusion events give rise to the aneuploidy observed in some trypanosome hybrids. However, it is also possible that these cells are inviable and subsequentially removed from the population of the salivary glands. While this work demonstrated the initiation of the meiotic cycle, attempts to identify haploid cells were unsuccessful, which could indicate that haploid cells are not produced as part of the trypanosome meiotic cycle or that these cells are highly transient in nature. It therefore remains possible that fusion occurs between two haploid cells or two cells each possessing two haploid nuclei.

The work of Peacock *et al.*(2011) was undertaken using strains of *T. b. brucei* and *T. b. gambiense* Group 2 and therefore further work will be required in order to determine whether these genes are

also functionally expressed in *T. b. gambiense* Group 1, *T. congolense* and *T. vivax* and the other kinetoplastids. Genetic exchange in *Leishmania sp.* bears strong similarities to that of *T. brucei,* occurring during development in the vector with biparental inheritance of chromosomes and the observation of raised DNA contents in some but not all progeny lines (Akopyants *et al.*, 2009), suggesting *Leishmania sp.* may also make use of its ancestral complement of meiotic genes. *T. cruzi* in contrast appears to use a very different, non Mendelian system with fusion of diploid cells and random chromosome loss (Gaunt *et al.*, 2003). Given the ability of evolution to adapt existing systems to new purposes, it remains to be seen whether the classical meiotic genes have been co-opted for use by this alternative system of mating.

The development and widespread use of microsatellite markers has provided the clearest evidence as to the role and frequency of mating in the populations of kinetoplastids by providing hypervariable, selectively neutral markers that are sensitive enough to identify and type infections without the requirement for laboratory adaptation and growth. From these markers the picture that is emerging is one of complexity and variation, with population structures ranging from near panmixia (*T. b. rhodesiense*) and frequent inbreeding (*T. congolense*, *Leishmania sp.*); epidemic (*T. b. brucei*) and finally strict or effective clonality (*T. b. gambiense* Group 1, *T. b. rhodesiense, T. cruzi*, *Leishmania sp.*, *T. vivax*). For *T. b. rhodesiense*, it is now clear that the non obligatory nature of genetic exchange in *T. brucei*, and the unknown factors controlling the frequency of genetic exchange, can have significant effects on the dynamics of *T. b. rhodesiense* populations, giving rise to both effective clonality and frequent genetic exchange (Chapter 5). With this observation of a variable population structure for *T. brucei* further laboratory and population based investigations into the other species, most notably *T. congolense* and *T. vivax*, are required in order to determine whether the existing results are representative of the African trypanosomes as a whole.

From the studies presented here it is apparent that further research is sorely needed in each species, albeit for very different reasons. Chapters 3 and 4 demonstrated the difficulty in working with *T. vivax*, a fact that has long been noted in the literature (Gardiner, 1989) and is a strong contributor to the lack of knowledge about this species. While the prevalence of this species was over 80% in both the 2007 and 2009 equine populations the single locus microsatellite genotyping was successful in less than 10% of the animals identified as *T. vivax* positive, despite the use of nested PCR protocols and collection of 2 ml of whole blood during sampling in 2009. These failures could have arisen due to a number of factors, the simplest being that the parasitaemia in infected animals is low enough to prevent amplification with single copy markers. If this is the case then it would suggest that *T. vivax* in The Gambia is characterised by either very low parasitaemia infections or that the hosts in the region are subject to much higher levels of exposure to *T. vivax* than *T. brucei* or *T. congolense*. It is also possible that the high prevalence of *T. vivax* but low genotyping success resulted from false positives. These could potentially arise through cross reactivity with host DNA,

although sequencing of the amplified material suggests this is not the case, or due to the presence of residual trypanosome DNA in the host originating from challenge that did not lead to infection. Alternatively the use of a Kenyan *T. vivax* isolate in the design of the microsatellite markers may have led to the introduction of null alleles which could not be amplified with the primers utilised here. Molecular techniques such as those employed here are commonly utilised in the examination of field samples and it is therefore important that these issues are resolved in order to obtain accurate estimations of the prevalence of *T. vivax*. Further assessment of infections in equines not brought to clinic, cattle, tsetse and biting flies from the region will shed further light on the epidemiology of *T. vivax* in addition to the existence and potential role of these low parasitaemia infections.

The need for further investigations into *T. congolense* arise for a very different reason from that of *T. vivax*, namely that genetic exchange may be a frequent occurrence in the species with populations defined by a high level of inbreeding. If genetic exchange is truly occurring on a regular basis it should be possible to confirm its existence in the species, and through use of the tools already developed in the analysis of *T. brucei*, gain insight into the processes shared with, and different from, mating in *T. brucei*. Adaptation of the fluorescent tagging techniques already employed in *T. brucei* should be capable of not only confirming the occurrence of genetic exchange but will allow for the rapid identification of the developmental stage in which it occurs. For *T. congolense* these studies may be aided by the development of culture techniques for each of the life cycle stages (Coustou et al., 2010) which may eliminate the requirement for transmission through tsetse flies.

The extensive homozygote excess in *T. congolense* and low number of repeated MLGs, as observed here and in the earlier 2007 population (Morrison *et al.*, 2009b) suggest that genetic exchange is a frequent occurrence in this species, with the frequency of mixed infections in the tsetse controlling the relative rates of selfing and outcrossing. While our knowledge of the processes controlling genetic exchange are limited, at present, modelling approaches may be able to shed light on the frequency of genetic exchange, outcrossing, selfing and mixed tsetse transmissions required in order to observe the population structures described in this thesis. Such approaches would allow for the generation of testable hypotheses regarding the population dynamics of trypanosomes which could be compared to field populations of interest.

From the studies of Uganda and Malawi the most pressing questions are those of the frequency of genetic exchange in other foci of *T. b. rhodesiense* and the mechanisms controlling its frequency. Of the *T. b. rhodesiense* populations investigated to date that of Malawi is the first to show evidence of frequent genetic exchange and shows stark differences to those of Uganda, most notably of a high level of intra-population diversity and only two repeated MLGs within the

population. The Ugandan populations, in contrast to Malawi are much more closely related to one another, with clonal expansion of select genotypes common. This raises the possibility once again that *T. b. rhodesiense* may possess mechanisms of self incompatibility or of distinguishing self from non self during tsetse transmission (Gibson *et al.*, 1997), limiting the potential for successful inbreeding and requiring a sufficient level of diversity and mixed genotype tsetse infections in order for detectable levels of genetic exchange to occur. Repeated clonal expansion of genotypes, while potentially favouring the fittest genotypes within the population could repeatedly select for a small number of mating types, limiting the potential for genetic exchange until further diversity is introduced or evolves. The founder effects observed in Soroti and Kaberamiado could further enhance this selective force through the establishment of new foci essentially composed of a single self incompatible genotype, effectively preventing mating in the long term. That selfing has been observed during single strain transmission (Peacock *et al.*, 2009) may instead indicate that the dynamics of the mating process in *T. brucei* may act to limit but not completely block selfing in *T. brucei*.

T. congolense, in contrast appears to allow for high levels of inbreeding, which suggests the two species may have taken different approaches to the role of genetic exchange. Elucidating the factors that control the frequency of genetic exchange however is likely to be a difficult process, if only due to the potential for multiple interacting variables and our current lack of knowledge concerning the mechanism of genetic exchange itself. The observation of effective or strict clonality in *T. cruzi*, *T. b. gambiense* Group 1 and *T. vivax* would appear to represent one extreme of the spectrum and could have arisen through simple chance mutations which were sufficient to disrupt genetic exchange. In order to be maintained within the environment it is likely that such a mutation would have needed to occur upon a high fitness genotype, capable of outcompeting mating strains long enough to become established as a new species or sub-species. Alternatively it is possible that a situation similar to that which may be present in Uganda, of low population diversity and self incompatibility has favoured such a mutation by removing the selective pressure to maintain a functional meiotic system.

As the factors controlling meiosis remain unclear at present it is also possible that environmental variables, such as the vector or host species, have played a role in the evolution of clonality. The Salivarian trypanosomes are capable of infecting a wide range of tsetse species, however, it is possible that only some are conducive to mating. Long term association of a species, such as *T. vivax*, with a vector incapable of supporting genetic exchange could have led to loss of mating in the species by removing the selective pressure to maintain functional copies of the genes required for meiosis. Over shorter time frames vector imposed constraints could lead to temporary clonality within individual foci. This may explain the differences observed between the population structures identified in Uganda and Malawi if the vectors of the two foci differ enough to respectively prevent

and facilitate genetic exchange. In Uganda however the related sub-species *T. b. brucei* has been observed to possess an epidemic population structure, with genetic exchange masked by clonal expansion (MacLeod *et al.*, 2000). If the tsetse species present within a focus are responsible for determining whether genetic exchange is occurring, this would suggest that either the two sub-species respond differently to the same vector or that they may be spread by different vectors with that transmitting *T. b. rhodesiense* not conducive to genetic exchange.

The kinetoplastids, with their diverse range of reproduction strategies appear to inhabit the middle ground between clonality and sexuality, with the many species positioned at different positions along the spectrum. *T. brucei*, the archetype with its non-obligatory mating system appears to be playing the system, by attempting to get the best of both worlds. By possessing the potential for sexual reproduction trypanosomes retain the ability to shuffle their genetic material into new combinations by first meiotic recombination and subsequently by combining material from two parental cells, allowing for the mixing of lineages which can bring together traits that would be forever isolated in strictly clonal organisms. The shift towards clonality through the presence of the non-obligatory mating system, while reducing the frequency with which the genome is mixed with that of the local population may fulfil the important function of ensuring the continuation of the fittest genotypes. Under such a system it could be envisaged that within a population individual genotypes could be maintained over long periods of time, with occasional mating events giving rise to new combinations which could be selected against; maintained at low frequency; or selected for, becoming the new dominant genotype through clonal expansion. With the correct balance of conditions multiple short term lineages could thus be maintained within a population, with a level of self recognition or self incompatibility limiting the potential for inbreeding. That *T. cruzi* retains the ability to undergo even rare genetic exchange attests to the advantages offered by even occasional mixing of the gene pool and further investigation will be required in order to determine whether *T. b. gambiense* Group 1 and *T. vivax* also possess the capability for rare mating events.

The final section of this thesis has focused upon the genomic analysis of three *T. brucei* laboratory lines and genome wide reconstruction of the chromosomal haplotypes of TREU 927, the *T. brucei* reference strain (Chapter 6). Genome sequence data are increasingly taking on a central role in molecular biology be it merely for simple gene sequence retrieval, identification of mutations following the application of selective pressures or for genome wide association studies. At present however there are no publically available resources collecting together the genomic variation of the Salivarian trypanosomes and even the sequence describing the genome of the reference strain, TREU 927 is devoid of internal SNP variation. The work here, in examining three common laboratory strains, including TREU 927, provides an initial insight into genome wide variation, both between the three strains and within each strain. The data generated form a valuable resource into trypanosome sequence variation and its distribution within the genome of these strains.

However with the widespread use of numerous laboratory strains, each with their own distinct origin, genetic background, complement of SNPs and unique laboratory phenotypes it will be necessary to add further sequencing to this resource in order to understand the effect of strain specific SNP combinations.

The second half of the genomics studies focused upon the reconstruction of the chromosomal haplotypes of TREU 927, the *T. brucei* genome reference strain. Combined with the SNP data and analysis this provides a high density SNP map for TREU 927, allowing for the identification of the sequence of individual alleles of features of interest and in doing so makes it possible to easily identify cis acting variation that may give rise to allele specific effects across the entire genome. Combined with the earlier microsatellite map (MacLeod *et al.*, 2005b) this SNP map allows for fine tuning of quantitative trait loci mapping through the genotyping of individual SNPs in regions identified as contributing to phenotypes of interest. The genomics work, however, has focused solely upon SNP variation and as such is only the first step in analysing the trypanosome genomes as other sources of variation also play significant roles in controlling the biology and observable phenotypes of organisms. Future work will therefore need to investigate these, in particular the role of indels, genomic rearrangements and copy number variation. In focusing upon the megabase chromosomes this work has also excluded analysis of the kinetoplast, intermediate and mini-chromosomes which will need to be included in future work before we can fully understand genomic variation within the Salivarian trypanosomes.

While this application of whole genome sequencing was not employed in the analysis of field samples the future of population studies is undoubtedly that of population genomics, utilising high throughput techniques in order to type populations with thousands of markers across the entire genome. In order for such genomics studies to be undertaken with trypanosomes a number of barriers will first need to be overcome, predominantly focused on the sampling of trypanosomes from natural populations. The primary issues currently faced are separation of parasite and host DNA; collection of sufficient parasite DNA for sequencing and the presence of mixed trypanosome infections when sampling from animals. While infections of laboratory animals can reach extremely high parasitaemias naturally occurring infections are typically at far lower parasitaemias, dependent upon both host and parasite species. Further in natural infections the number of trypanosomes per ml of blood is typically far lower than the number of host white blood cells per ml of blood which coupled to the differing genome sizes of host and parasite the total concentration of trypanosome DNA is likely to be many orders of magnitude smaller than that of the host.

The second barrier to population genomics, collection of sufficient material is directly related to the first issue as even if pure trypanosome DNA can be separated it is unclear how much blood would need to be sampled to provide the material required for whole genome sequencing. This is of

particular concern with species associated with the lowest of parasitaemias, namely *T. b. gambiense* Group 1 and *T. vivax.* The failure to genotype the majority of the *T. vivax* positive samples with the single copy microsatellite markers (Chapters 3 and 4) emphasises the difficulty that genome sequencing would face. In face of these low parasitaemias parasite concentration methods will be required in order to isolate all the trypanosomes present in large volumes of blood. The most promising technique currently available is mini anion exchange centrifugation technique (Lumsden *et al.*, 1979; Büscher *et al.*, 2009), which should filter out the majority of host DNA from samples, however this brings with it its own issues, notably difficulties in use under field conditions, and total DNA yield. Additional studies will therefore be required in order to determine the size of the columns required in order to obtain sufficient quantities of trypanosome DNA from field infections and whether host DNA is sufficiently excluded.

The issue of mixed infections presents a further barrier that may hinder future population genomic studies. Mixed genotype infections of a single species are particularly difficult to identify in SNP studies as the majority of SNPs are represented by only two alleles present as either their respective homozygous forms or the combined heterozygous SNP. This is in contrast to microsatellite alleles where multiple alleles are common; facilitating the identification of mixed genotype infections and panels of these markers may be required in order to screen samples for mixed infections prior to sequencing. A further issue is the role of mixed species infections when sampling from non-human hosts, which may further complicate matters if reads align to multiple reference sequences.

If these difficulties, and those more widely associated with the collection of samples from the field, can be overcome then trypanosome genomic studies will be well placed to substantially increase our knowledge of these parasites. Population genomic studies will vastly increase the power available to determine the frequency of mating and the relationships between trypanosomes within a population by sampling the majority of genomic variation, allowing for even highly related individuals to be distinguished from one another. By sampling diversity across the entire genome it will be possible to examine linkage disequilibrium along the length of any given chromosome, hampered at present by the employment of only low number of genetic markers which are typically spread over different chromosomes. The ability to test for linkage along chromosomes allows for the detection of recombination events, a key marker of meiosis and thus genetic exchange and with sufficient sampling there will come the ability to calculate population level haplotype blocks and identify the location of recombination hotspots within the genome.

Beyond the investigation of genetic exchange genomics studies will play a significant role in many laboratory and field studies. As the cost drops it will be feasible to sequence all lines of interest in order to identify novel variations, such as following selection for particular phenotypes. Field studies will also benefit greatly from the increasing availability of genome sequencing. Genome

wide association studies will allow for the identification of important variation in not only the parasites but in hosts and vectors in order to determine the gene variants responsible for clearance of, control of, or susceptibility to infection. The field of phylogenetics will likewise benefit, by increasing the power of the techniques used, allowing for identification of more subtle evolutionary relationships between individuals. This will in turn allow for comparative studies of chromosomes and regions within them, making it possible to investigate their individual evolutionary histories. Finally by comparing the nucleotide diversity of regions of interest to the genome as a whole it will also be possible to identify regions under selective pressure, through the presence of highly conserved sequences between species and geographically separated isolates or through alterations to population level heterozygosity due to purifying or diversifying selection.

The work in this thesis, in focusing upon the population genetics of the African Salivarian trypanosomes has demonstrated the variability of the mating strategies of employed by *T. b. rhodesiense, T. congolense* and *T. vivax* and shown that even within a single sub-species there is a range in the frequency of genetic exchange. Such a variable mating system is likely to have had a considerable impact on the evolutionary history of these species, further complicated by the founder effects associated with spread of the disease and potential for highly structured populations controlled by the interactions of vector, host, parasite and environment. That the role of genetic exchange in determining the epidemiology of HAT and Nagana warrants further study due to the clinical implications of such a system goes without question. The positioning of the trypanosomes in the boundary between clonality and sexuality however means that the species may be ideally suited to furthering our knowledge of the evolutionary role and selective pressures associated with genetic exchange and recombination, a role which in and of itself is worthy of additional study. The future role of genome sequencing in such studies is unquestionable and as these technologies become available it will be necessary to ensure the knowledge derived from prior studies continues to be employed as population genetics continues to transform into population genomics.

Appendices

Species / clade	Primer ID	5’ – 3’ Primer Sequence
T. brucei	TBR-1	GAATATTAAACAATGCGCAG
	TBR-2	CCATTTATTAGCTTTGTTGC
T. congolense Forest	TCF-1	GGACACGCCAGAAGGTACTT
	TCF-2	GTTCTCGCACCAAATCCAAC
T. congolense Savannah	TCS-1	CGAGAACGGGCACTTTGCGA
	TCS-2	GGACAAACAAATCCCGCACA
T. vivax	TVW-1	CTGAGTGCTCCATGTGCCAC
	TVW-2	CCACCAGAACACCAACCTGA

Appendix 1

Primer sequences for the species specific primers used in identifying the presence of trypanosome infections in field samples, originally published by Masiga *et al.*(1992).

Microsatellite ID	Genome contig (size)	Primer ID	5’ – 3’ Sequence
TV3	Tviv1699d03.p1k	TVN3A	ATGTTTATACGCCGCTTGA
	TVIV.0.298087	TVN3B	GTTATGCTCGTCACTTGTT
	(2,222 bp)	MTV3A	GAACATACCATTATATTCACTGC
		TV3B	TTGGTGGTGTGCTGTCATT
TV4	Tviv1189b11.p1k	TVN4A	CTTCTCCATCTGTGCATGAG
	TVIV.0.311716	TVN4B	TAGCATCAGATAGGCGTATTAT
	(2,340 bp)	MTV4A	TGTTGTTGTTGTAGTGAAGGT
		TV4B	TATACGTGGCAGACAGATAA
TV6	Tviv1381c08.q1k	TVN6A	TGTTCTACATGCTTGGAGGT
	TVIV.0.2778	TVN6B	TATGCGTTCACAACACTCTG
	(2,548 bp)	MTV6A	CGTTCTGGGTGAAGTGACA
		TV6B	ACAGTACGGACAAGGACAC
TV14	Tviv1738g11.q1k	TVN14A	GACGCAATCACATGAACATA
	TVIV.0.24886	TVN14B	CCTCCAGCTACGATCAAGT
	(3,017 bp)	MTV14A	GGTGTGCCATAATATACAGCA
		TV14B	CTGGATTGAAGCACACTAAC
TV17	Tviv1224g12.q1kw	TVN17A	CCAGATGTATAAACAGAGACTT
	TVIV.0.49233	TVN17B	CTACCTGTTCCTTCGCATT
	(3,211 bp)	MTV17A	CACCTAACAGAGTACACCAA
		TV17B	ATCTCAGTTCATTACATACATCT
TV24	Tviv1975d12.q1k	TV24A	AGTGTCGTCAAGGTACTGA
	TVIV.0.43146	TV24B	TCTGTTAATGAGGTCCACACT
	(4,037 bp)	MTVN24A	TTCTCGTACCCACTCTGAA
		TVN24B	TTGGTGAACAGATTGAAGC
TV31	Tviv1905b03.p1k	TVN31A	GTCATATCAGAAGAGTGGTGT
	TVIV.0.253449	TVN31B	ACGAGATACGTCGGTGGAT
	(35,758 bp)	MTV31A	CGACGGTAAGTTAGTGACA
		TV31B	TGAATGAGGAACATGGAGTTG
TV49	Tviv827e12.q1k	TV49A	CAGTCAAGCCTAATGATCCTC
	TVIV.0.221992	TV49B	GTGAGGTTACGAGGACAGA
	(23,506 bp)	MTVN49A	TTCTTTGCTGCCTTGTACTG
		TVN49B	GGCAAGTAACTGTCGTAGA

Appendix 2

Primer details for the genotyping of *T. vivax* microsatellites, originally published in Duffy *et al.*(2009). Sequences denoted in italics represent the second round primers for use in nested PCR. Presence of M at the start of primer ID indicate these primers bore a 5' FAM or HEX modification.

Microsatellite ID	Primer ID	5’ – 3’ Sequence
TCM-1	TCM1A	TACAAATGACTGTAGAGCGGC
	TCM1B	CTGTGTGTATAATGATTCATTCG
	TCM1C	CTAGAAGCGAGTAACAGCC
	TCM1D	AAGGGTTCGTACCACAGCCC
TCM-2	TCM2A	GGTAAGACAAAGTTGTGGGTG
	TCM2B	ATGTGACCGATGCTCCGAAC
	TCM2C	CAGTCATGTATATGTTTGTG
	TCM2D	CCTGAAATGGGTCTACTGAG
TCM-3	TCM3A	TCTATTGTTCACGTCTCGTG
	TCM3B	ACTCATTGCATAAAGGCTAG
	TCM3C	CATGCTCTTAGGTTCCATCGG
	TCM3D	AGCATCCGACATTGAAACGAC
TCM-4	TCM4A	CTTAACGCTGCTTCAGTAGC
	TCM4B	AGTACACACGACTTCACCTCC
	TCM4C	GTCTCTTTCCGCACAGTGAC
	TCM4D	GGGGGAAGATATTAAAGACAC
TCM5	TCM5A	CAATGGTTCAATAAGCGCACC
	TCM5B	AAGGCAAGTAAGTTACGC
	TCM5C	CTTCCACGAGTCCCTAATCGAC
	TCM5D	TTGCTCACTGTCAAGGCGTGC
TCM-6	TCM6A	GAATGCGAGACCTGCTTCTTGG
	TCM6B	CATTTAGACTCTCACTTTCCG
	TCM6C	AACCACCACTTCCGTGCACCGG
	TCM6D	CCATGAGCTTTATGCGACCTCTAC
TCM-7	TCM7A	GTGTAGTTTGTTATACTTCG
	TCM7B	GTTAAATACTTGTGAGAGCCAGC
	TCM7C	TCATAGAGGCAAGTGCGTAGC
	TCM7D	CCAGAATAAGAATACTTACTGC

Appendix 3

Primer details for the genotyping of *T. congolense* microsatellites, originally published by Morrison *et al.*(2009b).

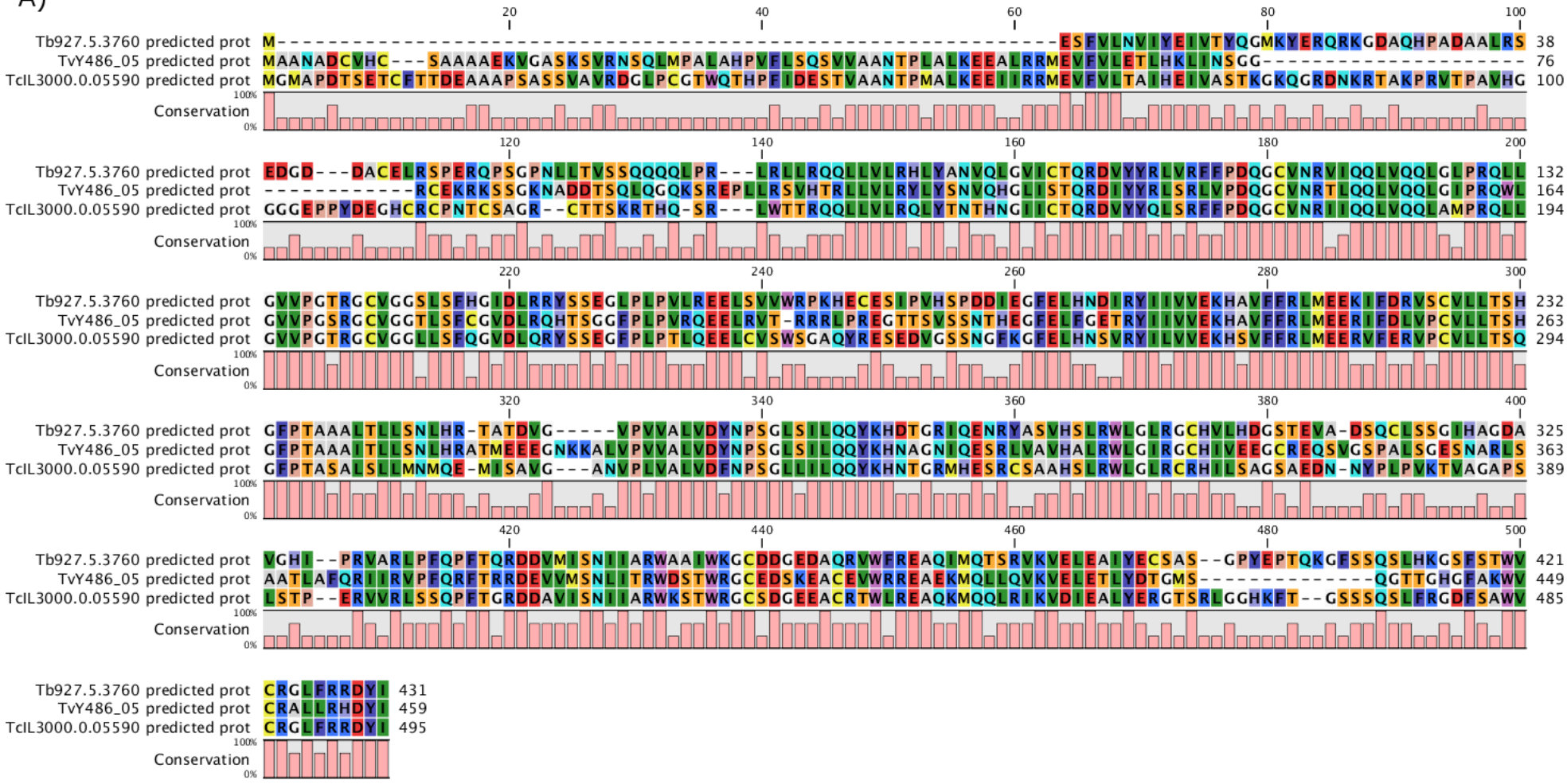
Microsatellite ID	Chromosome	Primer ID	5’ – 3’ Sequence
Ch1/18	1	CH1/18-C	TATAATGCGTTTGTGAGAAT
		Ch1/18-D	GAAGGGAGGGAACAGAAGCAGGG
		CH1/18-A	TGTGAGAATGGTACTCACGCGCTG
		CH1/18-B	ACAACGTTAGCACACAATTCCTGTG
Ch2/PLC	2	CH2/PLC-G2	TTAAGTGGACGACGAAATAACAACA
		CH2/PLC-H4	TTCAAACACCGTCCCCCTCAATAAT
		CH2/PLC-G	CAACGACGTTGGAAGAGTGTGAAC
		CH2/PLC-H3	CCACTGACCTTTCATTTGATCGCTTTC
Ch3/5L5	3	CH3/5L5-AA	GAGCGTACATTGCAGGTAGTGCGTAGCG
		CH3/5L5-B	GGAAACTGCTTAAACTTGCGTGAG
		CH3/5L5-A	GTACGTGGTTAACCACAACCTACT
		CH3/5L5-BB	GTATTTTTCATGGCACACAACATAT
Ch3/IJ15/1	3	Ch3/IJ15/1-C	AGGCTTAGACGAGTGTCAGG
		Ch3/IJ15/1-D	GTAAATAGACACAGTGAAACCG
		Ch3/IJ15/1-A	GTTAGGTTACGCAAGTCAGT
		Ch3/IJ15/1-B	GAAACACTCAGTTCCACACC
Ch4/M12C12	4	CH4/M12C12-B	TACCCTCATCAAGTGGTCG
		CH4/M12C12-C	AAAACCTCATCCAGTCGCACTGG
		CH4/M12C12-A	TGGACACACAGAAGCCTACCG
		CH4/M12C12-D	AGTGTGGTGGTGCGTGCAAACTTGG
Ch5/JS2	5	CH5/JS2-C	AGTAATGGGAATGAGCGTCACCAG
		CH5/JS2-D	GATCTTCGCTTACACAAGCGGTAC
		CH5/JS2-A	GATTGGCGCAACAACTTTCACATACG
		CH5/JS2-B	CCCTTTCTTCCTTGGCCATTGTTTTACTAT
Ch9/4	9	Ch9/4-C	CATCGATGAGAAGTACACTG
		Ch9/4-D	AACAGACTAGGAAAGTATAC
		Ch9/4-A	GTGGAGGAGTGCTGATGA
		Ch9/4-B	ATGTAAGATATTAGAGCAGTAAA

Appendix 4

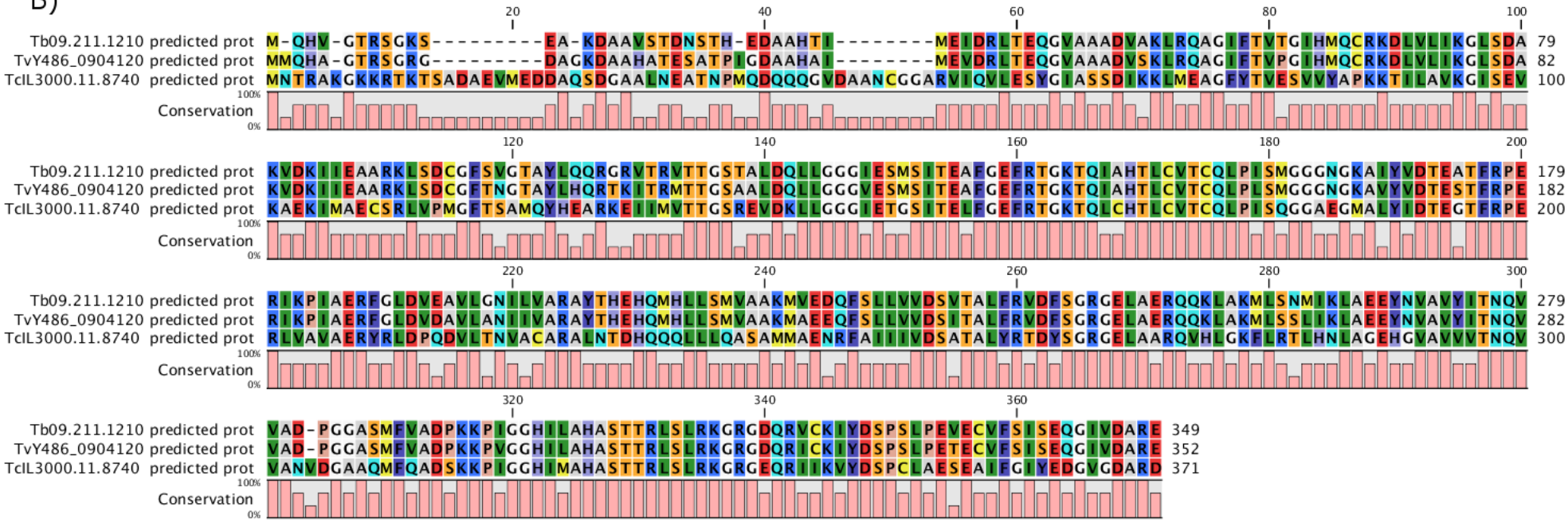
Primer details for the genotyping of *T. brucei* microsatellites, originally published in (MacLeod *et al.*, 2005b).

Appendix 5

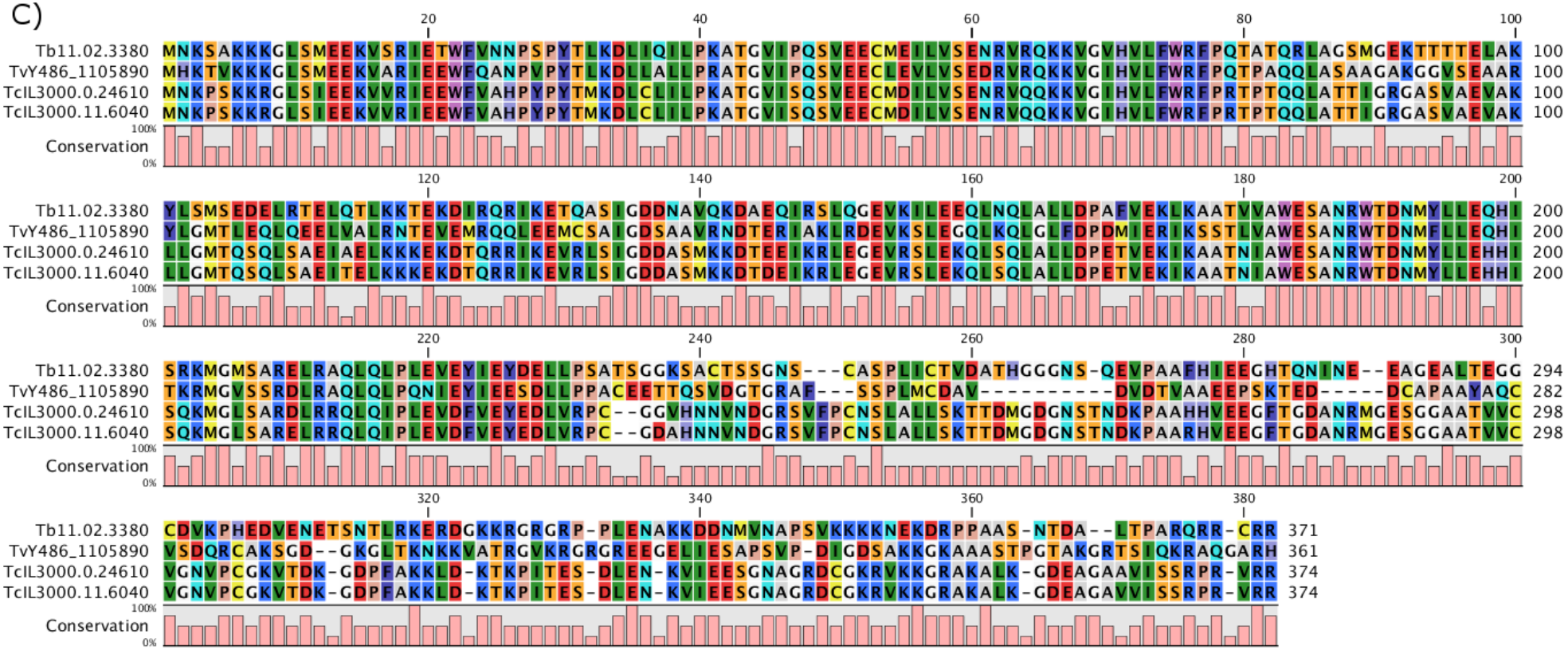
A)

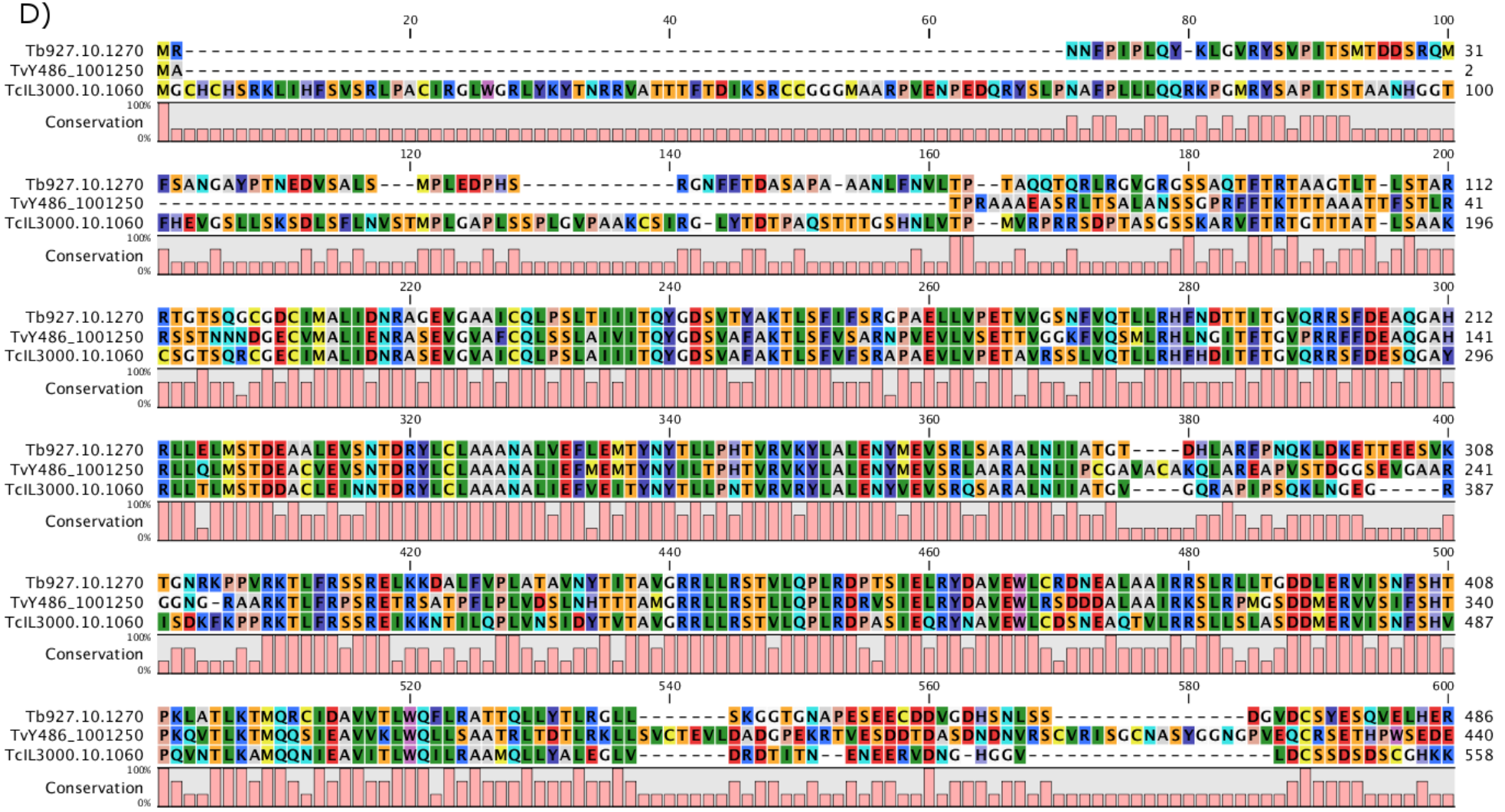

Tb927.5.3760 predicted prot
TvY486_05 predicted prot
TcIL3000.0.05590 predicted prot
Conservation
100%
0%

B)

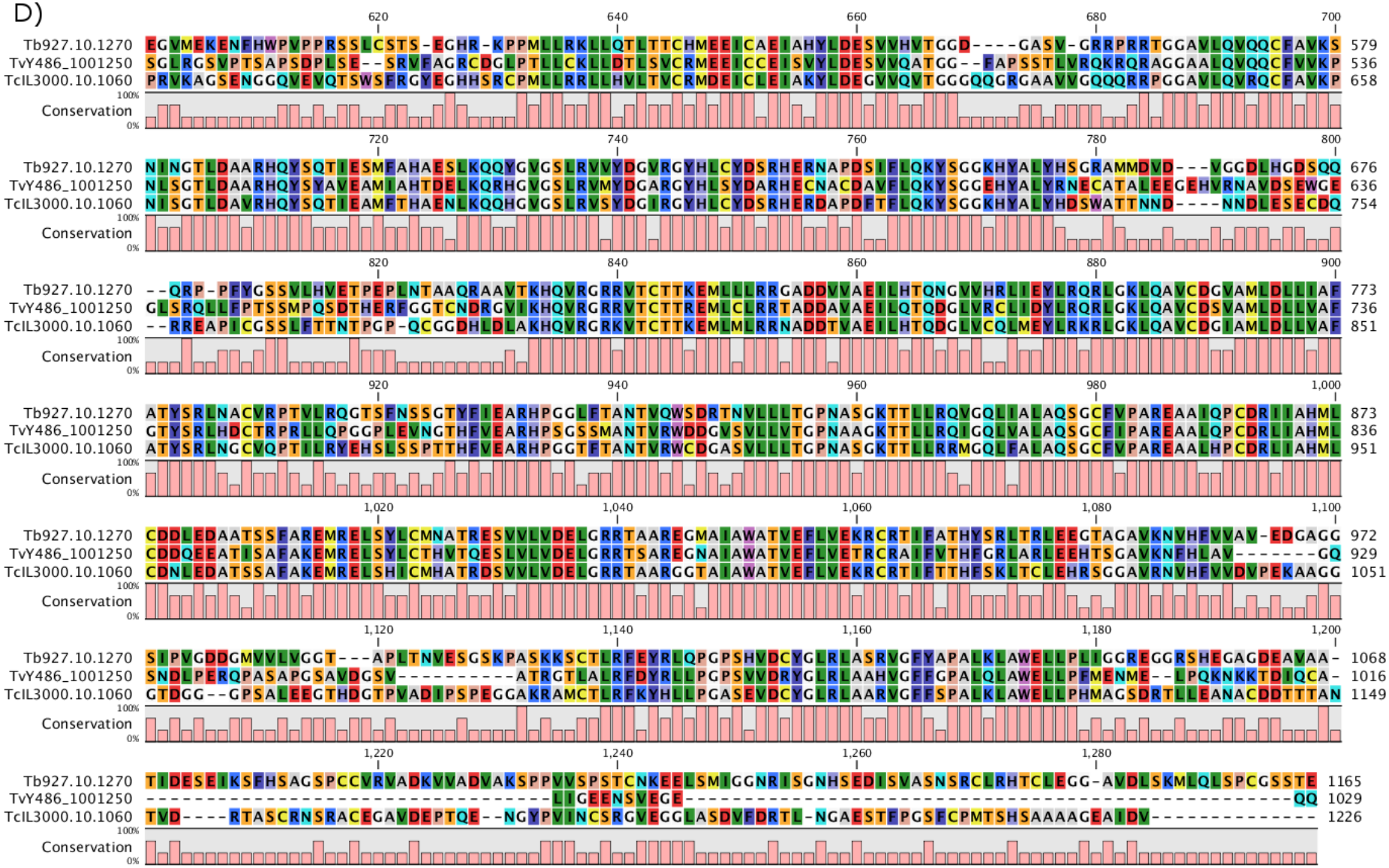

20
40
60
80
100
Tb09.211.1210 predicted prot M-QHV-GTRSGKS---------EA-KDAAVSTDNSTH-EDAAHTI--------MEIDRLTEQGVAAADVAKLRQAGIFTVTGIHMQCRKDLVLIKGLSDA 79
TvY486_0904120 predicted prot MMQHA-GTRSGRG---------DAGKDAAHATESATPIGDAAHAI--------MEVDRLTEQGVAAADVSKLRQAGIFTVPGIHMQCRKDLVLIKGLSDA 82
TcIL3000.11.8740 predicted prot MNTRAKGKKRTKTSADAEVMEDDAQSDGAALNEATNPMQDQQQGVDAANCGGARVIQVLESYGIASSDIKKLMEAGFYTVESVVYAPKKTILAVKGISEV 100
100%
Conservation
0%
120
140
160
180
200
Tb09.211.1210 predicted prot KVDKIIEAARKLSDCGFSVGTAYLQQRGRVTRVTTGSTALDQLLGGGIESMSITEAFGEFRTGKTQIAHTLCVTCQLPISMGGGNGKAIYVDTEATFRPE 179
TvY486_0904120 predicted prot KVDKIIEAARKLSDCGFTNGTAYLHQRTKITRMTTGSAALDQLLGGGVESMSITEAFGEFRTGKTQIAHTLCVTCQLPLSMGGGNGKAVYVDTESTFRPE 182
TcIL3000.11.8740 predicted prot KAEKIMAECSRLVPMGFTSAMQYHEARKEIIMVTTGSREVDKLLGGGIETGSITELFGEFRTGKTQLCHTLCVTCQLPISQGGAEGMALYIDTEGTFRPE 200
100%
Conservation
0%
220
240
260
280
300
Tb09.211.1210 predicted prot RIKPIAERFGLDVEAVLGNILVARAYTHEHQMHLLSMVAAKMVEDQFSLLVVDSVTALFRVDFSGRGELAERQQKLAKMLSNMIKLAEEYNVAVYITNQV 279
TvY486_0904120 predicted prot RIKPIAERFGLDVDAVLANIIVARAYTHEHQMHLLSMVAAKMAEEQFSLLVVDSITALFRVDFSGRGELAERQQKLAKMLSSLIKLAEEYNVAVYITNQV 282
TcIL3000.11.8740 predicted prot RLVAVAERYRLDPQDVLTNVACARALNTDHQQQLLLQASAMMAENRFAIIIVDSATALYRTDYSGRGELAARQVHLGKFLRTLHNLAGEHGVAVVVTNQV 300
100%
Conservation
0%
320
340
360
Tb09.211.1210 predicted prot VAD-PGGASMFVADPKKPIGGHILAHASTTRLSLRKGRGDQRVCKIYDSPSLPEVECVFSISEQGIVDARE 349
TvY486_0904120 predicted prot VAD-PGGASMFVADPKKPVGGHILAHASTTRLSLRKGRGDQRICKIYDSPSLPETECVFSISEQGIVDARE 352
TcIL3000.11.8740 predicted prot VANVDGAAQMFQADSKKPIGGHIMAHASTTRLSLRKGRGEQRIIKVYDSPCLAESEAIFGIYEDGVGDARD 371
100%
Conservation
0%

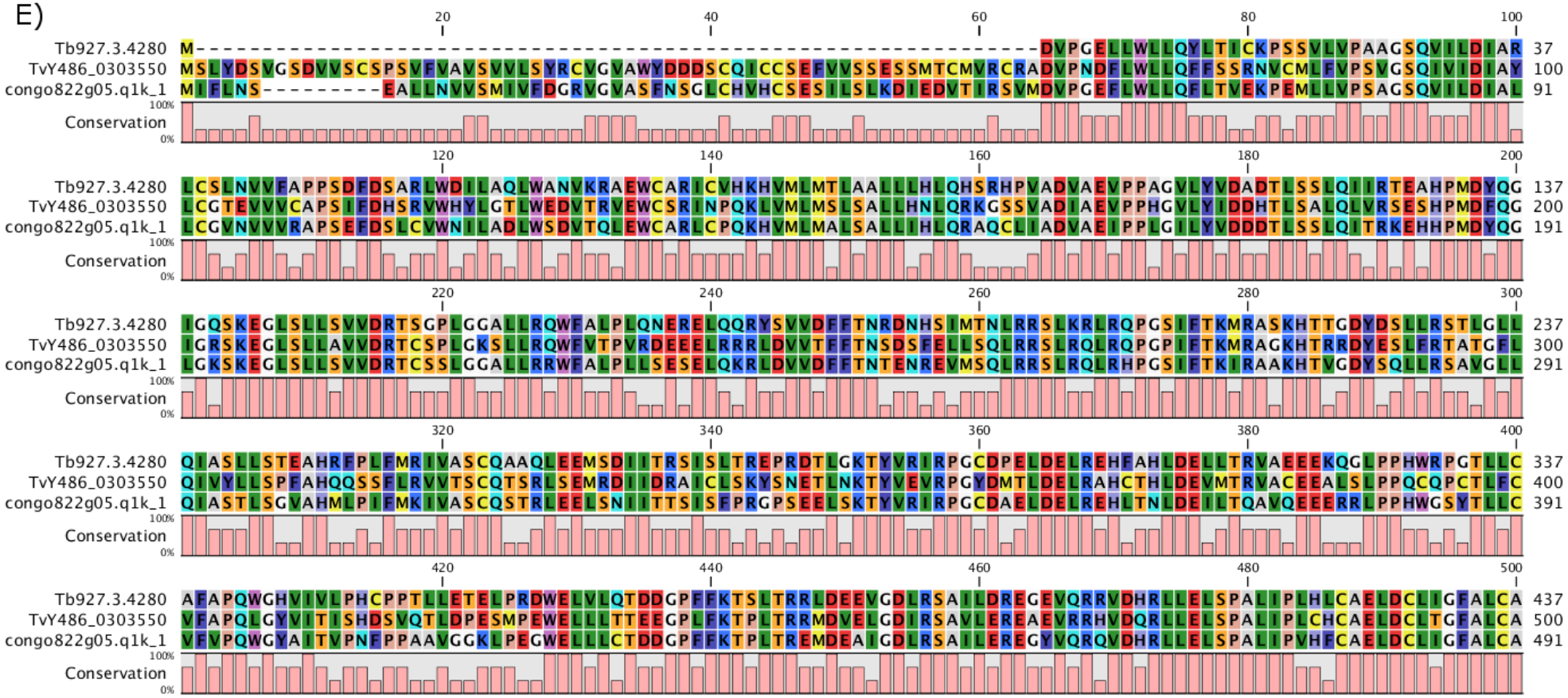
C)


20
40
60
80
100
Tb11.02.3380 MNKSAKKKGLSMEEKVSRIETWFVNNPSPYTLKDLIQILPKATGVIPQSVEECMEILVSENRVRQKKVGVHVLFWRFPQTATQRLAGSMGEKTTTTELAK 100
TvY486_1105890 MHKTVKKKGLSMEEKVARIEEWFQANPVPYTLKDLLALLPRATGVIPQSVEECLEVLVSEDRVRQKKVGIHVLFWRFPQTPAQQLASAAGAKGGVSEAAR 100
TcIL3000.0.24610 MNKPSKKRGLSIEEKVVRIEEWFVAHPYPYTMKDLCLILPKATGVISQSVEECMDILVSENRVQQKKVGIHVLFWRFPRTPTQQLATTIGRGASVAEVAK 100
TcIL3000.11.6040 MNKPSKKRGLSIEEKVVRIEEWFVAHPYPYTMKDLCLILPKATGVISQSVEECMDILVSENRVQQKKVGIHVLFWRFPRTPTQQLATTIGRGASVAEVAK 100
100%
Conservation
0%
120
140
160
180
200
Tb11.02.3380 YLSMSEDELRTELQTLKKTEKDIRQRIKETQASIGDDNAVQKDAEQIRSLQGEVKILEEQLNQLALLDPAFVEKLKAATVVAWESANRWTDNMYLLEQHI 200
TvY486_1105890 YLGMTLEQLQEELVALRNTEVEMRQQLEEMCSAIGDSAAVRNDTERIAKLRDEVKSLEGQLKQLGLFDPDMIERIKSSTLVAWESANRWTDNMFLLEQHI 200
TcIL3000.0.24610 LLGMTQSQLSAEIAELKKKEKDTQRRIKEVRLSIGDDASMKKDTEEIKRLEGEVRSLEKQLSQLALLDPETVEKIKAATNIAWESANRWTDNMYLLEHHI 200
TcIL3000.11.6040 LLGMTQSQLSAEITELKKKEKDTQRRIKEVRLSIGDDASMKKDTDEIKRLEGEVRSLEKQLSQLALLDPETVEKIKAATNIAWESANRWTDNMYLLEHHI 200
100%
Conservation
0%
220
240
260
280
300
Tb11.02.3380 SRKMGMSARELRAQLQLPLEVEYIEYDELLPSATSGGKSACTSSGNS---CASPLICTVDATHGGGNS-QEVPAAFHIEEGHTQNINE--EAGEALTEGG 294
TvY486_1105890 TKRMGVSSRDLRAQLQLPQNIEYIEESDLLPPACEETTQSVDGTGRAF---SSPLMCDAV----------DVDTVAAEEPSKTED-----DCAPAAYAQC 282
TcIL3000.0.24610 SQKMGLSARDLRRQLQIPLEVDFVEYEDLVRPC--GGVHNNVNDGRSVFPCNSLALLSKTTDMGDGNSTNDKPAAHHVEEGFTGDANRMGESGGAATVVC 298
TcIL3000.11.6040 SQKMGLSARELRRQLQIPLEVDFVEYEDLVRPC--GDAHNNVNDGRSVFPCNSLALLSKTTDMGDGNSTNDKPAARHVEEGFTGDANRMGESGGAATVVC 298
100%
Conservation
0%
320
340
360
380
Tb11.02.3380 CDVKPHEDVENETSNTLRKERDGKKRGRGRP-PLENAKKDDNMVNAPSVKKKKNEKDRPPAAS-NTDA--LTPARQRR-CRR 371
TvY486_1105890 VSDQRCAKSGD--GKGLTKNKKVATRGVKRGRGREEGELIESAPSVP-DIGDSAKKGKAAASTPGTAKGRTSIQKRAQGARH 361
TcIL3000.0.24610 VGNVPCGKVTDK-GDPFAKKLD-KTKPITES-DLEN-KVIEESGNAGRDCGKRVKKGRAKALK-GDEAGAAVISSRPR-VRR 374
TcIL3000.11.6040 VGNVPCGKVTDK-GDPFAKKLD-KTKPITES-DLEN-KVIEESGNAGRDCGKRVKKGRAKALK-GDEAGAVVISSRPR-VRR 374
100%
Conservation
0%

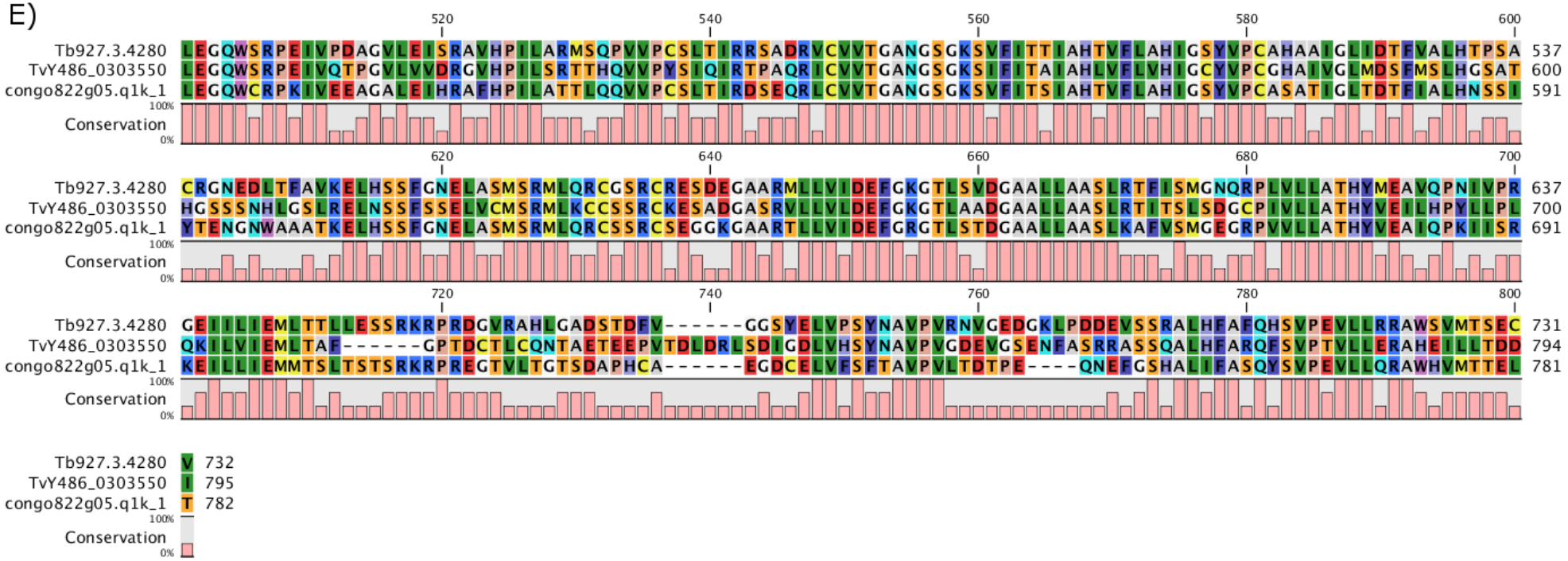

D)
Tb927.10.1270
TvY486_1001250
TcIL3000.10.1060
Conservation
100%
0%

D)

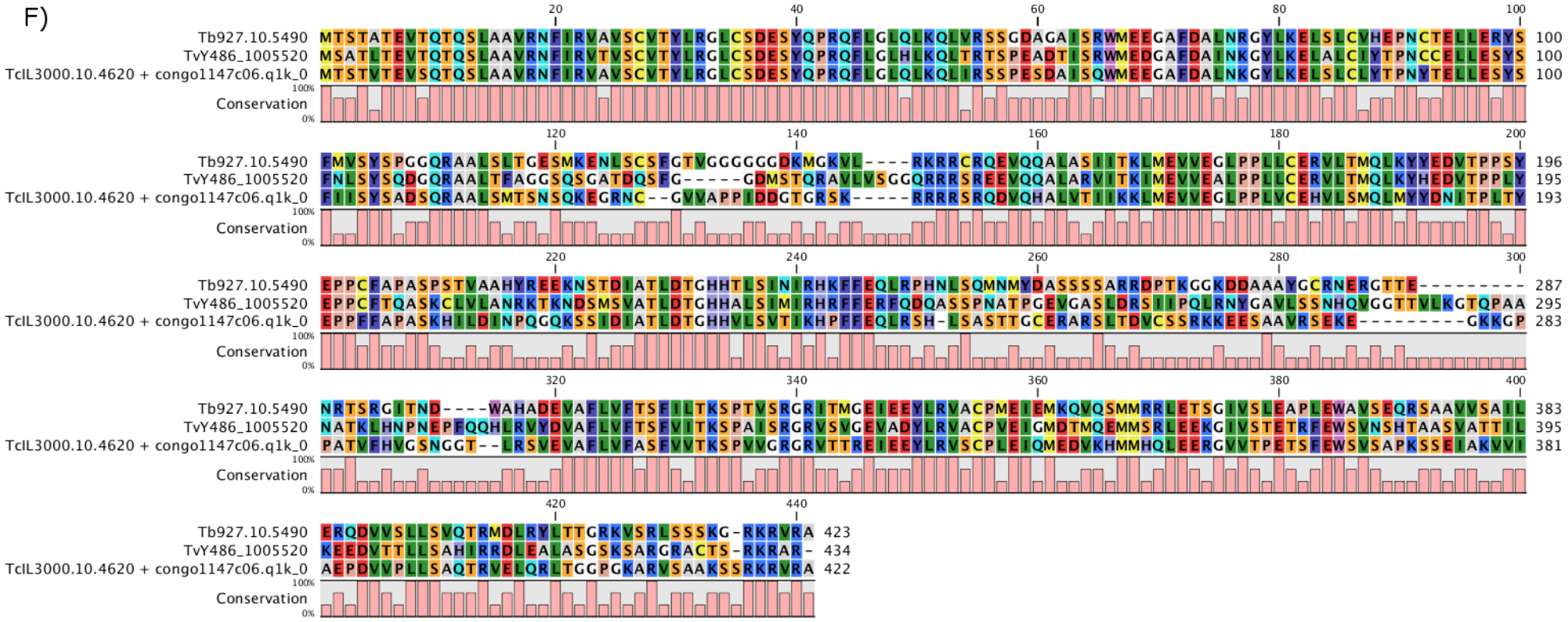

620 640 660 680 700
Tb927.10.1270 EGVMEKENFHWPVPPRSSLCSTS-EGHR-KPPMLLRKLLQTLTTCHMEEICAEIAHYLDESVVHVTGGD----GASV-GRRPRRTGGAVLQVQQCFAVKS 579
TvY486_1001250 SGLRGSVPTSAPSDPLSE--SRVFAGRCDGLPTLLCKLLDTLSVCRMEEICCEISVYLDESVVQATGG--FAPSSTLVRQKRQRAGGAALQVQQCFVVKP 536
TcIL3000.10.1060 PRVKAGSENGGQVEVQTSWSFRGYEGHHSRCPMLLRRLLHVLTVCRMDEICLEIAKYLDEGVVQVTGGGQQGRGAAVVGQQQRRPGGAVLQVRQCFAVKP 658
Conservation
720 740 760 780 800
Tb927.10.1270 NINGTLDAARHQYSQTIESMFAHAESLKQQYGVGSLRVVYDGVRGYHLCYDSRHERNAPDSIFLQKYSGGKHYALYHSGRAMMDVD---VGGDLHGDSQQ 676
TvY486_1001250 NLSGTLDAARHQYSYAVEAMIAHTDELKQRHGVGSLRVMYDGARGYHLSYDARHECNACDAVFLQKYSGGEHYALYRNECATALEEGEHVRNAVDSEWGE 636
TcIL3000.10.1060 NISGTLDAVRHQYSQTIEAMFTHAENLKQQHGVGSLRVSYDGIRGYHLCYDSRHERDAPDFTFLQKYSGGKHYALYHDSWATTNND----NNDLESECDQ 754
Conservation
820 840 860 880 900
Tb927.10.1270 --QRP-PFYGSSVLHVETPEPLNTAAQRAAVTKHQVRGRRVTCTTKEMLLLRRGADDVVAEILHTQNGVVHRLIEYLRQRLGKLQAVCDGVAMLDLLIAF 773
TvY486_1001250 GLSRQLLFPTSSMPQSDTHERFGGTCNDRGVIKHQVRGRRVTCTTREMLCLRRTADDAVAEILQTQDGLVRCLIDYLRQRLGKLQAVCDSVAMLDLLVAF 736
TcIL3000.10.1060 --RREAPICGSSLFTTNTPGP-QCGGDHLDLAKHQVRGRKVTCTTKEMLMLRRNADDTVAEILHTQDGLVCQLMEYLRKRLGKLQAVCDGIAMLDLLVAF 851
Conservation
920 940 960 980 1,000
Tb927.10.1270 ATYSRLNACVRPTVLRQGTSFNSSGTYFIEARHPGGLFTANTVQWSDRTNVLLLTGPNASGKTTLLRQVGQLIALAQSGCFVPAREAAIQPCDRIIAHML 873
TvY486_1001250 GTYSRLHDCTRPRLLQPGGPLEVNGTHFVEARHPSGSSMANTVRWDDGVSVLLVTGPNAAGKTTLLRQIGQLVALAQSGCFIPAREAALQPCDRLIAHML 836
TcIL3000.10.1060 ATYSRLNGCVQPTILRYEHSLSSPTTHFVEARHPGGTFTANTVRWCDGASVLLLTGPNASGKTTLLRRMGQLFALAQSGCFVPAREAALHPCDRLIAHML 951
Conservation
1,020 1,040 1,060 1,080 1,100
Tb927.10.1270 CDDLEDAATSSFAREMRELSYLCMNATRESVVLVDELGRRTAAREGMAIAWATVEFLVEKRCRTIFATHYSRLTRLEEGTAGAVKNVHFVVAV-EDGAGG 972
TvY486_1001250 CDDQEEATISAFAKEMRELSYLCTHVTQESLVLVDELGRRTSAREGNAIAWATVEFLVETRCRAIFVTHFGRLARLEEHTSGAVKNFHLAV-------GQ 929
TcIL3000.10.1060 CDNLEDATSSAFAKEMRELSHICMHATRDSVVLVDELGRRTAARGGTAIAWATVEFLVEKRCRTIFTTHFSKLTCLEHRSGGAVRNVHFVVDVPEKAAGG 1051
Conservation
1,120 1,140 1,160 1,180 1,200
Tb927.10.1270 SIPVGDDGMVVLVGGT---APLTNVESGSKPASKKSCTLRFEYRLQPGPSHVDCYGLRLASRVGFYAPALKLAWELLPLIGGREGGRSHEGAGDEAVAA- 1068
TvY486_1001250 SNDLPERQPASAPGSAVDGSV----------ATRGTLALRFDYRLLPGPSVVDRYGLRLAAHVGFFGPALQLAWELLPFMENME--LPQKNKKTDIQCA- 1016
TcIL3000.10.1060 GTDGG--GPSALEEGTHDGTPVADIPSPEGGAKRAMCTLRFKYHLLPGASEVDCYGLRLAARVGFFSPALKLAWELLPHMAGSDRTLLEANACDDTTTAN 1149
Conservation
1,220 1,240 1,260 1,280
Tb927.10.1270 TIDESEIKSFHSAGSPCCVRVADKVVADVAKSPPVVSPSTCNKEELSMIGGNRISGNHSEDISVASNSRCLRHTCLEGG-AVDLSKMLQLSPCGSSTE 1165
TvY486_1001250 ----------LIGEENSVEGE----------QQ 1029
TcIL3000.10.1060 TVD----RTASCRNSRACEGAVDEPTQE--NGYPVINCSRGVEGGLASDVFDRTL-NGAESTFPGSFCPMTSHSAAAAGEAIDV---------- 1226
Conservation


E)
Tb927.3.4280
TvY486_0303550
congo822g05.q1k_1
Conservation
100%
0%

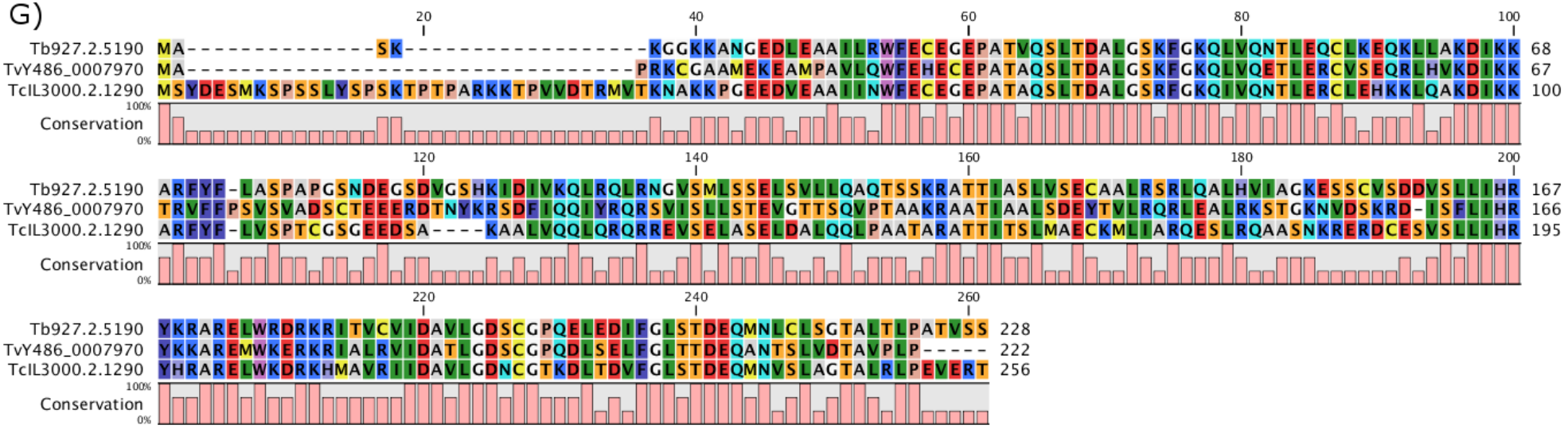
E)


520
540
560
580
600
Tb927.3.4280 LEGQWSRPEIVPDAGVLEISRAVHPILARMSQPVVPCSLTIRRSADRVCVVTGANGSGKSVFITTIAHTVFLAHIGSYVPCAHAAIGLIDTFVALHTPSA 537
TvY486_0303550 LEGQWSRPEIVQTPGVLVVDRGVHPILSRTTHQVVPYSIQIRTPAQRICVVTGANGSGKSIFITAIAHLVFLVHIGCYVPCGHAIVGLMDSFMSLHGSAT 600
congo822g05.q1k_1 LEGQWCRPKIVEEAGALEIHRAFHPILATTLQQVVPCSLTIRDSEQRLCVVTGANGSGKSVFITSIAHTVFLAHIGSYVPCASATIGLTDTFIALHNSSI 591
100%
Conservation
0%
620
640
660
680
700
Tb927.3.4280 CRGNEDLTFAVKELHSSFGNELASMSRMLQRCGSRCRESDEGAARMLLVIDEFGKGTLSVDGAALLAASLRTFISMGNQRPLVLLATHYMEAVQPNIVPR 637
TvY486_0303550 HGSSSNHLGSLRELNSSFSSELVCMSRMLKCCSSRCKESADGASRVLLVLDEFGKGTLAADGAALLAASLRTITSLSDGCPIVLLATHYVEILHPYLLPL 700
congo822g05.q1k_1 YTENGNWAAATKELHSSFGNELASMSRMLQRCSSRCSEGGKGAARTLLVIDEFGRGTLSTDGAALLAASLKAFVSMGEGRPVVLLATHYVEAIQPKIISR 691
100%
Conservation
0%
720
740
760
780
800
Tb927.3.4280 GEIILIEMLTTLLESSRKRPRDGVRAHLGADSTDFV------GGSYELVPSYNAVPVRNVGEDGKLPDDEVSSRALHFAFQHSVPEVLLRRAWSVMTSEC 731
TvY486_0303550 QKILVIEMLTAF-------GPTDCTLCQNTAETEEPVTDLDRLSDIGDLVHSYNAVPVGDEVGSENFASRRASSQALHFARQFSVPTVLLERAHEILLTDD 794
congo822g05.q1k_1 KEILLIEMMTSLTSTSRKRPREGTVLTGTSDAPHCA------EGDCELVFSFTAVPVLTDTPE----QNEFGSHALIFASQYSVPEVLLQRAWHVMTTEL 781
100%
Conservation
0%
Tb927.3.4280 V 732
TvY486_0303550 I 795
congo822g05.q1k_1 T 782
100%
Conservation
0%

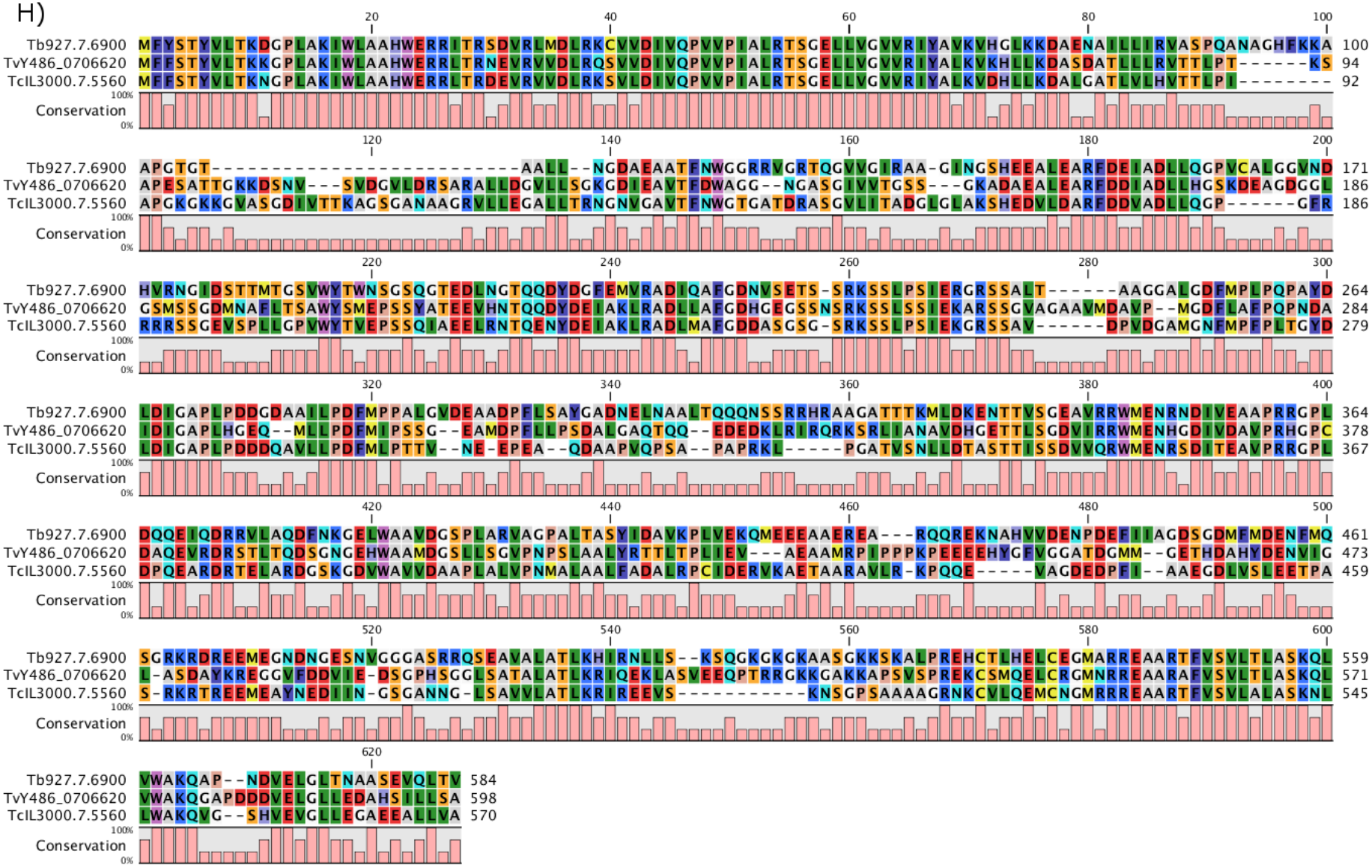
F)


Tb927.10.5490
TvY486_1005520
TcIL3000.10.4620 + congo1147c06.q1k_0
Conservation
100%
0%

G)


20
40
60
80
100
Tb927.2.5190 MA----------SK----------KGGKKANGEDLEAAILRWFECEGEPATVQSLTDALGSKFGKQLVQNTLEQCLKEQKLLAKDIKK 68
TvY486_0007970 MA----------PRKCGAAMEKEAMPAVLQWFEHECEPATAQSLTDALGSKFGKQLVQETLERCVSEQRLHVKDIKK 67
TcIL3000.2.1290 MSYDESMKSPSSLYSPSKTPTPARKKTPVVDTRMVTKNAKKPGEEDVEAAIINWFECEGEPATAQSLTDALGSRFGKQIVQNTLERCLEHKKLQAKDIKK 100
100%
Conservation
0%
120
140
160
180
200
Tb927.2.5190 ARFYF-LASPAPGSNDEGSDVGSHKIDIVKQLRQLRNGVSMLSSELSVLLQAQTSSKRATTIASLVSECAALRSRLQALHVIAGKESSCVSDDVSLLIHR 167
TvY486_0007970 TRVFFPSVSVADSCTEEERDTNYKRSDFIQQIYRQRSVISLLSTEVGTTSQVPTAAKRAATIAALSDEYTVLRQRLEALRKSTGKNVDSKRD-ISFLIHR 166
TcIL3000.2.1290 ARFYF-LVSPTCGSGEEDSA----KAALVQQLQRQRREVSELASELDALQQLPAATARATTITSLMAECKMLIARQESLRQAASNKRERDCESVSLLIHR 195
100%
Conservation
0%
220
240
260
Tb927.2.5190 YKRARELWRDRKRITVCVIDAVLGDSCGPQELEDIFGLSTDEQMNLCLSGTALTLPATVSS 228
TvY486_0007970 YKKAREMWKERKRIALRVIDATLGDSCGPQDLSELFGLTTDEQANTSLVDTAVPLP----- 222
TcIL3000.2.1290 YHRARELWKDRKHMAVRIIDAVLGDNCGTKDLTDVFGLSTDEQMNVSLAGTALRLPEVERT 256
100%
Conservation
0%

H)


Tb927.7.6900
TvY486_0706620
TcIL3000.7.5560
Conservation
100%
0%

Appendix 5

Amino acid sequence alignments for the eight core meiotic genes in *T. brucei*, *T. vivax* and *T. congolense*. Sequence alignments for *T. brucei*, *T. vivax* and *T. congolense* respectively for the eight core meiotic genes identified by Schurko and Logsdon (2008) are shown for *spo11* (a), *dmc1* (b), *mnd1* (c), *msh4* (d), *msh5* (e), *hop1* (f), *hop2* (g) and *rec8* (h). *T. congolense* sequences were identified by Morrison *et al.* (2009b). Amino acids are colour coded based upon side chain properties with similar amino acids grouped together. Conservation indicates direct identity between the respective sequences with 100% representing presence of the same amino acid at a position in every sequence in the alignment and minimal conservation (25% for *mnd1*, 33% for each other alignment) indicating at unique amino acid in each sequence at the indicated position. For *mnd1* two *T. congolense* sequences, TcIL3000.0.24610 and TcIL3000.11.6040, with sequence identity of 98% to one another are present within the genome sequence. Alignment of *T. congolense hop1* was performed utilising a sequence combined sequences of TcIL3000.10.4620 and congo1147c06.q1k_0.

Appendix 6

Sample	Date	Species[a]	Locus and allele sizes[b]												MLG[c]
			TV6		TV14		TV17		TV24		TV31		TV49		
59	Mar 06	D	279	293	211	229	143	206	122	140	476	506	103	138	8
72	Mar 06	D	279	301	211	211	143	206	140	157	476	476	103	103	9
102	Mar 06	H	279	279	211	211	143	206	140	152	476	506	103	138	4
103	Mar 06	H	279	279	211	229	143	206	122	140	476	506	103	138	7
148	Mar 06	H	279	302	211	211	143	206	140	157	476	476	103	103	9
1067	Aug 06	H	279	302	211	229	143	143	140	152	476	506	103	138	2
1068	Aug 06	D	279	279	211	211	143	206	140	152	476	506	103	138	4
1081	Aug 06	H	279	293	211	229	143	206	122	140	476	506	103	138	8
1096	Aug 06	H	279	279	211	211	143	206	140	152	476	506	103	138	4
1100	Aug 06	D	279	279	211	211	143	206	140	152	476	506	103	138	4
c2	Aug 06	C	279	293	211	229	143	206	122	140	476	506	103	138	8
2002	Jan 07	D	279	293	211	229	143	206	122	140	476	506	103	138	8
2005	Jan 07	H	279	293	211	229	143	206	122	140	476	506	103	138	8
2029	Jan 07	H	279	293	211	229	143	206	122	140	476	506	103	138	8
2037	Jan 07	H	279	302	211	211	143	206	140	157	476	476	103	103	9
2047	Jan 07	H	279	293	211	211	143	206	152	157	476	506	103	138	5
2058	Jan 07	H	279	279	211	211	143	206	140	157	476	476	103	103	10
2062	Jan 07	H	279	279	211	211	143	206	140	152	476	506	103	138	4
2069	Jan 07	H	279	293	211	229	143	206	122	140	476	506	103	138	8
2078	Jan 07	H	279	279	211	211	143	206	140	152	506	506	103	138	3
2081	Jan 07	D	279	293	211	229	143	206	122	140	476	506	103	138	8
3003	Jan 07	C	279	293	211	229	143	206	122	140	476	506	103	138	8
3035	Jan 07	C	279	279	211	211	143	206	152	157	476	506	103	138	6
3036	Jan 07	C	279	293	211	229	143	206	122	140	476	506	103	138	8
3037	Jan 07	C	279	293	211	229	143	206	122	140	476	506	103	138	8
3038	Jan 07	C	279	302	211	211	143	206	140	157	476	476	103	103	9
3039	Jan 07	C	279	293	211	229	143	206	122	140	476	506	103	138	8
3040	Jan 07	C	279	293	211	229	143	206	122	140	476	506	103	138	8
3041	Jan 07	C	279	279	211	211	143	206	140	157	476	476	103	103	10

3042	Jan 07	C	279	293	211	229	143	206	122	140	476	506	103	138	8
3043	Jan 07	C	279	293	211	229	143	206	122	140	476	506	1003	138	8
ILRAD V34	-	-	279	293	211	211	160	185	140	136	476	476	138	117	1

Appendix 6

Collection details (sampling date and host species) and genotype of the 31 fully genotyped *Trypanosoma vivax* samples plus the control sample, ILRAD V34. Samples were genotyped at six microsatellite loci specific to *T. vivax* and allele sizes determined by comparison to labelled size standards following size-separation on a capillary-based sequencer.

[a] Species: D, donkey; H, horse; C, cattle.

[b] Allele size given in bp.

[c] MLG, multilocus genotype, where each multilocus genotype represents a unique combination of microsatellite alleles across the six loci examined for this study.

Appendix 7

ID	Date of collection	Sampling Location	PCV	TP	Microscopic infection status	Cholesterol	Species	Trypanosome species present based the presence of 177bp PCR repeat		
								T. brucei	*T. congolense*	*T. vivax*
GAM047	27/03/09	Banni	22	8.4	Negative	4.8	Horse	Positive	Negative	Positive
GAM048	27/03/09	Banni	29	6.7	Negative	6.9	Horse	Positive	Negative	Positive
GAM049	27/03/09	Banni	30	8.0	Negative	5.6	Horse	Positive	Negative	Positive
GAM050	27/03/09	Banni	39	7.6	Negative	5.9	Horse	Negative	Negative	Positive
GAM051	27/03/09	Banni	35	8.0	Negative	4.1	Donkey	Negative	Negative	Positive
GAM052	27/03/09	Banni	34	8.4	Negative	5.0	Donkey	Positive	Negative	Positive
GAM053	27/03/09	Banni	34	8.2	Negative	4.3	Donkey	Positive	Negative	Positive
GAM054	27/03/09	Banni	39	7.2	Negative	3.1	Donkey	Negative	Negative	Positive
GAM055	27/03/09	Banni	32	7.4	Negative	3.3	Donkey	Positive	Negative	Positive
GAM0.56	27/03/09	Banni	32	8.4	Negative	4.6	Donkey	Negative	Negative	Positive
GAM057	27/03/09	Banni	33	7.2	Negative	2.9	Donkey	Positive	Negative	Positive
GAM058	27/03/09	Banni	30	8.7	Negative	3.6	Donkey	Positive	Negative	Positive
GAM059	27/03/09	Banni	35	6.7	Negative	2.9	Donkey	Positive	Negative	Positive
GAM060	27/03/09	Banni	36	6.8	Negative	4.2	Donkey	Negative	Negative	Positive
GAM061	27/03/09	Banni	39	9.0	Negative	2.9	Donkey	Positive	Positive	Positive
GAM062	27/03/09	Banni	42	7.2	Negative	3.1	Donkey	Positive	Positive	Negative
GAM072	28/03/09	Missera	38	8.6	Negative	2.8	Donkey	Positive	Negative	Positive
GAM073	28/03/09	Missera	30	7.2	Negative	2.9	Horse	Positive	Negative	Positive
GAM074	28/03/09	Missera	33	8.4	Negative	6.0	Donkey	Negative	Negative	Positive
GAM075	28/03/09	Missera	39	8.0	Negative	4.8	Horse	Positive	Negative	Positive
GAM076	28/03/09	Missera	28	8.7	Negative	4.1	Donkey	Negative	Positive	Positive
GAM077	28/03/09	Missera	30	9.2	Negative	4.7	Donkey	Positive	Negative	Positive
GAM078	28/03/09	Missera	40	8.6	Negative	4.4	Donkey	Negative	Negative	Positive
GAM079	28/03/09	Missera	34	8.4	Negative	4.9	Donkey	Negative	Negative	Positive
GAM080	28/03/09	Missera	36	8.7	Negative	5.9	Donkey	Negative	Negative	Negative
GAM081	28/03/09	Missera	34	8.7	Negative	4.7	Donkey	Positive	Positive	Positive
GAM082	28/03/09	Missera	34	6.8	Negative	7.2	Horse	Positive	Negative	Positive

GAM083	28/03/09	Missera	32	9.4	Negative	3.6	Donkey	Positive	Negative	Positive
GAM084	28/03/09	Brikamabah	36	8.0	Negative	3.9	-	Positive	Negative	Positive
GAM085	28/03/09	Brikamabah	42	8.2	Negative	4.8	-	Negative	Negative	Positive
GAM086	28/03/09	Brikamabah	30	8.6	Negative	3.9	Horse	Positive	Negative	Positive
GAM087	28/03/09	Brikamabah	39	6.8	Negative	4.5	Horse	Positive	Negative	Positive
GAM088	28/03/09	Brikamabah	46	8.0	Negative	6.7	Horse	Positive	Negative	Positive
GAM089	28/03/09	Brikamabah	30	7.4	Negative	3.2	Horse	Negative	Negative	Positive
GAM090	28/03/09	Brikamabah	33	9.0	Negative	4.3	Horse	Positive	Negative	Positive
GAM091	28/03/09	Brikamabah	31	>12	Negative	2.9	Horse	Negative	Negative	Positive
GAM095	29/03/09	Missera Torben	25	9.4	Negative	2.7	Horse	Positive	Positive	Positive
GAM096	29/03/09	M'Fana	26	7.9	Negative	Lo	Horse	Positive	Positive	Negative
GAM098	29/03/09	M'Fana	26	9.2	Negative	5.1	Horse	Positive	Negative	Negative
GAM099	29/03/09	M'Fana	23	9.2	Negative	3.6	Donkey	Positive	Negative	Positive
GAM100	29/03/09	M'Fana	25	8.6	Positive	4.2	Donkey	Positive	Positive	Positive
GAM101	29/03/09	M'Fana	39	8.6	Negative	5.6	-	Positive	Negative	Positive
GAM102	30/03/09	Wasu	15	6.6	Positive	3.2	Horse	Negative	Positive	Positive
GAM103	30/03/09	Wasu	30	8.4	Negative	2.6	Horse	Negative	Negative	Positive
GAM104	30/03/09	Wasu	39	7.8	Negative	4.4	Horse	Positive	Negative	Positive
GAM105	30/03/09	Wasu	11	8.4	Negative	2.9	Horse	Positive	Positive	Positive
GAM106	30/03/09	Wasu	20	7.2	Positive	2.9	Horse	Negative	Positive	Positive
GAM107	30/03/09	Wasu	32	7.2	Negative	2.9	Horse	Positive	Positive	Positive
GAM108	30/03/09	Wasu	29	7.0	Positive	3.9	Horse	Negative	Positive	Positive
GAM109	30/03/09	Wasu	20	9.2	Positive	3.1	Horse	Positive	Negative	Positive
GAM110	30/03/09	Wasu	36	8.2	Positive	4.3	Horse	Positive	Positive	Negative
GAM111	30/03/09	Wasu	33	7.8	Negative	4.6	-	Negative	Positive	Positive
GAM112	30/03/09	Wasu	30	8.6	Positive	4.0	Horse	Positive	Positive	Positive
GAM113	30/03/09	Wasu	28	7.6	Negative	3.3	Horse	Positive	Negative	Positive
GAM114	30/03/09	Wasu	31	6.2	Negative	4.4	-	Positive	Negative	Positive
ECG001	03/04/09	Sambel Kunda	22	8.6	Negative	4.2	Horse	Negative	Negative	Positive
X001	04/04/09	Brikamabah	32	8.8	Negative	4.6	Horse	Positive	Negative	Positive
X002	04/04/09	Brikamabah	32	7.6	Negative	4.3	Horse	Negative	Negative	Positive
X003	04/04/09	Brikamabah	30	8.0	Negative	4.2	Horse	Negative	Positive	Positive

X004	04/04/09	Brikamabah	36	8.4	Negative	4	Horse	Negative	Negative	Negative
X005	04/04/09	Brikamabah	19	8.2	Positive	4.8	Horse	Negative	Positive	Positive
X006	04/04/09	Brikamabah	28	8.7	Negative	4.9	Horse	Negative	Negative	Positive
X007	04/04/09	Brikamabah	24	8.0	Positive	3.6	Horse	Negative	Positive	Positive
X008	04/04/09	Brikamabah	29	7.2	Negative	4	-	Negative	Negative	Positive
X009	04/04/09	Brikamabah	10	8.2	Negative	4.1	Horse	Negative	Negative	Positive
X010	04/04/09	Brikamabah	25	6.2	Positive	3.5	Horse	Negative	Negative	Negative
X011	04/04/09	Brikamabah	39	7.7	Negative	4.9	-	Negative	Positive	Positive
X012	04/04/09	Brikamabah	33	7.8	Negative	4.7	Horse	Negative	Negative	Positive
X013	04/04/09	Brikamabah	27	10.2	Negative	4.6	Horse	Negative	Negative	Positive
X014	04/04/09	Brikamabah	15	7.4	Negative	4.2	Horse	Negative	Negative	Positive
X015	04/04/09	Brikamabah	18	8.6	Negative	3.4	-	Positive	Negative	Positive
X016	04/04/09	Brikamabah	48	9.6	Negative	5.4	-	Negative	Positive	Negative
X017	04/04/09	Brikamabah	48	7.4	Negative	4.2	-	Negative	Negative	Positive
X018	05/04/09	M'fana	29	9.2	Negative	4.2	Horse	Negative	Positive	Positive
X019	05/04/09	M'fana	32	6.2	Negative	4.3	-	Negative	Negative	Positive
X020	05/04/09	M'fana	30	6.8	Negative	2.9	-	Negative	Negative	Negative
X021	05/04/09	M'fana	42	8.6	Negative	4.8	-	Positive	Negative	Positive
X022	05/04/09	M'fana	31	6.8	Negative	3.7	-	Negative	Negative	Positive
X023	05/04/09	M'fana	37	8.2	Negative	3.5	-	Negative	Positive	Positive
X024	05/04/09	M'fana	31	7.0	Negative	3.1	-	Negative	Positive	Positive
X025	05/04/09	M'fana	32	8.5	Negative	4.1	-	Positive	Positive	Positive
X026	05/04/09	M'fana	19	8.0	Negative	3.8	-	Negative	Positive	Positive
X027	05/04/09	M'fana	20	8.2	Negative	5.3	-	Positive	Negative	Positive
GAM154	07/04/09	Touba	38	7.0	Negative	4.5	Horse	Positive	Negative	Positive
GAM163	07/04/09	Touba	32	7.2	Positive	5	Horse	Negative	Positive	Positive
GAM167	07/04/09	Kununku	18	8.6	Negative	3.4	Horse	Positive	Negative	Positive
GAM115	08/04/09	Jahally	18	9.6	Negative	3.9	Horse	Negative	Negative	Positive
GAM116	08/04/09	Jahally	20	9.8	Negative	4.2	Horse	Negative	Negative	Positive
GAM117	08/04/09	Jahally	22	10.0	Negative	4.8	Horse	Negative	Positive	Positive
GAM118	08/04/09	Jahally	24	7.8	Negative	6.9	Horse	Negative	Negative	Positive
GAM119	08/04/09	Jahally	42	7.8	Negative	4.4	Horse	Negative	Negative	Positive
GAM120	08/04/09	Jahally	18	10	Positive	4.3	Donkey	Negative	Positive	Positive

GAM121	08/04/09	Jahally	30	7.4	Negative	6.3	Donkey	Negative	Negative	Positive
GAM122	08/04/09	Jahally	24	8.6	Negative	4.5	-	Negative	Negative	Positive
GAM123	08/04/09	Jahally	34	7.0	Negative	4.2	Horse	Negative	Negative	Positive
GAM124	08/04/09	Jahally	16	9.8	-	3.6	Donkey	Positive	Negative	Positive
GAM125	08/04/09	Jahally	20	7.6	-	4.5	Horse	Negative	Positive	Positive
GAM126	08/04/09	Madina	30	7.0	Negative	5.4	Horse	Positive	Negative	Positive
GAM127	08/04/09	Madina	41	7.6	Negative	4	Horse	Negative	Negative	Negative
GAM128	08/04/09	Madina	33	7.8	Negative	5.1	Horse	Negative	Negative	Positive
GAM129	08/04/09	Madina	21	7.6	Negative	3.6	Donkey	Negative	Positive	Positive
GAM130	08/04/09	Madina	22	9.4	Positive	4.2	Horse	Positive	Positive	Positive
GAM131	08/04/09	Madina	31	6.2	Negative	4.4	-	Negative	Positive	Positive
GAM132	08/04/09	Madina	28	7.6	Positive	6.4	-	Positive	Negative	Positive
GAM133	09/04/09	Kerewan	30	8.8	Negative	5.4	Donkey	Positive	Negative	Positive
GAM134	09/04/09	Kerewan	28	6.6	Negative	6.4	Donkey	Positive	Negative	Positive
GAM135	09/04/09	Kerewan	25	9.0	Negative	7.1	Donkey	Positive	Negative	Positive
GAM136	09/04/09	Kerewan	24	7.8	Negative	4.3	Donkey	Negative	Negative	Positive
GAM137	09/04/09	Kerewan	31	8.8	Negative	5.9	Donkey	Positive	Negative	Positive
GAM138	09/04/09	Kerewan	19	9.2	Positive	5.1	Donkey	Negative	Positive	Positive
GAM139	09/04/09	Kerewan	25	7.2	Negative	Lo	Horse	Negative	Negative	Positive
GAM140	09/04/09	Kerewan	27	7.4	Negative	4.1	Donkey	Negative	Positive	Positive
GAM141	09/04/09	Kerewan	26	10.2	Negative	5.2	Donkey	Negative	Positive	Negative
GAM142	09/04/09	Kerewan	28	6.6	Negative	3.9	Horse	Negative	Negative	Negative
GAM143	09/04/09	Kerewan	36	7.6	Negative	4.7	Horse	Positive	Positive	Positive
GAM144	09/04/09	Kerewan	37	7.8	Negative	4.8	Horse	Negative	Negative	Negative
GAM145	09/04/09	Kerewan	8	8.2	Positive	4.9	Donkey	Positive	Positive	Positive
GAM146	09/04/09	Kerewan	20	8.0	Positive	4.3	Donkey	Positive	Positive	Positive
X028	15/04/09	Kerrtamim	27	8.2	Negative	5	Horse	Negative	Negative	Positive
ECG004	15/04/09	Kerrtamim	39	7.0	Positive	4.4	Horse	Negative	Negative	Negative
X029	15/04/09	Kerrtamim	32	7.2	Negative	5	Horse	Positive	Negative	Positive
X030	15/04/09	Kerrtamim	25	8.6	Negative	4.3	Horse	Negative	Negative	Positive
ECG005	15/04/09	Kerrtamim	27	6.8	Positive	5.1	Horse	Positive	Positive	Positive
X031	15/04/09	Kerrtamim	31	6.8	Negative	5.7	Horse	Positive	Negative	Positive
X032	15/04/09	Kerrtamim	38	7.4	Negative	5.4	Horse	Negative	Negative	Positive

X033	15/04/09	Kerrtamim	38	7.8	Negative	5.3	Horse	Positive	Negative	Positive
X034	15/04/09	Thirty Mile	21	8.4	Negative	3.6	Horse	Positive	Negative	Positive
X035	15/04/09	Thirty Mile	-	-	-	3.7	-	Positive	Positive	Negative
X036	15/04/09	Thirty Mile	27	9.6	Negative	3.5	Donkey	Negative	Positive	Positive
X037	15/04/09	Thirty Mile	31	8.4	Positive	4.3	Horse	Negative	Positive	Positive
X038	15/04/09	Thirty Mile	21	7.7	Negative	3.6	Horse	Negative	Negative	Positive
X039	15/04/09	Thirty Mile	20	8.6	Positive	3.2	Horse	Positive	Negative	Positive
X040	15/04/09	Thirty Mile	19	7.2	Negative	3.6	-	Negative	Negative	Positive
X041	15/04/09	Thirty Mile	27	-	Negative	3.2	-	Positive	Positive	Positive
ECG006	17/04/09	Sambel Kunda	15	9.2	Negative	2.7	Mule	Positive	Positive	Positive
X042	22/04/09	Sintu Alhagie	20	7.4	-	3.4	-	Positive	Negative	Positive
ECG007	22/04/09	Sintu Alhagie	30	9.2	-	5	Horse	Positive	Negative	Positive
X043	22/04/09	Sintu Alhagie	32	7.4	-	4.7	Horse	Negative	Negative	Positive
X044	22/04/09	Sintu Alhagie	30	7.4	-	4.2	Horse	Positive	Negative	Positive
X045	22/04/09	Sintu Alhagie	40	7.8	-	4	Horse	Negative	Positive	Positive
X046	22/04/09	Sintu Alhagie	38	8.0	-	4.3	Horse	Positive	Negative	Positive
X047	22/04/09	Sintu Alhagie	36	10.2	-	4.7	-	Negative	Negative	Negative
X048	22/04/09	Sintu Alhagie	38	9.8	-	3.8	Horse	Positive	Positive	Positive
X049	22/04/09	Sintu Alhagie	41	7.2	-	4.3	-	Negative	Negative	Positive
X050	22/04/09	Sintu Alhagie	46	6.8	-	4.2	Horse	Positive	Positive	Positive
X051	22/04/09	Sintu Alhagie	38	8.8	-	4.7	-	Negative	Negative	Positive
X052	22/04/09	Sintu Alhagie	26	9.6	-	4.2	Donkey	Positive	Negative	Negative
X053	22/04/09	Sintu Alhagie	30	7.4	-	4	-	Negative	Negative	Negative
X054	25/04/09	Brikamabah	29	7.6	Negative	5.9	-	Positive	Positive	Positive
X055	25/04/09	Brikamabah	29	9.4	Negative	4.1	-	Negative	Positive	Positive
X056	25/04/09	Brikamabah	23	6.7	Positive	4.1	-	Negative	Positive	Positive
X057	25/04/09	Brikamabah	33	8.4	Negative	4.2	-	Negative	Positive	Positive
X058	25/04/09	Brikamabah	37	6.0	Negative	3.9	-	Positive	Negative	Positive
X059	25/04/09	Brikamabah	34	6.6	Negative	4.2	-	Negative	Negative	Positive
X060	25/04/09	Brikamabah	35	8.4	Negative	4.5	-	Positive	Negative	Positive
X061	25/04/09	Brikamabah	40	7.4	Negative	4.2	-	Positive	Negative	Positive
X062	25/04/09	Brikamabah	74	9.0	Negative	4.2	-	Negative	Negative	Positive
X063	25/04/09	Brikamabah	30	8.6	Negative	4	-	Negative	Negative	Positive

X064	25/04/09	Brikamabah	10	6.4	Negative	4.3	-	Positive	Positive	Positive
X065	25/04/09	Brikamabah	30	7.8	Negative	4.3	-	Positive	Positive	Positive
X066	25/04/09	Brikamabah	25	7.0	Positive	5	-	Negative	Positive	Positive
X067	25/04/09	Brikamabah	13	8.6	Positive	5.2	-	Negative	Positive	Positive
X068	25/04/09	Brikamabah	40	7.6	Negative	4.2	-	Negative	Positive	Positive
X069	26/04/09	M'fana	42	7.2	Negative	6.3	-	Negative	Negative	Negative
X070	26/04/09	M'fana	25	8.2	Negative	4.3	-	Positive	Positive	Positive
X071	26/04/09	M'fana	31	8.0	Negative	4.4	-	Positive	Negative	Negative
X072	26/04/09	M'fana	39	6.7	Negative	5.1	-	Negative	Positive	Positive
X073	26/04/09	M'fana	33	7.4	Positive	3.8	-	Negative	Positive	Positive
X074	26/04/09	M'fana	31	7.5	Positive	4.4	-	Negative	Positive	Positive
X075	26/04/09	M'fana	34	8.2	Negative	5	-	Negative	Positive	Positive
X076	26/04/09	M'fana	36	8.8	Negative	5.2	-	Negative	Negative	Negative
X077	26/04/09	M'fana	20	9.4	Negative	5.2	-	Negative	Positive	Positive
X078	27/04/09	Wasu	40	7.2	Negative	4.4	-	Negative	Negative	Positive
X079	27/04/09	Wasu	32	7.8	Negative	3.7	-	Negative	Negative	Negative
X080	27/04/09	Wasu	34	7.4	Negative	4.8	-	Negative	Negative	Positive
X081	27/04/09	Wasu	36	7.4	Negative	3.9	-	Negative	Negative	Positive
X082	27/04/09	Wasu	38	7.4	Negative	4.2	-	Negative	Negative	Negative
X083	27/04/09	Wasu	20	7.4	Negative	4.2	-	Negative	Positive	Positive
X084	27/04/09	Wasu	28	6.4	Negative	3.6	-	Negative	Negative	Positive
X085	27/04/09	Wasu	29	7.0	Negative	4.9	-	Positive	Negative	Positive
X086	27/04/09	Wasu	29	7.8	Negative	4.5	-	Positive	Negative	Positive
X087	28/04/09	Jareng	-	-	-	4	-	Negative	Positive	Positive
X088	28/04/09	Jareng	-	-	-	2.9	-	Negative	Positive	Positive
X089	28/04/09	Jareng	-	-	-	3	-	Negative	Negative	Positive
X090	30/04/09	Sambel Kunda	30	8.0	Positive	3.9	-	Positive	Positive	Positive
X091	30/04/09	Kerewan	22	6.6	Negative	3.3	-	Negative	Positive	Positive
X092	30/04/09	Kerewan	40	7.4	Negative	4	-	Positive	Negative	Positive
X093	30/04/09	Kerewan	19	10.2	Positive	3.8	-	Positive	Positive	Positive
X094	30/04/09	Kerewan	38	9.4	Negative	3.8	-	Negative	Negative	Positive
X095	30/04/09	Kerewan	20	-	Negative	3.9	-	Negative	Negative	Positive
X096	30/04/09	Kerewan	33	-	Negative	4.5	-	Negative	Negative	Negative

X097	30/04/09	Kerewan	19	-	-	3.8	-	Negative	Positive	Positive
X098	30/04/09	Missera	43	7.8	-	3.6	-	Negative	Negative	Positive
X099	30/04/09	Missera	22	9.6	-	3.8	-	Negative	Negative	Positive
X100	30/04/09	Missera	42	9.8	-	3.8	-	Negative	Negative	Positive
X101	03/05/09	M'fana	-	-	-	3.6	-	Negative	Positive	Negative
X102	03/05/09	M'fana	-	-	-	5.1	-	Negative	Negative	Positive
X103	03/05/09	M'fana	-	-	-	3.4	-	Negative	Negative	Positive

Appendix 7

Sample collection details, field results and PCR based speciation results for the 2009 Gambian samples, collected from equines brought to clinics run by the GHDT, The Gambia. - = Data not available. Missing sample data arose as a result of record keeping errors during clinics and typically only became apparent during collation of all the data following conclusion of each day's clinic.

5’ **CTGAGTGCTCCATGTGCCAC**GTTGGCACGCTCCACTGTCTAGCGTGACGC
GATGGCCCGTGCACTGTCCCGCACCCCTTCCCCACTCCCTTTCGCACCTC
TCGCTCCGGCCGTGCGCCTTCT**TCAGGTTGGTGTTCTGGTGG** 3’

Appendix 8

The amplified *T. vivax* specific sequence with the sequences of TVW 1 and TVW 2 (Masiga *et al.*, 1992) highlighted in bold at the start and end of the sequence respectively

	Locus and allele sizes[a]											
Sample	TV6		TV14		TV24		TV17		TV31		TV49	
Gam102	0	0	211	211	152	152	206	206	506	506	0	0
Gam132	279	279	211	211	152	157	143	206	476	476	103	138
X066	279	279	211	229	122	122	143	206	476	476	103	138
X054	302	302	229	229	140	140	185	185	476	476	103	103
ECG007	279	279	211	211	140	152	206	206	506	506	103	138
X049	293	293	211	211	0	0	143	143	476	476	103	103
X102	0	0	211	211	157	157	143	143	476	476	4	4
X098	279	279	211	229	152	152	143	143	506	506	103	103
Gam093	279	279	211	211	152	152	143	206	476	476	103	138
X067	279	279	211	229	157	157	143	185	476	476	103	138
X088	293	302	211	229	152	157	143	143	476	476	103	138

Appendix 9

Genotyping results of the eleven genotyped *T. vivax* samples from the Gambia 2009 population following genotyping at six microsatellite markers. [a] Allele size given in bp. 0 indicates missing data at this locus for a given sample. [b] MLG, multilocus genotype, where each multilocus genotype represents a unique combination of microsatellite alleles across the six loci examined for this study, with numbers a continuation of those identified in the Gambia 2006 population.

Appendix 10

Sample	Locus and allele sizes[a] TCM 1		TCM 2		TCM 3		TCM 4		TCM 5		TCM 6		TCM 7		MLG[b]
ECG005	176	10	181	181	185	185	156	156	156	156	169	180	179	195	1
ECG006	170	10	181	181	182	185	153	153	161	161	175	185	176	185	2
GAM076	173	176	183	183	182	182	153	153	156	156	185	185	176	179	3
GAM095	176	180	185	185	175	185	156	156	161	153	175	185	176	176	4
GAM106	176	176	185	185	185	185	160	163	161	161	180	185	179	182	5
GAM107	176	180	177	179	171	175	156	156	153	153	180	185	179	179	6
GAM108	176	176	191	191	171	175	153	153	156	156	169	175	176	182	7
GAM120	173	173	181	181	175	182	156	156	161	161	185	185	157	185	8
GAM130	178	10	183	183	185	191	156	156	161	165	185	185	176	192	9
GAM146	176	176	181	181	163	191	156	156	161	161	180	185	179	179	10
X003	180	180	177	177	175	175	156	156	153	153	180	180	176	176	11
X005	173	180	191	191	171	171	153	153	156	156	175	188	176	179	12
X023	170	10	181	181	182	185	153	153	156	156	180	185	176	185	13
X026	176	180	181	181	171	171	156	156	161	161	169	180	179	182	14
X037	176	180	181	181	171	175	156	156	170	170	169	188	185	185	15
X048	170	176	177	181	171	171	156	156	161	165	169	185	176	185	16
X056	170	170	181	181	185	185	153	153	156	161	185	185	157	176	17
X057	170	170	181	181	185	185	153	153	156	156	185	185	157	176	18
X065	176	200	177	177	171	171	153	153	156	156	180	185	176	179	19
X067	176	176	177	191	148	148	156	156	156	156	185	185	185	185	20
X070	178	180	194	194	148	185	153	153	156	153	169	185	176	185	21
X073	176	176	181	181	171	175	156	156	148	148	185	195	176	176	22
X083	176	180	177	181	163	163	156	160	156	165	180	185	179	179	23

X088	176	176	183	183	185	185	156	156	156	156	173	173	176	179	24
X091	176	176	181	181	148	171	156	156	153	156	185	185	179	185	25
X101	176	176	177	177	148	171	156	160	156	165	185	185	179	185	26
GAM061	176	176	0	0	171	171	0	0	153	153	0	0	182	182	
GAM096	176	176	177	177	163	163	156	156	156	175	0	0	179	179	
GAM100	180	180	0	0	163	200	156	156	156	165	173	173	176	182	
GAM105	176	180	0	0	171	175	0	0	156	165	180	185	179	179	
GAM131	0	0	177	177	148	171	156	160	156	165	185	185	179	185	
GAM141	176	176	177	177	163	175	156	156	165	165	0	0	179	179	
X024	180	180	191	191	171	175	0	0	156	156	173	175	176	182	
X025	173	173	177	177	0	0	0	0	161	161	180	180	185	185	
X055	170	170	0	0	0	0	0	0	156	156	169	180	179	179	
X064	173	176	181	181	171	175	156	156	156	165	188	188	0	0	
X066	0	0	191	191	171	171	153	153	0	0	175	188	176	179	
X074	170	170	0	0	175	175	156	156	0	0	185	188	179	185	
X077	180	180	0	0	175	175	156	156	153	153	175	175	176	176	
X087	0	0	181	181	171	191	156	156	153	167	169	185	179	192	
X093	170	10	181	181	0	0	153	153	161	153	180	180	176	185	

Appendix 10

Genotyping results of the 41 genotyped *T. congolense* samples from The Gambia, 2009 population following genotyping at 7 microsatellite markers. [a] Allele size given in bp. 0 indicates missing data at this locus for a given sample. [b] MLG, multilocus genotype, where each multilocus genotype represents a unique combination of microsatellite alleles across the six loci examined for this study.

Appendix 11

Sample	Population	Focus of origin	Collection period	Locus and allele sizes[a]														MLG[b]
				Ch1/18		Ch2/PLC		Ch3/IJ15/I		Ch3/5L5		Ch4/M12C12		Ch5/JS2		Ch9/4		
	Malawi																	
NKK/T/022	2003	Malawi	2002-2003	153	162	136	136	132	145	111	111	98	107	98	98	140	140	1
NKK/T/025		Malawi	2002-2003	153	162	136	136	132	145	111	111	98	107	98	98	140	140	1
NKK/T/042		Malawi	2002-2003	162	162	136	136	145	145	111	111	98	107	98	98	148	148	2
NKK/T/037		Malawi	2002-2003	162	162	136	136	120	120	111	111	98	107	169	169	140	148	4
NKK/T/030		Malawi	2002-2003	162	162	136	136	120	145	111	111	98	107	169	169	140	148	5
NKK/T/058		Malawi	2002-2003	162	162	136	136	120	145	111	111	98	107	169	169	140	148	5
NKK/T/054		Malawi	2002-2003	162	162	136	143	115	145	111	111	98	107	98	98	148	148	6
NKK/T/053		Malawi	2002-2003	162	162	136	148	145	145	111	111	98	107	98	98	140	148	7
NKK/T/006		Malawi	2002-2003	162	162	136	148	132	132	111	111	98	107	98	98	148	148	8
NKK/T/021		Malawi	2002-2003	153	162	136	124	132	132	111	111	107	107	98	98	140	140	9
NKK/T/010		Malawi	2002-2003	153	162	136	136	132	145	111	111	107	107	98	98	140	140	10
NKK/T/005		Malawi	2002-2003	162	162	136	136	132	145	111	111	107	107	169	169	140	140	11
NKK/T/044		Malawi	2002-2003	162	162	136	143	145	145	111	111	107	107	98	98	140	148	12
NKK/T/027		Malawi	2002-2003	162	162	136	136	132	145	111	120	98	107	90	98	140	148	14
NKK/T/039		Malawi	2002-2003	153	162	136	136	115	145	111	120	98	107	98	98	140	148	15
NKK/T/012		Malawi	2002-2003	162	162	136	136	132	132	111	120	98	107	98	98	140	140	16
NKK/T/035		Malawi	2002-2003	162	162	136	136	132	145	111	120	98	107	98	98	140	140	17
NKK/T/009		Malawi	2002-2003	162	162	136	136	132	145	111	120	98	107	98	169	140	140	18
NKK/T/038		Malawi	2002-2003	162	162	136	136	115	132	111	120	98	107	98	169	140	148	19
NKK/T/028		Malawi	2002-2003	162	162	136	136	115	115	111	120	98	107	169	169	140	148	20
NKK/T/007		Malawi	2002-2003	145	162	136	136	132	145	111	120	107	107	98	98	140	140	32
NKK/T/026		Malawi	2002-2003	162	162	136	136	132	132	111	120	107	107	98	98	144	144	33
NKK/T/056		Malawi	2002-2003	162	162	136	136	115	145	111	120	107	107	169	169	140	148	34
NKK/T/002		Malawi	2002-2003	162	162	136	136	132	132	111	111	98	107	98	169	0	0	
NKK/T/057		Malawi	2002-2003	162	162	136	136	0	0	111	120	107	107	98	98	0	0	
NKK/T/003		Malawi	2002-2003	162	162	136	136	115	115	111	111	98	107	169	169	0	0	
NKK/T/049		Malawi	2002-2003	162	162	136	143	0	0	111	120	98	107	98	169	148	148	

NKK/T/040		Malawi	2002-2003	162	162	136	136	115	145	111	120	98	107	169	169	0	0	
SER020	Soroti 2003	Soroti	2002-2003	162	176	136	151	141	141	111	120	98	107	98	98	144	161	21
SER041		Soroti	2002-2003	162	176	136	151	141	141	111	120	98	107	98	98	144	161	21
SER062		Soroti	2002-2003	162	176	151	151	141	141	111	120	98	107	98	98	144	161	28
SER077		Soroti	2002-2003	162	176	151	166	141	141	111	120	98	107	90	98	144	161	29*
SER003		Soroti	2002-2003	162	176	151	166	141	141	111	120	98	107	98	98	133	161	30
SER006		Soroti	2002-2003	162	176	151	166	141	141	111	120	98	107	98	98	144	161	31*
SER007		Soroti	2002-2003	162	176	151	166	141	141	111	120	98	107	98	98	144	161	31*
SER034		Soroti	2002-2003	162	176	151	166	141	141	111	120	98	107	98	98	144	161	31*
SER064		Soroti	2002-2003	162	176	151	166	141	141	111	120	98	107	98	98	144	161	31*
SER066		Soroti	2002-2003	162	176	151	166	141	141	111	120	98	107	98	98	144	161	31*
SER067		Soroti	2002-2003	162	176	151	166	141	141	111	120	98	107	98	98	144	161	31*
SER079		Soroti	2002-2003	162	176	151	166	141	141	111	120	98	107	98	98	144	161	31*
SER057		Soroti	2002-2003	162	176	136	151	141	141	111	120	107	107	98	98	144	161	35
SER016		Soroti	2002-2003	162	176	143	151	141	141	111	120	107	107	98	98	144	161	36
SER001		Soroti	2002-2003	162	176	151	151	141	141	111	120	107	107	98	98	144	161	42
SER027		Soroti	2002-2003	162	176	151	151	141	141	111	120	107	107	98	98	144	161	42
SER047		Soroti	2002-2003	162	176	151	151	141	141	111	120	107	107	98	98	144	161	42
SER093		Soroti	2002-2003	162	176	151	151	141	141	111	120	107	107	98	98	144	161	42
SER098		Soroti	2002-2003	162	176	151	151	141	141	111	120	107	107	98	98	144	161	42
SER105		Soroti	2002-2003	162	176	151	151	141	141	111	120	107	107	98	98	144	161	42
SER111		Soroti	2002-2003	162	176	151	151	141	141	111	120	107	107	98	98	144	161	42
SER058		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	90	98	144	144	43
SER042		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	90	98	144	161	44
SER024		Soroti	2002-2003	162	162	151	166	141	141	111	120	107	107	98	98	144	161	45
SER044		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	144	46
SER060		Soroti	2002-2003	162	176	151	166	115	141	111	120	107	107	98	98	144	161	47
SER002		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER008		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER009		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER010		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER011		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49

SER012	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER013	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER014	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER017	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER018	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER022	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER023	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER025	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER026	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER028	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER029	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER030	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER031	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER032	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER033	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER036	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER037	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER039	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER040	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER045	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER048	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER049	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER051	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER053	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER055	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER056	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER061	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER063	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER065	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER068	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER069	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER071	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER072	Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49

SER073		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER074		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER075		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER076		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER078		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER080		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER082		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER083		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER084		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER085		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER086		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER087		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	161	49
SER004		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	98	144	164	50
LIRI003		Soroti	2002-2003	162	162	151	166	141	141	111	120	107	107	98	108	144	155	51
SER108		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	98	124	144	161	52
SER059		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	134	98	98	144	161	56
SER054		Soroti	2002-2003	0	0	151	166	141	141	111	120	107	107	98	98	144	161	
SER070		Soroti	2002-2003	0	0	151	166	141	141	111	120	107	107	98	98	144	161	
SER050		Soroti	2002-2003	162	176	0	0	141	141	111	120	107	107	98	98	144	161	
SER081		Soroti	2002-2003	162	176	151	166	141	141	111	120	107	107	0	0	144	161	
LIRI001	Tororo 2003	Tororo	2002-2003	162	162	124	151	122	141	111	120	98	107	98	98	144	161	13
LIRI033		Tororo	2002-2003	162	176	151	151	115	138	111	120	98	107	90	98	144	150	22
LIRI029		Tororo	2002-2003	162	176	151	151	122	122	111	120	98	107	90	98	144	161	23
LIRI005		Tororo	2002-2003	162	176	151	151	122	141	111	120	98	107	90	98	144	161	24
LIRI018		Tororo	2002-2003	162	176	151	151	122	141	111	120	98	107	90	98	144	161	24
LIRI030		Tororo	2002-2003	162	176	151	151	122	141	111	120	98	107	90	98	144	161	24
LIRI031		Tororo	2002-2003	162	176	151	151	122	141	111	120	98	107	90	98	144	161	24
LIRI032		Tororo	2002-2003	162	176	151	151	122	141	111	120	98	107	90	98	144	161	24
LIRI008		Tororo	2002-2003	162	162	151	151	122	141	111	120	98	107	98	98	144	161	25
LIRI022		Tororo	2002-2003	162	162	151	151	122	141	111	120	98	107	98	98	144	161	25
LIRI016		Tororo	2002-2003	162	176	151	151	141	141	111	120	98	107	98	98	144	155	26
LIRI002		Tororo	2002-2003	162	176	151	151	122	141	111	120	98	107	98	98	144	161	27

LIRI004		Tororo	2002-2003	162	176	151	151	122	141	111	120	98	107	98	98	144	161	27
LIRI012		Tororo	2002-2003	162	176	151	151	122	141	111	120	98	107	98	98	144	161	27
LIRI028		Tororo	2002-2003	162	176	151	151	122	141	111	120	98	107	98	98	144	161	27
LIRI025		Tororo	2002-2003	162	176	151	166	141	141	111	120	98	107	90	98	144	161	29*
LIRI015		Tororo	2002-2003	162	176	151	166	141	141	111	120	98	107	98	98	144	161	31*
LIRI011		Tororo	2002-2003	162	162	151	151	122	141	111	120	107	107	90	98	144	161	39
LIRI024		Tororo	2002-2003	162	176	151	151	122	122	111	120	107	107	98	98	144	161	41
LIRI010		Tororo	2002-2003	162	162	151	151	122	141	111	120	107	152	90	98	144	161	57
LIRI023		Tororo	2002-2003	162	162	151	151	122	141	111	120	107	152	90	98	144	161	57
LIRI007		Tororo	2002-2003	162	176	151	151	122	141	120	120	98	107	90	98	144	161	58
LIRI014		Tororo	2002-2003	162	162	151	151	122	141	120	120	98	107	98	98	144	155	59
LIRI019		Tororo	2002-2003	162	162	151	151	141	141	120	120	98	107	98	108	144	155	60
LIRI009		Tororo	2002-2003	162	176	151	166	122	141	120	120	98	107	90	98	144	161	61
LIRI013		Tororo	2002-2003	162	176	151	166	122	141	120	120	98	107	98	98	144	161	62
LIRI027		Tororo	2002-2003	162	162	151	151	0	0	111	120	98	107	90	98	0	0	
LIRI026		Tororo	2002-2003	162	176	151	151	0	0	111	120	98	107	90	98	144	161	
LIRI021		Tororo	2002-2003	162	176	151	151	0	0	111	120	107	122	90	98	144	164	
	Tororo																	
K3192	pre 1990	Tororo focus	1988-1990	162	162	143	157	138	141	111	120	107	107	108	108	144	155	37
Mela80		Tororo focus	1988-1990	162	219	148	151	122	141	111	120	107	107	98	98	150	155	38
K3442		Tororo focus	1988-1990	162	176	151	151	120	141	111	120	107	107	98	98	144	161	40
K3438		Tororo focus	1988-1990	162	176	151	166	122	141	111	120	107	107	98	98	144	161	48
K3198		Tororo focus	1988-1990	162	162	157	157	138	141	111	120	107	107	98	98	144	155	53
K3440		Tororo focus	1988-1990	162	176	157	157	120	141	111	120	107	107	98	98	144	161	54
K3448		Tororo focus	1988-1990	162	176	157	157	120	141	111	120	107	107	98	98	144	161	54
K2340		Tororo focus	1988-1990	162	162	157	157	138	141	111	120	107	107	108	108	144	155	55
EA2498[c]		Tororo focus	1977	162	162	143	151	138	141	120	120	107	107	98	108	144	155	63
K3183		Tororo focus	1988-1990	162	162	148	157	138	141	120	120	107	107	98	98	144	155	64
K2976		Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	98	98	144	155	65
K3176		Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	98	98	144	155	65
K3180		Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	98	98	144	155	65
K3196		Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	98	98	144	155	65
UgA90		Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	98	98	144	155	65

UgL	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	98	98	144	155	65
K3199	Tororo focus	1988-1990	162	162	157	157	120	154	120	120	107	107	98	108	144	155	66
K3203	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	98	108	144	155	67
K3205	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	98	108	144	155	67
K3206	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	98	108	144	155	67
UgM	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	98	108	144	155	67
K3188	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	108	108	144	144	68
K3189	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	108	108	144	144	68
K2344	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	108	108	144	155	69
K2427	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	108	108	144	155	69
K2428	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	108	108	144	155	69
K2556	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	108	108	144	155	69
K3185	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	108	108	144	155	69
K3190	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	108	108	144	155	69
K3186	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	108	108	144	155	69
K2381	Tororo focus	1988-1990	162	162	157	157	141	141	120	120	107	107	108	108	144	155	70
K2382	Tororo focus	1988-1990	162	162	157	157	141	141	120	120	107	107	108	108	144	155	70
UgC90	Tororo focus	1988-1990	162	162	157	166	115	138	120	120	107	107	98	10	155	161	71
K3200	Tororo focus	1988-1990	162	162	157	166	138	141	120	120	107	107	98	98	144	155	72
U89/8	Tororo focus	1988-1990	162	162	157	166	138	141	120	120	107	107	98	98	144	155	72
UgI	Tororo focus	1988-1990	162	162	157	166	138	141	120	120	107	107	98	98	144	155	72
UgJ	Tororo focus	1988-1990	162	162	157	166	138	141	120	120	107	107	98	98	144	155	72
UTAR 3[c]	Tororo focus	1981	162	162	157	166	136	141	120	120	107	107	98	108	144	155	73
1042[c]	Tororo focus	1961	162	162	157	166	138	141	120	120	107	107	98	108	144	155	74
1301	Tororo focus	1988-1990	162	162	157	166	138	141	120	120	107	107	98	108	144	155	74
K3194	Tororo focus	1988-1990	162	162	157	166	138	141	120	120	107	107	98	108	144	155	74
K3206	Tororo focus	1988-1990	162	162	157	166	138	141	120	120	107	107	98	108	144	155	74
EATRO 795[c]	Tororo focus	1964	162	162	157	166	138	141	120	120	107	107	98	108	144	155	74
K3445	Tororo focus	1988-1990	224	237	148	157	136	138	111	111	107	107	0	0	150	161	
K2380	Tororo focus	1988-1990	162	162	0	0	141	141	120	120	0	0	108	108	0	0	
MA66	Tororo focus	1988-1990	162	162	0	0	138	141	120	120	107	107	0	0	144	155	
K3110	Tororo focus	1988-1990	162	162	157	157	122	141	120	120	107	107	86	86	0	0	
K2350	Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	108	108	0	0	

MAWERA66		Tororo focus	1988-1990	162	162	151	151	138	141	120	120	107	107	0	0	144	155	
K116		Tororo focus	1988-1990	162	162	157	157	138	141	120	120	107	107	0	0	144	155	
LIL037	Kaberamaido 2009	Kaberamaido	2009-2010	162	176	151	151	141	141	111	142	107	107	98	98	144	161	75
LIL050		Kaberamaido	2009-2010	162	176	166	166	141	141	111	111	107	107	98	98	144	161	76
LIL076		Kaberamaido	2009-2010	162	176	151	151	138	141	111	111	107	107	98	98	144	161	77
LIL080		Kaberamaido	2009-2010	162	176	151	151	138	141	111	111	107	107	98	98	144	161	77
LIL001		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL002		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL004		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL006		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL007		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL009		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL013		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL014		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL019		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL020		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL022		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL023		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL024		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL039		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL040		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL041		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL042		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL045		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL048		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL062		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL064		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL066		Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78

LIL067	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL068	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL070	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL071	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL074	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL075	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL078A2	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL081	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL088	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL090	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL091	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL092	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL093	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL094	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL096	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL138	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	144	161	78
LIL031	Kaberamaido	2009-2010	162	176	151	151	141	141	111	142	107	107	98	98	0	0	
LIL015	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	0	0	
LIL021	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	0	0	
LIL025	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	0	0	
LIL027	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	0	0	
LIL034	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	0	0	
LIL036	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	0	0	
LIL060	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	0	0	
LIL061	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	0	0	
LIL072	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	0	0	
LIL073	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	0	0	
LIL082	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	0	0	

LIL087	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	107	107	98	98	0	0
LIL026	Kaberamaido	2009-2010	162	176	151	151	138	141	111	111	107	107	98	98	0	0
LIL084	Kaberamaido	2009-2010	162	176	0	0	141	141	111	111	107	107	98	98	144	161
LIL078A1	Kaberamaido	2009-2010	162	176	151	151	141	141	111	111	0	0	98	98	144	161
LIL003	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	144	161
LIL008	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	144	161
LIL038	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	144	161
LIL047	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	144	161
LIL049	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	144	161
LIL058	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	144	161
LIL059	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	144	161
LIL079	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	144	161
LIL005	Kaberamaido	2009-2010	162	176	151	151	138	141	0	0	107	107	98	98	144	161
LIL011	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	0	0
LIL017	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	0	0
LIL028	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	0	0
LIL029	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	0	0
LIL030	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	0	0
LIL033	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	0	0
LIL035	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	0	0
LIL043	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	0	0
LIL044	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	0	0
LIL046	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	0	0
LIL063	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	0	0
LIL065	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	0	0
LIL085	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	0	0
LIL086	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	107	107	98	98	0	0
LIL083	Kaberamaido	2009-2010	162	176	151	151	138	141	0	0	107	107	98	98	0	0

LIL095	Kaberamaido	2009-2010	162	176	151	151	138	141	0	0	107	107	98	98	0	0
LIL069	Kaberamaido	2009-2010	162	176	151	151	138	141	0	0	107	107	0	0	144	161
LIL016	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	0	0	98	98	0	0
LIL057	Kaberamaido	2009-2010	162	176	151	151	141	141	0	0	0	0	98	98	0	0

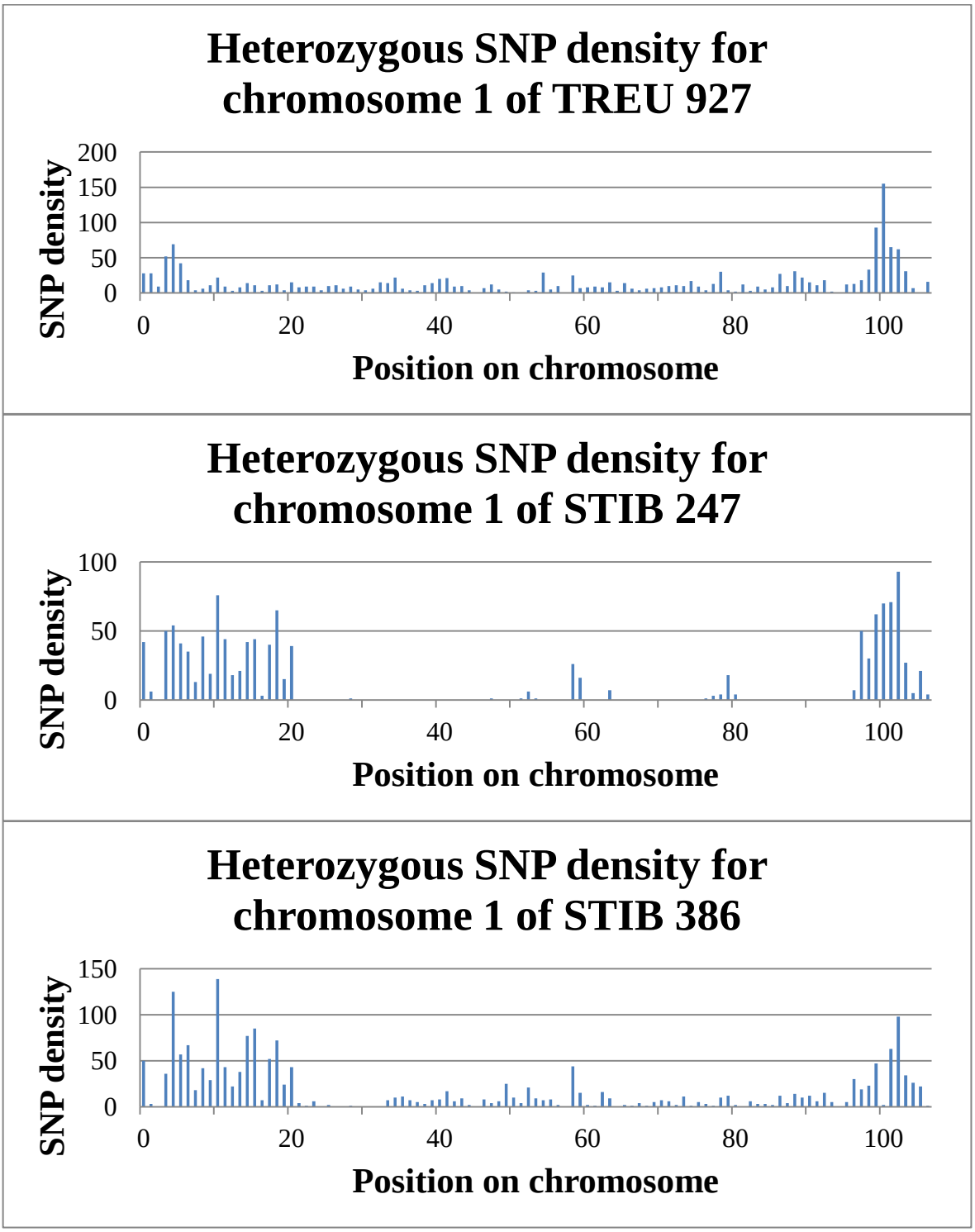
Appendix 11

Genotypes of the 277 *T. b. rhodesiense* samples used in the analysis of this sub-species in Uganda and Malawi, as presented in Chapter 5. Samples are grouped by origin population with identical genotypes within a population grouped together. The Tororo pre-1991 population encompasses samples from the wider Tororo focus, which includes districts of western Kenya in addition to a small number of isolates originating from earlier than 1988.

[a] Allele size given in bp. 0 indicates missing data at this locus for a given sample. [b] MLG, multilocus genotype, where each multilocus genotype represents a unique combination of microsatellite alleles across the seven loci examined for this study. [c] Isolates of the Tororo pre-1991 focus isolated prior to the 1988-1990 period.

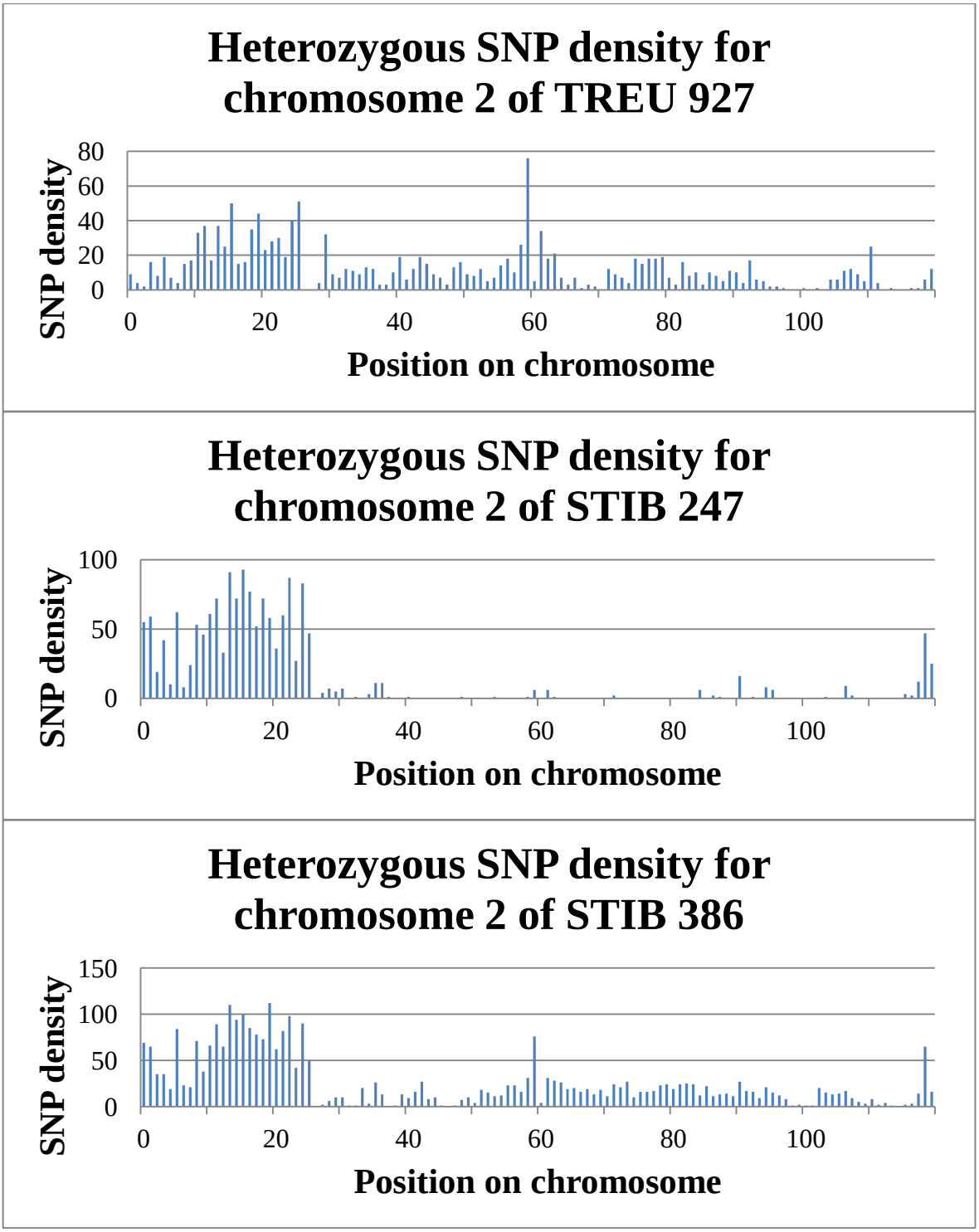
* These MLGs were identified within both the Tororo 2003 and Soroti 2003 populations.

Appendix 12



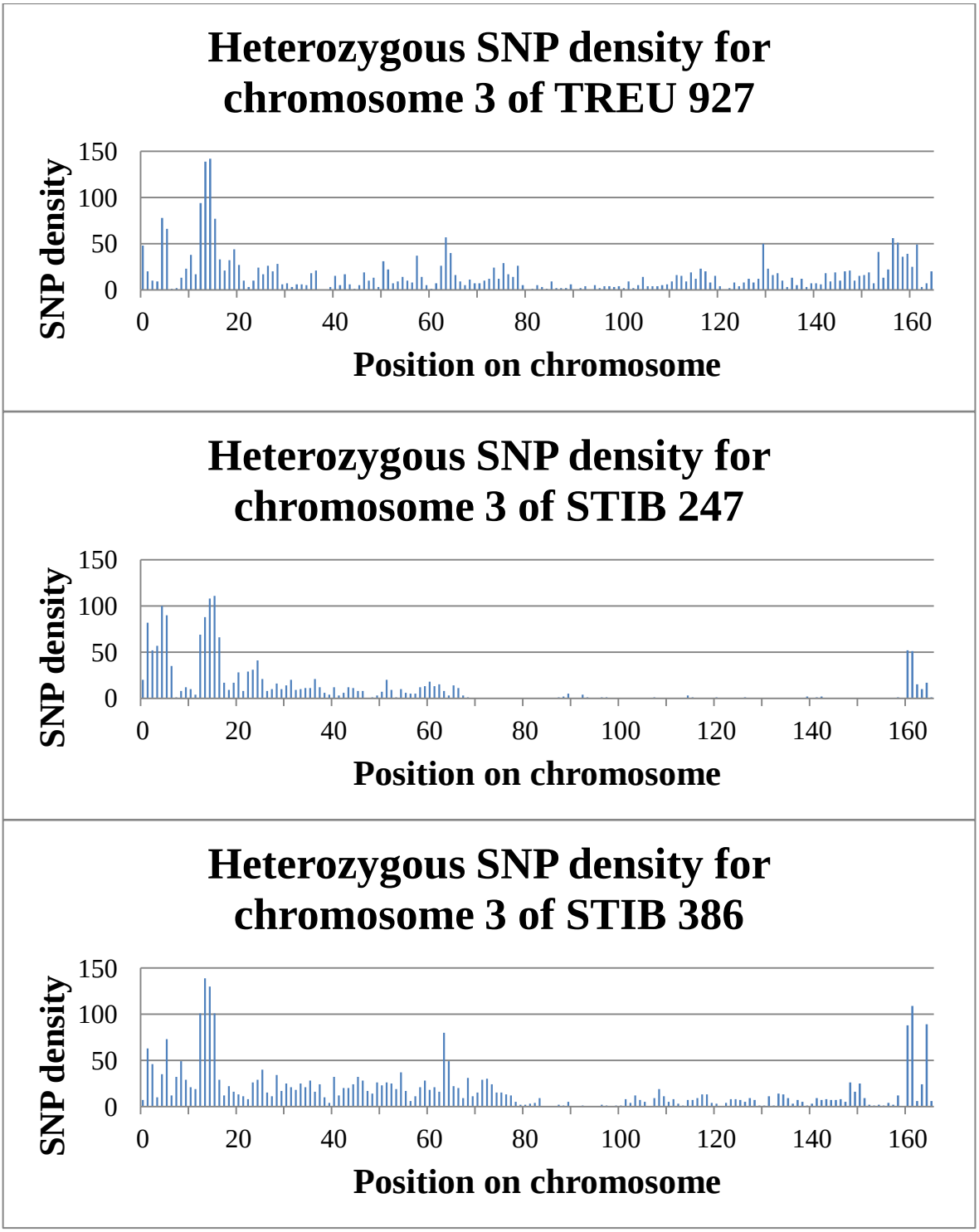
Appendix 12

Per chromosome heterozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of heterozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



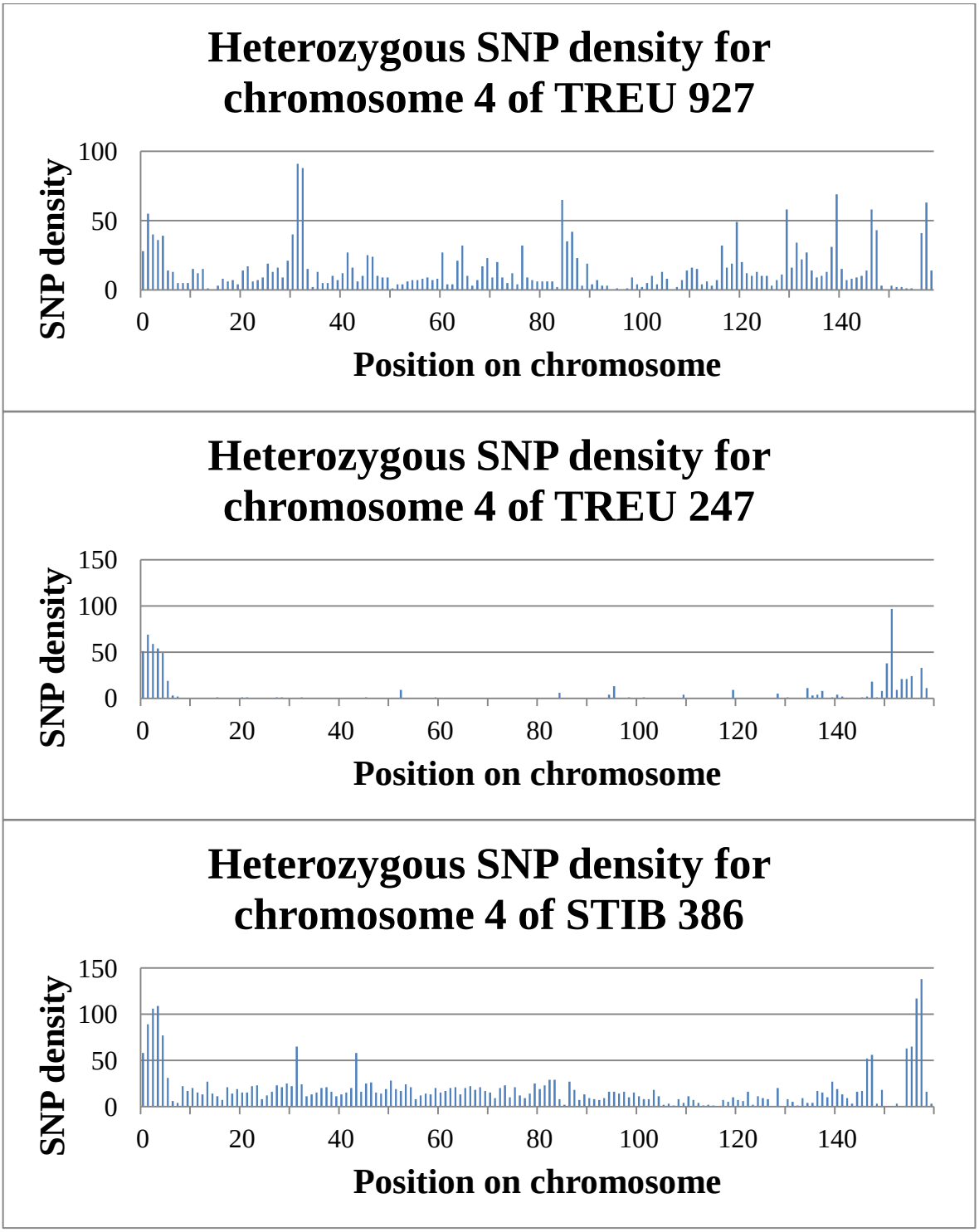
Appendix 12

Per chromosome heterozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of heterozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



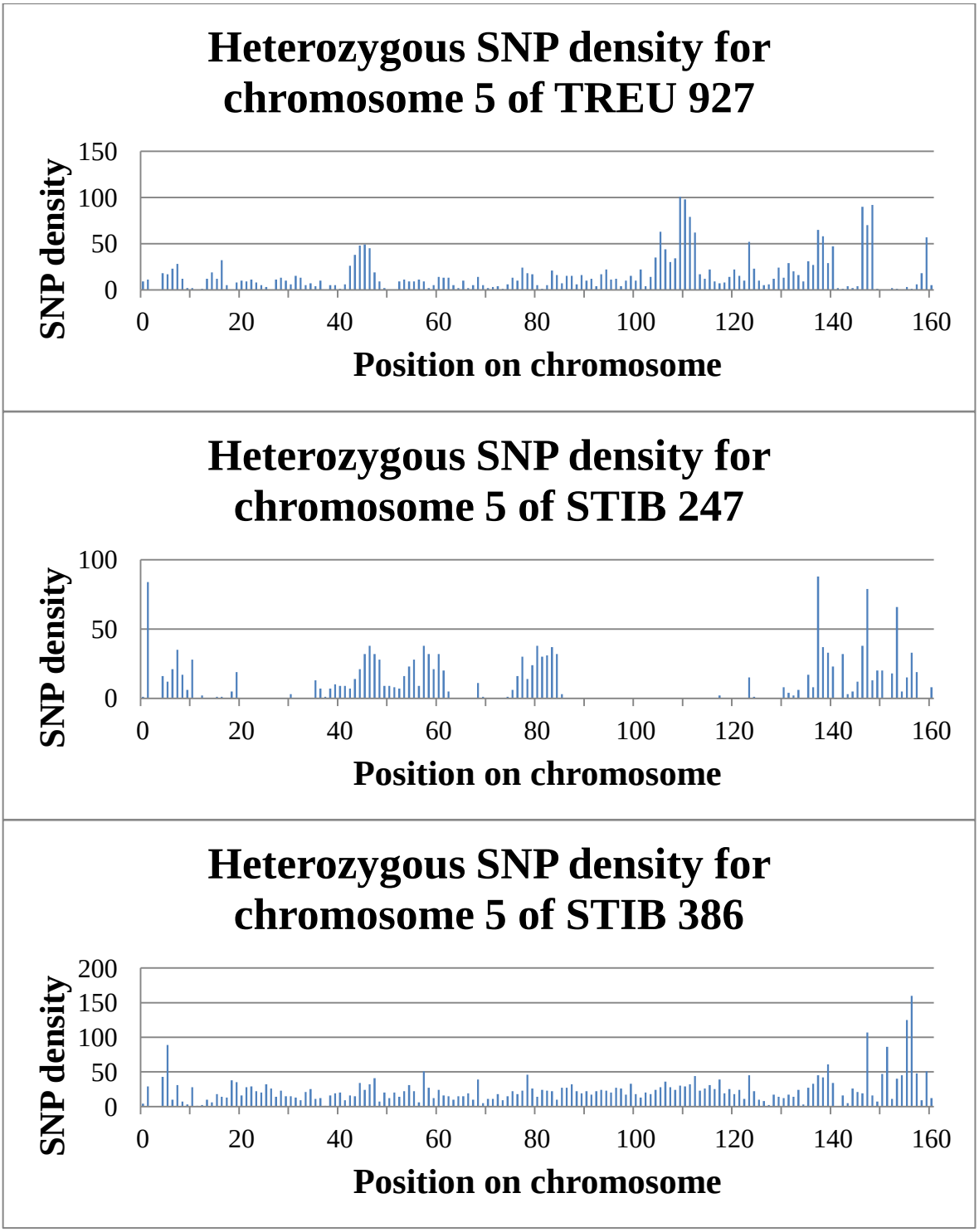
Appendix 12

Per chromosome heterozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of heterozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



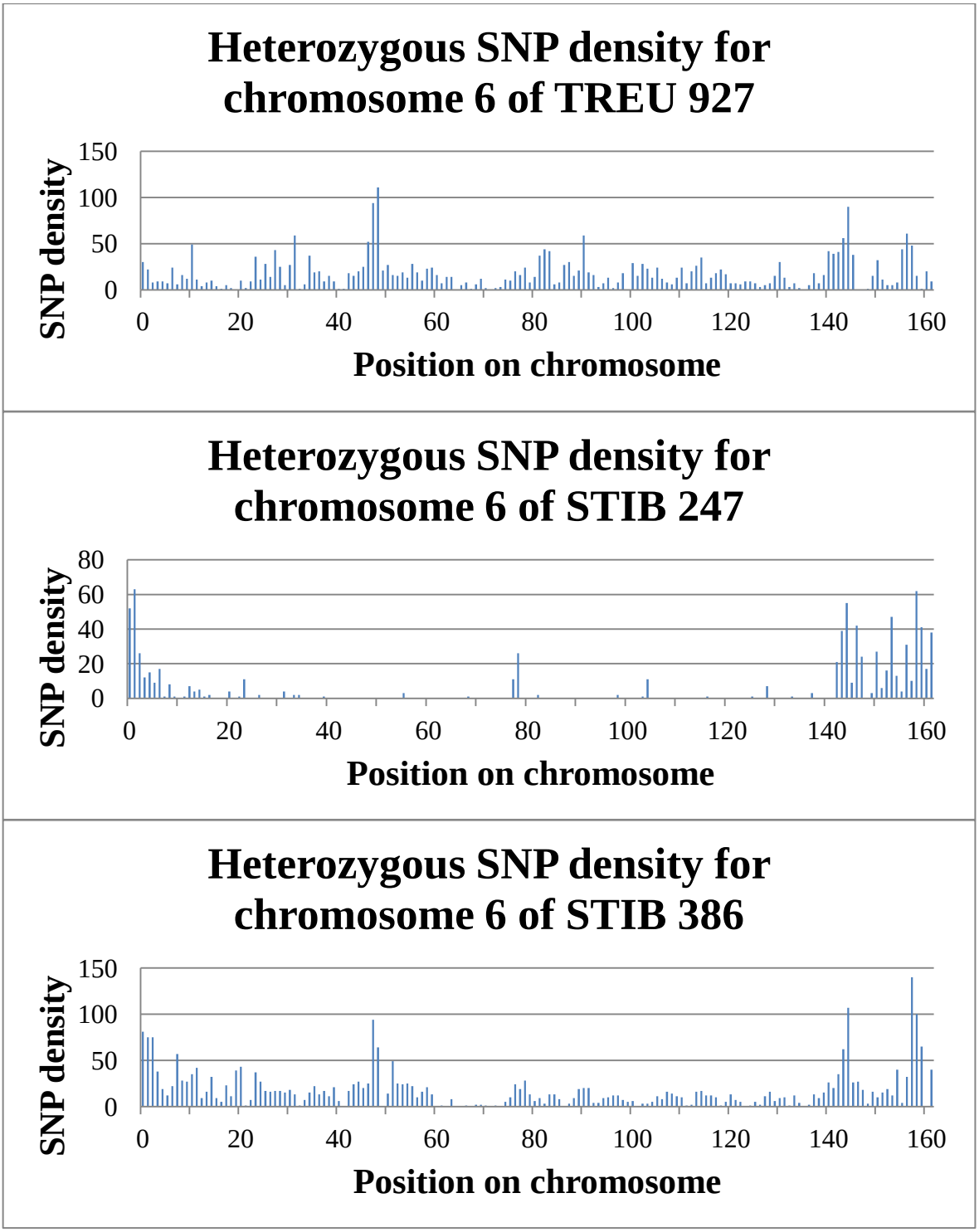
Appendix 12

Per chromosome heterozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of heterozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



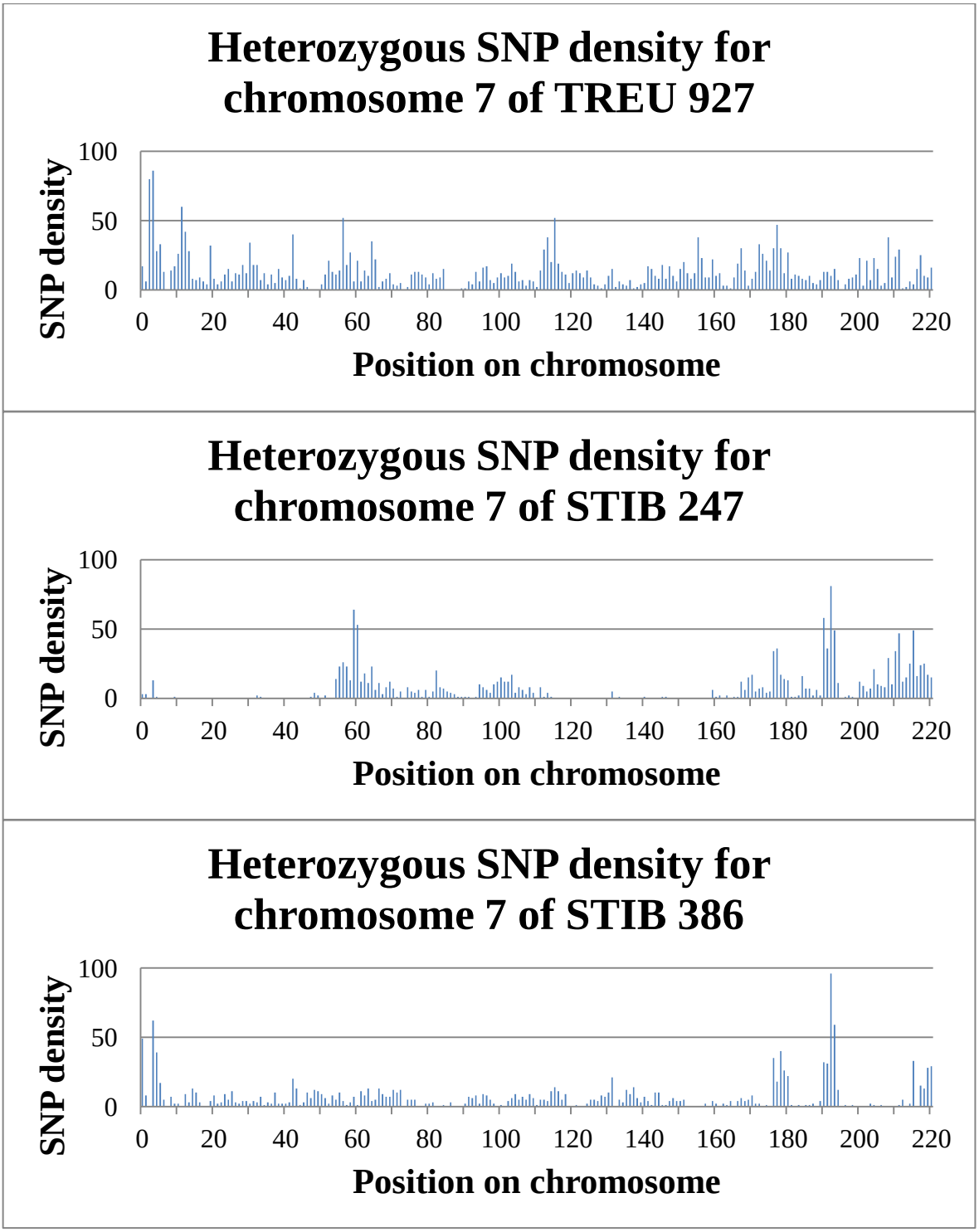
Appendix 12

Per chromosome heterozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of heterozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



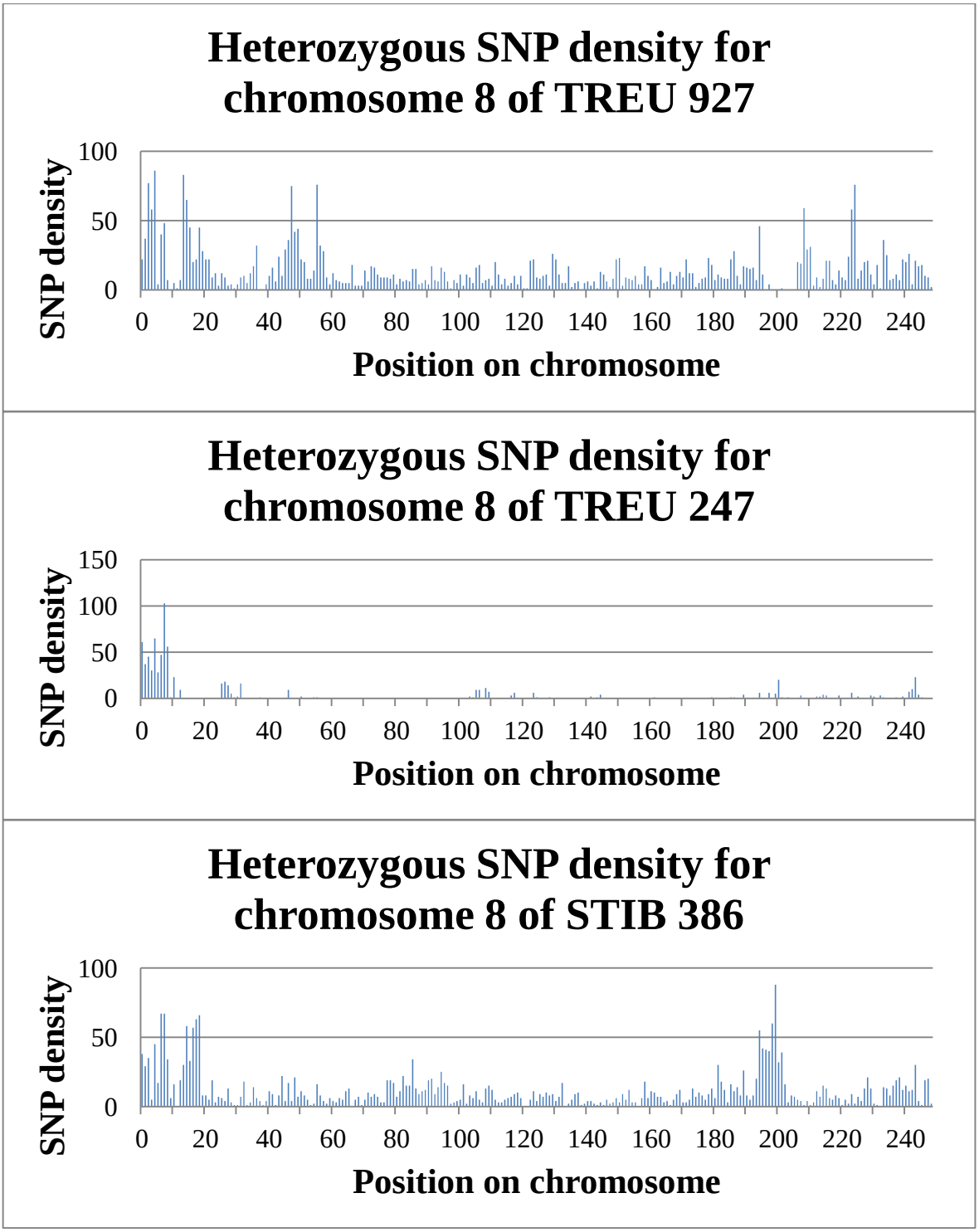
Appendix 12

Per chromosome heterozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of heterozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



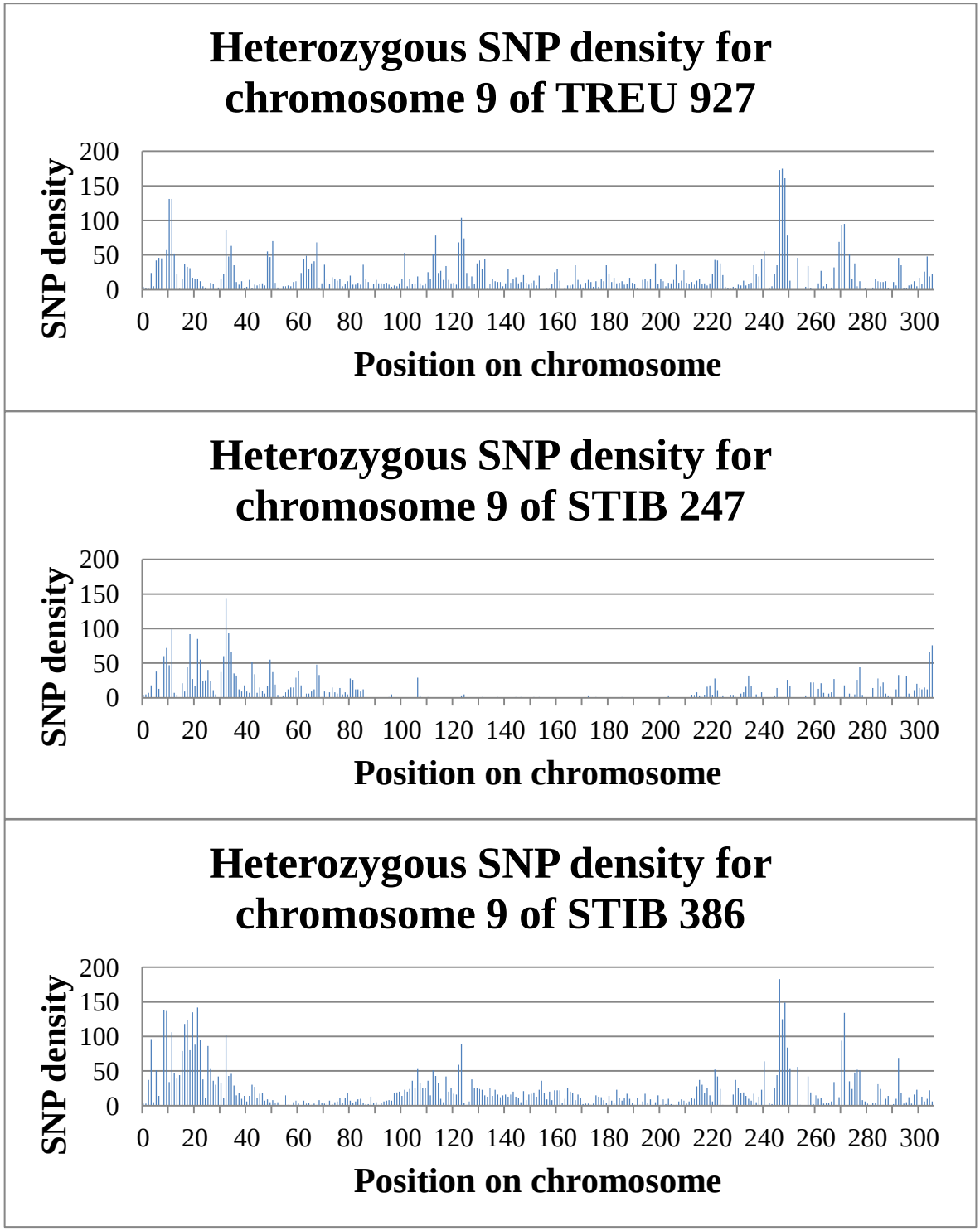
Appendix 12

Per chromosome heterozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of heterozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



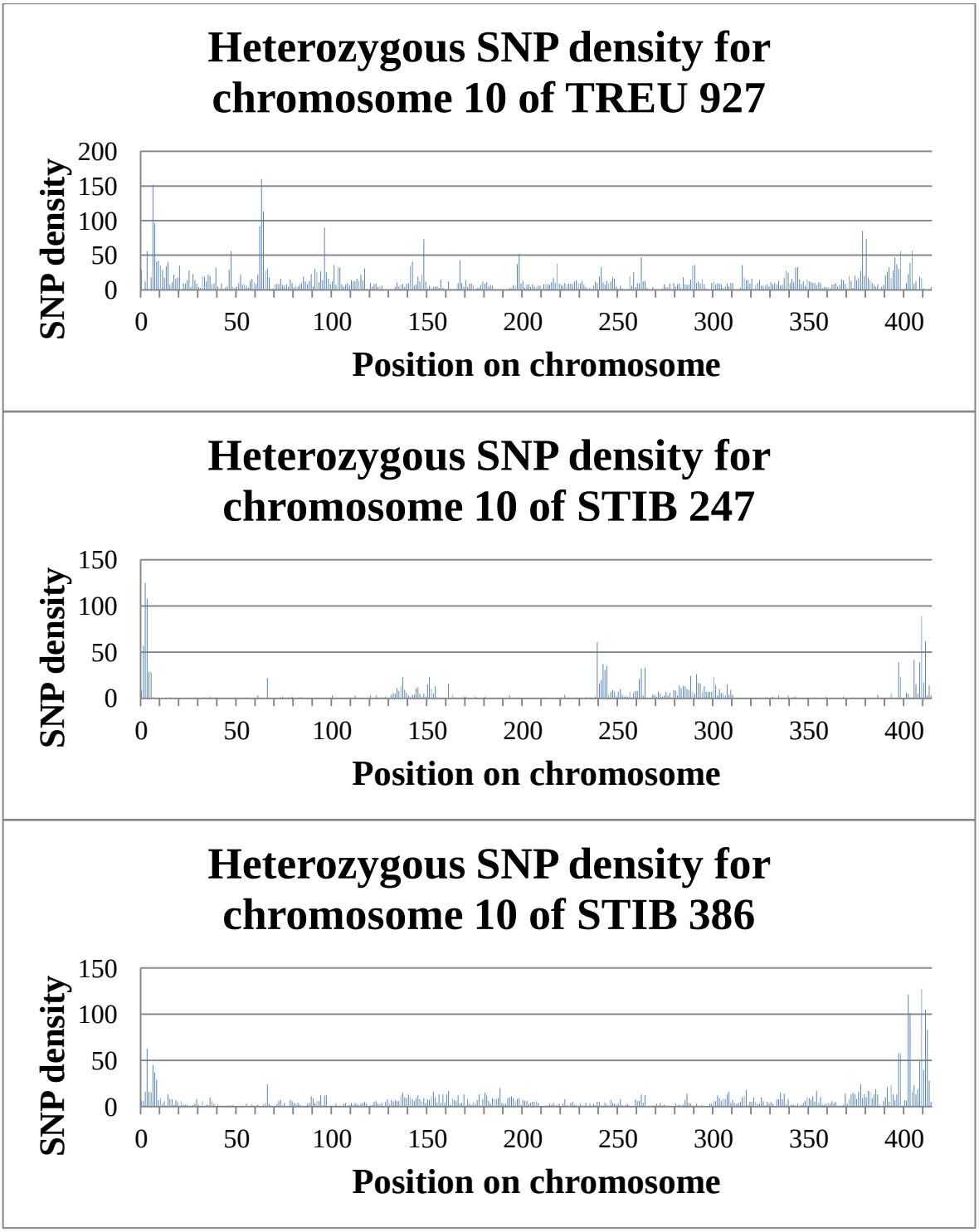
Appendix 12

Per chromosome heterozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of heterozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



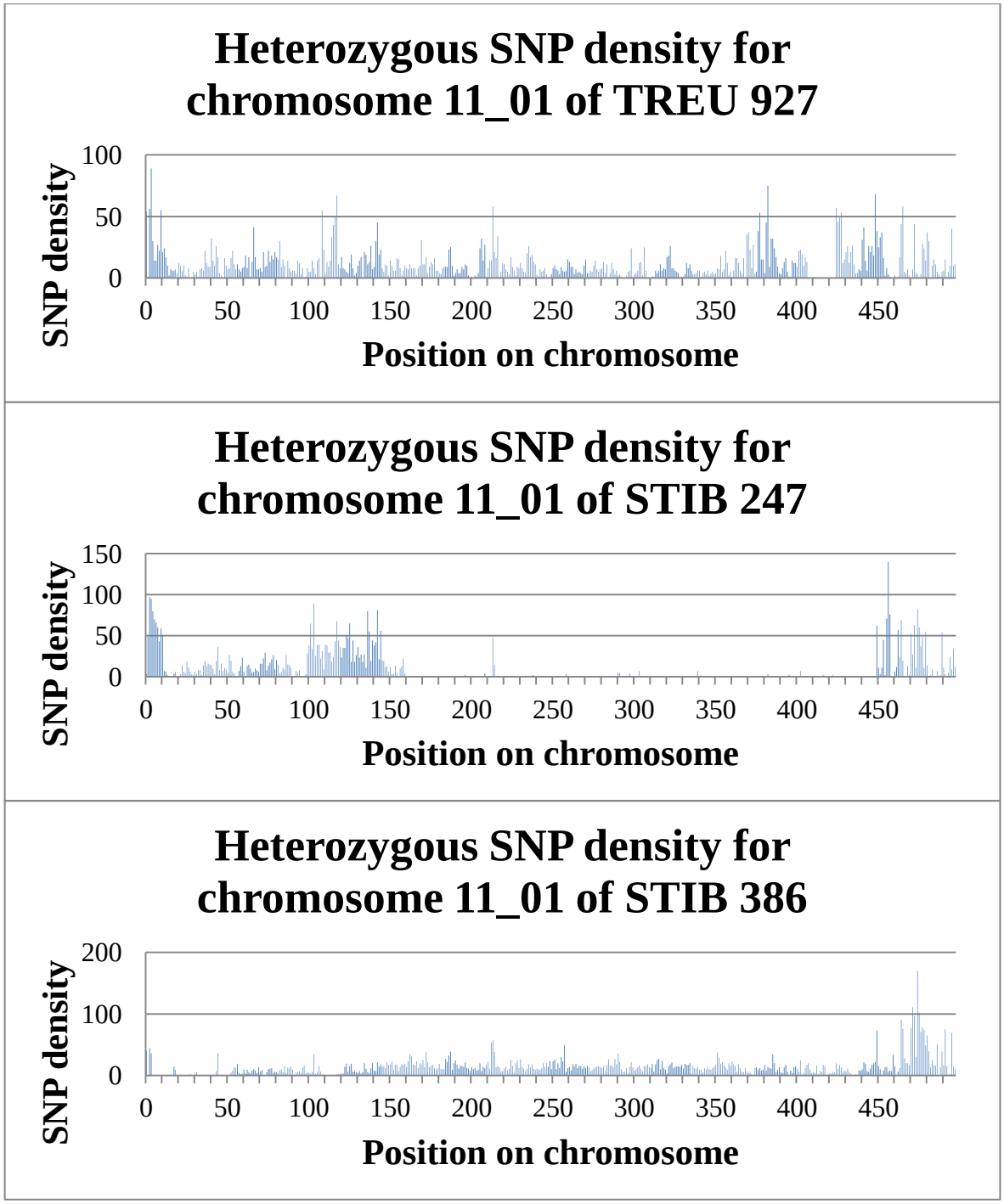
Appendix 12

Per chromosome heterozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of heterozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



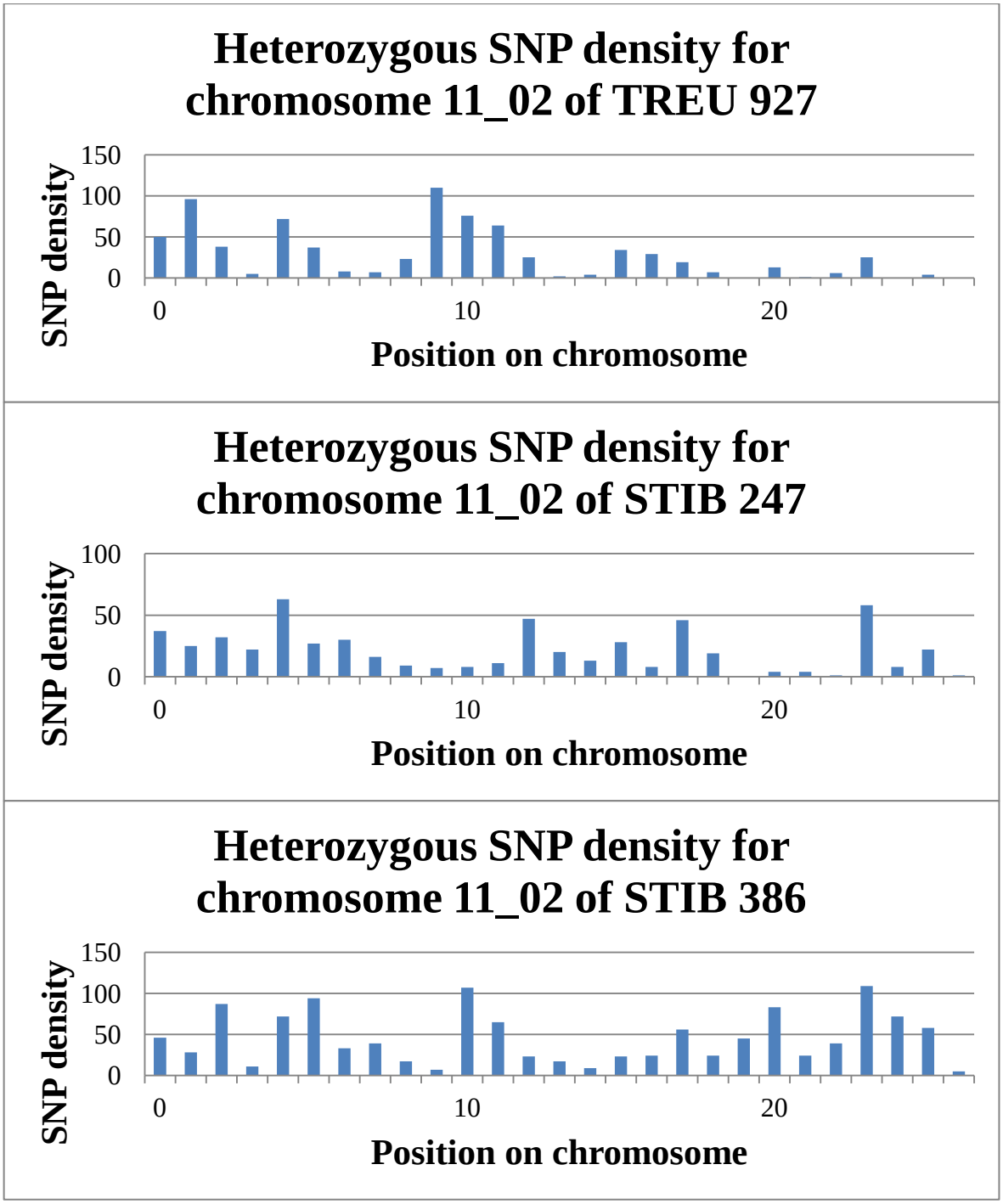
Appendix 12

Per chromosome heterozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of heterozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



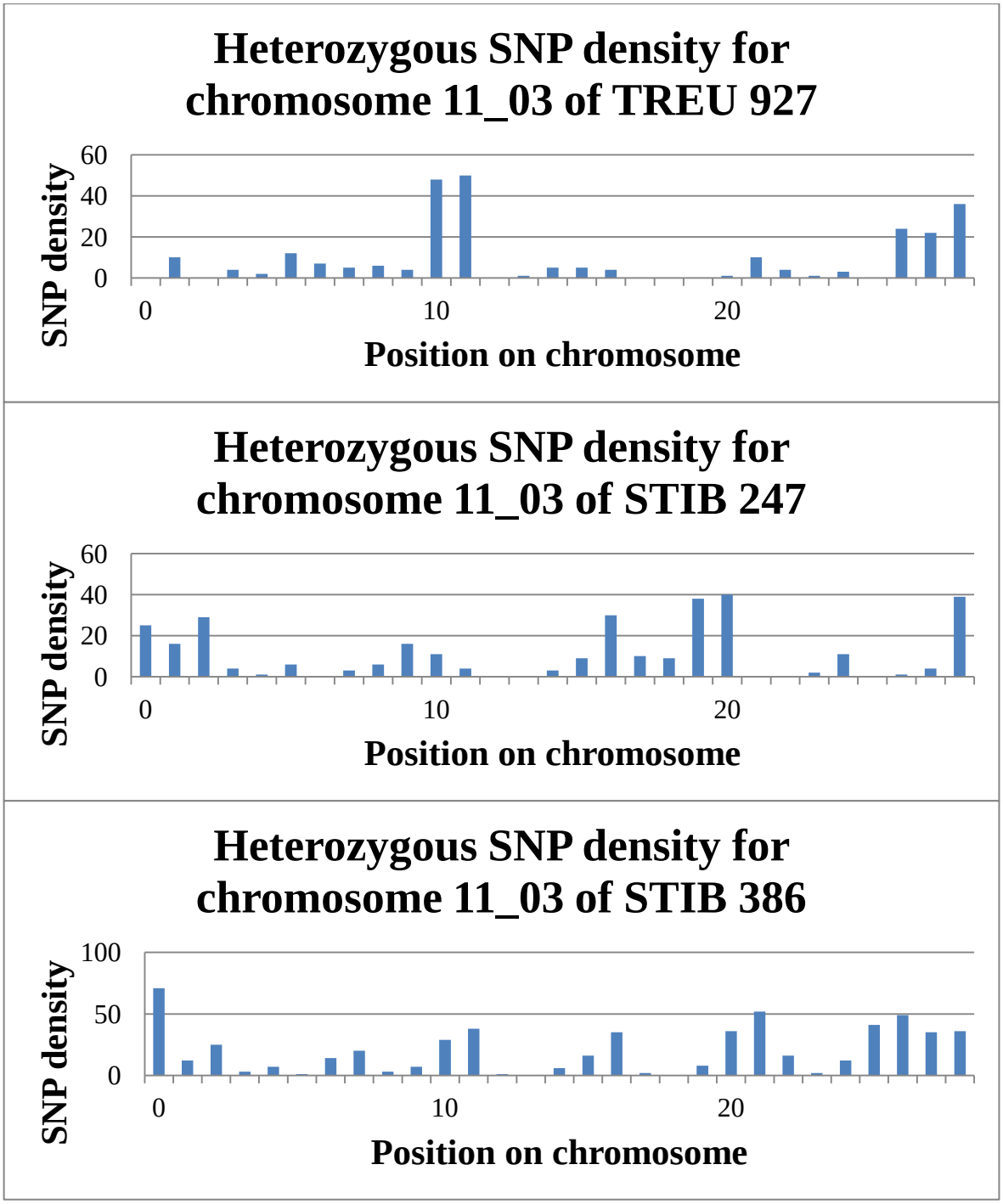
Appendix 12

Per chromosome heterozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of heterozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



Appendix 12

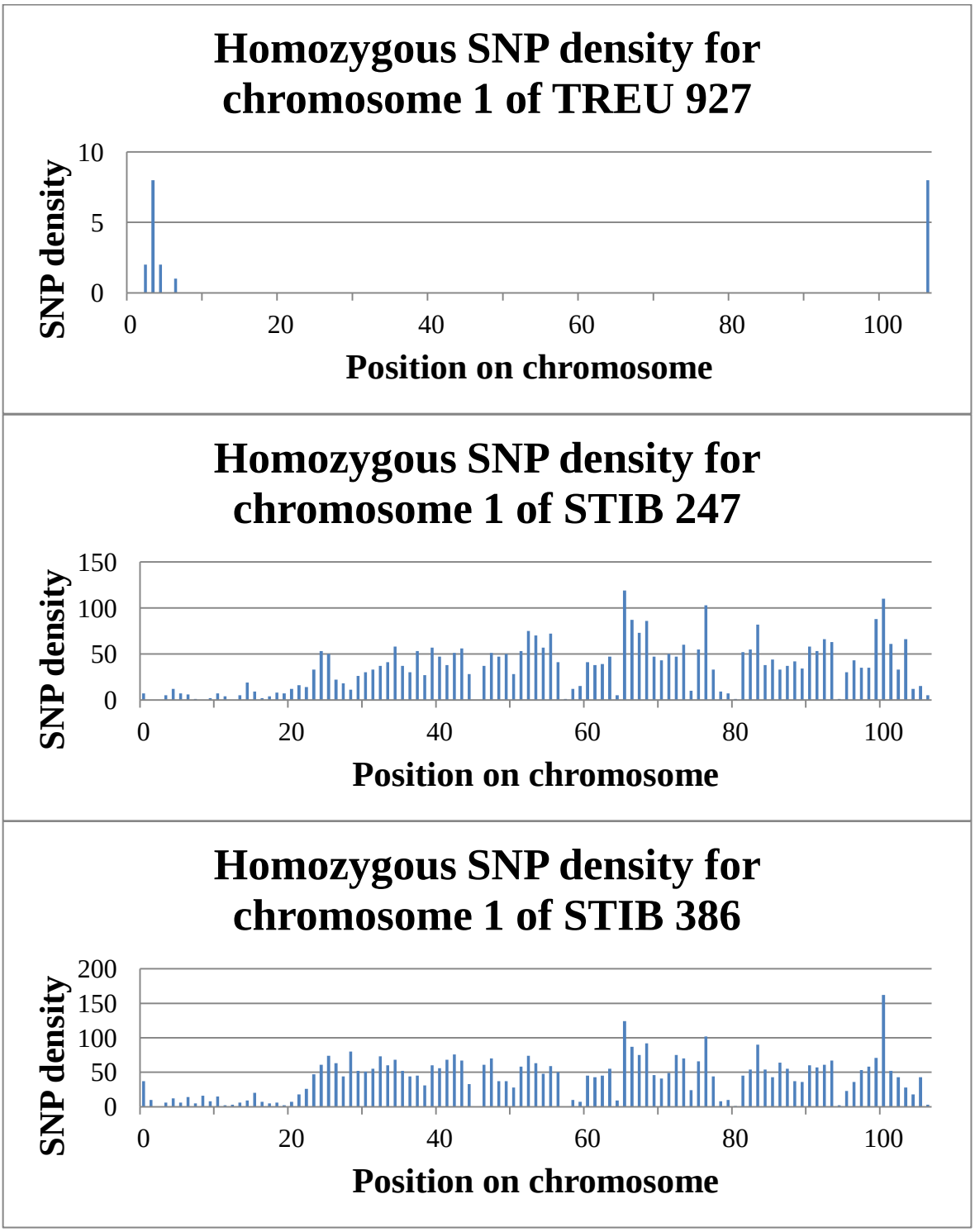
Per chromosome heterozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of heterozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



Appendix 12

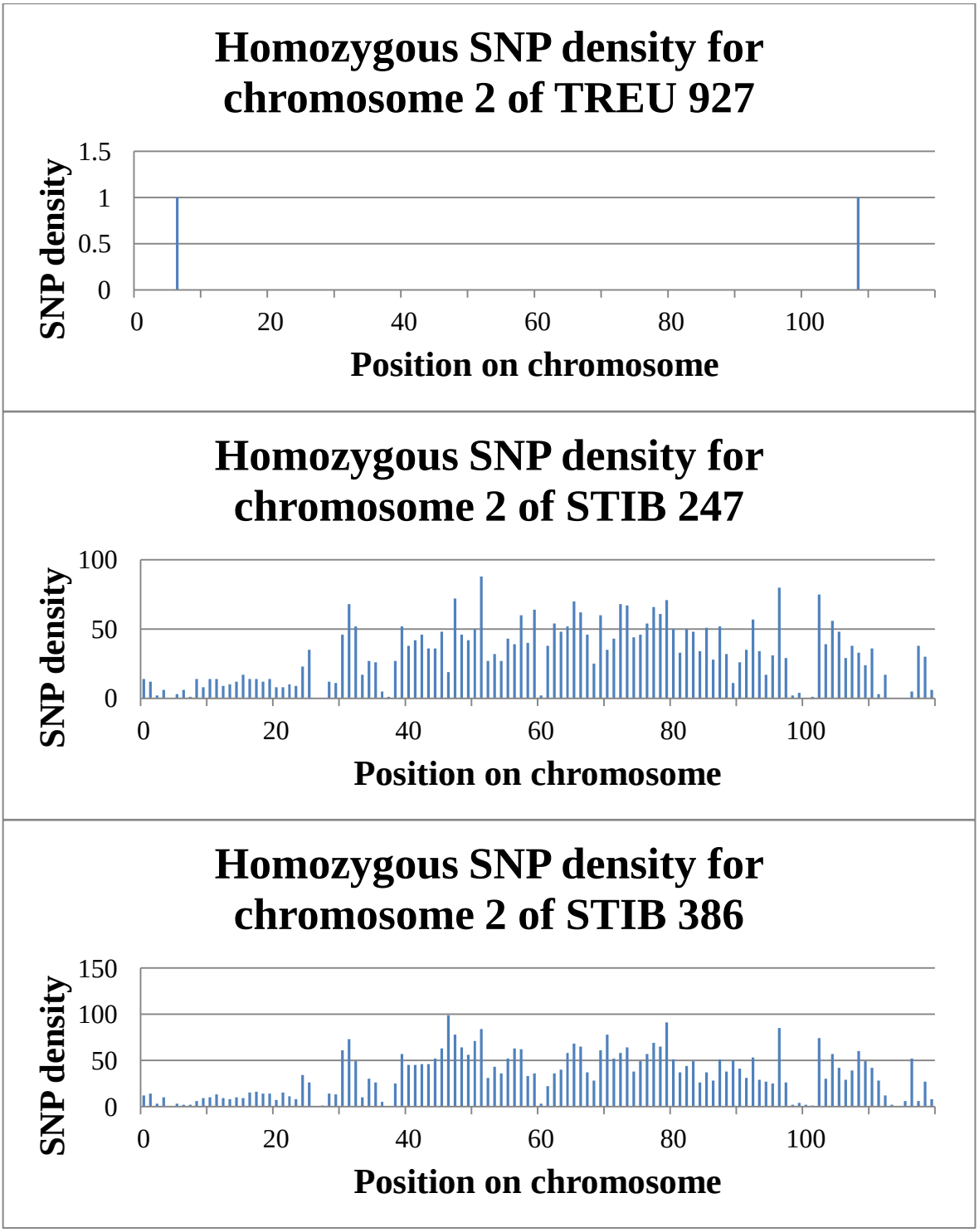
Per chromosome heterozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of heterozygous SNPs per 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.

Appendix 13



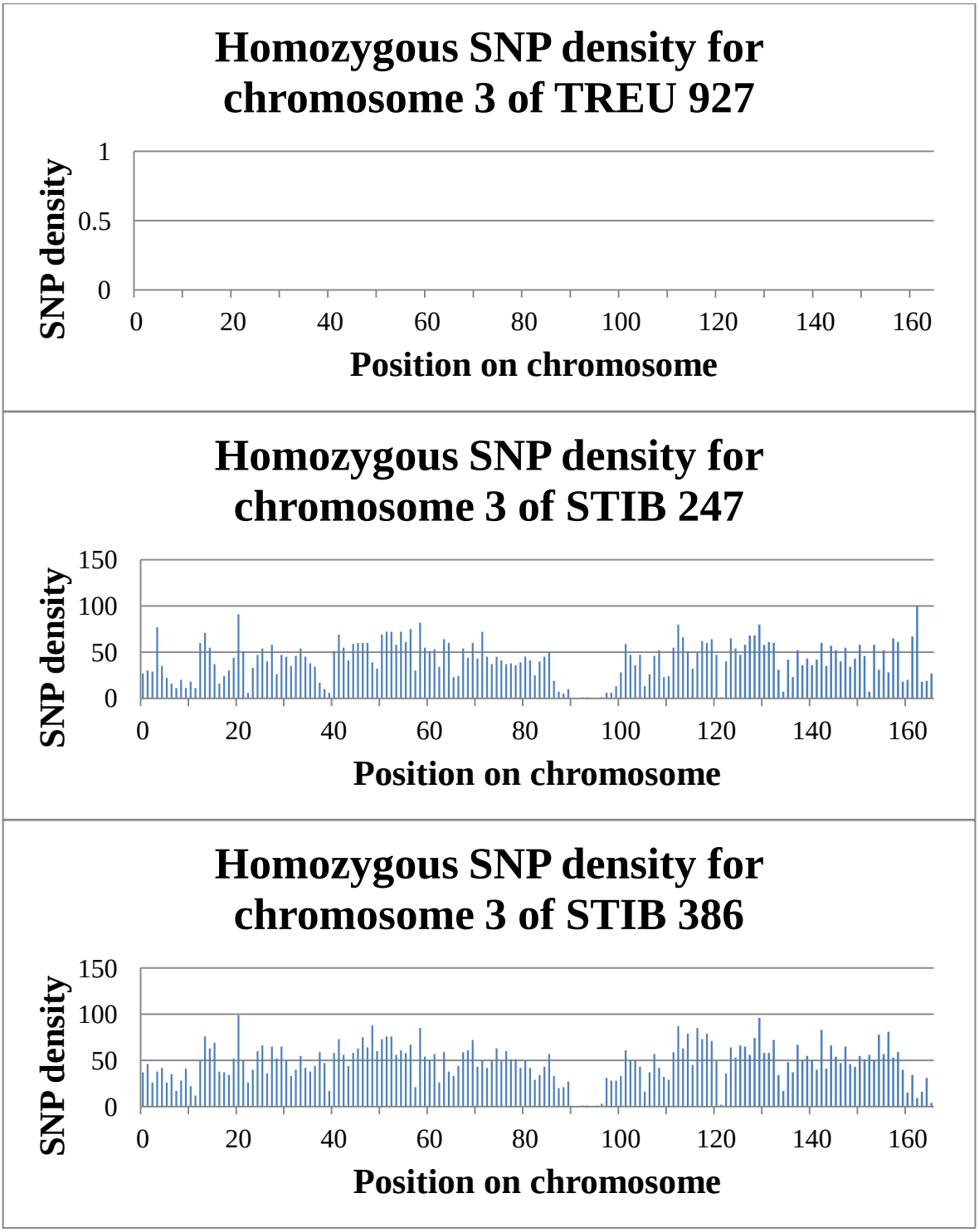
Appendix 13

Per chromosome homozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of homozygous SNPs compared to the TREU 927 reference sequence within a 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



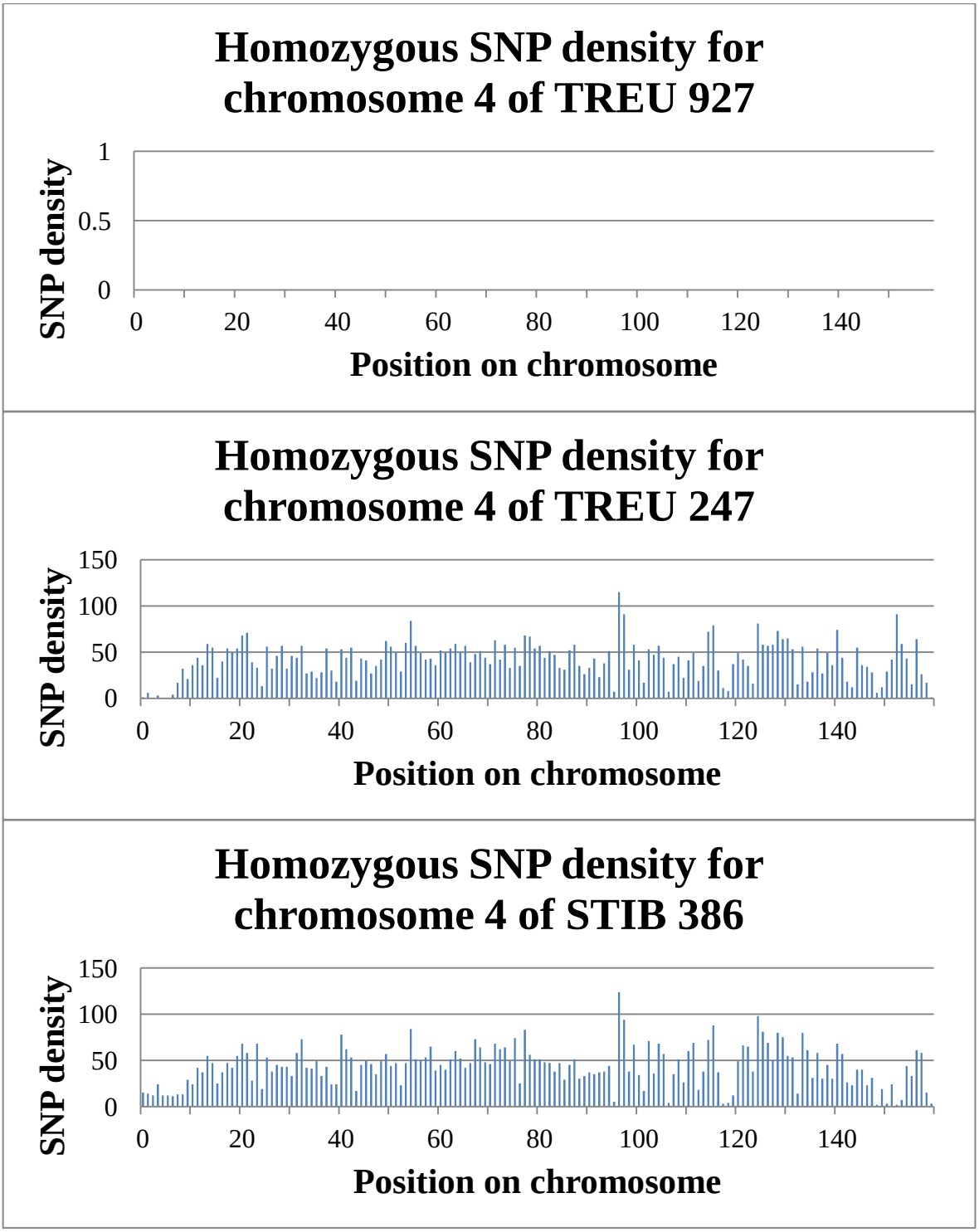
Appendix 13

Per chromosome homozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of homozygous SNPs compared to the TREU 927 reference sequence within a 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



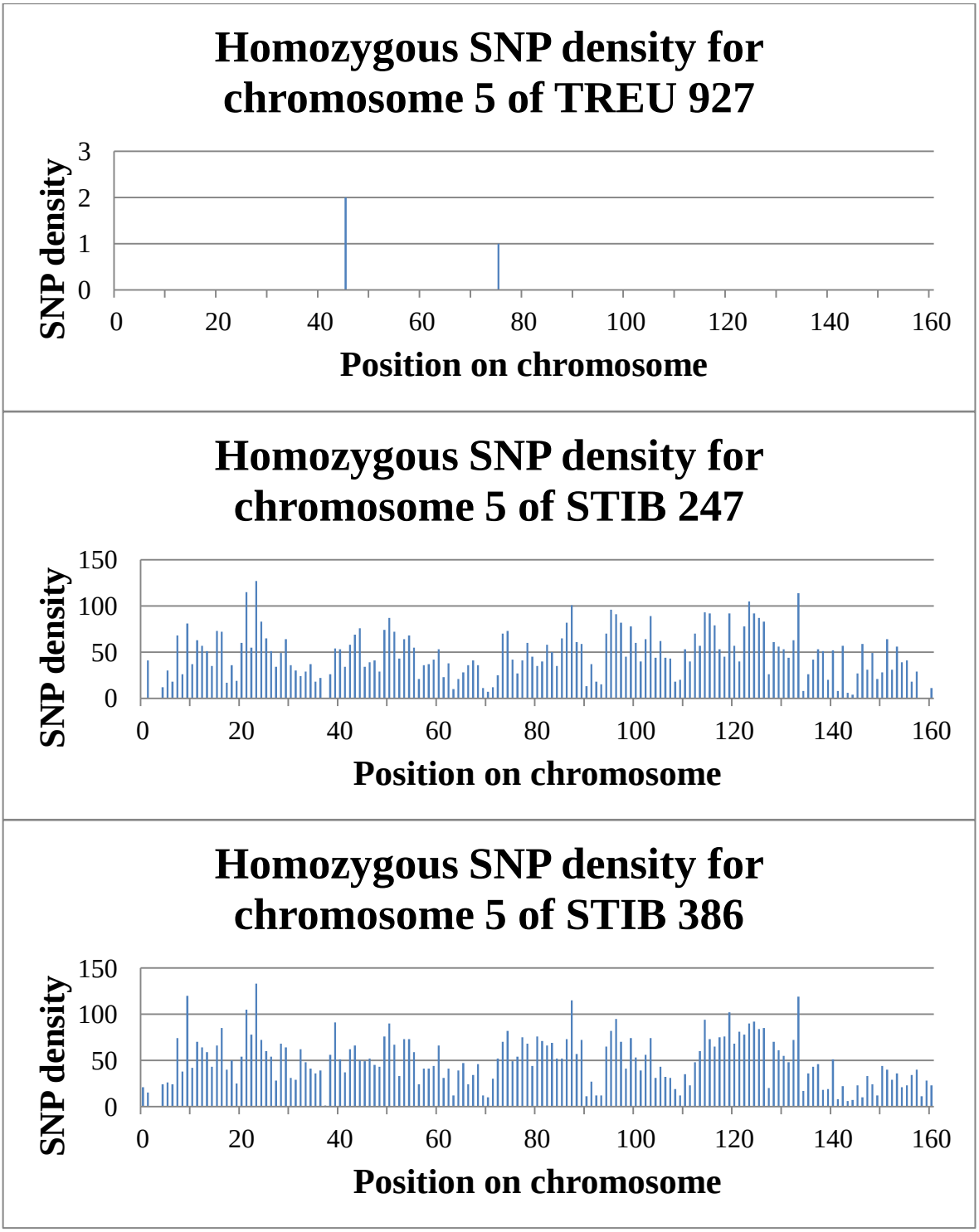
Appendix 13

Per chromosome homozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of homozygous SNPs compared to the TREU 927 reference sequence within a 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



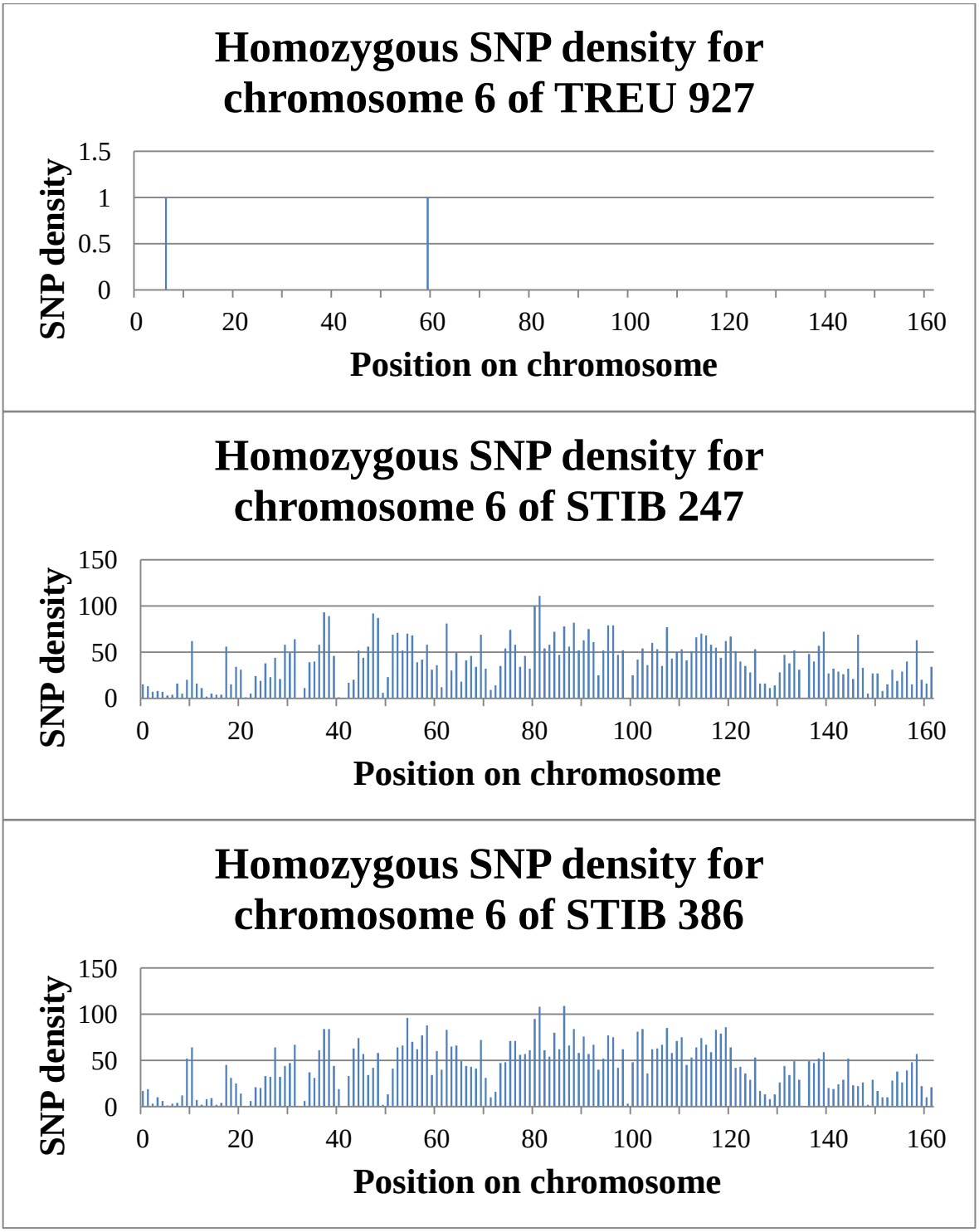
Appendix 13

Per chromosome homozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of homozygous SNPs compared to the TREU 927 reference sequence within a 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



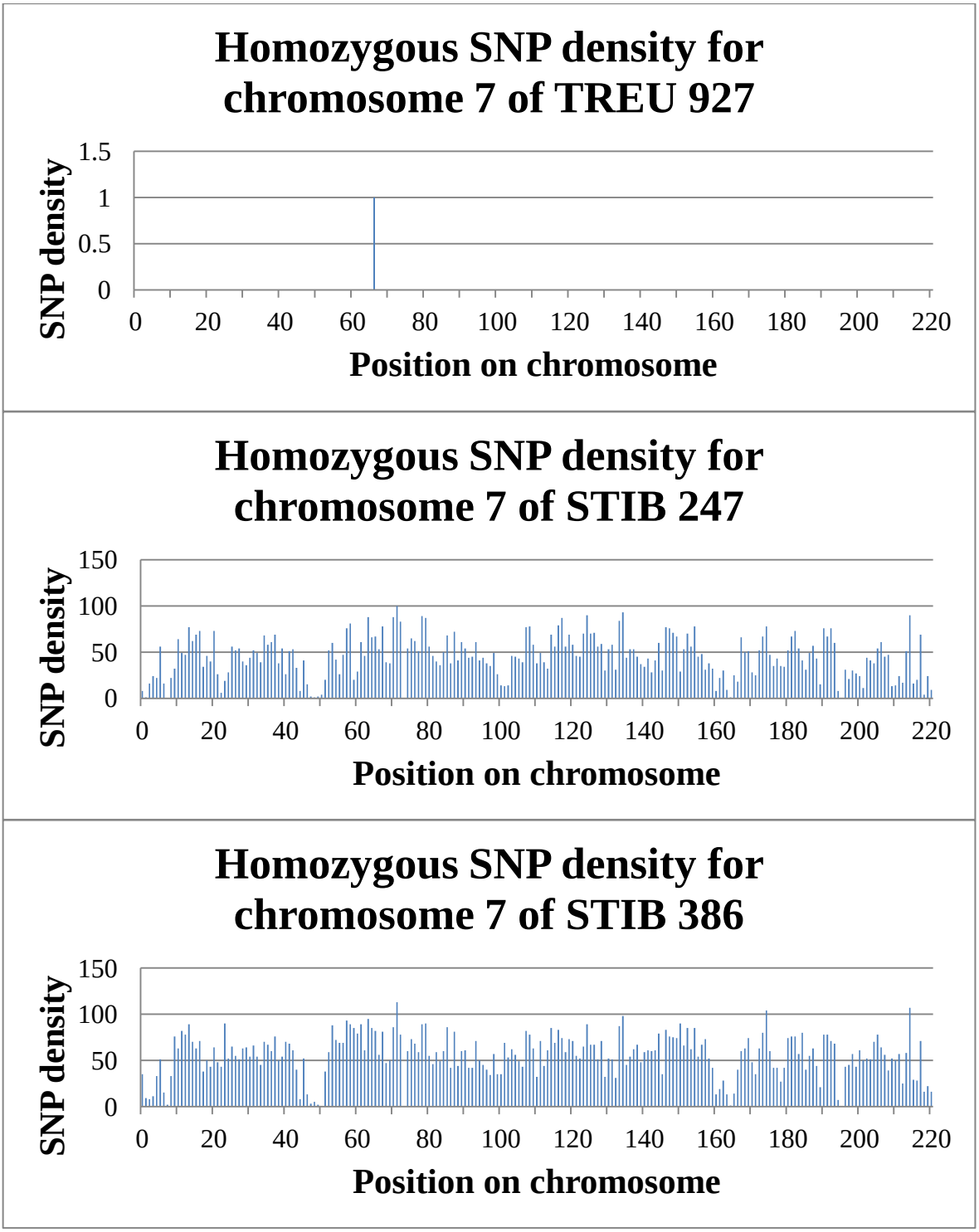
Appendix 13

Per chromosome homozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of homozygous SNPs compared to the TREU 927 reference sequence within a 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



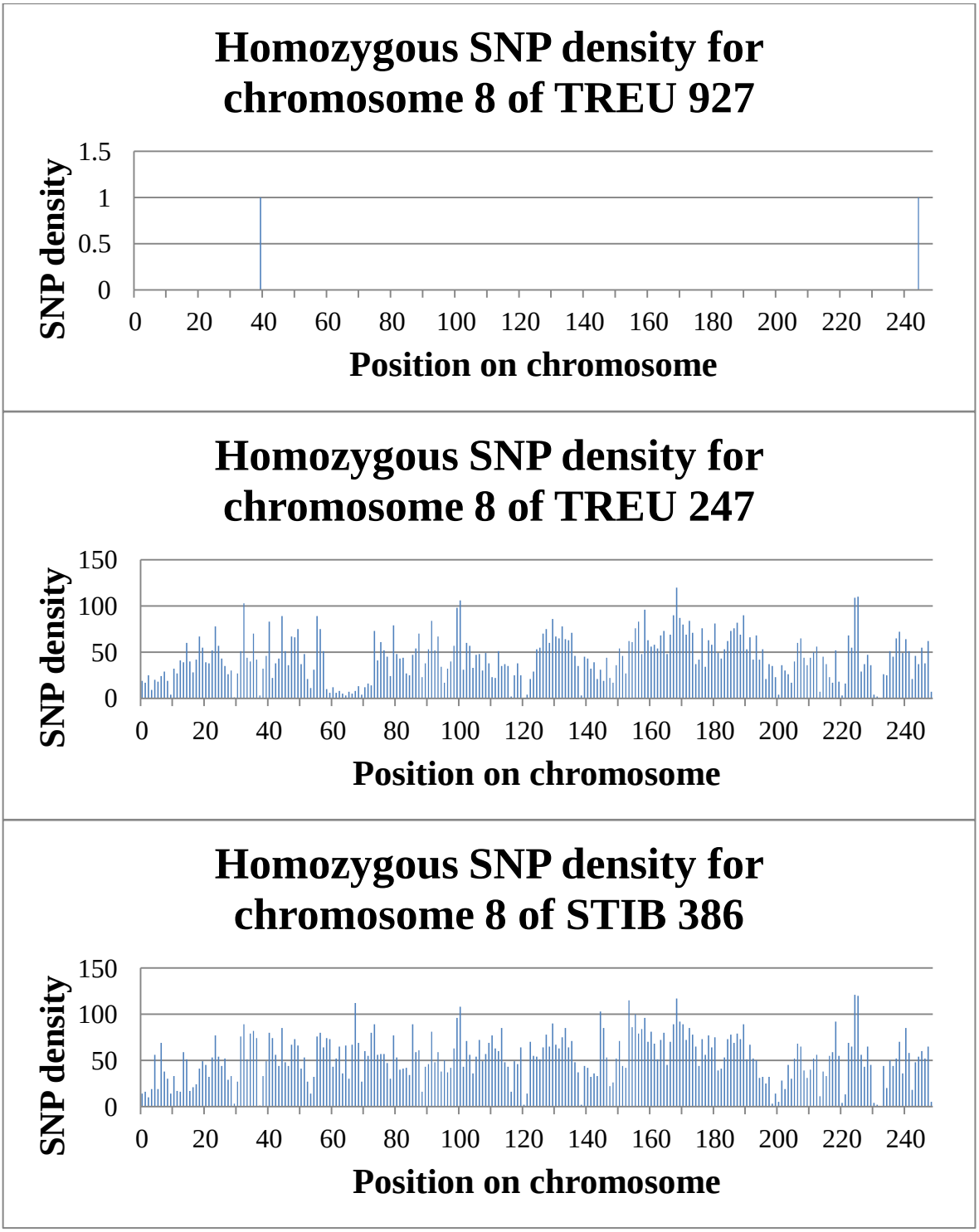
Appendix 13

Per chromosome homozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of homozygous SNPs compared to the TREU 927 reference sequence within a 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



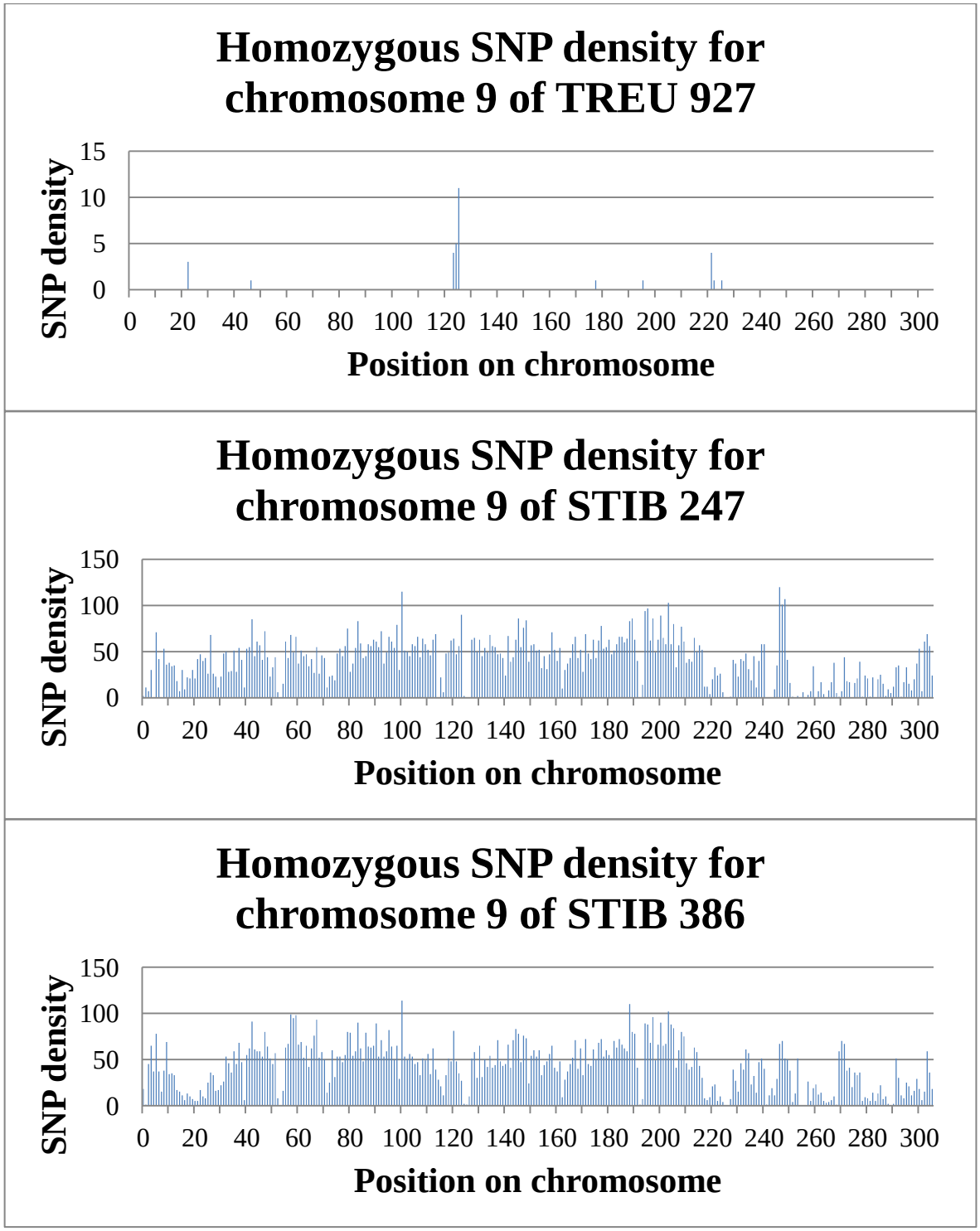
Appendix 13

Per chromosome homozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of homozygous SNPs compared to the TREU 927 reference sequence within a 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



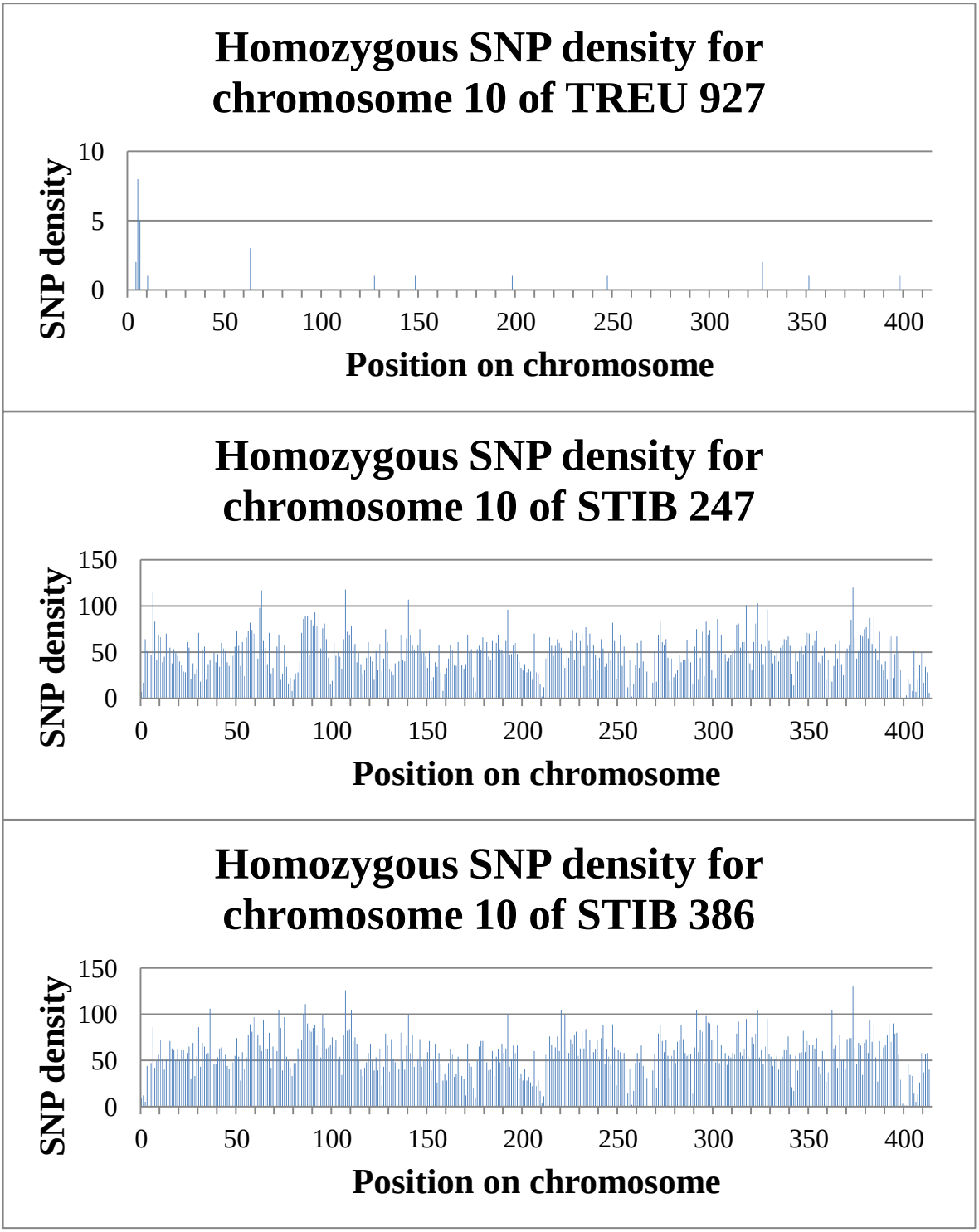
Appendix 13

Per chromosome homozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of homozygous SNPs compared to the TREU 927 reference sequence within a 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



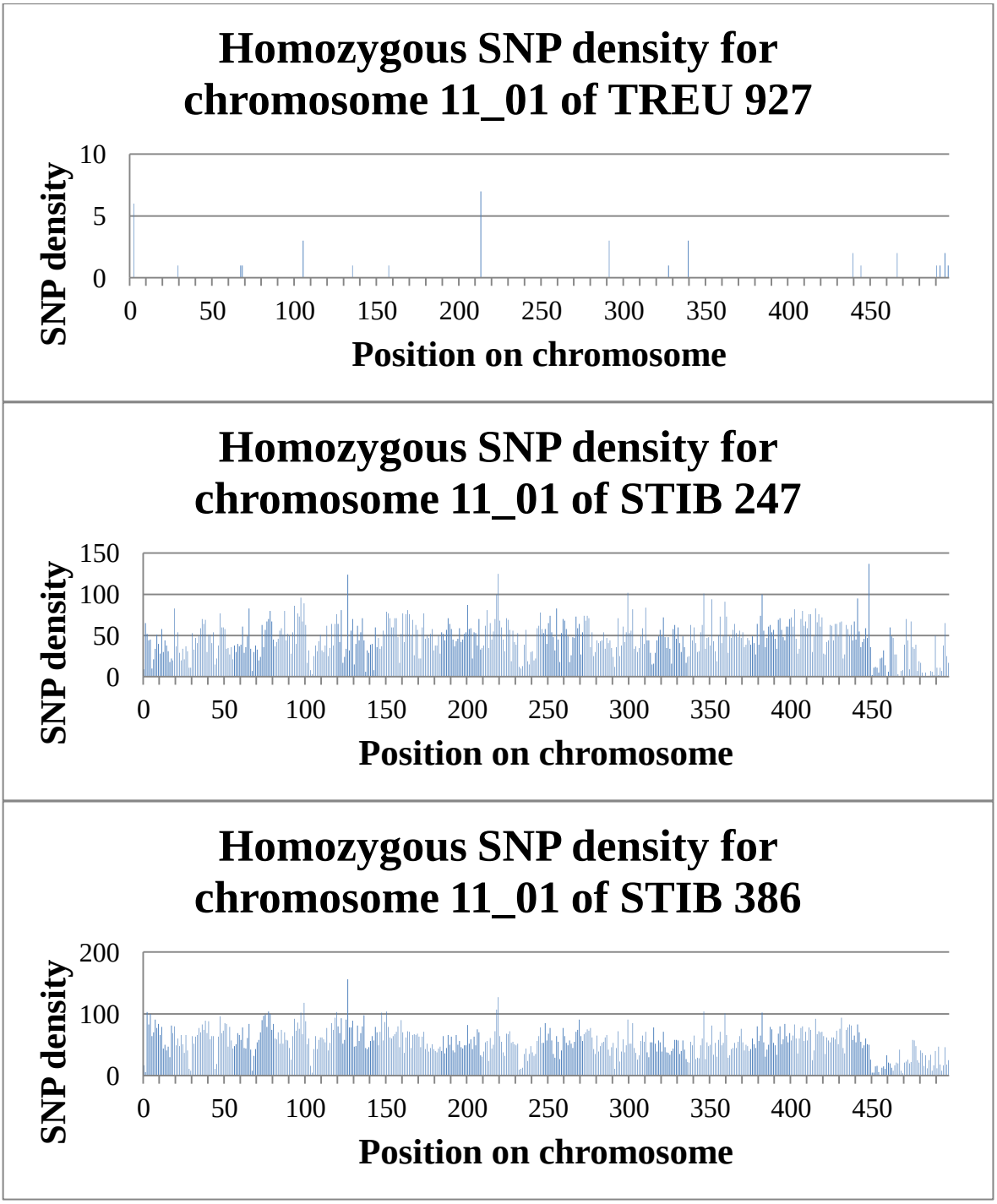
Appendix 13

Per chromosome homozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of homozygous SNPs compared to the TREU 927 reference sequence within a 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



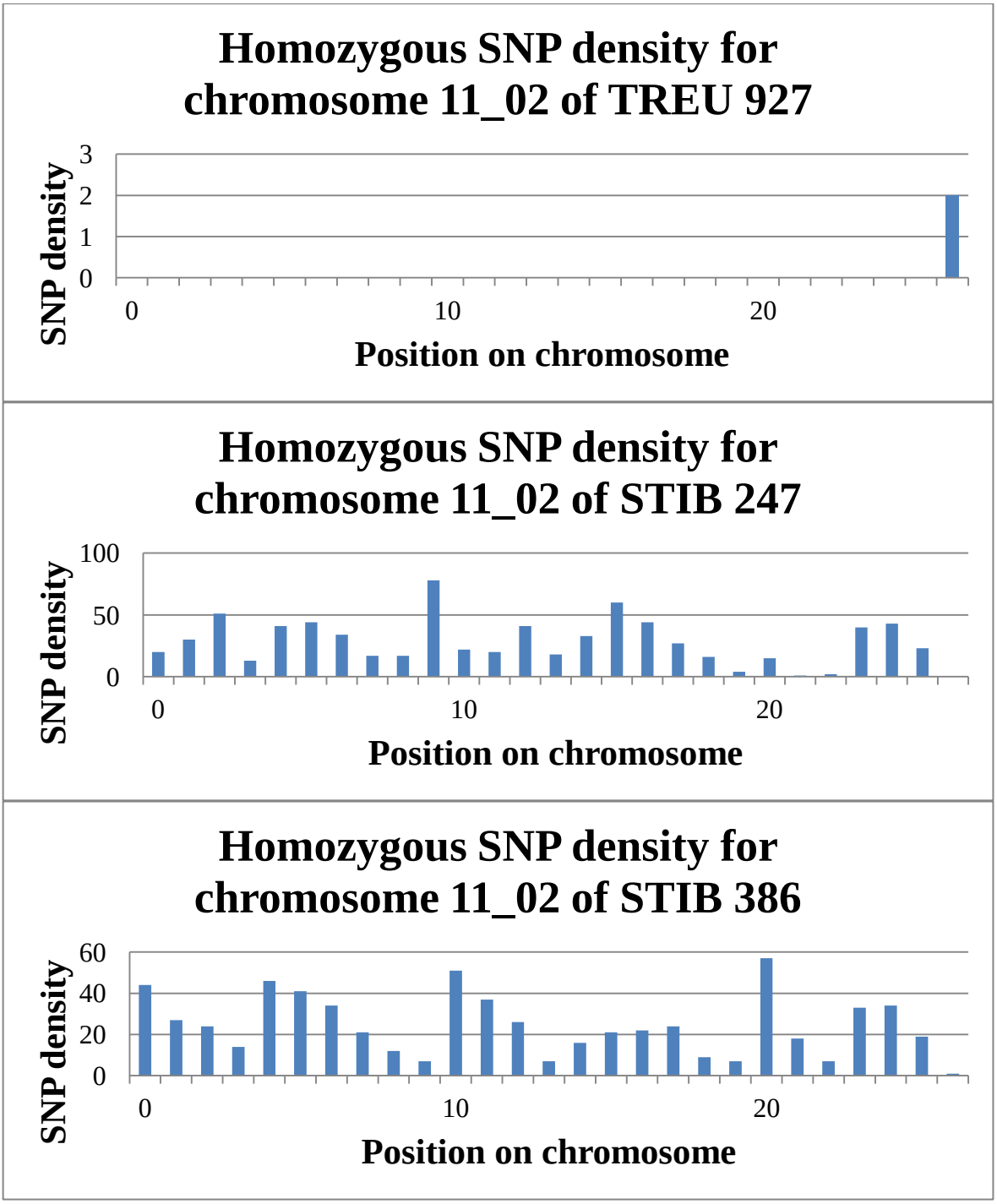
Appendix 13

Per chromosome homozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of homozygous SNPs compared to the TREU 927 reference sequence within a 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



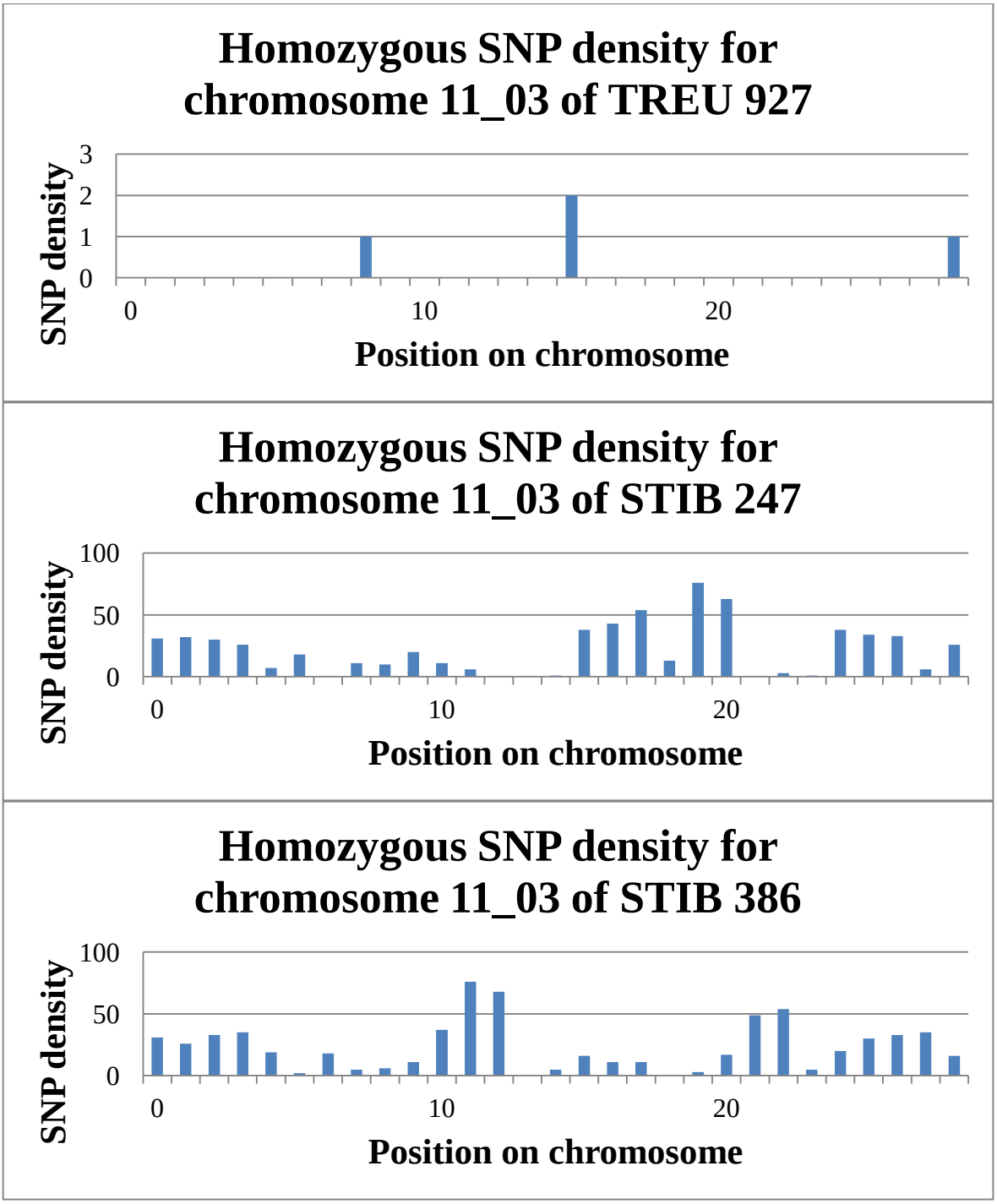
Appendix 13

Per chromosome homozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of homozygous SNPs compared to the TREU 927 reference sequence within a 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



Appendix 13

Per chromosome homozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of homozygous SNPs compared to the TREU 927 reference sequence within a 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.



Appendix 13

Per chromosome homozygous SNP densities for TREU 927, STIB 247 and STIB 386. SNP density was calculated as the number of homozygous SNPs compared to the TREU 927 reference sequence within a 10 kb region. Each division along the X axis covers 10 such regions and therefore represents a distance of 100 kb.

Bibliography

Adams, E.R., Malele, I.I., Msangi, A.R., Gibson, W.C., (2006). Trypanosome identification in wild tsetse populations in Tanzania using generic primers to amplify the ribosomal RNA ITS-1 region. Acta Trop 100, 103-109.

Adams, E.R., Hamilton, P.B., (2008). New molecular tools for the identification of trypanosome species. Future Microbiol 3, 167-176.

Adams, E.R., Hamilton, P.B., Malele, I.I., Gibson, W.C., (2008). The identification, diversity and prevalence of trypanosomes in field caught tsetse in Tanzania using ITS-1 primers and fluorescent fragment length barcoding. Infect. Genet. Evol 8, 439-444.

Adams, E.R., Hamilton, P.B., Rodrigues, A.C., Malele, I.I., Delespaux, V., Teixeira, M.M.G., Gibson, W., (2009). New *Trypanosoma (Duttonella) vivax* genotypes from tsetse flies in East Africa. Parasitology 1-10.

Agbo, E.C., Majiwa, P.A., Claassen, E.J., Roos, M.H., (2001). Measure of molecular diversity within the *Trypanosoma brucei* subspecies *Trypanosoma brucei brucei* and *Trypanosoma brucei gambiense* as revealed by genotypic characterization. Exp. Parasitol 99, 123-131.

Agbo, E.C., Clausen, P.-H., Büscher, P., Majiwa, P.A.O., Claassen, E., te Pas, M.F.W., (2003). Population genetic structure and cladistic analysis of *Trypanosoma brucei* isolates. Infect. Genet. Evol 3, 165-174.

Akopyants, N.S., Kimblin, N., Secundino, N., Patrick, R., Peters, N., Lawyer, P., Dobson, D.E., Beverley, S.M., Sacks, D.L., (2009).Demonstration of genetic exchange during cyclical development of *Leishmania* in the sand fly vector. Science 324, 265-268.

Alam, M.Z., Kuhls, K., Schweynoch, C., Sundar, S., Rijal, S., Shamsuzzaman, A.K.M., Raju, B.V.S., Salotra, P., Dujardin, J.-C., Schönian, G., (2009). Multilocus microsatellite typing (MLMT) reveals genetic homogeneity of *Leishmania donovani* strains in the Indian subcontinent. Infect. Genet. Evol 9, 24-31.

Al-Jawabreh, A., Diezmann, S., Müller, M., Wirth, T., Schnur, L.F., Strelkova, M.V., Kovalenko, D.A., Razakov, S.A., Schwenkenbecher, J., Kuhls, K., Schönian, G., (2008).Identification of geographically distributed sub-populations of *Leishmania (Leishmania) major* by microsatellite analysis.BMC Evol.Biol 8, 183.

Allsopp, B.A., Newton, S.D., (1985).Characterization of *Trypanosoma (Duttonella) vivax* by isoenzyme analysis. Int. J. Parasitol 15, 265-270.

Allsopp, R., (2001). Options for vector control against trypanosomiasis in Africa. Trends Parasitol 17, 15-19.

de Almeida, P.J., Ndao, M., Van Meirvenne, N., Geerts, S., (1997). Diagnostic evaluation of PCR in goats experimentally infected with *Trypanosoma vivax*. Acta Trop 66, 45-50.

Anderson, T.J., Haubold, B., Williams, J.T., Estrada-Franco, J.G., Richardson, L., Mollinedo, R., Bockarie, M., Mokili, J., Mharakurwa, S., French, N., Whitworth, J., Velez, I.D., Brockman, A.H., Nosten, F., Ferreira, M.U., Day, K.P., (2000a). Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Mol. Biol. Evol 17, 1467-1482.

Anderson, T.J., Paul, R.E., Donnelly, C.A., Day, K.P., (2000b). Do malaria parasites mate non-randomly in the mosquito midgut? Genet. Res 75, 285-296.

Anderson, N.E., Mubanga, J., Fevre, E.M., Picozzi, K., Eisler, M.C., Thomas, R., Welburn, S.C., (2011). Characterisation of the wildlife reservoir community for human and animal trypanosomiasis in the luangwa valley, zambia. PLoS Negl Trop Dis 5, e1211.

Annan, Z., Durand, P., Ayala, F.J., Arnathau, C., Awono-Ambene, P., Simard, F., Razakandrainibe, F.G., Koella, J.C., Fontenille, D., Renaud, F., (2007).Population genetic structure of *Plasmodium falciparum* in the two main African vectors, Anopheles gambiae and Anopheles funestus.Proc. Natl. Acad. Sci. U.S.A 104, 7987-7992.

Ansorge, W.J., (2009). Next-generation DNA sequencing techniques. N Biotechnol 25, 195-203.

Anthony, T.G., Conway, D.J., Cox-Singh, J., Matusop, A., Ratnam, S., Shamsul, S., Singh, B., (2005).Fragmented population structure of *Plasmodium falciparum* in a region of declining endemicity. J. Infect. Dis 191, 1558-1564.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M., Depledge, D.P., Fischer, S., Gajria, B., Gao, X., Gardner, M.J., Gingle, A., Grant, G., Harb, O.S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., Kissinger, J.C., Kraemer, E., Li, W., Logan, F.J., Miller, J.A., Mitra, S., Myler, P.J., Nayak, V., Pennington, C., Phan, I., Pinney, D.F., Ramasamy, G., Rogers, M.B., Roos, D.S., Ross, C., Sivam, D., Smith, D.F., Srinivasamoorthy, G., Stoeckert, C.J., Jr, Subramanian, S., Thibodeau, R., Tivey, A., Treatman, C., Velarde, G., Wang, H., (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res 38, D457-462.

Ayala, F.J., (1993). *Trypanosoma* and *Leishmania* have clonal population structures of epidemiological significance. Biol. Res 26, 47-63.

Ayala, F.J., (1998. Is sex better? Parasites say “no”. Proc. Natl. Acad. Sci. U.S.A 95, 3346-3348.

Babiker, H.A., Walliker, D., (1997). Current views on the population structure of *Plasmodium falciparum*: Implications for control. Parasitol.Today (Regul. Ed.) 13, 262-267.

Baker, N., Alsford, S., Horn, D., (2011). Genome-wide RNAi screens in African trypanosomes identify the nifurtimox activator NTR and the eflornithine transporter AAT6. Mol. Biochem. Parasitol 176, 55-57.

Balber, A.E., (1972). *Trypanosoma brucei*: fluxes of the morphological variants in intact and X-irradiated mice. Exp. Parasitol 31, 307-319.

Balloux, F., Lehmann, L., de Meeûs, T., (2003).The population genetics of clonal and partially clonal diploids. Genetics 164, 1635-1644.

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

Balmer, O., Caccone, A., (2008).Multiple-strain infections of *Trypanosoma brucei* across Africa. Acta Trop 107, 275-279.

Balmer, O., Beadell, J.S., Gibson, W., Caccone, A., (2011).Phylogeography and Taxonomy of *Trypanosoma brucei*. PLoS Negl Trop Dis 5, e961.

Bañuls, A.L., Guerrini, F., Le Pont, F., Barrera, C., Espinel, I., Guderian, R., Echeverria, R., Tibayrenc, M., (1997).Evidence for hybridization by multilocus enzyme electrophoresis and random amplified polymorphic DNA between *Leishmania braziliensis* and *Leishmania panamensis/guyanensis* in Ecuador. J. Eukaryot. Microbiol 44, 408-411.

Bañuls, A.L., Hide, M., Tibayrenc, M., (1999a). Molecular epidemiology and evolutionary genetics of *Leischmania* parasites. Int. J. Parasitol 29, 1137-1147.

Bañuls, A.L., Jonquieres, R., Guerrini, F., Le Pont, F., Barrera, C., Espinel, I., Guderian, R., Echeverria, R., Tibayrenc, M., (1999b). Genetic analysis of leishmania parasites in Ecuador: are *Leishmania (Viannia) panamensis* and *Leishmania (V.) Guyanensis* distinct taxa? Am. J. Trop. Med. Hyg 61, 838-845.

Barrett, M.P., Gilbert, I.H., (2006). Targeting of toxic compounds to the trypanosome's interior. Adv. Parasitol 63, 125-183.

Barrett, M.P., Boykin, D.W., Brun, R., Tidwell, R.R., (2007). Human African trypanosomiasis: pharmacological re-engagement with a neglected disease. Br. J. Pharmacol 152, 1155-1171.

Barry, J.D., McCulloch, R., (2001). Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. Adv. Parasitol 49, 1-70.

Barton, N.H., Charlesworth, B., (1998). Why sex and recombination? Science 281, 1986-1990.

Bastien, P., Blaineau, C., Pages, M., (1992). *Leishmania*: sex, lies and karyotype. Parasitol.Today (Regul. Ed.) 8, 174-177.

Batchelor, N.A., Atkinson, P.M., Gething, P.W., Picozzi, K., Fèvre, E.M., Kakembo, A.S.L., Welburn, S.C., (2009). Spatial predictions of Rhodesian Human African Trypanosomiasis (sleeping sickness) prevalence in Kaberamaido and Dokolo, two newly affected districts of Uganda. PLoS Negl Trop Dis 3, e563.

Batista, J.S., Riet-Correa, F., Teixeira, M.M.G., Madruga, C.R., Simões, S.D.V., Maia, T.F., (2007). Trypanosomiasis by *Trypanosoma vivax* in cattle in the Brazilian semiarid: Description of an outbreak and lesions in the nervous system. Vet. Parasitol 143, 174-181.

Batista, J.S., Oliveira, A.F., Rodrigues, C.M.F., Damasceno, C.A.R., Oliveira, I.R.S., Alves, H.M., Paiva, E.S., Brito, P.D., Medeiros, J.M.F., Rodrigues, A.C., Teixeira, M.M.G., (2009). Infection by *Trypanosoma vivax* in goats and sheep in the Brazilian semiarid region: from acute disease outbreak to chronic cryptic infection. Vet. Parasitol 165, 131-135.

Batista, J.S., Rodrigues, C.M., García, H.A., Bezerra, F.S., Olinda, R.G., Teixeira, M.M., Soto-Blanco, B., (2011). Association of *Trypanosoma vivax* in extracellular sites with central nervous system lesions and changes in cerebrospinal fluid in experimentally infected goats. Vet. Res 42, 63.

Baton, L.A., Ranford-Cartwright, L.C., (2005). Spreading the seeds of million-murdering death: metamorphoses of malaria in the mosquito. Trends Parasitol 21, 573-580.

Belli, A.A., Miles, M.A., Kelly, J.M., (1994). A putative *Leishmania panamensis/Leishmania braziliensis* hybrid is a causative agent of human cutaneous leishmaniasis in Nicaragua. Parasitology 109 (Pt 4), 435-442.

Bengaly, Z., Sidibe, I., Boly, H., Sawadogo, L., Desquesnes, M., (2002).Comparative pathogenicity of three genetically distinct *Trypanosoma congolense*-types in inbred Balb/c mice. Vet. Parasitol 105, 111-118.

Berberof, M., Pérez-Morga, D., Pays, E., (2001).A receptor-like flagellar pocket glycoprotein specific to *Trypanosoma brucei gambiense*. Mol. Biochem. Parasitol 113, 127-138.

Berrang-Ford, L., Odiit, M., Maiso, F., Waltner-Toews, D., McDermott, J., (2006). Sleeping sickness in Uganda: revisiting current and historical distributions. Afr Health Sci 6, 223-231.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., et al., (2005).The genome of the African trypanosome *Trypanosoma brucei*. Science 309, 416-422.

Bingle, L.E., Eastlake, J.L., Bailey, M., Gibson, W.C., (2001).A novel GFP approach for the analysis of genetic exchange in trypanosomes allowing the in situ detection of mating events. Microbiology (Reading, Engl.) 147, 3231-3240.
Birdsell, J.A., Wills, C., (2003). The evolutionary origin and maintenance of sexual recombination: a review of contemporary models, in: Evolutionary Biology. Springer, pp. 27-138.
Biteau, Bringaud, Gibson, Truc, Baltz, (2000). Characterization of Trypanozoon isolates using a repeated coding sequence and microsatellite markers. Mol. Biochem. Parasitol 105, 187-202.
Blaineau, C., Bastien, P., Pagès, M., (1992). Multiple forms of chromosome I, II and V in a restricted population of *Leishmania infantum* contrasting with monomorphism in individual strains suggest haploidy or automixy. Mol. Biochem. Parasitol 50, 197-204.
Van Bogaert, I., Haemers, A., (1989). Eflornithine.A new drug in the treatment of sleeping sickness. Pharm Weekbl Sci 11, 69-75.
Bogliolo, A.R., Lauria-Pires, L., Gibson, W.C., (1996). Polymorphisms in *Trypanosoma cruzi*: evidence of genetic recombination. Acta Trop 61, 31-40.
Borst, P., Fase-Fowler, F., Gibson, W.C., (1981).Quantitation of genetic differences between *Trypanosoma brucei gambiense, rhodesiense* and *brucei* by restriction enzyme analysis of kinetoplast DNA. Mol. Biochem. Parasitol 3, 117-131.
Borst, P., van der Ploeg, M., van Hoek, J.F., Tas, J., James, J., (1982). On the DNA content and ploidy of trypanosomes. Mol. Biochem. Parasitol 6, 13-23.
Boschetti, C., Pouchkina-Stantcheva, N., Hoffmann, P., Tunnacliffe, A., (2011). Foreign genes and novel hydrophilic protein genes participate in the desiccation response of the bdelloid rotifer *Adineta ricciae*. J. Exp. Biol 214, 59-68.
Bossard, G., Boulange, A., Holzmuller, P., Thévenon, S., Patrel, D., Authie, E., (2010). Serodiagnosis of bovine trypanosomosis based on HSP70/BiP inhibition ELISA. Vet. Parasitol 173, 39-47.
Van den Bossche, P., Chitanga, S., Masumu, J., Marcotty, T., Delespaux, V., (2011). Virulence in *Trypanosoma congolense* Savannah subgroup.A comparison between strains and transmission cycles. Parasite Immunol 33, 456-460.
Boussau, B., Daubin, V., (2010). Genomes as documents of evolutionary history.Trends Ecol. Evol.(Amst.) 25, 224-232.
Bouyer, J., Balenghien, T., Ravel, S., Vial, L., Sidibé, I., Thévenon, S., Solano, P., De Meeûs, T., (2009). Population sizes and dispersal pattern of tsetse flies: rolling on the river? Mol. Ecol 18, 2787-2797.
Brenière, S.F., Braquemond, P., Solari, A., Agnèse, J.F., Tibayrenc, M., (1991). An isoenzyme study of naturally occurring clones of *Trypanosoma cruzi* isolated from both sides of the West Andes highland. Trans. R. Soc. Trop. Med. Hyg 85, 62-66.
Brookfield, J.F., (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. Mol. Ecol 5, 453-455.
Brun, R., Hecker, H., Lun, Z.R., (1998). *Trypanosoma evansi* and *T. equiperdum*: distribution, biology, treatment and phylogenetic relationship (a review). Vet. Parasitol 79, 95-107.
Bucheton, B., Macleod, A., Jamonneau, V., (2011). Human host determinants influencing the outcome of *T. b. gambiense* infections. Parasite Immunol 33, 438-447.
Büscher, P., Mumba Ngoyi, D., Kaboré, J., Lejon, V., Robays, J., Jamonneau, V., Bebronne, N., Van der Veken, W., Biéler, S., (2009). Improved Models of Mini Anion Exchange Centrifugation Technique (mAECT) and Modified Single Centrifugation (MSC) for sleeping sickness diagnosis and staging. PLoS Negl Trop Dis 3, e471.
Burri, C., Keiser, J., (2001). Pharmacokinetic investigations in patients from northern Angola refractory to melarsoprol treatment. Trop. Med. Int. Health 6, 412-420.
Burri, C., (2010). Chemotherapy against human African trypanosomiasis: is there a road to success? Parasitology 137, 1987-1994.
Cabrera, L., De Witte, J., Victor, B., Vermeiren, L., Zimic, M., Brandt, J., Geysen, D., (2009). Specific detection and identification of African trypanosomes in bovine peripheral blood by means of a PCR-ELISA assay. Vet. Parasitol 164, 111-117.
Callejas, S., Leech, V., Reitter, C., Melville, S., (2006). Hemizygous subtelomeres of an African trypanosome chromosome may account for over 75% of chromosome length. Genome Res 16, 1109-1118.
Camara, M., Camara, O., Ilboudo, H., Sakande, H., Kaboré, J., N'Dri, L., Jamonneau, V., Bucheton, B., (2010). Sleeping sickness diagnosis: use of buffy coats improves the sensitivity of the mini anion exchange centrifugation test. Trop. Med. Int. Health 15, 796-799.
Capewell, P., (2011). Human Serum Resistance in *Trypanosoma brucei*.Thesis, (PhD).University of Glasgow.
Carrasco, H.J., Frame, I.A., Valente, S.A., Miles, M.A., (1996).Genetic exchange as a possible source of genomic diversity in sylvatic populations of *Trypanosoma cruzi*. Am. J. Trop. Med. Hyg 54, 418-424.

Catley, A., Irungu, P., Simiyu, K., Dadye, J., Mwakio, W., Kiragu, J., Nyamwaro, S.O., (2002). Participatory investigations of bovine trypanosomiasis in Tana River District, Kenya. Med. Vet. Entomol 16, 55-66.

Chakraborty, R., De Andrade, M., Daiger, S.P., Budowle, B., (1992). Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. Ann. Hum. Genet 56, 45-57.

Chappuis, F., Loutan, L., Simarro, P., Lejon, V., Büscher, P., (2005). Options for field diagnosis of human african trypanosomiasis. Clin.Microbiol. Rev 18, 133-146.

Chargui, N., Amro, A., Haouas, N., Schönian, G., Babba, H., Schmidt, S., Ravel, C., Lefebvre, M., Bastien, P., Chaker, E., Aoun, K., Zribi, M., Kuhls, K., (2009).Population structure of Tunisian *Leishmania infantum* and evidence for the existence of hybrids and gene flow between genetically different populations. Int. J. Parasitol 39, 801-811.

Checchi, F., Piola, P., Ayikoru, H., Thomas, F., Legros, D., Priotto, G., (2007). Nifurtimox plus Eflornithine for Late-Stage Sleeping Sickness in Uganda: A Case Series. PLoS Negl Trop Dis 1, e64.

Checchi, F., Filipe, J.A.N., Barrett, M.P., Chandramohan, D., (2008a). The natural progression of gambiense sleeping sickness: what is the evidence? PLoS Negl Trop Dis 2, e303.

Checchi, F., Filipe, J.A.N., Haydon, D.T., Chandramohan, D., Chappuis, F., (2008b).Estimates of the duration of the early and late stage of gambiense sleeping sickness. BMC Infect. Dis 8, 16.

Cherenet, T., Sani, R.A., Panandam, J.M., Nadzr, S., Speybroeck, N., van den Bossche, P., (2004). Seasonal prevalence of bovine trypanosomosis in a tsetse-infested zone and a tsetse-free zone of the Amhara Region, north-west Ethiopia. Onderstepoort J. Vet. Res 71, 307-312.

Cherenet, T., Sani, R.A., Speybroeck, N., Panandam, J.M., Nadzr, S., Van den Bossche, P., (2006).A comparative longitudinal study of bovine trypanosomiasis in tsetse-free and tsetse-infested zones of the Amhara Region, northwest Ethiopia. Vet. Parasitol 140, 251-258.

Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B., Christen, R., (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics 7, 439.

Chisi, J.E., Misiri, H., Zverev, Y., Nkhoma, A., Sternberg, J.M., (2004). Anaemia in human African trypanosomiasis caused by *Trypanosoma brucei rhodesiense*. East Afr Med J 81, 505-508.

Cibulskis, R.E., (1988). Origins and organization of genetic diversity in natural populations of *Trypanosoma brucei*. Parasitology 96 (Pt 2), 303-322.

Clayton, C., Shapira, M., (2007).Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol. Biochem. Parasitol 156, 93-101.

Conway, D.J., Roper, C., Oduola, A.M., Arnot, D.E., Kremsner, P.G., Grobusch, M.P., Curtis, C.F., Greenwood, B.M., (1999).High recombination rate in natural populations of *Plasmodium falciparum*.Proc. Natl. Acad. Sci. U.S.A 96, 4506-4511.

Conway, D.J., (2007). Molecular epidemiology of malaria.Clin.Microbiol. Rev 20, 188-204.

Cooper, A., Tait, A., Sweeney, L., Tweedie, A., Morrison, L., Turner, C.M.R., MacLeod, A., (2008). Genetic analysis of the human infective trypanosome *Trypanosoma brucei gambiense*: chromosomal segregation, crossing over, and the construction of a genetic map. Genome Biol 9, R103.

Cooper, A.C., (2009). Linkage mapping and genetic analysis of *Trypanosoma brucei*.Thesis, (PhD).University of Glasgow.

Cordon-Obras, C., Berzosa, P., Ndong-Mabale, N., Bobuakasi, L., Buatiche, J.N., Ndongo-Asumu, P., Benito, A., Cano, J., (2009).*Trypanosoma brucei gambiense* in domestic livestock of Kogo and Mbini foci (Equatorial Guinea). Trop. Med. Int. Health 14, 535-541.

Cortez, A.P., Ventura, R.M., Rodrigues, A.C., Batista, J.S., Paiva, F., Añez, N., Machado, R.Z., Gibson, W.C., Teixeira, M.M.G., (2006).The taxonomic and phylogenetic relationships of *Trypanosoma vivax* from South America and Africa. Parasitology 133, 159-169.

Courtin, D., Berthier, D., Thevenon, S., Dayo, G.-K., Garcia, A., Bucheton, B., (2008). Host genetics in African trypanosomiasis. Infect. Genet. Evol 8, 229-238.

Coustou, V., Guegan, F., Plazolles, N., Baltz, T., (2010). Complete in vitro life cycle of *Trypanosoma congolense*: development of genetic tools. PLoS Negl Trop Dis 4, e618.

Cox, F.E.G. (Ed.), (1993). Modern parasitology: A textbook of parasitology. Wiley-Blackwell, Oxford, UK.

Cox, A.P., Tosas, O., Tilley, A., Picozzi, K., Coleman, P., Hide, G., Welburn, S.C., (2010). Constraints to estimating the prevalence of trypanosome infections in East African zebu cattle.Parasit Vectors 3, 82.

Cruz, A.K., Titus, R., Beverley, S.M., (1993). Plasticity in chromosome number and testing of essential genes in *Leishmania* by targeting. Proc. Natl. Acad. Sci. U.S.A 90, 1599-1603.

Cuglovici, D.A., Bartholomeu, D.C., Reis-Cunha, J.L., Carvalho, A.U., Ribeiro, M.F.B., (2010). Epidemiologic aspects of an outbreak of *Trypanosomavivax* in a dairy cattle herd in Minas Gerais state, Brazil. Vet Parasitol 169, 320-326.

Cupolillo, E., Grimaldi, G., Jr, Momen, H., (1997). Genetic diversity among *Leishmania (Viannia)* parasites. Ann Trop Med Parasitol 91, 617-626.

Dayo, G.-K., Bengaly, Z., Messad, S., Bucheton, B., Sidibe, I., Cene, B., Cuny, G., Thevenon, S., (2010).Prevalence and incidence of bovine trypanosomosis in an agro-pastoral area of southwestern Burkina Faso. Res. Vet. Sci 88, 470-477.

Deborggraeve, S., Büscher, P., (2010). Molecular diagnostics for sleeping sickness: what is the benefit for the patient? Lancet Infect Dis 10, 433-439.

Deborggraeve, S., Lejon, V., Ekangu, R.A., Mumba Ngoyi, D., Pati Pyana, P., Ilunga, M., Mulunda, J.P., Büscher, P., (2011). Diagnostic accuracy of PCR in gambiense sleeping sickness diagnosis, staging and post-treatment follow-up: a 2-year longitudinal study. PLoS Negl Trop Dis 5, e972.

Degen, R., Pospichal, H., Enyaru, J., Jenni, L., (1995). Sexual compatibility among *Trypanosoma brucei* isolates from an epidemic area in southeastern Uganda. Parasitol. Res 81, 253-257.

Delafosse, A., Thébaud, E., Desquesnes, M., Michaux, Y., (2006).Epidemiology of *Trypanosoma vivax* infection in cattle in the tse-tse free area of Lake Chad. Prev. Vet. Med 74, 108-119.

Delgado, O., Cupolillo, E., Bonfante-Garrido, R., Silva, S., Belfort, E., Grimaldi Júnior, G., Momen, H., (1997). Cutaneous leishmaniasis in Venezuela caused by infection with a new hybrid between *Leishmania (Viannia) braziliensis* and *L. (V.) guyanensis*. Mem. Inst. Oswaldo Cruz 92, 581-582.

Dempster, A.P., Laird, N.M., Rubin, D.B., (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm.J Roy. Stat. Soc. B 39, 1-38.

Desjeux, P., Dedet, J.P., (1989). Isoenzyme characterization of 112 *Leishmania* isolates from French Guiana. Trans. R. Soc. Trop. Med. Hyg 83, 610-612.

Desquesnes, M., (1997).Evaluation of a simple PCR technique for the diagnosis of *Trypanosoma vivax* infection in the serum of cattle in comparison to parasitological techniques and antigen-enzyme-linked immuno sorbent assay. Acta Trop 65, 139-148.

Desquesnes, M., Dávila, A.M.R., (2002). Applications of PCR-based tools for detection and identification of animal trypanosomes: a review and perspectives. Vet. Parasitol 109, 213-231.

Desquesnes, M., Dia, M.L., (2003).Mechanical transmission of *Trypanosoma congolense* in cattle by the African tabanid Atylotus agrestis. Exp. Parasitol 105, 226-231.

Desquesnes, M., Dia, M.L., (2004).Mechanical transmission of *Trypanosoma vivax* in cattle by the African tabanid Atylotus fuscipes. Vet. Parasitol 119, 9-19.

Dhollander, S., Jallow, A., Mbodge, K., Kora, S., Sanneh, M., Gaye, M., Bos, J., Leak, S., Berkvens, D., Geerts, S., (2006). Equine trypanosomosis in the Central River Division of The Gambia: a study of veterinary gate-clinic consultation records. Prev. Vet. Med 75, 152-162.

Dickin, S.K., Gibson, W.C., (1989). Hybridisation with a repetitive DNA probe reveals the presence of small chromosomes in *Trypanosoma vivax*. Mol. Biochem. Parasitol 33, 135-142.

Dirie, M.F., Murphy, N.B., Gardiner, P.R., (1993a). DNA fingerprinting of *Trypanosoma vivax* isolates rapidly identifies intraspecific relationships. J. Eukaryot. Microbiol 40, 132-134.

Dirie, M.F., Otte, M.J., Thatthi, R., Gardiner, P.R., (1993b). Comparative studies of *Trypanosoma (Duttonella) vivax* isolates from Colombia. Parasitology 106 (Pt 1), 21-29.

Dixon, M.W.A., Thompson, J., Gardiner, D.L., Trenholme, K.R., (2008). Sex in *Plasmodium*: a sign of commitment. Trends Parasitol 24, 168-175.

DLS-ITC, (1993).National Livestock Census. Department of Livestock Services (Abuko) and The International Trypanotolerance Centre, The Gambia.

Docampo, R., Moreno, S.N.J., (2003). Current chemotherapy of human African trypanosomiasis.Parasitol.Res 90 Supp 1, S10-13.

Donelson, J.E., (2003). Antigenic variation and the African trypanosome genome. Acta Trop 85, 391-404.

DOP-DOSA, (2002).Gambian Agriculture Census 2001/2002. Department of Planning / Department of State for Agriculture, Banjul, The Gambia.

Duffy, C.W., Morrison, L.J., Black, A., Pinchbeck, G.L., Christley, R.M., Schoenefeld, A., Tait, A., Turner, C.M.R., MacLeod, A., (2009). *Trypanosoma vivax* displays a clonal population structure. Int. J. Parasitol 39, 1475-1483.

Dujardin, J.C., Bañuls, A.L., Llanos-Cuentas, A., Alvarez, E., DeDoncker, S., Jacquet, D., Le Ray, D., Arevalo, J., Tibayrenc, M., (1995).Putative *Leishmania* hybrids in the Eastern Andean valley of Huanuco, Peru. Acta Trop 59, 293-307.

Dukes, P., Gibson, W.C., Gashumba, J.K., Hudson, K.M., Bromidge, T.J., Kaukus, A., Asonganyi, T., Magnus, E., (1992). Absence of the LiTat 1.3 (CATT antigen) gene in *Trypanosoma brucei* gambiense stocks from Cameroon. Acta Trop 51, 123-134.

Durand, P., Michalakis, Y., Cestier, S., Oury, B., Leclerc, M.C., Tibayrenc, M., Renaud, F., (2003).Significant linkage disequilibrium and high genetic diversity in a population of *Plasmodium falciparum* from an area (Republic of the Congo) highly endemic for malaria. Am. J. Trop. Med. Hyg 68, 345-349.

Eisler, M.C., Lessard, P., Masake, R.A., Moloo, S.K., Peregrine, A.S., (1998). Sensitivity and specificity of antigen-capture ELISAs for diagnosis of *Trypanosoma congolense* and *Trypanosoma vivax* infections in cattle. Vet. Parasitol 79, 187-201.

El-Sayed, N.M., Hegde, P., Quackenbush, J., Melville, S.E., Donelson, J.E., (2000). The African trypanosome genome. Int. J. Parasitol 30, 329-345.

El-Sayed, N.M.A., Ghedin, E., Song, J., MacLeod, A., Bringaud, F., Larkin, C., Wanless, D., Peterson, J., Hou, L., Taylor, S., Tweedie, A., Biteau, N., Khalak, H.G., Lin, X., Mason, T., Hannick, L., Caler, E., Blandin, G., Bartholomeu, D., Simpson, A.J., Kaul, S., Zhao, H., Pai, G., Van Aken, S., Utterback, T., Haas, B., Koo, H.L., Umayam, L., Suh, B., Gerrard, C., Leech, V., Qi, R., Zhou, S., Schwartz, D., Feldblyum, T., Salzberg, S., Tait, A., Turner, C.M.R., Ullu, E., White, O., Melville, S., Adams, M.D., Fraser, C.M., Donelson, J.E., (2003). The sequence and analysis of *Trypanosoma brucei* chromosome II. Nucleic Acids Res 31, 4856-4863.

El-Sayed, N.M., Myler, P.J., Bartholomeu, D.C., Nilsson, D., Aggarwal, G., et al., (2005). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 309, 409-415.

Engstler, M., Pfohl, T., Herminghaus, S., Boshart, M., Wiegertjes, G., Heddergott, N., Overath, P., (2007). Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. Cell 131, 505-515.

Enwezor, F.N.C., Umoh, J.U., Esievo, K.A.N., Halid, I., Zaria, L.T., Anere, J.I., (2009). Survey of bovine trypanosomosis in the Kachia Grazing Reserve, Kaduna State, Nigeria. Vet. Parasitol 159, 121-125.

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

Enyaru, J.C., Matovu, E., Akol, M., Sebikali, C., Kyambadde, J., Schmidt, C., Brun, R., Kaminsky, R., Ogwal, L.M., Kansiime, F., (1998).Parasitological detection of *Trypanosoma brucei gambiense* in serologically negative sleeping-sickness suspects from north-western Uganda. Ann Trop Med Parasitol 92, 845-850.

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

Enyaru, J.C.K., Matovu, E., Nerima, B., Akol, M., 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. Ann. N. Y. Acad. Sci 1081, 311-319.

Epstein, D.J., (2009). Cis-regulatory mutations in human disease. Brief Funct Genomic Proteomic 8, 310-316.

Evanno, G., Regnaut, S., Goudet, J., (2005).Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study.Mol. Ecol 14, 2611-2620.

Evans, D.A., Kennedy, W.P., Elbihari, S., Chapman, C.J., Smith, V., Peters, W., (1987).Hybrid formation within the genus *Leishmania*? Parassitologia 29, 165-173.

Fairlamb, A.H., (2003). Chemotherapy of human African trypanosomiasis: current and future prospects. Trends Parasitol 19, 488-494.

Fasogbon, A.I., Knowles, G., Gardiner, P.R., (1990). A comparison of the isoenzymes of *Trypanosoma (Duttonella) vivax* isolates from East and West Africa. Int. J. Parasitol 20, 389-394.

Faye, D., Pereira de Almeida, P.J., Goossens, B., Osaer, S., Ndao, M., Berkvens, D., Speybroeck, N., Nieberding, F., Geerts, S., (2001). Prevalence and incidence of trypanosomosis in horses and donkeys in the Gambia. Vet. Parasitol 101, 101-114.

Fenn, K., Matthews, K.R., (2007). The cell biology of *Trypanosoma brucei* differentiation.Curr.Opin. Microbiol 10, 539-546.

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

Fèvre, E.M., Coleman, P.G., Welburn, S.C., Maudlin, I., (2004). Reanalyzing the 1900-1920 sleeping sickness epidemic in Uganda. Emerging Infect. Dis 10, 567-573.

Fèvre, E.M., Picozzi, K., Fyfe, J., Waiswa, C., Odiit, M., Coleman, P.G., Welburn, S.C., (2005).A burgeoning epidemic of sleeping sickness in Uganda. Lancet 366, 745-747.

Fèvre, E.M., Wissmann, B.V., Welburn, S.C., Lutumba, P., (2008). The burden of human african trypanosomiasis. PLoS Negl Trop Dis 2, e333.

Flicek, P., Birney, E., (2009). Sense from sequence reads: methods for alignment and assembly. Nat. Methods 6, S6-S12.

Galiza, G.J.N., Garcia, H.A., Assis, A.C.O., Oliveira, D.M., Pimentel, L.A., Dantas, A.F.M., Simões, S.V.D., Teixeira, M.M.G., Riet-Correa, F., (2011).High mortality and lesions of the central nervous system in Trypanosomosis by *Trypanosoma vivax* in Brazilian hair sheep. Vet Parasitol.

Gao, H., Williamson, S., Bustamante, C.D., (2007). A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. Genetics 176, 1635-1651.

Gardiner, P.R., Wilson, A.J., (1987). *Trypanosoma (Duttonefla) vivax*.Parasitol.Today (Regul. Ed.) 3, 49-52.

Gardiner, P.R., (1989). Recent studies of the biology of *Trypanosoma vivax*. Adv. Parasitol 28, 229-317.

Garside, L., Bailey, M., Gibson, W., (1994).DNA content and molecular karyotype of trypanosomes of the subgenus Nannomonas. Acta Trop 57, 21-28.
Gashumba, J.K., Baker, R.D., Godfrey, D.G., (1988). *Trypanosoma congolense*: the distribution of enzymic variants in east and west Africa. Parasitology 96 (Pt 3), 475-486.
Gashumba, J.K., Komba, E.K., Truc, P., Allingham, R.M., Ferris, V., Godfrey, D.G., (1994). The persistence of genetic homogeneity among *Trypanosoma brucei rhodesiense* isolates from patients in north-west Tanzania. Acta Trop 56, 341-348.
Gaunt, M.W., Yeo, M., Frame, I.A., Stothard, J.R., Carrasco, H.J., Taylor, M.C., Mena, S.S., Veazey, P., Miles, G.A.J., Acosta, N., de Arias, A.R., Miles, M.A., (2003).Mechanism of genetic exchange in American trypanosomes. Nature 421, 936-939.
Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., Moazami-Goudarzi, K., Foulley, J.-L., (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. BMC Genomics 10, 550.
Gelanew, T., Kuhls, K., Hurissa, Z., Weldegebreal, T., Hailu, W., Kassahun, A., Abebe, T., Hailu, A., Schönian, G., (2010). Inference of population structure of *Leishmania donovani* strains isolated from different Ethiopian visceral leishmaniasis endemic areas. PLoS Negl Trop Dis 4, e889.
Gibson, W., Mehlitz, D., Lanham, S.M., Godfrey, D.G., (1978).The identification of *Trypanosoma bruceigambiense* in Liberian pigs and dogs by isoenzymes and by resistance to human plasma. Tropenmed Parasitol 29, 335-345.
Gibson, W.C., de C Marshall, T.F., Godfrey, D.G., (1980). Numerical analysis of enzyme polymorphism: a new approach to the epidemiology and taxonomy of trypanosomes of the subgenus Trypanozoon. Adv. Parasitol 18, 175-246.
Gibson, W.C., Osinga, K.A., Michels, P.A., Borst, P., (1985). Trypanosomes of subgenus Trypanozoon are diploid for housekeeping genes. Mol. Biochem. Parasitol 16, 231-242.
Gibson, W.C., (1986). Will the real *Trypanosoma b. gambiense* please stand up.Parasitol.Today (Regul. Ed.) 2, 255-257.
Gibson, W.C., Borst, P., (1986). Size-fractionation of the small chromosomes of Trypanozoon and Nannomonas trypanosomes by pulsed field gradient gel electrophoresis. Mol. Biochem. Parasitol 18, 127-140.
Gibson, W.C., (1989). Analysis of a genetic cross between *Trypanosoma brucei rhodesiense* and *T. b. brucei*. Parasitology 99 Pt 3, 391-402.
Gibson, W., Garside, L., (1991). Genetic exchange in *Trypanosoma brucei brucei*: variable chromosomal location of housekeeping genes in different trypanosome stocks. Mol. Biochem. Parasitol 45, 77-89.
Gibson, W., Garside, L., Bailey, M., (1992). Trisomy and chromosome size changes in hybrid trypanosomes from a genetic cross between *Trypanosoma brucei rhodesiense* and *T. b. brucei*. Mol. Biochem. Parasitol 51, 189-199.
Gibson, W., Whittington, H., (1993). Genetic exchange in *Trypanosoma brucei*: selection of hybrid trypanosomes by introduction of genes conferring drug resistance. Mol. Biochem. Parasitol 60, 19-26.
Gibson, W., Bailey, M., (1994). Genetic exchange in *Trypanosoma brucei*: evidence for meiosis from analysis of a cross between drug-resistant transformants. Mol. Biochem. Parasitol 64, 241-252.
Gibson, W., Kanmogne, G., Bailey, M., (1995).A successful backcross in *Trypanosoma brucei*. Mol. Biochem. Parasitol 69, 101-110.
Gibson, W., Winters, K., Mizen, G., Kearns, J., Bailey, M., (1997). Intraclonal mating in *Trypanosoma brucei* is associated with out-crossing. Microbiology (Reading, Engl.) 143 (Pt 3), 909-920.
Gibson, W., (2002). Will the real *Trypanosoma brucei rhodesiense* please step forward? Trends Parasitol 18, 486-490.
Gibson, W., Backhouse, T., Griffiths, A., (2002). The human serum resistance associated gene is ubiquitous and conserved in *Trypanosoma brucei rhodesiense* throughout East Africa. Infect. Genet. Evol 1, 207-214.
Gibson, W., (2003). Species concepts for trypanosomes: from morphological to molecular definitions? Kinetoplastid Biol Dis 2, 10.
Gibson, W., Peacock, L., Ferris, V., Williams, K., Bailey, M., (2006).Analysis of a cross between green and red fluorescent trypanosomes.Biochem. Soc. Trans 34, 557-559.
Gibson, W., (2007).Resolution of the species problem in African trypanosomes. Int. J. Parasitol 37, 829-838.
Gibson, W., Peacock, L., Ferris, V., Williams, K., Bailey, M., (2008).The use of yellow fluorescent hybrids to indicate mating in *Trypanosoma brucei*.Parasit Vectors 1, 4.
Gladyshev, E.A., Meselson, M., Arkhipova, I.R., (2008).Massive horizontal gene transfer in *bdelloid rotifers*. Science 320, 1210-1213.
Gladyshev, E., Meselson, M., (2008).Extreme resistance of *bdelloid rotifers* to ionizing radiation.Proc. Natl. Acad. Sci. U.S.A 105, 5139-5144.
Gondwe, N., Marcotty, T., Vanwambeke, S., De Pus, C., Mulumba, M., Van den Bossche, P., (2009). Distribution and Density of Tsetse Flies (Glossinidae: Diptera) at the Game/People/Livestock

Interface of the Nkhotakota Game Reserve Human Sleeping Sickness Focus in Malawi. Ecohealth 6, 260-265.
Goodhead, I., Archibald, A., Amwayi, P., Brass, A., Gibson, J., Hall, N., Hughes, M.A., Limo, M., Iraqi, F., Kemp, S.J., Noyes, H.A., (2010).A comprehensive genetic analysis of candidate genes regulating response to *Trypanosoma congolense* infection in mice. PLoS Negl Trop Dis 4, e880.
Gottesdiener, K., Garciá-Anoveros, J., Lee, M.G., Van der Ploeg, L.H., (1990).Chromosome organization of the protozoan *Trypanosoma brucei*. Mol. Cell. Biol 10, 6079-6083.
De Greef, C., Hamers, R., (1994). The serum resistance-associated (SRA) gene of *Trypanosoma brucei rhodesiense* encodes a variant surface glycoprotein-like protein. Mol. Biochem. Parasitol 68, 277-284.
De Greef, C., Imberechts, H., Matthyssens, G., Van Meirvenne, N., Hamers, R., (1989). A gene expressed only in serum-resistant variants of *Trypanosoma brucei rhodesiense*. Mol. Biochem. Parasitol 36, 169-176.
Green, H.P., Del Pilar Molina Portela, M., St Jean, E.N., Lugli, E.B., Raper, J., (2003). Evidence for a *Trypanosoma brucei* lipoprotein scavenger receptor. J. Biol. Chem 278, 422-427.
Hadany, L., Comeron, J.M., (2008). Why are sex and recombination so common? Ann. N. Y. Acad. Sci 1133, 26-43.
Haile, S., Papadopoulou, B., (2007).Developmental regulation of gene expression in trypanosomatid parasitic protozoa.Curr.Opin. Microbiol 10, 569-577.
Hajduk, S.L., Moore, D.R., Vasudevacharya, J., Siqueira, H., Torri, A.F., Tytler, E.M., Esko, J.D., (1989). Lysis of Trypanosoma brucei by a toxic subspecies of human high density lipoprotein.J. Biol. Chem 264, 5210-5217.
Halkett, F., Simon, J.-C., Balloux, F., (2005).Tackling the population genetics of clonal and partially clonal organisms.Trends Ecol. Evol.(Amst.) 20, 194-201.
Hall, N., Berriman, M., Lennard, N.J., Harris, B.R., Hertz-Fowler, C., Bart-Delabesse, E.N., Gerrard, C.S., Atkin, R.J., Barron, A.J., Bowman, S., Bray-Allen, S.P., Bringaud, F., Clark, L.N., Corton, C.H., Cronin, A., Davies, R., Doggett, J., Fraser, A., Grüter, E., Hall, S., Harper, A.D., Kay, M.P., Leech, V., Mayes, R., Price, C., Quail, M.A., Rabbinowitsch, E., Reitter, C., Rutherford, K., Sasse, J., Sharp, S., Shownkeen, R., MacLeod, A., Taylor, S., Tweedie, A., Turner, C.M.R., Tait, A., Gull, K., Barrell, B., Melville, S.E., (2003). The DNA sequence of chromosome I of an African trypanosome: gene content, chromosome organisation, recombination and polymorphism. Nucleic Acids Res 31, 4864-4873.
Hamilton, P.B., Stevens, J.R., Gaunt, M.W., Gidley, J., Gibson, W.C., (2004). Trypanosomes are monophyletic: evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA. Int. J. Parasitol 34, 1393-1404.
Hamilton, P.B., Adams, E.R., Malele, I.I., Gibson, W.C., (2008). A novel, high-throughput technique for species identification reveals a new species of tsetse-transmitted trypanosome related to the *Trypanosoma brucei* subgenus, Trypanozoon. Infect. Genet. Evol 8, 26-33.
Hanotte, O., Ronin, Y., Agaba, M., Nilsson, P., Gelhaus, A., Horstmann, R., Sugimoto, Y., Kemp, S., Gibson, J., Korol, A., Soller, M., Teale, A., (2003).Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle.Proc. Natl. Acad. Sci. U.S.A 100, 7443-7448.
Hartl, D.L., Volkman, S.K., Nielsen, K.M., Barry, A.E., Day, K.P., Wirth, D.F., Winzeler, E.A., (2002). The paradoxical population genetics of *Plasmodium falciparum*. Trends Parasitol 18, 266-272.
Hassold, T., Hall, H., Hunt, P., (2007). The origin of human aneuploidy: where we have been, where we are going. Hum. Mol. Genet 16 Spec No. 2, R203-208.
Hastings, I.M., Wedgwood-Oppenheim, B., (1997).Sex, strains and virulence.Parasitol.Today (Regul. Ed.) 13, 375-383.
Hedrick, P.W., (2005). Genetics of Populations, 3rd ed. Jones and Bartlett, Sudbury, Massachusetts.
Hide, G., Welburn, S.C., Tait, A., Maudlin, I., (1994). Epidemiological relationships of *Trypanosoma brucei* stocks from south east Uganda: evidence for different population structures in human infective and non-human infective isolates. Parasitology 109 (Pt 1), 95-111.
Hide, G., Angus, S.D., Holmes, P.H., Maudlin, I., Welburn, S.C., (1998). *Trypanosoma brucei*: comparison of circulating strains in an endemic and an epidemic area of a sleeping sickness focus. Exp. Parasitol 89, 21-29.
Hide, G., Tilley, A., Welburn, S.C., Maudlin, I., Tait, A., (2000). *Trypanosoma brucei*: identification of trypanosomes with genotypic similarity to human infective isolates in tsetse isolated from a region free of human sleeping sickness. Exp. Parasitol 96, 67-74.
Hide, G., Tait, A., (2009). Molecular epidemiology of African sleeping sickness. Parasitology 136, 1491-1500.
Hill, E.W., O'Gorman, G.M., Agaba, M., Gibson, J.P., Hanotte, O., Kemp, S.J., Naessens, J., Coussens, P.M., MacHugh, D.E., (2005). Understanding bovine trypanosomiasis and trypanotolerance: the promise of functional genomics. Vet. Immunol. Immunopathol 105, 247-258.

Hoare, C., (1972). The Trypanosomes of Mammals.Blackwell Scientific Publications, Oxford, UK.
Holm, S.H., Beech, J.P., Barrett, M.P., Tegenfeldt, J.O., (2011).Separation of parasites from human blood using deterministic lateral displacement. Lab Chip 11, 1326-1332.
Hope, M., MacLeod, A., Leech, V., Melville, S., Sasse, J., Tait, A., Turner, C.M., (1999). Analysis of ploidy (in megabase chromosomes) in *Trypanosoma brucei* after genetic exchange. Mol. Biochem. Parasitol 104, 1-9.
Hörandl, E., (2009. A combinational theory for maintenance of sex. Heredity 103, 445-457.
Hughes, A., Piontkivska, H., (2003a). Phylogeny of Trypanosomatidae and Bodonidae (Kinetoplastida) based on 18S rRNA: evidence for paraphyly of *Trypanosoma* and six other genera. Mol. Biol. Evol 20, 644-652.
Hughes, A., Piontkivska, H., (2003b). Molecular phylogenetics of Trypanosomatidae: contrasting results from 18S rRNA and protein phylogenies. Kinetoplastid Biol Dis 2, 15.
Hunt, R., Sauna, Z.E., Ambudkar, S.V., Gottesman, M.M., Kimchi-Sarfaty, C., (2009). Silent (synonymous) SNPs: should we care about them? Methods Mol. Biol 578, 23-39.
Hutchinson, O.C., Fèvre, E.M., Carrington, M., Welburn, S.C., (2003). Lessons learned from the emergence of a new *Trypanosoma brucei rhodesiense* sleeping sickness focus in Uganda. Lancet Infect Dis 3, 42-45.
Hutchinson, O.C., Picozzi, K., Jones, N.G., Mott, H., Sharma, R., Welburn, S.C., Carrington, M., (2007). Variant Surface Glycoprotein gene repertoires in *Trypanosoma brucei* have diverged to become strain-specific. BMC Genomics 8, 234.
Iles, M.M., (2008). What can genome-wide association studies tell us about the genetics of common disease? PLoS Genet 4, e33.
Imboden, M.A., Laird, P.W., Affolter, M., Seebeck, T., (1987). Transcription of the intergenic regions of the tubulin gene cluster of *Trypanosoma brucei*: evidence for a polycistronic transcription unit in a eukaryote. Nucleic Acids Res 15, 7357-7368.
International HapMap Consortium, (2003). The International HapMap Project. Nature 426, 789-796.
International HapMap Consortium, (2005). A haplotype map of the human genome. Nature 437, 1299-1320.
Iraqi, F., Clapcott, S.J., Kumari, P., Haley, C.S., Kemp, S.J., Teale, A.J., (2000.Fine mapping of trypanosomiasis resistance loci in murine advanced intercross lines.Mamm. Genome 11, 645-648.
Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., et al., (2005).The genome of the kinetoplastid parasite, Leishmania major. Science 309, 436-442.
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., Hertz-Fowler, C., (2010). The genome sequence of *Trypanosoma brucei gambiense*, causative agent of chronic human african trypanosomiasis. PLoS Negl Trop Dis 4, e658.
Jaitner, J., Corr, N., Dempfle, L., (2003). Ownership pattern and management practices of cattle herds in the Gambia: implications for a breeding programme. Trop Anim Health Prod 35, 179-187.
Jamonneau, V., N'Guessan, P., N'Dri, L., Simarro, P., Truc, P., (2000).Exploration of the distribution of *Trypanosoma brucei ssp*. in West Africa, by multilocus enzyme electrophoresis. Ann Trop Med Parasitol 94, 643-649.
Jamonneau, V., Garcia, A., Ravel, S., Cuny, G., Oury, B., Solano, P., N'Guessan, P., N'Dri, L., Sanon, R., Frézil, J.L., Truc, P., (2002). Genetic characterization of *Trypanosoma brucei gambiense* and clinical evolution of human African trypanosomiasis in Côte d'Ivoire. Trop. Med. Int. Health 7, 610-621.
Jamonneau, V., Ravel, S., Garcia, A., Koffi, M., Truc, P., Laveissière, C., Herder, S., Grébaut, P., Cuny, G., Solano, P., (2004). Characterization of *Trypanosoma brucei s.l.* infecting asymptomatic sleeping-sickness patients in Côte d'Ivoire: a new genetic group? Ann Trop Med Parasitol 98, 329-337.
Jenni, L., Marti, S., Schweizer, J., Betschart, B., Le Page, R.W., Wells, J.M., Tait, A., Paindavoine, P., Pays, E., Steinert, M., (1986). Hybrid formation between African trypanosomes during cyclical transmission. Nature 322, 173-175.
Johnson, P.J., Kooter, J.M., Borst, P., (1987). Inactivation of transcription by UV irradiation of *T. brucei* provides evidence for a multicistronic transcription unit including a VSG gene. Cell 51, 273-281.
Jones, T.W., Dávila, A.M., (2001). *Trypanosoma vivax*--out of Africa. Trends Parasitol 17, 99-101.
Judson, O.P., Normark, B.B., (1996). Ancient asexual scandals.Trends Ecol. Evol.(Amst.) 11, 41-46.
Kaboré, J., Koffi, M., Bucheton, B., Macleod, A., Duffy, C., Ilboudo, H., Camara, M., De Meeûs, T., Belem, A.M.G., Jamonneau, V., (2011).First evidence that parasite infecting apparent aparasitemic serological suspects in human African trypanosomiasis are *Trypanosoma brucei gambiense* and are similar to those found in patients. Infect Genet Evol 11, 1250-1255.
Kalu, A., Oboegbulem, S., Uzoukwu, M., (2001).Trypanosomosis in small ruminants maintained by low riverine tsetse population in central Nigeria.Small Rumin. Res 40, 109-115.
Kalus, A.A., Fredericks, L.P., Hacker, B.M., Dommisch, H., Presland, R.B., Kimball, J.R., Dale, B.A., (2009). Association of a genetic polymorphism (-44 C/G SNP) in the human DEFB1 gene with

expression and inducibility of multiple beta-defensins in gingival keratinocytes. BMC Oral Health 9, 21.
Kanmogne, G.D., Asonganyi, T., Gibson, W.C., (1996a).Detection of *Trypanosoma brucei gambiense*, in serologically positive but aparasitaemic sleeping-sickness suspects in Cameroon, by PCR. Ann Trop Med Parasitol 90, 475-483.
Kanmogne, G.D., Stevens, J.R., Asonganyi, T., Gibson, W.C., (1996b). Characterization of *Trypanosoma brucei gambiense* isolates using restriction fragment length polymorphisms in 5 variant surface glycoprotein genes. Acta Trop 61, 239-254.
Karimuribo, E.D., Morrison, L.J., Black, A., Turner, C.M.R., Kambarage, D.M., Ballingall, K.T., (2011). Analysis of host genetic factors influencing African trypanosome species infection in a cohort of Tanzanian Bos indicus cattle. Vet. Parasitol 179, 35-42.
Kazibwe, A.J.N., Nerima, B., de Koning, H.P., Mäser, P., Barrett, M.P., Matovu, E., (2009). Genotypic status of the TbAT1/P2 adenosine transporter of *Trypanosoma brucei gambiense* isolates from Northwestern Uganda following melarsoprol withdrawal. PLoS Negl Trop Dis 3, e523.
Kemp, S.J., Hanotte, O., Agaba, M., Noyes, H.A., Gibson, J., Archibald, A., Rennie, C., Hulme, H., Brass, A., (2008). Genomics approaches to study the biology underlying resistance to trypanosomiasis--some unexpected lessons. Dev Biol (Basel) 132, 89-91.
Kennedy, P.G.E., (2004). Human African trypanosomiasis of the CNS: current issues and challenges. J. Clin. Invest 113, 496-504.
Kennedy, P.G.E., (2006a). Human African trypanosomiasis-neurological aspects. J. Neurol 253, 411-416.
Kennedy, P.G.E., (2006b). Diagnostic and neuropathogenesis issues in human African trypanosomiasis. Int. J. Parasitol 36, 505-512.
Kennedy, P.G.E., (2008). The continuing problem of human African trypanosomiasis (sleeping sickness). Ann. Neurol 64, 116-126.
Kieft, R., Capewell, P., Turner, C.M.R., Veitch, N.J., MacLeod, A., Hajduk, S., (2010).Mechanism of *Trypanosoma brucei gambiense* (group 1) resistance to human trypanosome lytic factor. Proc. Natl. Acad. Sci. U.S.A 107, 16137-16141.
Kierstein, S., Noyes, H., Naessens, J., Nakamura, Y., Pritchard, C., Gibson, J., Kemp, S., Brass, A., (2006).Gene expression profiling in a mouse model for African trypanosomiasis. Genes Immun 7, 667-679.
Kilgour, V., Godfrey, D.G., (1977).The persistence in the field of two characteristic isoenzyme patterns in Nigerian *Trypanosoma vivax*. Ann Trop Med Parasitol 71, 387-389.
Kilgour, V., Godfrey, D.G., Na'isa, B.K., (1975).Isoenzymes of two aminotransferases among *Trypanosoma vivax* in Nigerian cattle. Ann Trop Med Parasitol 69, 329-335.
Kinabo, L.D., (1993). Pharmacology of existing drugs for animal trypanosomiasis. Acta Trop 54, 169-183.
Koffi, M., Solano, P., Barnabé, C., de Meeûs, T., Bucheton, B., Cuny, G., Jamonneau, V., (2007). Genetic characterisation of *Trypanosoma brucei s.l.* using microsatellite typing: new perspectives for the molecular epidemiology of human African trypanosomiasis. Infect. Genet. Evol 7, 675-684.
Koffi, M., De Meeûs, T., Bucheton, B., Solano, P., Camara, M., Kaba, D., Cuny, G., Ayala, F.J., Jamonneau, V., (2009).Population genetics of *Trypanosoma brucei gambiense*, the agent of sleeping sickness in Western Africa. Proc. Natl. Acad. Sci. U.S.A 106, 209-214.
Komar, A.A., (2007. Silent SNPs: impact on gene function and phenotype. Pharmacogenomics 8, 1075-1080.
Komba, E.K., Kibona, S.N., Ambwene, A.K., Stevens, J.R., Gibson, W.C., (1997). Genetic diversity among *Trypanosoma brucei rhodesiense* isolates from Tanzania. Parasitology 115 (Pt 6), 571-579.
de Koning, H.P., Jarvis, S.M., (1999). Adenosine transporters in bloodstream forms of *Trypanosoma brucei brucei*: substrate recognition motifs and affinity for trypanocidal drugs. Mol. Pharmacol 56, 1162-1170.
de Koning, H.P., (2001). Transporters in African trypanosomes: role in drug action and resistance. Int. J. Parasitol 31, 512-522.
de Koning, D.-J., Carlborg, O., Haley, C.S., (2005). The genetic dissection of immune response using gene-expression studies and genome mapping. Vet. Immunol. Immunopathol 105, 343-352.
Kotlyar, S., (2010). Recommendations for control of East african sleeping sickness in Uganda. J Glob Infect Dis 2, 43-48.
Kozarewa, I., Turner, D.J., (2011). 96-Plex Molecular Barcoding for the Illumina Genome Analyzer. Methods Mol Biol 733, 279-298.
Krafsur, E.S., (2009). Tsetse flies: genetics, evolution, and role as vectors. Infect. Genet. Evol 9, 124-141.
Kuboki, N., Inoue, N., Sakurai, T., Di Cello, F., Grab, D.J., Suzuki, H., Sugimoto, C., Igarashi, I., (2003. Loop-mediated isothermal amplification for detection of African trypanosomes. J. Clin. Microbiol 41, 5517-5524.
Kuepfer, I., Hhary, E.P., Allan, M., Edielu, A., Burri, C., Blum, J.A., (2011).Clinical presentation of *T.b. rhodesiense* sleeping sickness in second stage patients from Tanzania and Uganda. PLoS Negl Trop Dis 5, e968.

Kuhls, K., Keilonat, L., Ochsenreither, S., Schaar, M., Schweynoch, C., Presber, W., Schönian, G., (2007). Multilocus microsatellite typing (MLMT) reveals genetically isolated populations between and within the main endemic regions of visceral leishmaniasis. Microbes Infect 9, 334-343.

Kuhls, K., Chicharro, C., Cañavate, C., Cortes, S., Campino, L., Haralambous, C., Soteriadou, K., Pratlong, F., Dedet, J.-P., Mauricio, I., Miles, M., Schaar, M., Ochsenreither, S., Radtke, O.A., Schönian, G., (2008). Differentiation and gene flow among European populations of *Leishmania infantum* MON-1. PLoS Negl Trop Dis 2, e261.

Lai, D.-H., Hashimi, H., Lun, Z.-R., Ayala, F.J., Lukes, J., (2008). Adaptations of *Trypanosoma brucei* to gradual loss of kinetoplast DNA: *Trypanosoma equiperdum* and *Trypanosoma evansi* are petite mutants of *T. brucei*. Proc. Natl. Acad. Sci. U.S.A 105, 1999-2004.

Lapinski, J., Tunnacliffe, A., (2003). Anhydrobiosis without trehalose in *bdelloid rotifers*. FEBS Lett 553, 387-390.

Lejon, V., Ngoyi, D.M., Boelaert, M., Büscher, P., (2010). A CATT Negative Result after Treatment for Human African Trypanosomiasis Is No Indication for Cure. PLoS Negl Trop Dis 4, e590.

Leonard, G., Soanes, D.M., Stevens, J.R., (2011). Resolving the question of trypanosome monophyly: A comparative genomics approach using whole genome data sets with low taxon sampling. Infect. Genet. Evol 11, 955-959.

Leperre, P., Claxton, J.R., (1994). Comparative study of trypanosomosis in zebu and N'Dama cattle in The Gambia. Trop Anim Health Prod 26, 139-145.

Lewis, M.D., Llewellyn, M.S., Gaunt, M.W., Yeo, M., Carrasco, H.J., Miles, M.A., (2009). Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in *Trypanosoma cruzi* populations and expose contrasts between natural and experimental hybrids. Int. J. Parasitol 39, 1305-1317.

Li, H., Durbin, R., (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

Li, H., Homer, N., (2010). A survey of sequence alignment algorithms for next-generation sequencing. Brief. Bioinformatics 11, 473-483.

Liang, X.-hai, Haritan, A., Uliel, S., Michaeli, S., (2003).Trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. Eukaryotic Cell 2, 830-840.

Llewellyn, M.S., Miles, M.A., Carrasco, H.J., Lewis, M.D., Yeo, M., Vargas, J., Torrico, F., Diosque, P., Valente, V., Valente, S.A., Gaunt, M.W., (2009). Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. PLoS Pathog 5, e1000410.

de Luca D'oro, G.M., Gardenal, C.N., Perret, B., Crisci, J.V., Montamat, E.E., (1993). Genetic structure of *Trypanosoma cruzi* populations from Argentina estimated from enzyme polymorphism. Parasitology 107 (Pt 4), 405-410.

Lukes, J., Mauricio, I.L., Schönian, G., Dujardin, J.-C., Soteriadou, K., Dedet, J.-P., Kuhls, K., Tintaya, K.W.Q., Jirků, M., Chocholová, E., Haralambous, C., Pratlong, F., Oborník, M., Horák, A., Ayala, F.J., Miles, M.A., (2007).Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy.Proc. Natl. Acad. Sci. U.S.A 104, 9375-9380.

Lumsden, W.H., Kimber, C.D., Evans, D.A., Doig, S.J., (1979). *Trypanosoma brucei*: Miniature anion-exchange centrifugation technique for detection of low parasitaemias: Adaptation for field use. Trans. R. Soc. Trop. Med. Hyg 73, 312-317.

Macgregor, P., Matthews, K.R., (2010). New discoveries in the transmission biology of sleeping sickness parasites: applying the basics. J Mol Med 88, 865-871.

Machado, R.L.D., Povoa, M.M., Calvosa, V.S.P., Ferreira, M.U., Rossit, A.R.B., dos Santos, E.J.M., Conway, D.J., (2004). Genetic structure of *Plasmodium falciparum* populations in the Brazilian Amazon region. J. Infect. Dis 190, 1547-1555.

MacLean, L., Chisi, J.E., Odiit, M., Gibson, W.C., Ferris, V., Picozzi, K., Sternberg, J.M., (2004). Severity of human african trypanosomiasis in East Africa is associated with geographic location, parasite genotype, and host inflammatory cytokine response profile. Infect. Immun 72, 7040-7044.

Maclean, L., Odiit, M., Macleod, A., Morrison, L., Sweeney, L., Cooper, A., Kennedy, P.G.E., Sternberg, J.M., (2007).Spatially and genetically distinct African Trypanosome virulence variants defined by host interferon-gamma response. J. Infect. Dis 196, 1620-1628.

MacLean, L.M., Odiit, M., Chisi, J.E., Kennedy, P.G.E., Sternberg, J.M., (2010).Focus-specific clinical profiles in human African Trypanosomiasis caused by *Trypanosoma brucei rhodesiense*. PLoS Negl Trop Dis 4, e906.

MacLeod, A., Turner, C.M., Tait, A., (1999). A high level of mixed *Trypanosoma brucei* infections in tsetse flies detected by three hypervariable minisatellites. Mol. Biochem. Parasitol 102, 237-248.

MacLeod, A., Tweedie, A., Welburn, S.C., Maudlin, I., Turner, C.M., Tait, A., (2000). Minisatellite marker analysis of *Trypanosoma brucei*: reconciliation of clonal, panmictic, and epidemic population genetic structures. Proc. Natl. Acad. Sci. U.S.A 97, 13442-13447.

MacLeod, A., Tait, A., Turner, C.M., (2001a). The population genetics of *Trypanosoma brucei* and the origin of human infectivity. Philos. Trans. R. Soc. Lond., B, Biol. Sci 356, 1035-1044.

MacLeod, A., Turner, C.M., Tait, A., (2001b). The detection of geographical substructuring of *Trypanosoma brucei* populations by the analysis of minisatellite polymorphisms. Parasitology 123, 475-482.

MacLeod, A., Welburn, S., Maudlin, I., Turner, C.M., Tait, A., (2001c).Evidence for multiple origins of human infectivity in *Trypanosoma brucei* revealed by minisatellite variant repeat mapping. J. Mol. Evol 52, 290-301.

MacLeod, A., Tweedie, A., McLellan, S., Hope, M., Taylor, S., Cooper, A., Sweeney, L., Turner, C.M.R., Tait, A., (2005a). Allelic segregation and independent assortment in *T. brucei* crosses: proof that the genetic system is Mendelian and involves meiosis. Mol. Biochem. Parasitol 143, 12-19.

MacLeod, A., Tweedie, A., McLellan, S., Taylor, S., Hall, N., Berriman, M., El-Sayed, N.M., Hope, M., Turner, C.M.R., Tait, A., (2005b). The genetic map and comparative analysis with the physical map of *Trypanosoma brucei*. Nucleic Acids Res 33, 6688-6693.

Madanitsa, M., Chisi, J., Ngwira, B., (2009). The epidemiology of trypanosomiasis in Rumphi district, Malawi: a ten year retrospective study. Malawi Med J 21, 22-27.

Madruga, C.R., Araújo, F.R., Cavalcante-Goes, G., Martins, C., Pfeifer, I.B., Ribeiro, L.R., Kessler, R.H., Soares, C.O., Miguita, M., Melo, E.P.S., Almeida, R.F.C., Lima, M.M.S.C., (2006).The development of an enzyme-linked immunosorbent assay for *Trypanosoma vivax* antibodies and its use in epidemiological surveys.Mem. Inst. Oswaldo Cruz 101, 801-807.

Magez, S., Caljon, G., Tran, T., Stijlemans, B., Radwanska, M., (2010).Current status of vaccination against African trypanosomiasis. Parasitology 137, 2017-2027.

Magnus, E., Vervoort, T., Van Meirvenne, N., (1978).A card-agglutination test with stained trypanosomes (C.A.T.T.) for the serological diagnosis of *T. b. gambiense* trypanosomiasis. Ann Soc Belg Med Trop 58, 169-176.

Magona, J.W., Mayende, J.S.P., Walubengo, J., (2002). Comparative evaluation of the antibody-detection ELISA technique using microplates precoated with denatured crude antigens from *Trypanosoma congolense* or *Trypanosoma vivax*. Trop Anim Health Prod 34, 295-308.

Magona, J.W., Walubengo, J., Odimin, J.T., (2008). Acute haemorrhagic syndrome of bovine trypanosomosis in Uganda. Acta Trop 107, 186-191.

Mahama, C.I., Desquesnes, M., Dia, M.L., Losson, B., De Deken, R., Speybroeck, N., Geerts, S., (2005). A longitudinal epidemiological survey of bovine trypanosomosis and its vectors in the White Volta river basin of Northern Ghana. Vet. Parasitol 128, 201-208.

Malele, I., Craske, L., Knight, C., Ferris, V., Njiru, Z., Hamilton, P., Lehane, S., Lehane, M., Gibson, W., (2003). The use of specific and generic primers to identify trypanosome infections of wild tsetse flies in Tanzania by PCR. Infect. Genet. Evol 3, 271-279.

Manolio, T.A., Brooks, L.D., Collins, F.S., (2008). A HapMap harvest of insights into the genetics of common disease. J. Clin. Invest 118, 1590-1605.

Manolio, T.A., Collins, F.S., (2009). The HapMap and genome-wide association studies in diagnosis and therapy. Annu. Rev. Med 60, 443-456.

Mark Welch, D., Meselson, M., (2000).Evidence for the evolution of *bdelloid rotifers* without sexual reproduction or genetic exchange. Science 288, 1211-1215.

Mark Welch, D.B., Cummings, M.P., Hillis, D.M., Meselson, M., (2004a). Divergent gene copies in the asexual class *Bdelloidea (Rotifera)* separated before the *bdelloid* radiation or within *bdelloid* families. Proc. Natl. Acad. Sci. U.S.A 101, 1622-1625.

Mark Welch, J.L., Mark Welch, D.B., Meselson, M., (2004b).Cytogenetic evidence for asexual evolution of *bdelloid rotifers*. Proc. Natl. Acad. Sci. U.S.A 101, 1618-1621.

Martínez-Calvillo, S., Vizuet-de-Rueda, J.C., Florencio-Martínez, L.E., Manning-Cela, R.G., Figueroa-Angulo, E.E., (2010).Gene expression in trypanosomatid parasites. J. Biomed. Biotechnol 2010, 525241.

Martín-Sánchez, J., Gállego, M., Barón, S., Castillejo, S., Morillas-Marquez, F., (2006). Pool screen PCR for estimating the prevalence of *Leishmania infantum* infection in sandflies (Diptera: Nematocera, Phlebotomidae). Trans. R. Soc. Trop. Med. Hyg 100, 527-532.

Masake, R.A., Nantulya, V.M., Akol, G.W., Musoke, A.J., (1984). Cerebral trypanosomiasis in cattle with mixed *Trypanosoma congolense* and *T. brucei brucei* infections. Acta Trop 41, 237-246.

Mäser, P., Lüscher, A., Kaminsky, R., (2003). Drug transport and drug resistance in African trypanosomes. Drug Resist. Updat 6, 281-290.

Masiga, D.K., Smyth, A.J., Hayes, P., Bromidge, T.J., Gibson, W.C., (1992). Sensitive detection of trypanosomes in tsetse flies by DNA amplification. Int. J. Parasitol 22, 909-918.

Masiga, D.K., McNamara, J.J., Gibson, W.C., (1996). A repetitive DNA sequence specific for *Trypanosoma (Nannomonas) godfreyi*. Vet. Parasitol 62, 27-33.

Maslov, D.A., Lukes, J., Jirku, M., Simpson, L., (1996). Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: implications for the evolution of parasitism in the trypanosomatid protozoa. Mol. Biochem. Parasitol 75, 197-205.

Masumu, J., Marcotty, T., Geysen, D., Geerts, S., Vercruysse, J., Dorny, P., den Bossche, P.V., (2006). Comparison of the virulence of *Trypanosoma congolense* strains isolated from cattle in a trypanosomiasis endemic area of eastern Zambia. Int. J. Parasitol 36, 497-501.

Masumu, J., Marcotty, T., Geerts, S., Vercruysse, J., Van den Bossche, P., (2009). Cross-protection between *Trypanosoma congolense* strains of low and high virulence. Vet. Parasitol 163, 127-131.

Mathieu-Daudé, F., Tibayrenc, M., (1994). Isozyme variability of *Trypanosoma brucei s.l.*: genetic, taxonomic, and epidemiological significance. Exp. Parasitol 78, 1-19.

Mathieu-Daudé, F., Stevens, J., Welsh, J., Tibayrenc, M., McClelland, M., (1995). Genetic diversity and population structure of *Trypanosoma brucei*: clonality versus sexuality. Mol. Biochem. Parasitol 72, 89-101.

Matovu, E., Seebeck, T., Enyaru, J.C., Kaminsky, R., (2001).Drug resistance in *Trypanosoma brucei spp.*, the causative agents of sleeping sickness in man and nagana in cattle. Microbes Infect 3, 763-770.

Matthews, K.R., (2005). The developmental cell biology of *Trypanosoma brucei*. J. Cell. Sci 118, 283-290.

Mattioli, R.C., Zinsstag, J., Pfister, K., (1994). Frequency of trypanosomosis and gastrointestinal parasites in draught donkeys in The Gambia in relation to animal husbandry. Trop Anim Health Prod 26, 102-108.

Maynard Smith, J., (1978). The Evolution of Sex.Cambridge University Press, Cambridge, UK.

Maynard Smith, J., Smith, N.H., O'Rourke, M., Spratt, B.G., (1993). How clonal are bacteria? Proc. Natl. Acad. Sci. U.S.A 90, 4384-4388.

McNamara, J.J., Mohammed, G., Gibson, W.C., (1994). *Trypanosoma (Nannomonas) godfreyi sp*.from tsetse flies in The Gambia: biological and biochemical characterization. Parasitology 109 (Pt 4), 497-509.

Meade, K.G., O'Gorman, G.M., Hill, E.W., Narciandi, F., Agaba, M., Kemp, S.J., O'Farrelly, C., MacHugh, D.E., (2009).Divergent antimicrobial peptide (AMP) and acute phase protein (APP) responses to *Trypanosoma congolense* infection in trypanotolerant and trypanosusceptible cattle. Mol. Immunol 47, 196-204.

de Meeûs, T., Lehmann, L., Balloux, F., (2006). Molecular epidemiology of clonal diploids: a quick overview and a short DIY (do it yourself) notice. Infect. Genet. Evol 6, 163-170.

de Meeûs, T., McCoy, K.D., Prugnolle, F., Chevillon, C., Durand, P., Hurtrez-Boussès, S., Renaud, F., (2007). Population genetics and molecular epidemiology or how to "débusquer la bête". Infect. Genet. Evol 7, 308-332.

Mehlitz, D., Zillmann, U., Scott, C.M., Godfrey, D.G., (1982). Epidemiological studies on the animal reservoir of Gambiense sleeping sickness. Part III. Characterization of trypanozoon stocks by isoenzymes and sensitivity to human serum. Tropenmed Parasitol 33, 113-118.

Melville, S.E., (1997). Parasite genome analysis. Genome research in *Trypanosoma brucei*: chromosome size polymorphism and its relevance to genome mapping and analysis. Trans. R. Soc. Trop. Med. Hyg 91, 116-120.

Melville, S.E., Leech, V., Gerrard, C.S., Tait, A., Blackwell, J.M., (1998). The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers. Mol. Biochem. Parasitol 94, 155-173.

Melville, S.E., Gerrard, C.S., Blackwell, J.M., (1999). Multiple causes of size variation in the diploid megabase chromosomes of African tyrpanosomes. Chromosome Res 7, 191-203.

Melville, S.E., Leech, V., Navarro, M., Cross, G.A., (2000). The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* stock 427. Mol. Biochem. Parasitol 111, 261-273.

Metzker, M.L., (2005). Emerging technologies in DNA sequencing. Genome Res 15, 1767-1776.

Metzker, M.L., (2010). Sequencing technologies - the next generation. Nat. Rev. Genet 11, 31-46.

Mihok, S., Maramba, O., Munyoki, E., Kagoiya, J., (1995). Mechanical transmission of *Trypanosoma spp*. by African Stomoxyinae (Diptera: Muscidae). Trop. Med. Parasitol 46, 103-105.

Milet, J., Nuel, G., Watier, L., Courtin, D., Slaoui, Y., Senghor, P., Migot-Nabias, F., Gaye, O., Garcia, A., (2010).Genome wide linkage study, using a 250K SNP map, of *Plasmodium falciparum* infection and mild malaria attack in a Senegalese population. PLoS ONE 5, e11616.

Miller, J.R., Koren, S., Sutton, G., (2010).Assembly algorithms for next-generation sequencing data. Genomics 95, 315-327.

Miruk, A., Hagos, A., Yacob, H.T., Asnake, F., Basu, A.K., (2008). Prevalence of bovine trypanosomosis and trypanocidal drug sensitivity studies on *Trypanosoma congolense* in Wolyta and Dawero zones of southern Ethiopia. Vet. Parasitol 152, 141-147.

Mori, Y., Notomi, T., (2009). Loop-mediated isothermal amplification (LAMP): a rapid, accurate, and cost-effective diagnostic method for infectious diseases. J. Infect. Chemother 15, 62-69.

Morrison, A.C., Ferro, C., Morales, A., Tesh, R.B., Wilson, M.L., (1993). Dispersal of the sand fly *Lutzomyia longipalpis* (Diptera: Psychodidae) at an endemic focus of visceral leishmaniasis in Colombia. J. Med. Entomol 30, 427-435.

Morrison, L.J., McCormack, G., Sweeney, L., Likeufack, A.C.L., Truc, P., Turner, C.M., Tait, A., MacLeod, A., (2007). Use of multiple displacement amplification to increase the detection and genotyping of trypanosoma species samples immobilized on FTA filters. Am. J. Trop. Med. Hyg 76, 1132-1137.
Morrison, L.J., Mallon, M.E., Smith, H.V., MacLeod, A., Xiao, L., Tait, A., (2008a). The population structure of the *Cryptosporidium parvum* population in Scotland: a complex picture. Infect. Genet. Evol 8, 121-129.
Morrison, L.J., Tait, A., McCormack, G., Sweeney, L., Black, A., Truc, P., Likeufack, A.C.L., Turner, C.M., MacLeod, A., (2008b). *Trypanosoma brucei gambiense* Type 1 populations from human patients are clonal and display geographical genetic differentiation. Infect. Genet. Evol 8, 847-854.
Morrison, L.J., Marcello, L., McCulloch, R., (2009a). Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity. Cell. Microbiol 11, 1724-1734.
Morrison, L.J., Tweedie, A., Black, A., Pinchbeck, G.L., Christley, R.M., Schoenefeld, A., Hertz-Fowler, C., MacLeod, A., Turner, C.M.R., Tait, A., (2009b).Discovery of mating in the major African livestock pathogen *Trypanosoma congolense*. PLoS ONE 4, e5564.
Morrison, L.J., (2011). Parasite-driven pathogenesis in *Trypanosoma brucei* infections.Parasite Immunol 33, 448-55.
Mu, J., Awadalla, P., Duan, J., McGee, K.M., Joy, D.A., McVean, G.A.T., Su, X.-zhuan, (2005). Recombination hotspots and population structure in *Plasmodium falciparum*. PLoS Biol 3, e335.
Mu, J., Myers, R.A., Jiang, H., Liu, S., Ricklefs, S., Waisberg, M., Chotivanich, K., Wilairatana, P., Krudsood, S., White, N.J., Udomsangpetch, R., Cui, L., Ho, M., Ou, F., Li, H., Song, J., Li, G., Wang, X., Seila, S., Sokunthea, S., Socheat, D., Sturdevant, D.E., Porcella, S.F., Fairhurst, R.M., Wellems, T.E., Awadalla, P., Su, X.-zhuan, (2010). *Plasmodium falciparum* genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. Nat. Genet 42, 268-271.
Muller, H.J., (1964). The relation of recombination to mutational advance.Mutat. Res 106, 2-9.
Mumba, D., Bohorquez, E., Messina, J., Kande, V., Taylor, S.M., Tshefu, A.K., Muwonga, J., Kashamuka, M.M., Emch, M., Tidwell, R., Büscher, P., Meshnick, S.R., (2011. Prevalence of human african trypanosomiasis in the democratic republic of the congo. PLoS Negl Trop Dis 5, e1246.
Murray, M., Murray, P.K., McIntyre, W.I., (1977).An improved parasitological technique for the diagnosis of African trypanosomiasis. Trans. R. Soc. Trop. Med. Hyg 71, 325-326.
Murray, M., Morrison, W.I., Whitelaw, D.D., (1982). Host susceptibility to African trypanosomiasis: trypanotolerance. Adv. Parasitol 21, 1-68.
Musunuru, K., Kathiresan, S., (2008). HapMap and mapping genes for cardiovascular disease. Circ Cardiovasc Genet 1, 66-71.
Mzilahowa, T., McCall, P.J., Hastings, I.M., (2007).“Sexual” population structure and genetics of the malaria agent *P. falciparum*. PLoS ONE 2, e613.
Naessens, J., Teale, A.J., Sileghem, M., (2002).Identification of mechanisms of natural resistance to African trypanosomiasis in cattle. Vet. Immunol. Immunopathol 87, 187-194.
Naessens, J., Leak, S.G.A., Kennedy, D.J., Kemp, S.J., Teale, A.J., (2003). Responses of bovine chimaeras combining trypanosomosis resistant and susceptible genotypes to experimental infection with *Trypanosoma congolense*. Vet. Parasitol 111, 125-142.
Naessens, J., (2006). Bovine trypanotolerance: A natural ability to prevent severe anaemia and haemophagocytic syndrome? Int. J. Parasitol 36, 521-528.
Nair, S., Williams, J.T., Brockman, A., Paiphun, L., Mayxay, M., Newton, P.N., Guthmann, J.-P., Smithuis, F.M., Hien, T.T., White, N.J., Nosten, F., Anderson, T.J.C., (2003). A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. Mol. Biol. Evol 20, 1526-1536.
Nair, S., Nash, D., Sudimack, D., Jaidee, A., Barends, M., Uhlemann, A.-C., Krishna, S., Nosten, F., Anderson, T.J.C., (2007).Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. Mol. Biol. Evol 24, 562-573.
Neafsey, D.E., Schaffner, S.F., Volkman, S.K., Park, D., Montgomery, P., Milner, D.A., Jr, Lukens, A., Rosen, D., Daniels, R., Houde, N., Cortese, J.F., Tyndall, E., Gates, C., Stange-Thomann, N., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Ferreira, M.U., Moraes, S. do L., Dash, A.P., Chitnis, C.E., Wiegand, R.C., Hartl, D.L., Birren, B.W., Lander, E.S., Sabeti, P.C., Wirth, D.F., (2008). Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. Genome Biol 9, R171.
Neiman, M., Meirmans, S., Meirmans, P.G., (2009). What can asexual lineage age tell us about the maintenance of sex? Ann. N. Y. Acad. Sci 1168, 185-200.
Nganga, J.K., Soller, M., Iraqi, F.A., (2010). High resolution mapping of trypanosomosis resistance loci Tir2 and Tir3 using F12 advanced intercross lines with major locus Tir1 fixed for the susceptible allele. BMC Genomics 11, 394.
Van Nieuwenhove, S., Schechter, P.J., Declercq, J., Boné, G., Burke, J., Sjoerdsma, A., (1985).Treatment of gambiense sleeping sickness in the Sudan with oral DFMO (DL-alpha-difluoromethylornithine), an inhibitor of ornithine decarboxylase; first field trial. Trans. R. Soc. Trop. Med. Hyg 79, 692-698.

Nimpaye, H., Njiokou, F., Njine, T., Njitchouang, G.R., Cuny, G., Herder, S., Asonganyi, T., Simo, G., (2011).*Trypanosoma vivax*, *T. congolense* "forest type" and *T. simiae*: prevalence in domestic animals of sleeping sickness foci of Cameroon. Parasite 18, 171-179.

Nishant, K.T., Rao, M.R.S., (2006). Molecular features of meiotic recombination hot spots. Bioessays 28, 45-56.

Niu, T., (2004.Algorithms for inferring haplotypes.Genet. Epidemiol 27, 334-347.

Njiokou, F., Nkinin, S.W., Grébaut, P., Penchenier, L., Barnabé, C., Tibayrenc, M., Herder, S., (2004a). An isoenzyme survey of *Trypanosoma brucei s.l.* from the Central African subregion: population structure, taxonomic and epidemiological considerations. Parasitology 128, 645-653.

Njiokou, F., Simo, G., Mbida Mbida, A., Truc, P., Cuny, G., Herder, S., (2004b). A study of host preference in tsetse flies using a modified heteroduplex PCR-based method. Acta Trop 91, 117-120.

Njiokou, F., Simo, G., Nkinin, S.W., Laveissière, C., Herder, S., (2004c).Infection rate of *Trypanosoma brucei s.l.*, *T. vivax*, *T. congolense* "forest type", and *T. simiae* in small wild vertebrates in south Cameroon. Acta Trop 92, 139-146.

Njiru, Z.K., Makumi, J.N., Okoth, S., Ndungu, J.M., Gibson, W.C., (2004a). Identification of trypanosomes in *Glossina pallidipes* and *G. longipennis* in Kenya. Infect. Genet. Evol 4, 29-35.

Njiru, Z.K., Ndung'u, K., Matete, G., Ndungu, J.M., Gibson, W.C., (2004b). Detection of *Trypanosoma brucei rhodesiense* in animals from sleeping sickness foci in East Africa using the serum resistance associated (SRA) gene. Acta Trop 90, 249-254.

Njiru, Z.K., Mikosza, A.S.J., Matovu, E., Enyaru, J.C.K., Ouma, J.O., Kibona, S.N., Thompson, R.C.A., Ndung'u, J.M., (2008a). African trypanosomiasis: sensitive and rapid detection of the sub-genus Trypanozoon by loop-mediated isothermal amplification (LAMP) of parasite DNA. Int. J. Parasitol 38, 589-599.

Njiru, Z.K., Mikosza, A.S.J., Armstrong, T., Enyaru, J.C., Ndung'u, J.M., Thompson, A.R.C., (2008b).Loop-Mediated Isothermal Amplification (LAMP) Method for Rapid Detection of *Trypanosoma brucei rhodesiense*. PLoS Negl Trop Dis 2, e147.

Njiru, Z.K., Ouma, J.O., Bateta, R., Njeru, S.E., Ndungu, K., Gitonga, P.K., Guya, S., Traub, R., (2011a).Loop-mediated isothermal amplification test for *Trypanosoma vivax* based on satellite repeat DNA. Vet Parasitol 180, 358-62.

Njiru, Z.K., Traub, R., Ouma, J.O., Enyaru, J.C., Matovu, E., (2011b). Detection of Group 1 *Trypanosoma brucei gambiense* by loop-mediated isothermal amplification. J. Clin. Microbiol 49, 1530-1536.

Njiru, Z.K., (2011). Rapid and sensitive detection of human African trypanosomiasis by loop-mediated isothermal amplification combined with a lateral-flow dipstick. Diagn.Microbiol. Infect. Dis 69, 205-209.

Nolder, D., Roncal, N., Davies, C.R., Llanos-Cuentas, A., Miles, M.A., (2007).Multiple hybrid genotypes of *Leishmania (viannia)* in a focus of mucocutaneous Leishmaniasis. Am. J. Trop. Med. Hyg 76, 573-578.

Noyes, H.A., Alimohammadian, M.H., Agaba, M., Brass, A., Fuchs, H., Gailus-Durner, V., Hulme, H., Iraqi, F., Kemp, S., Rathkolb, B., Wolf, E., de Angelis, M.H., Roshandel, D., Naessens, J., (2009). Mechanisms controlling anaemia in *Trypanosoma congolense* infected mice. PLoS ONE 4, e5170.

Noyes, H., Brass, A., Obara, I., Anderson, S., Archibald, A.L., Bradley, D.G., Fisher, P., Freeman, A., Gibson, J., Gicheru, M., Hall, L., Hanotte, O., Hulme, H., McKeever, D., Murray, C., Oh, S.J., Tate, C., Smith, K., Tapio, M., Wambugu, J., Williams, D.J., Agaba, M., Kemp, S.J., (2011). Genetic and expression analysis of cattle identifies candidate genes in pathways responding to *Trypanosoma congolense* infection. Proc. Natl. Acad. Sci. U.S.A 108, 9304-9309.

O'Gorman, G.M., Park, S.D.E., Hill, E.W., Meade, K.G., Mitchell, L.C., Agaba, M., Gibson, J.P., Hanotte, O., Naessens, J., Kemp, S.J., MacHugh, D.E., (2006). Cytokine mRNA profiling of peripheral blood mononuclear cells from trypanotolerant and trypanosusceptible cattle infected with *Trypanosoma congolense*. Physiol. Genomics 28, 53-61.

O'Gorman, G.M., Park, S.D.E., Hill, E.W., Meade, K.G., Coussens, P.M., Agaba, M., Naessens, J., Kemp, S.J., MacHugh, D.E., (2009). Transcriptional profiling of cattle infected with *Trypanosoma congolense* highlights gene expression signatures underlying trypanotolerance and trypanosusceptibility. BMC Genomics 10, 207.

Oberle, M., Balmer, O., Brun, R., Roditi, I., (2010).Bottlenecks and the maintenance of minor genotypes during the life cycle of *Trypanosoma brucei*. PLoS Pathog 6, e1001023.

Oli, M.W., Cotlin, L.F., Shiflett, A.M., Hajduk, S.L., (2006). Serum resistance-associated protein blocks lysosomal targeting of trypanosome lytic factor in *Trypanosoma brucei*. Eukaryotic Cell 5, 132-139.

Oliveira, R.P., Broude, N.E., Macedo, A.M., Cantor, C.R., Smith, C.L., Pena, S.D., (1998). Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites.Proc. Natl. Acad. Sci. U.S.A 95, 3776-3780.

Orjuela-Sánchez, P., Karunaweera, N.D., da Silva-Nunes, M., da Silva, N.S., Scopel, K.K.G., Gonçalves, R.M., Amaratunga, C., Sá, J.M., Socheat, D., Fairhust, R.M., Gunawardena, S., Thavakodirasah, T., Galapaththy, G.L.N., Abeysinghe, R., Kawamoto, F., Wirth, D.F., Ferreira, M.U., (2010). Single-

nucleotide polymorphism, linkage disequilibrium and geographic structure in the malaria parasite *Plasmodium vivax*: prospects for genome-wide association studies. BMC Genet 11, 65.
Osório, A.L.A.R., Madruga, C.R., Desquesnes, M., Soares, C.O., Ribeiro, L.R.R., Costa, S.C.G. da, (2008).*Trypanosoma (Duttonella) vivax*: its biology, epidemiology, pathogenesis, and introduction in the New World--a review. Mem. Inst. Oswaldo Cruz 103, 1-13.
Otte, M.J., Abuabara, J.Y., (1991). Transmission of South American *Trypanosoma vivax* by the neotropical horsefly Tabanus nebulosus. Acta Trop 49, 73-76.
Otto, S.P., (2009).The evolutionary enigma of sex.Am. Nat 174 Suppl 1, S1-S14.
Paindavoine, P., Pays, E., Laurent, M., Geltmeyer, Y., Le Ray, D., Mehlitz, D., Steinert, M., (1986a). The use of DNA hybridization and numerical taxonomy in determining relationships between *Trypanosoma brucei* stocks and subspecies. Parasitology 92 (Pt 1), 31-50.
Paindavoine, P., Zampetti-Bosseler, F., Pays, E., Schweizer, J., Guyaux, M., Jenni, L., Steinert, M., (1986b). Trypanosome hybrids generated in tsetse flies by nuclear fusion. EMBO J 5, 3631-3636.
Paindavoine, P., Zampetti-Bosseler, F., Coquelet, H., Pays, E., Steinert, M., (1989). Different allele frequencies in *Trypanosoma brucei brucei* and *Trypanosoma brucei gambiense* populations. Mol. Biochem. Parasitol 32, 61-71.
Paszkiewicz, K., Studholme, D.J., (2010). De novo assembly of short sequence reads. Brief. Bioinformatics 11, 457-472.
Paul, R.E., Packer, M.J., Walmsley, M., Lagog, M., Ranford-Cartwright, L.C., Paru, R., Day, K.P., (1995).Mating patterns in malaria parasite populations of Papua New Guinea. Science 269, 1709-1711.
Paul, R.E., Day, K.P., (1998). Mating Patterns of *Plasmodium falciparum*.Parasitol.Today (Regul. Ed.) 14, 197-202.
Peacock, L., Ferris, V., Bailey, M., Gibson, W., (2007). Dynamics of infection and competition between two strains of *Trypanosoma brucei brucei* in the tsetse fly observed using fluorescent markers. Kinetoplastid Biol Dis 6, 4.
Peacock, L., Ferris, V., Bailey, M., Gibson, W., (2008). Fly transmission and mating of *Trypanosoma brucei brucei* strain 427. Mol. Biochem. Parasitol 160, 100-106.
Peacock, L., Ferris, V., Bailey, M., Gibson, W., (2009). Intraclonal mating occurs during tsetse transmission of *Trypanosoma brucei*. Parasit Vectors 2, 43.
Peacock, L., Ferris, V., Sharma, R., Sunter, J., Bailey, M., Carrington, M., Gibson, W., (2011).Identification of the meiotic life cycle stage of *Trypanosoma brucei* in the tsetse fly.Proc. Natl. Acad. Sci. U.S.A 108, 3671-3676.
Peakall, R., Smouse, P.E., (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6, 288 - 295.
Peiffer, D.A., Gunderson, K.L., (2009). Design of tag SNP whole genome genotyping arrays. Methods Mol. Biol 529, 51-61.
Pépin, J., Mpia, B., (2005). Trypanosomiasis relapse after melarsoprol therapy, Democratic Republic of Congo, 1982-2001. Emerging Infect. Dis 11, 921-927.
Peregrine, A.S., Mamman, M., (1993). Pharmacology of diminazene: a review. Acta Trop 54, 185-203.
Pereira de Almeida, P.J., Ndao, M., Goossens, B., Osaer, S., (1998). PCR primer evaluation for the detection of trypanosome DNA in naturally infected goats. Vet. Parasitol 80, 111-116.
Pettersson, E., Lundeberg, J., Ahmadian, A., (2009). Generations of sequencing technologies. Genomics 93, 105-111.
Picozzi, K., Fèvre, E.M., Odiit, M., Carrington, M., Eisler, M.C., Maudlin, I., Welburn, S.C., (2005). Sleeping sickness in Uganda: a thin line between two fatal diseases. BMJ 331, 1238-1241.
Pinchbeck, G.L., Morrison, L.J., Tait, A., Langford, J., Meehan, L., Jallow, S., Jallow, J., Jallow, A., Christley, R.M., (2008). Trypanosomosis in The Gambia: prevalence in working horses and donkeys detected by whole genome amplification and PCR, and evidence for interactions between trypanosome species. BMC Vet. Res 4, 7.
Piontkivska, H., Hughes, A.L., (2005). Environmental kinetoplastid-like 18S rRNA sequences and phylogenetic relationships among Trypanosomatidae: paraphyly of the genus Trypanosoma. Mol. Biochem. Parasitol 144, 94-99.
Pires, S.F., DaRocha, W.D., Freitas, J.M., Oliveira, L.A., Kitten, G.T., Machado, C.R., Pena, S.D.J., Chiari, E., Macedo, A.M., Teixeira, S.M.R., (2008). Cell culture and animal infection with distinct *Trypanosoma cruzi* strains expressing red and green fluorescent proteins. Int. J. Parasitol 38, 289-297.
Van der Ploeg, L.H., Valerio, D., De Lange, T., Bernards, A., Borst, P., Grosveld, F.G., (1982). An analysis of cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant surface glycoproteins are clustered in the genome. Nucleic Acids Res 10, 5905-5923.
Van der Ploeg, L.H., Cornelissen, A.W., Barry, J.D., Borst, P., (1984). Chromosomes of kinetoplastida.EMBO J 3, 3109-3115.

Pouchkina-Stantcheva, N.N., McGee, B.M., Boschetti, C., Tolleter, D., Chakrabortee, S., Popova, A.V., Meersman, F., Macherel, D., Hincha, D.K., Tunnacliffe, A., (2007).Functional divergence of former alleles in an ancient asexual invertebrate. Science 318, 268-271.

Priotto, G., Kasparian, S., Ngouama, D., Ghorashian, S., Arnold, U., Ghabri, S., Karunakara, U., (2007). Nifurtimox-eflornithine combination therapy for second-stage *Trypanosoma brucei gambiense* sleeping sickness: a randomized clinical trial in Congo. Clin. Infect. Dis 45, 1435-1442.

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

Pumpaibool, T., Arnathau, C., Durand, P., Kanchanakhan, N., Siripoon, N., Suegorn, A., Sitthi-Amorn, C., Renaud, F., Harnyuttanakorn, P., (2009). Genetic diversity and population structure of *Plasmodium falciparum* in Thailand, a low transmission country. Malar.J 8, 155.

Radwanska, M., Claes, F., Magez, S., Magnus, E., Perez-Morga, D., Pays, E., Büscher, P., (2002). Novel primer sequences for polymerase chain reaction-based detection of *Trypanosoma brucei gambiense*. Am. J. Trop. Med. Hyg 67, 289-295.

Radwanska, M., Guirnalda, P., De Trez, C., Ryffel, B., Black, S., Magez, S., (2008). Trypanosomiasis-induced B cell apoptosis results in loss of protective anti-parasite antibody responses and abolishment of vaccine-induced memory responses. PLoS Pathog 4, e1000078.

Ragoussis, J., (2009). Genotyping technologies for genetic research. Annu Rev Genomics Hum Genet 10, 117-133.

Ramesh, M.A., Malik, S.-B., Logsdon, J.M., (2005).A phylogenomic inventory of meiotic genes; evidence for sex in Giardia and an early eukaryotic origin of meiosis.Curr. Biol 15, 185-191.

Rana, N.A., Ebenezer, N.D., Webster, A.R., Linares, A.R., Whitehouse, D.B., Povey, S., Hardcastle, A.J., (2004). Recombination hotspots and block structure of linkage disequilibrium in the human genome exemplified by detailed analysis of PGM1 on 1p31. Hum. Mol. Genet 13, 3089-3102.

Raper, J., Fung, R., Ghiso, J., Nussenzweig, V., Tomlinson, S., (1999). Characterization of a novel trypanosome lytic factor from human serum. Infect. Immun 67, 1910-1916.

Ravel, C., Cortes, S., Pratlong, F., Morio, F., Dedet, J.-P., Campino, L., (2006). First report of genetic hybrids between two very divergent Leishmania species: *Leishmania infantum* and *Leishmania major*. Int. J. Parasitol 36, 1383-1388.

Razakandrainibe, F.G., Durand, P., Koella, J.C., De Meeüs, T., Rousset, F., Ayala, F.J., Renaud, F., (2005).“Clonal” population structure of the malaria agent *Plasmodium falciparum* in high-infection regions. Proc. Natl. Acad. Sci. U.S.A 102, 17388-17393.

Read, A.F., Day, K.P., (1992). The genetic structure of malaria parasite populations.Parasitol.Today (Regul. Ed.) 8, 239-242.

Rennie, C., Hulme, H., Fisher, P., Halp, L., Agaba, M., Noyes, H.A., Kemp, S.J., Brass, A., (2008).A systematic, data-driven approach to the combined analysis of microarray and QTL data. Dev Biol (Basel) 132, 293-299.

Ricci, C., Melone, G., Santo, N., Caprioli, M., (2003).Morphological response of a *bdelloid rotifer* to desiccation. J. Morphol 257, 246-253.

Rice, W.R., Friberg, U., (2007). Genomic clues to an ancient asexual scandal.Genome Biol 8, 232.

Rich, S.M., Licht, M.C., Hudson, R.R., Ayala, F.J., (1998). Malaria’s Eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. U.S.A 95, 4425-4430.

Richner, D., Schweizer, J., Betschart, B., Jenni, L., (1989). Characterization of West African *Trypanosoma (Trypanozoon) brucei* isolates from man and animals using isoenzyme analysis and DNA hybridization. Parasitol. Res 76, 80-85.

Rickman, L.R., Robson, J., (1970a). The blood incubation infectivity test: a simple test which may serve to distinguish *Trypanosoma brucei* from *T. rhodesiense*. Bull. World Health Organ 42, 650-651.

Rickman, L.R., Robson, J., (1970b). The testing of proven *Trypanosoma brucei* and *T. rhodesiense* strains by the blood incubation infectivity test. Bull. World Health Organ 42, 911-916.

Rifkin, M.R., (1978). Identification of the trypanocidal factor in normal human serum: high density lipoprotein. Proc. Natl. Acad. Sci. U.S.A 75, 3450-3454.

Robays, J., Nyamowala, G., Sese, C., Betu Ku Mesu Kande, V., Lutumba, P., Van der Veken, W., Boelaert, M., (2008). High failure rates of melarsoprol for sleeping sickness, Democratic Republic of Congo. Emerging Infect. Dis 14, 966-967.

Roberts, L.W., Wellde, B.T., Reardon, M.J., Onyango, F.K., (1989). Mechanical transmission of *Trypanosoma brucei rhodesiense* by Glossina morsitans morsitans (Diptera:Glossinidae). Ann Trop Med Parasitol 83 Suppl 1, 127-131.

Rodgers, J., (2010). Trypanosomiasis and the brain. Parasitology 137, 1995-2006.

Rodrigues, A.C., Neves, L., Garcia, H.A., Viola, L.B., Marcili, A., Da Silva, F.M., Sigauque, I., Batista, J.S., Paiva, F., Teixeira, M.M.G., (2008). Phylogenetic analysis of *Trypanosoma vivax* supports the separation of South American/West African from East African isolates and a new *T. vivax*-like genotype infecting a nyala antelope from Mozambique. Parasitology 135, 1317-1328.

Rogers, M.E., Bates, P.A., (2007). Leishmania manipulation of sand fly feeding behavior results in enhanced transmission. PLoS Pathog 3, e91.
Rougeron, V., De Meeûs, T., Hide, M., Waleckx, E., Bermudez, H., Arevalo, J., Llanos-Cuentas, A., Dujardin, J.-C., De Doncker, S., Le Ray, D., Ayala, F.J., Bañuls, A.-L., (2009). Extreme inbreeding in *Leishmania braziliensis*. Proc. Natl. Acad. Sci. U.S.A 106, 10224-10229.
Rougeron, V., Bañuls, A.-L., Carme, B., Simon, S., Couppié, P., Nacher, M., Hide, M., DE Meeûs, T., (2011). Reproductive strategies and population structure in *Leishmania*: substantial amount of sex in *Leishmania Viannia guyanensis*. Mol Ecol 20, 3116-3127.
Rowlands, G.J., Mulatu, W., Authié, E., d' Ieteren, G.D., Leak, S.G., Nagda, S.M., Peregrine, A.S., (1993).Epidemiology of bovine trypanosomiasis in the Ghibe valley, southwest Ethiopia. 2. Factors associated with variations in trypanosome prevalence, incidence of new infections and prevalence of recurrent infections. Acta Trop 53, 135-150.
Roy Chowdhury, A., Bakshi, R., Wang, J., Yildirir, G., Liu, B., Pappas-Brown, V., Tolun, G., Griffith, J.D., Shapiro, T.A., Jensen, R.E., Englund, P.T., (2010). The killing of African trypanosomes by ethidium bromide. PLoS Pathog 6, e1001226.
Sanchez, G., Wallace, A., Muñoz, S., Venegas, J., Ortiz, S., Solari, A., (1993). Characterization of *Trypanosoma cruzi* populations by several molecular markers supports a clonal mode of reproduction. Biol. Res 26, 167-176.
Schultz, L., Wapling, J., Mueller, I., Ntsuke, P.O., Senn, N., Nale, J., Kiniboro, B., Buckee, C.O., Tavul, L., Siba, P.M., Reeder, J.C., Barry, A.E., (2010). Multilocus haplotypes reveal variable levels of diversity and population structure of *Plasmodium falciparum* in Papua New Guinea, a region of intense perennial transmission. Malar.J 9, 336.
Schurko, A.M., Logsdon, J.M., (2008). Using a meiosis detection toolkit to investigate ancient asexual "scandals" and the evolution of sex. Bioessays 30, 579-589.
Schweizer, J., Pospichal, H., Hide, G., Buchanan, N., Tait, A., Jenni, L., (1994). Analysis of a new genetic cross between two East African Trypanosoma brucei clones. Parasitology 109 (Pt 1), 83-93.
Schwenkenbecher, J.M., Wirth, T., Schnur, L.F., Jaffe, C.L., Schallig, H., Al-Jawabreh, A., Hamarsheh, O., Azmi, K., Pratlong, F., Schönian, G., (2006). Microsatellite analysis reveals genetic structure of *Leishmania tropica*. Int. J. Parasitol 36, 237-246.
Seidl, A., Dávila, A.M., Silva, R.A., (1999).Estimated financial impact of *Trypanosoma vivax* on the Brazilian pantanal and Bolivian lowlands.Mem. Inst. Oswaldo Cruz 94, 269-272.
Seridi, N., Amro, A., Kuhls, K., Belkaid, M., Zidane, C., Al-Jawabreh, A., Schönian, G., (2008).Genetic polymorphism of Algerian *Leishmania infantum* strains revealed by multilocus microsatellite analysis. Microbes Infect 10, 1309-1315.
Sharma, R., Gluenz, E., Peacock, L., Gibson, W., Gull, K., Carrington, M., (2009).The heart of darkness: growth and form of Trypanosoma brucei in the tsetse fly. Trends Parasitol 25, 517-524.
Shendure, J., Ji, H., (2008). Next-generation DNA sequencing. Nat. Biotechnol 26, 1135-1145.
Siepel, A., (2009). Phylogenomics of primates and their ancestral populations. Genome Res 19, 1929-1941.
Da Silva, A.S., Garcia Perez, H.A., Costa, M.M., França, R.T., De Gasperi, D., Zanette, R.A., Amado, J.A., Lopes, S.T.A., Teixeira, M.M.G., Monteiro, S.G., (2011). Horses naturally infected by *Trypanosoma vivax* in southern Brazil. Parasitol. Res 108, 23-30.
Simarro, P.P., Jannin, J., Cattand, P., (2008). Eliminating human African trypanosomiasis: where do we stand and what comes next? PLoS Med 5, e55.
Simarro, P.P., Diarra, A., Ruiz Postigo, J.A., Franco, J.R., Jannin, J.G., (2011). The Human African Trypanosomiasis Control and Surveillance Programme of the World Health Organization 2000–2009: The Way Forward. PLoS Negl Trop Dis 5, e1007.
Simo, G., Njiokou, F., Mbida Mbida, J.A., Njitchouang, G.R., Herder, S., Asonganyi, T., Cuny, G., (2008). Tsetse fly host preference from sleeping sickness foci in Cameroon: epidemiological implications. Infect. Genet. Evol 8, 34-39.
Simo, G., Njitchouang, G.R., Njiokou, F., Cuny, G., Asonganyi, T., (2011).*Trypanosoma brucei s.l.*: Microsatellite markers revealed high level of multiple genotypes in the mid-guts of wild tsetse flies of the Fontem sleeping sickness focus of Cameroon. Exp. Parasitol 128, 272-278.
Simpson, A.G.B., Stevens, J.R., Lukes, J., (2006). The evolution and diversity of kinetoplastid flagellates. Trends Parasitol 22, 168-174.
Simukoko, H., Marcotty, T., Phiri, I., Geysen, D., Vercruysse, J., Van den Bossche, P., (2007a).The comparative role of cattle, goats and pigs in the epidemiology of livestock trypanosomiasis on the plateau of eastern Zambia. Vet. Parasitol 147, 231-238.
Simukoko, H., Marcotty, T., Phiri, I., Vercruysse, J., Van den Bossche, P., (2007b).Heterogeneity in the trypanosomosis incidence in Zebu cattle of different ages and sex on the plateau of eastern Zambia. Acta Trop 103, 98-101.
Sinshaw, A., Abebe, G., Desquesnes, M., Yoni, W., (2006). Biting flies and *Trypanosoma vivax* infection in three highland districts bordering lake Tana, Ethiopia. Vet. Parasitol 142, 35-46.

Sloof, P., Bos, J.L., Konings, A.F., Menke, H.H., Borst, P., Gutteridge, W.E., Leon, W., (1983a). Characterization of satellite DNA in *Trypanosoma brucei* and *Trypanosoma cruzi*. J. Mol. Biol 167, 1-21.
Sloof, P., Menke, H.H., Caspers, M.P., Borst, P., (1983b). Size fractionation of *Trypanosoma brucei* DNA: localization of the 177-bp repeat satellite DNA and a variant surface glycoprotein gene in a mini-chromosomal DNA fraction. Nucleic Acids Res 11, 3889-3901.
Snow, W.F., Wacher, T.J., Rawlings, P., (1996). Observations on the prevalence of trypanosomosis in small ruminants, equines and cattle, in relation to tsetse challenge, in The Gambia. Vet. Parasitol 66, 1-11.
Solano, P., Ravel, S., de Meeûs, T., (2010). How can tsetse population genetics contribute to African trypanosomiasis control? Trends Parasitol 26, 255-263.
Souto, R.P., Zingales, B., (1993). Sensitive detection and strain classification of *Trypanosoma cruzi* by amplification of a ribosomal RNA sequence. Mol. Biochem. Parasitol 62, 45-52.
Souto, R.P., Fernandes, O., Macedo, A.M., Campbell, D.A., Zingales, B., (1996). DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. Mol. Biochem. Parasitol 83, 141-152.
Stenberg, P., Lundmark, M., Saura, A., (2003). MLGsim: A program for detecting clones using a simulation approach. Molecular Ecology Notes 3, 329 - 331.
Sterkers, Y., Lachaud, L., Crobu, L., Bastien, P., Pagès, M., (2011). FISH analysis reveals aneuploidy and continual generation of chromosomal mosaicism in *Leishmania major*. Cell. Microbiol 13, 274-283.
Sternberg, J., Tait, A., Haley, S., Wells, J.M., Le Page, R.W., Schweizer, J., Jenni, L., (1988). Gene exchange in African trypanosomes: characterisation of a new hybrid genotype. Mol. Biochem. Parasitol 27, 191-200.
Sternberg, J., Turner, C.M., Wells, J.M., Ranford-Cartwright, L.C., Le Page, R.W., Tait, A., (1989). Gene exchange in African trypanosomes: frequency and allelic segregation. Mol. Biochem. Parasitol 34, 269-279.
Sternberg, J., Tait, A., (1990). Genetic exchange in African trypanosomes. Trends Genet 6, 317-322.
Sternberg, J.M., Maclean, L., (2010). A spectrum of disease in human African trypanosomiasis: the host and parasite genetics of virulence. Parasitology 137, 2007-2015.
Stevens, J.R., Welburn, S.C., (1993). Genetic processes within an epidemic of sleeping sickness in Uganda. Parasitol. Res 79, 421-427.
Stevens, J.R., Tibayrenc, M., (1995). Detection of linkage disequilibrium in *Trypanosoma brucei* isolated from tsetse flies and characterized by RAPD analysis and isoenzymes. Parasitology 110 (Pt 2), 181-186.
Stevens, J.R., Tibayrenc, M., (1996).*Trypanosoma brucei s.l*: evolution, linkage and the clonality debate. Parasitology 112 (Pt 5), 481-488.
Stevens, J.R., Gibson, W., (1999).The molecular evolution of trypanosomes.Parasitol.Today (Regul. Ed.) 15, 432-437.
Stevens, J.R., Noyes, H.A., Dover, G.A., Gibson, W.C., (1999). The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. Parasitology 118 (Pt 1), 107-116.
Stijlemans, B., Vankrunkelsven, A., Caljon, G., Bockstal, V., Guilliams, M., Bosschaerts, T., Beschin, A., Raes, G., Magez, S., De Baetselier, P., (2010). The central role of macrophages in trypanosomiasis-associated anemia: rationale for therapeutical approaches. Endocr Metab Immune Disord Drug Targets 10, 71-82.
Stram, D.O., (2004). Tag SNP selection for association studies. Genet. Epidemiol 27, 365-374.
Sturm, N.R., Campbell, D.A., (2010). Alternative lifestyles: the population structure of *Trypanosoma cruzi*. Acta Trop 115, 35-43.
Sumba, A.L., Mihok, S., Oyieke, F.A., (1998). Mechanical transmission of *Trypanosoma evansi* and *T. congolense* by *Stomoxys niger* and *S. taeniatus* in a laboratory mouse model. Med. Vet. Entomol 12, 417-422.
Sun, C., Southard, C., Olopade, O.I., Di Rienzo, A., (2011). Differential allelic expression of c.1568C>A at UGT2B15 is due to variation in a novel cis-regulatory element in the 3'UTR. Gene 481, 24-28.
Tait, A., (1980). Evidence for diploidy and mating in trypanosomes. Nature 287, 536-538.
Tait, A., Babiker, E.A., Le Ray, D., (1984). Enzyme variation in *Trypanosoma brucei spp*. I. Evidence for the sub-speciation of *Trypanosoma brucei gambiense*. Parasitology 89 (Pt 2), 311-326.
Tait, A., Barry, J.D., Wink, R., Sanderson, A., Crowe, J.S., (1985). Enzyme variation in *T. brucei ssp*. II.Evidence for *T. b. rhodesiense* being a set of variants of *T. b. brucei*. Parasitology 90 (Pt 1), 89-100.
Tait, A., Buchanan, N., Hide, G., Turner, C.M., (1996). Self-fertilisation in *Trypanosoma brucei*. Mol. Biochem. Parasitol 76, 31-42.
Tait, A., Masiga, D., Ouma, J., MacLeod, A., Sasse, J., Melville, S., Lindegard, G., McIntosh, A., Turner, M., (2002). Genetic analysis of phenotype in *Trypanosoma brucei*: a classical approach to potentially complex traits. Philos. Trans. R. Soc. Lond., B, Biol. Sci 357, 89-99.

Tait, A., Macleod, A., Tweedie, A., Masiga, D., Turner, C.M.R., (2007). Genetic exchange in *Trypanosoma brucei*: evidence for mating prior to metacyclic stage development. Mol. Biochem. Parasitol 151, 133-136.

Talman, A.M., Domarle, O., McKenzie, F.E., Ariey, F., Robert, V., (2004). Gametocytogenesis: the puberty of *Plasmodium falciparum*. Malar.J 3, 24.

Targett, G.A., Wilson, V.C., (1973). The blood incubation infectivity test as a means of distinguishing between *Trypanosoma brucei brucei* and *T. brucei rhodesiense*. Int. J. Parasitol 3, 5-11.

Taylor, J.E., Rudenko, G., (2006). Switching trypanosome coats: what's in the wardrobe? Trends Genet 22, 614-620.

Teka, I.A., Kazibwe, A.J.N., El-Sabbagh, N., Al-Salabi, M.I., Ward, C.P., Eze, A.A., Munday, J.C., Mäser, P., Matovu, E., Barrett, M.P., de Koning, H.P., (2011). The diamidine diminazene aceturate is a substrate for the high-affinity pentamidine transporter: implications for the development of high resistance levels in trypanosomes. Mol. Pharmacol 80, 110-116.

Thekisoe, O.M.M., Omolo, J.D., Swai, E.S., Hayashida, K., Zhang, J., Sugimoto, C., Inoue, N., (2007a). Preliminary application and evaluation of loop-mediated isothermal amplification (LAMP) for detection of bovine theileriosis and trypanosomosis in Tanzania. Onderstepoort J. Vet. Res 74, 339-342.

Thekisoe, O.M.M., Kuboki, N., Nambota, A., Fujisaki, K., Sugimoto, C., Igarashi, I., Yasuda, J., Inoue, N., (2007b). Species-specific loop-mediated isothermal amplification (LAMP) for diagnosis of trypanosomosis. Acta Trop 102, 182-189.

Thekisoe, O.M.M., Bazie, R.S.B., Coronel-Servian, A.M., Sugimoto, C., Kawazu, S.-I., Inoue, N., (2009).Stability of Loop-Mediated Isothermal Amplification (LAMP) reagents and its amplification efficiency on crude trypanosome DNA templates. J. Vet. Med. Sci 71, 471-475.

Tibayrenc, M., Ward, P., Moya, A., Ayala, F.J., (1986). Natural populations of *Trypanosoma cruzi*, the agent of Chagas disease, have a complex multiclonal structure. Proc. Natl. Acad. Sci. U.S.A 83, 115-119.

Tibayrenc, M., Ayala, F.J., (1987).*Trypanosoma cruzi* populations: more clonal than sexual. Parasitol.Today (Regul. Ed.) 3, 189-190.

Tibayrenc, M., Kjellberg, F., Ayala, F.J., (1990). A clonal theory of parasitic protozoa: the population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. Proc. Natl. Acad. Sci. U.S.A 87, 2414-2418.

Tibayrenc, M., Kjellberg, F., Arnaud, J., Oury, B., Brenière, S.F., Dardé, M.L., Ayala, F.J., (1991). Are eukaryotic microorganisms clonal or sexual? A population genetics vantage.Proc. Natl. Acad. Sci. U.S.A 88, 5129-5133.

Tibayrenc, M., (1995).Population genetics of parasitic protozoa and other microorganisms. Adv. Parasitol 36, 47-115.

Tomlinson, S., Muranjan, M., Nussenzweig, V., Raper, J., (1997). Haptoglobin-related protein and apolipoprotein AI are components of the two trypanolytic factors in human serum. Mol. Biochem. Parasitol 86, 117-120.

Traoré-Leroux, T., Fumoux, F., Pinder, M., (1987). High density lipoprotein levels in the serum of trypanosensitive and trypanoresistant cattle. Changes during *Trypanosoma congolense* infection. Acta Trop 44, 315-323.

Truc, P., Bailey, J.W., Doua, F., Laveissière, C., Godfrey, D.G., (1994). A comparison of parasitological methods for the diagnosis of gambian trypanosomiasis in an area of low endemicity in Côte d'Ivoire. Trans. R. Soc. Trop. Med. Hyg 88, 419-421.

Turner, C.M., Barry, J.D., Maudlin, I., Vickerman, K., (1988). An estimate of the size of the metacyclic variable antigen repertoire of *Trypanosoma brucei rhodesiense*. Parasitology 97 (Pt 2), 269-276.

Turner, C.M., Sternberg, J., Buchanan, N., Smith, E., Hide, G., Tait, A., (1990). Evidence that the mechanism of gene exchange in *Trypanosoma brucei* involves meiosis and syngamy. Parasitology 101 Pt 3, 377-386.

Turner, C.M.R., McLellan, S., Lindergard, L.A.G., Bisoni, L., Tait, A., MacLeod, A., (2004). Human infectivity trait in *Trypanosoma brucei*: stability, heritability and relationship to sra expression. Parasitology 129, 445-454.

Ubeda, J.-M., Légaré, D., Raymond, F., Ouameur, A.A., Boisvert, S., Rigault, P., Corbeil, J., Tremblay, M.J., Olivier, M., Papadopoulou, B., Ouellette, M., (2008). Modulation of gene expression in drug resistant *Leishmania* is associated with gene amplification, gene deletion and chromosome aneuploidy. Genome Biol 9, R115.

Urdaneta, L., Lal, A., Barnabe, C., Oury, B., Goldman, I., Ayala, F.J., Tibayrenc, M., (2001).Evidence for clonal propagation in natural isolates of *Plasmodium falciparum* from Venezuela.Proc. Natl. Acad. Sci. U.S.A 98, 6725-6729.

Urwyler, S., Vassella, E., Van Den Abbeele, J., Renggli, C.K., Blundell, P., Barry, J.D., Roditi, I., (2005). Expression of procyclin mRNAs during cyclical transmission of *Trypanosoma brucei*. PLoS Pathog 1, e22.

Vanhollebeke, B., De Muylder, G., Nielsen, M.J., Pays, A., Tebabi, P., Dieu, M., Raes, M., Moestrup, S.K., Pays, E., (2008). A haptoglobin-hemoglobin receptor conveys innate immunity to *Trypanosoma brucei* in humans. Science 320, 677-681.
Vanhollebeke, B., Pays, E., (2010). The trypanolytic factor of human serum: many ways to enter the parasite, a single way to kill. Mol. Microbiol 76, 806-814.
Vickerman, K., (1969). On the surface coat and flagellar adhesion in trypanosomes. J. Cell. Sci 5, 163-193.
Vickerman, K., Luckins, A.G., (1969). Localization of variable antigens in the surface coat of *Trypanosoma brucei* using ferritin conjugated antibody. Nature 224, 1125-1126.
Vickerman, K., Preston, T.M., (1970). Spindle microtubules in the dividing nuclei of trypanosomes. J. Cell. Sci 6, 365-383.
Vickerman, K., (1985). Developmental cycles and biology of pathogenic trypanosomes. Br. Med. Bull 41, 105-114.
Vickerman, K., Tetley, L., Hendry, K.A., Turner, C.M., (1988). Biology of African trypanosomes in the tsetse fly. Biol. Cell 64, 109-119.
Vincent, I.M., Creek, D., Watson, D.G., Kamleh, M.A., Woods, D.J., Wong, P.E., Burchmore, R.J.S., Barrett, M.P., (2010). A molecular mechanism for eflornithine resistance in African trypanosomes. PLoS Pathog 6, e1001204.
Wahlund, S., (1928).Zusammensetzung von populationen und korrelationsers-chinungen von standpunkt der vererbungslehre aus betrachtet. Hereditas 65-106.
Wainwright, M., (2010). Dyes, trypanosomiasis and DNA: a historical and critical review. Biotech Histochem 85, 341-354.
Waiswa, C., Olaho-Mukani, W., Katunguka-Rwakishaya, E., (2003).Domestic animals as reservoirs for sleeping sickness in three endemic foci in south-eastern Uganda. Ann Trop Med Parasitol 97, 149-155.
Waiswa, C., Katunguka-Rwakishaya, E., (2004). Bovine trypanosomiasis in south-western Uganda: packed-cell volumes and prevalences of infection in the cattle. Ann Trop Med Parasitol 98, 21-27.
Walliker, D., (1991). Malaria parasites: Randomly interbreeding or "clonal" populations? Parasitol.Today (Regul. Ed.) 7, 232-235.
Wardrop, N.A., Atkinson, P.M., Gething, P.W., Fèvre, E.M., Picozzi, K., Kakembo, A.S.L., Welburn, S.C., (2010). Bayesian geostatistical analysis and prediction of Rhodesian human African trypanosomiasis. PLoS Negl Trop Dis 4, e914.
Wastling, S.L., Picozzi, K., Wamboga, C., VON Wissmann, B., Amongi-Accup, C., Wardrop, N.A., Stothard, J.R., Kakembo, A., Welburn, S.C., (2011). Latent *Trypanosoma brucei gambiense* foci in Uganda: a silent epidemic in children and adults? Parasitology 1-8.
Wastling, S.L., Welburn, S.C., (2011). Diagnosis of human sleeping sickness: sense and sensitivity. Trends Parasitol.
Welburn, S.C., Picozzi, K., Fèvre, E.M., Coleman, P.G., Odiit, M., Carrington, M., 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.
Welburn, S.C., Maudlin, I., Simarro, P.P., (2009).Controlling sleeping sickness - a review. Parasitology 136, 1943-1949.
Wellde, B.T., Reardon, M.J., Chumo, D.A., Kovatch, R.M., Waema, D., Wykoff, D.E., Mwangi, J., Boyce, W.L., Williams, J.S., (1989). Cerebral trypanosomiasis in naturally-infected cattle in the Lambwe Valley, south Nyanza, Kenya. Ann Trop Med Parasitol 83 Suppl 1, 151-160.
Wells, E.A., (1972). The importance of mechanical transmission in the epidemiology of nagana: a review. Trop Anim Health Prod 4, 74-88.
Wells, J.M., Prospero, T.D., Jenni, L., Le Page, R.W., (1987). DNA contents and molecular karyotypes of hybrid *Trypanosoma brucei*. Mol. Biochem. Parasitol 24, 103-116.
West, S.A., Reece, S.E., Read, A.F., (2001). Evolution of gametocyte sex ratios in malaria and related apicomplexan (protozoan) parasites. Trends Parasitol 17, 525-531.
WHO, (1998).Control and surveillance of African trypanosomiasis. Report of a WHO expert committee (No. 881), WHO Technical Report Series. World Health Organisation, Geneva.
WHO, (2006). Weekly epidemiological record (No. 8), Weekly Epidemiological Record. World Health Organisation, Geneva.
WHO, (2009).The Gambia Country Brief.World Health Organisation.
Wickstead, B., Ersfeld, K., Gull, K., (2004). The small chromosomes of *Trypanosoma brucei* involved in antigenic variation are constructed around repetitive palindromes. Genome Res 14, 1014-1024.
Widener, J., Nielsen, M.J., Shiflett, A., Moestrup, S.K., Hajduk, S., (2007). Hemoglobin is a co-factor of human trypanosome lytic factor. PLoS Pathog 3, 1250-1261.
Wilkinson, S.R., Kelly, J.M., (2009). Trypanocidal drugs: mechanisms, resistance and new targets. Expert Rev Mol Med 11, e31.

von Wissmann, B., Machila, N., Picozzi, K., Fèvre, E.M., deC Bronsvoort, B.M., Handel, I.G., Welburn, S.C., (2011). Factors associated with acquisition of human infective and animal infective trypanosome infections in domestic livestock in Western Kenya. PLoS Negl Trop Dis 5, e941.

Wootton, J.C., Feng, X., Ferdig, M.T., Cooper, R.A., Mu, J., Baruch, D.I., Magill, A.J., Su, X.-Z., (2002). Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. Nature 418, 320-323.

Xong, H.V., Vanhamme, L., Chamekh, M., Chimfwembe, C.E., Van Den Abbeele, J., Pays, A., Van Meirvenne, N., Hamers, R., De Baetselier, P., Pays, E., (1998). A VSG expression site-associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. Cell 95, 839-846.

Xu, Z., Kaplan, N.L., Taylor, J.A., (2007). Tag SNP selection for candidate gene association studies using HapMap and gene resequencing data. Eur. J. Hum. Genet 15, 1063-1070.

Yanowitz, J., (2010). Meiosis: making a break for it. Curr.Opin. Cell Biol 22, 744-751.

Young, C.J., Godfrey, D.G., (1983). Enzyme polymorphism and the distribution of *Trypanosoma congolense* isolates. Ann Trop Med Parasitol 77, 467-481.

Yun, O., Priotto, G., Tong, J., Flevaud, L., Chappuis, F., (2010). NECT is next: implementing the new drug combination therapy for *Trypanosoma brucei gambiense* sleeping sickness. PLoS Negl Trop Dis 4, e720.

Zhang, Q., Tibayrenc, M., Ayala, F.J., (1988).Linkage disequilibrium in natural populations of *Trypanosoma cruzi* (flagellate), the agent of Chagas' disease. J. Protozool 35, 81-85.

Zhou, X., Ren, L., Li, Y., Zhang, M., Yu, Y., Yu, J., (2010). The next-generation sequencing technology: a technology review and future perspective. Sci China Life Sci 53, 44-57.

Zingales, B., Andrade, S.G., Briones, M.R.S., Campbell, D.A., Chiari, E., Fernandes, O., Guhl, F., Lages-Silva, E., Macedo, A.M., Machado, C.R., Miles, M.A., Romanha, A.J., Sturm, N.R., Tibayrenc, M., Schijman, A.G., (2009). A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. Mem. Inst. Oswaldo Cruz 104, 1051-1054.