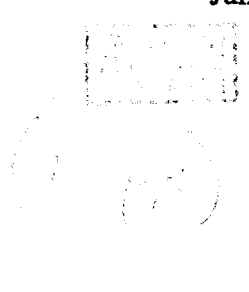
Genetic Analysis of *Trypanosoma brucei*

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Glasgow

by

Annette MacLeod
Wellcome Centre of Molecular Parasitology
University of Glasgow

June 1999



To my parents

Preface

Although the work contained in this thesis covers the same general area, the research envelops several distinct aspects within the overall theme. As such I have chosen to weight the bulk of the thesis into the central results sections. The introductory chapter (Chapter 1) is a brief introduction to the African trypanosome, *Trypanosoma brucei*, describing its sexual reproduction and population structure and presenting current typing systems for genome analysis, in particular the role of minisatellite loci as useful genetic markers. The first two results chapters (Chapters 3 and 4) describe the development of the tools for genome analysis, and the subsequent chapters contain the results of this analysis from laboratory crosses (Chapter 5) and field samples (Chapters 6-8). The discussion chapter (Chapter 9) attempts to draw general conclusions from these results and point to the way forward for future research.

Abstract

Many minisatellite loci show extremely high levels of variability in the length of their tandemly repeated sequences, due to a high rate of spontaneous mutation to new length alleles. Because of this hypervariability, minisatellites have been used extensively in many areas of biology, from forensic medicine to paternity testing. However such loci have been rarely studied in parasites, yet they can be of particular value in strain identification and the detection of recombination. In this study, three minisatellite loci, which vary extensively in both repeat copy number and sequence differences between repeat units, were identified and characterized from *Trypanosoma brucei*. These minisatellite markers along with four microsatellite loci, were chosen because they were all relatively small and so could be faithfully amplified by PCR from small quantities of DNA and, by using a nested PCR approach, from single trypanosomes. These markers were then used to investigate the nature and extent of sexual recombination in this parasite, using both laboratory crosses and population analysis.

Analysis of trypanosomes derived from laboratory crosses showed that minisatellite inheritance is in agreement with a Mendelian system and that such markers are particularly useful for the detection of cross and self-fertilization. Examination of the F1 hybrids from these crosses has identified some hybrids as being trisomic, but that, contrary to previous reports, triploidy is rare. The rate of recombination between homologous chromosomes was also examined and used to estimate the physical distance per centiMorgan (4.9-25kb/cM).

Although sexual recombination has been demonstrated to occur in laboratory experiments the extent of genetic exchange in natural populations remained to be elucidated. Analysis of a series of field samples isolated from tsetse flies indicated that a high proportion of tsetse flies harboured mixed *T. brucei* infections, a prerequisite for genetic exchange to occur in the field.

Minisatellite variant repeat PCR (MVR-PCR) was employed, to map the interspersion patterns of variant repeat units within a minisatellite locus, a ternary code for a number of different alleles was generated and from this the underlying mechanisms of mutation for one minisatellite (*MS42*) were inferred. This method of allele mapping was applied to a collection of field samples to study the relationship between *T. b. brucei* and *T. b. rhodesiense* populations and the extent of sexual recombination in natural populations in each sub-species. The analysis revealed that there is considerable sub-structuring in *T. brucei* populations, due to geographical barriers and host specificities. *T. b. rhodesiense* populations are distinct from *T. b. brucei* and a *T. b. rhodesiense*-specific marker has been identified for the Busoga (Uganda) focus. *T. b. rhodesiense* appears to have originated from a subset of genotypes in the local *T. b. brucei* population and the ability to infect humans has arisen independently in at least two different foci, *i.e. T. b. rhodesiense* is not a monophyletic sub-species. *T. b. rhodesiense* has a different (possibly clonal) population structure from *T. b. brucei*, with the latter having a population structure in which mating is a common feature.

Acknowledgements

Without doubt my greatest thanks goes to Andy and Mike, my supervisors, for their guidance, support and encouragement throughout the course of this project. I would especially like to thank Andy for undertaking the mammoth task of cloning hybrids.

I would like to thank everyone with whom I have worked in WUMP both past and present, especially Mercy, Michael, Des, Michelle, Kathryn, Aoibhinn and Hubert for their help and companionship.

I acknowledge with gratitude the gifts of field isolates from Ian Maudlin, Sue Welburn and Geoff Hide; the plasmid containing the minisatellite *MS42* from Mike Barrett; the sequences of microsatellites, JS2 and D2, from Jurgen Sasse and computer programs from John Armour.

Last but not least I would like to express my gratitude to my family and friends for their love and support and to Darren for his advice, encouragement, patience and, most of all, his computer.

Contents

List of Figures

List of Tables

Abbreviations

bp	base pair
BIIT	Blood Incubation Infectivity Test
BSA	Bovine Serum Albumin
CATT	Card Agglutination Test
DFMO	Difluoromethylornithine
DNA	Deoxyribonucleic Acid
EATRO	East African Trypanosomiasis Research Organisation
EDTA	Ethylenediamine Tetra-Acetic Acid (disodium salt)
EST	Expressed Sequence Tag
ET	Electrophoretic type
HAT	Human African Trypanosomiasis
HSR	Human Serum Resistant
HSS	Human Serum Sensitive
Kb	Kilobase
M	Molar
min	Minute
ml	millilitre
mRNA	messenger RNA
MVR	Minisatellite Variant Repeat
MVR-PCR	MVR mapping by PCR
ng	nanogram
ORF	Open Reading Frame
PARP	Procyclic Acidic Repetitive Protein
PBS	Phosphate Buffered Saline
PBSG	Phosphate Buffered Saline containing 1% Glucose
PCR	Polymerase Chain Reaction
PFGE	Pulsed Field Gel Electrophoresis
RAPD	Randomly Amplified Polymorphic DNA
RNA	Ribonucleic Acid
RFLP	Restriction Fragment Length Polymorphism
SDS	Sodium Dodecyl Sulphate
SLP	Single Locus Probe
STIB	Swiss Tropical Institute Basel
TREU	Trypanosomiasis Research Edinburgh University
UV	Ultraviolet
VSG	Variant Surface Glycoprotein

Publications

Some of this work has been published:

Barrett, M.P., MacLeod, A., Tovar, J., Sweetman, J.P., Tait, A., Le Page, R.W.F. and Melville, S.E. (1997). A single locus minisatellite sequence which distinguishes between *Trypanosoma brucei* isolates. *Molecular and Biochemical Parasitology* **86**:95-99.

MacLeod, A., Turner, C.M.R. and Tait, A. (1997). Detection of single copy gene sequences from single trypanosomes. *Molecular and Biochemical Parasitology* **84**:267-270.

MacLeod, A., Turner, C.M.R. and Tait, A. (1999) A high level of mixed *Trypanosoma brucei* infections in tsetse flies detected by three hypervariable minisatellites. *Molecular and biochemical Parasitology* **102**:237-248.

Hope, M., MacLeod, A., Melville, S., Tait, A. and Turner, C.M.R. (*in press*). Maintenance of diploidy (in megabase chromosomes) in *Trypanosoma brucei* after genetic exchange.

Chapter 1

Introduction

Trypanosomes

The order Kinetoplastida, genus *Trypanosoma*, encompasses a large range of single-celled parasitic protozoa found throughout the world in both vertebrate and invertebrate hosts. Although most of these species replicate in the host without causing any adverse affects, there are a few species which are associated with disease, the most noteworthy of these being *T. cruzi*, the aetiological agent of Chagas' disease which infects humans in South America and *T. brucei*, the African trypanosome which is transmitted by the tsetse fly and causes African sleeping sickness or trypanosomiasis. There were 45,000 reported cases of African sleeping sickness in 1997, although the real infection rate could be nearer 450,000, most of which are fatal unless treated [Molyneux, 1997; Barrett, 1999]. Trypanosomes also have an adverse effect on domestic livestock in sub-Saharan Africa as *T. brucei*, *T. congolense*, and *T. vivax* infect cattle causing the disease, nagana, which is economically important resulting in wasting, infertility and lethality. Because of trypanosomes' ability to cause disease in both humans and livestock, the threat of trypanosome-infected tsetse flies has prevented the exploitation of fertile land in tsetse infested areas of sub-Saharan Africa [Vickerman, 1997].

Within the genus *Trypanosoma* there are 472 different named species which infect a range of different hosts and are transmitted by a number of different vectors. The mammal-infective species have been classified into 2 groups depending on the different modes of infection: the Stercoraria and the Salivaria. There are 94 species in the Stercoraria, including *T. cruzi*, which develops in the hind gut of the insect vector (triatomid bug) and is excreted in the faeces, contaminating the bite or wound. Trypanosomes belonging to the Salivaria are transmitted in the saliva of the tsetse fly or mechanically by any biting fly, when the insect takes a blood meal (with the exception of *T. equiperdum* which is transmitted by venereal contact). There are 9 different species within the group Salivaria, most of which are pathogenic to the mammalian host, and can be separated into 4 subgenera, *Duttonella, Nannomonas, Pycnomonas and Trypanozoon* (Figure 1.1). The different species within the Salivaria differ in the hosts they infect, the diseases they cause and their development in the insect vector. For example, *T. brucei* and *T. congolense* develop in the tsetse midgut before migration to the mouthparts, *T. vivax* develops in the mouthparts alone and *T. evansi* is mechanically transmitted by tsetse flies (and biting flies)

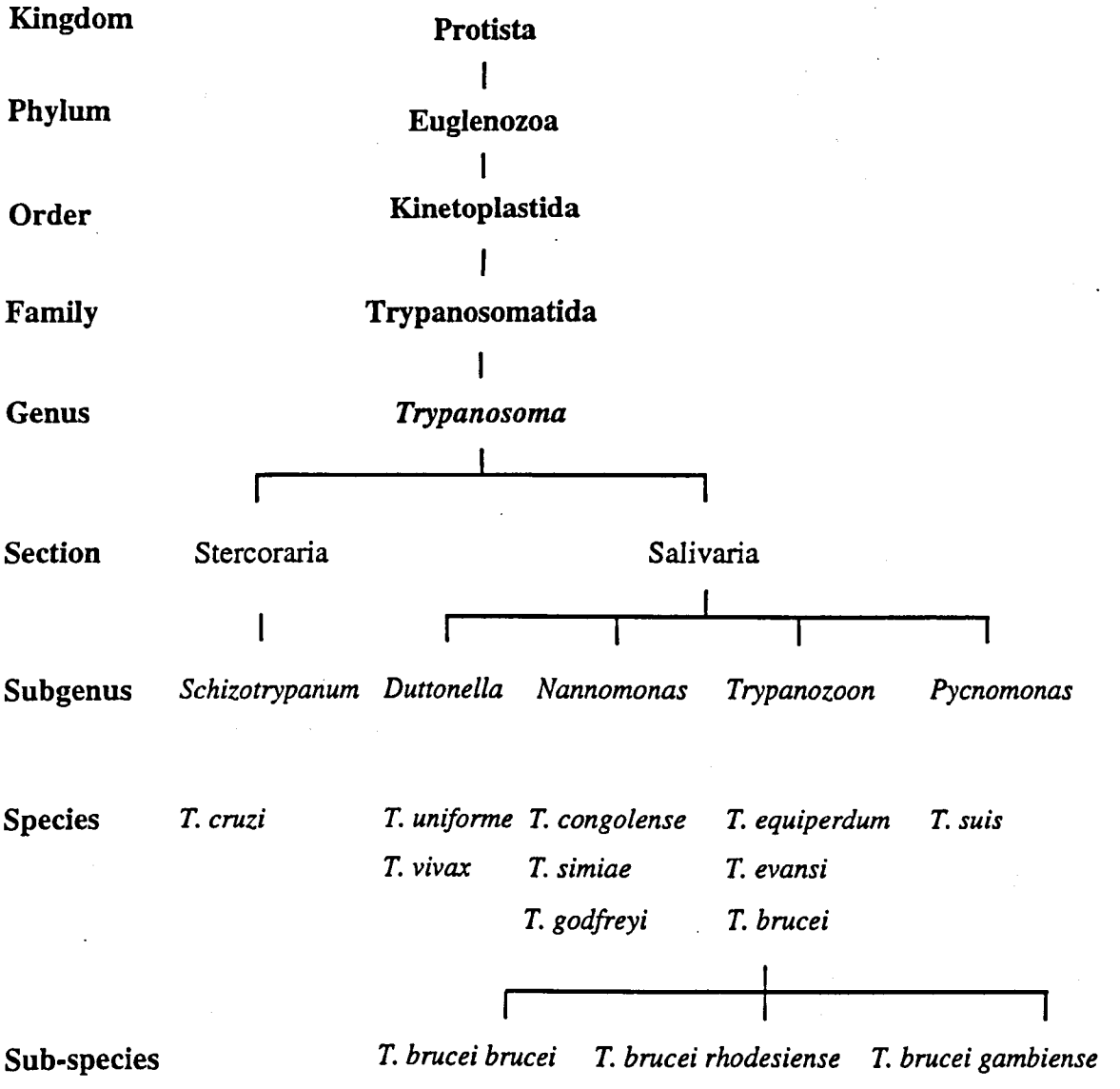


Figure 1.1. Classification of the salivarian trypanosomes.

with no cyclical development in the flies. Details of the different Salivarian species *i.e.* the diseases they cause, their hosts and vectors and their geographical distribution is summarised in Table 1.1 [adapted from Cox, 1993; McNamara *et al.*, 1994].

This study is concerned solely with the tsetse-transmitted *T. brucei* species of sub-Saharan Africa, which was named by Plimmer in honour of Bruce who, in 1895, discovered that trypanosomes were the aetiological agent for the disease (nagana) and were transmitted by tsetse flies [Vickerman, 1997]. The present method of classification identifies three morphologically identical subspecies of *T. brucei*: *T. brucei brucei, T. brucei rhodesiense* and *T. brucei gambiense* based on criteria of host specificities, geographical distribution and chronicity of disease [Hoare, 1972]. *T. b. brucei* is not infective to humans and is found throughout the tsetse region of Africa, *T. b. rhodesiense* is responsible for an acute disease in humans and is mainly found in East and Southern Africa, while *T. b. gambiense* causes a chronic sleeping sickness which is usually restricted to West and Central Africa.

Life Cycle of *T. brucei*

The life cycle of *T. brucei* is conducted in three different environments; the mammalian host, tsetse fly midgut and tsetse fly salivary glands. For each environment there are two life cycle stages, a proliferative stage which establishes an infection in that environment and a non-dividing form that is pre-adapted to the next environment (see Figure 1.2 adapted from Vickerman, (1985)). In this view of the life cycle no reference has been made to sexual reproduction as this is fully discussed in a later section.

A mammal becomes infected when the non-dividing metacyclic forms, in the tsetse fly's saliva, are injected into the dermis of the host as the infected fly takes a blood meal. A chancre (local swelling) develops at this site due to a local inflammatory reaction. The metacyclics, which enter the lymph and blood stream, transform into rapidly dividing long slender bloodstream forms which are dependent on the host for many nutrients. The bloodstream forms are completely covered by a protective variant surface glycoprotein (VSG) coat to which the mammalian host mounts an immune response, resulting in the parasite population falling dramatically. Long slender bloodstream forms have the ability to change the surface coat to antigenically different variants at a high frequency: approximately once in every 100 cell divisions [Turner and Barry, 1989]. The trypanosome expressing the new variant surface coat avoids the immune response and divides rapidly, causing a rise in the parasite population. This new variant then becomes the target for the immune response. By the time the host is prepared to eliminate one antigenic type of trypanosome, the next generation has already been established. The switching to different antigenic types within the mammalian host results in a chronic infection with characteristic waves of parasitaemia. The majority of trypanosomes are long slender forms; however another type of trypanosome, non-dividing short stumpy form, is also present in the bloodstream population, becoming more abundant at high

Species	Vector	Host	Disease	Distribution
T. b. brucei	Tsetse flies	Cattle, pigs, sheep, wild game	Nagana	Africa
T. b. rhodesiense	Tsetse flies	Humans, cattle, wild game	Sleeping sickness	E. Africa
T. b. gambiense	Tsetse flies	Humans, pigs, sheep	Sleeping sickness	W. Africa
T. evansi	Tabanid flies (and any biting fly)	Equines, cattle	Surra	N. Africa/ Middle East/ S.E. Asia
	Tabanid flies (and any biting fly)	Equines, cattle	Mal de Caderas	Central / S. America / S.E. Asia
T. equiperdum	Venereal disease	Equines	Dourine	Africa/ Middle East
T. suis	Tsetse flies	Pigs	Acute porcine trypanosomiasis	Central/ E. Africa
T. congolense	Tsetse flies	Cattle, sheep, goats, etc.	Nagana	Africa
T. simiae	Tsetse flies	Pigs	Acute porcine trypanosomiasis	Central/ E. Africa
T. uniforme	Tsetse flies	Cattle	None / Mild	Central/ E. Africa
T. vivax	Tsetse flies (and any biting fly)	Cattle, sheep, goats, etc.	Nagana	Africa
	Tabanid flies (and any biting fly)	Cattle	Huequera	S. America
T. godfreyi	Tsetse flies	Pigs, warthog	Acute porcine trypanosomiasis	W. Africa

Table 1.1. Salivarian trypanosomes [adapted from Cox, 1993; McNamara *et al.*, 1994].

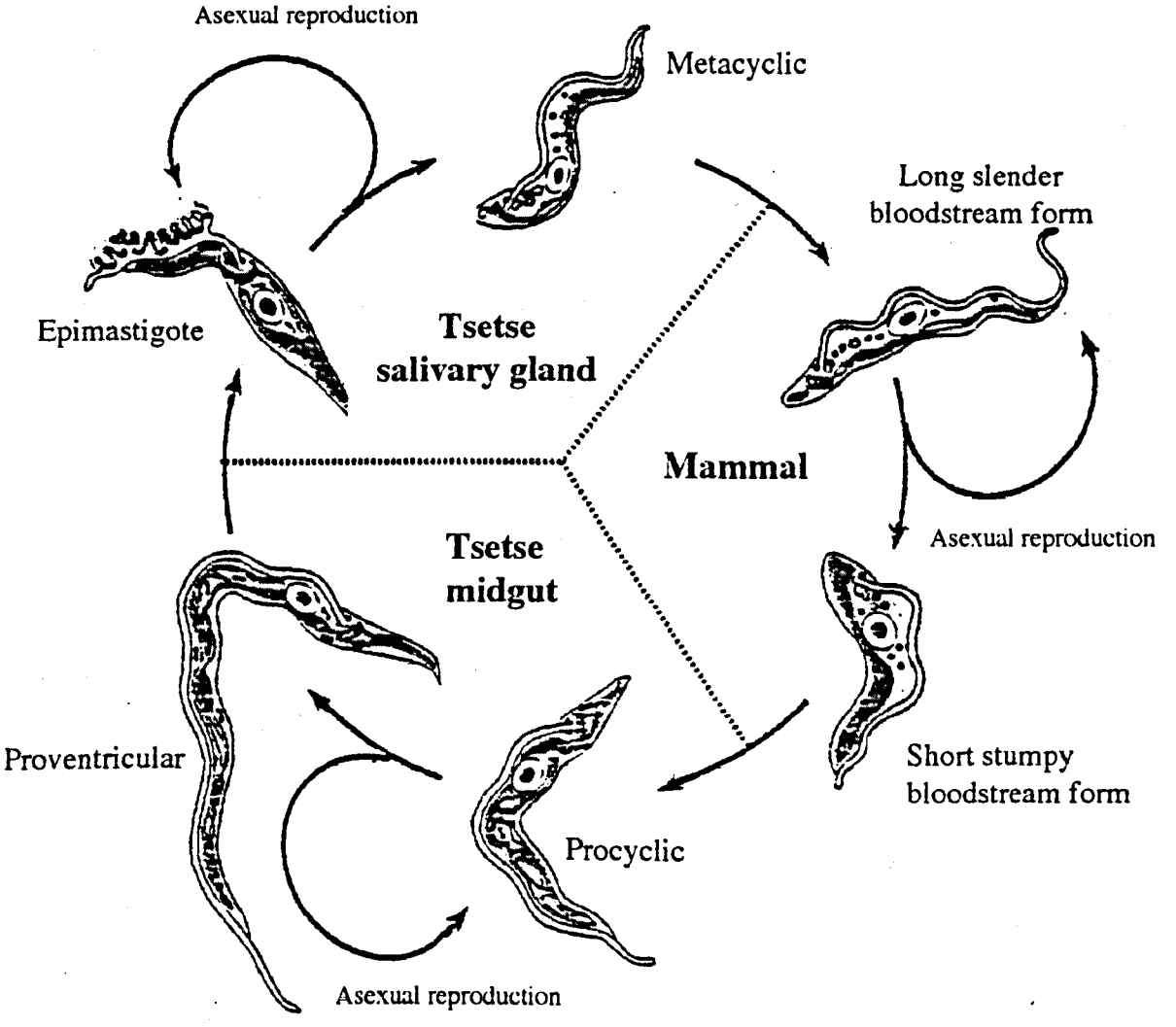


Figure 1.2. The life cycle of *T. brucei*. The life cycle consists of at least 6 stages, two life cycle stages for each of the three environments (mammailan host, tsetse salivary glands, tsetse midgut). In each environment one life cycle stage is able to undergo asexual reproduction, establishing a dense infection and the other stage is unable to divide but is pre-adapted to life in the next environment. This figure is adapted from Vickerman, (1985).

parasitaemia. These short stumpy forms are pre-adapted to being able to establish an infection in the next environment, the tsetse fly midgut. As a tsetse fly takes a blood meal from the infected mammalian host, the short stumpy forms are ingested and pass to the midgut where they transform to procyclics and establish an infection between the gut wall and the peritrophic membrane. Differentiation involves the loss of the VSG coat and the appearance of an invariant coat, specific to this life cycle stage, procyclin or procyclic acidic repetitive protein (PARP) [Roditi *et al.*, 1987].

Procyclics then transform to the non-dividing elongated proventricular form which travels though the fly's alimentary canal to the proboscis and then to the salivary glands. Some weeks (3-5) after the midgut infection has been established, the salivary glands become infected when the proventricular forms transform to dividing epimastigotes which attach to the epithelium of the salivary gland. The cycle is completed when the epimastigotes transform to the non-dividing metacyclics which are not attached to the epithelium and are pre-adapted to survival in the mammalian host, complete with a VSG coat. Metacyclic stage cells have a more limited repertoire of possible VSG coats for their surface called metacyclic variant antigen types or M-VATs [Barry *et al.*, 1983].

Trypanosomes have developed the ability to survive in two very different hosts by undergoing a number of morphological, metabolic and genetic changes. The most striking changes involve the surface coat and the two unique organelles, the kinetoplast and the glycosome. The large single copy organelle, the kinetoplast, consists of a network of several thousand circular DNA molecules, comprising two types, termed the minicircles and maxicircles based on their size [Vickerman, 1994]. The maxicircles are similar to the mitochondrial DNA of other eukaryotes and contain genes which are essential for kinetoplast biogenesis: genes coding for mitochondrial ribosomal RNAs and proteins involved in the electron transport chain and ATP synthesis [Simpson, 1987]. The minicircles are known to encode small RNA molecules which can facilitate the insertion/deletion of uridine nucleotides in maxicircle transcripts in a process known as RNA editing [Kable *et al.*, 1996]. The other organelle unique to the order Kinetoplastida is the glycosome, containing the enzymes involved in glycolysis, which are essential for survival in the blood of the mammalian host as glucose is the main food source [Michels *et al.*, 1997]. In the midgut of the tsetse fly glycolysis is less important as the main energy source is the amino acid, proline. In order to utilise this source the mitochondrial respiratory chain becomes activated.

Human African Trypanosomiasis

Trypanosomiasis causes immense human suffering. In the early phase of an infection the symptoms are non-specific: nausea, headache, fever and lethargy, which can easily be confused with other diseases, such as malaria. In the late phase, the parasites cross the blood brain barrier, causing neural damage which results in the classical symptoms of

sleeping sickness: severe neurological disturbance, irregular sleep patterns, coma and death. Without treatment, for the acute disease (*rhodesiense* sleeping sickness) death can occur in 6-12 months, and in 5-20 years for the chronic disease (*gambiense* sleeping sickness).

The observation of trypanosomes in the blood of patients is the main method of diagnosis. However, the waves of parasitaemia, as a result of antigenic variation, and the sequestering of the trypanosomes to tissues such as the heart and liver, can result in few parasites circulating in the blood, making positive diagnosis difficult. A number of kits have been developed to aid the detection of trypanosomes in blood, which include concentration techniques, such as the mini-anion exchange column and the microhaematocrit centrifugation techniques [Nantulya, 1991], although they have not raised the level of sensitivity enough to avoid generating false negative results from patients with low parasitaemias. The card agglutination test, which detects antibodies to *T. brucei* antigens, is a cheap sensitive technique, however, it is unable to distinguish between a current infection and one that has already been treated and cured [Truc *et al.*, 1994]. A sensitive technique allowing the early detection of infections is still desirable, as it would facilitate the control of the disease.

The treatment of this debilitating disease has changed little in the past 50 years. Suramin and pentamidine, developed in 1922 and 1937 respectively, remain the preferred drugs for treatment of early stage infections, while melarsoprol, introduced in 1949, is used in the treatment of late infections. All of these drugs are toxic to the mammalian host causing side effects ranging from vomiting to neurological disorders and, in the case of melarsoprol, causing encephalopathy and death in 5% of treated patients [Pepin and Milford, 1994]. The first new drug since the introduction of melarsoprol, DL-a-difluoromethylornithine (DFMO), was introduced in 1990. Originally developed as an anti-cancer agent, this drug inhibits ornithine decarboxylase and is highly effective at eliminating *T. b. gambiense* although it is not effective against all isolates of *T. b. rhodesiense* [WHO, 1995]. The main limitation of this drug is expense: at $500 per patient compared to $60 per patient for melarsoprol, few developing countries can afford the treatment, although the WHO has been trying to reduce the cost of production [WHO, 1995].

The improper use of melarsoprol and pentamidine has led to the emergence of resistant *T. brucei* strains [Scott *et al.*, 1996]. Laboratory experiments with resistant *T. b. rhodesiense* strains have given encouraging results for the use of combination therapy with DFMO and either suramin or melarsoprol for the treatment of the disease involving drug resistant *T. b. rhodesiense* strains [WHO, 1995].

Although there is an obvious need for new, cheap drugs, which do not cause severe side effects, the development of such a drug would be a non-profit making venture for a pharmaceutical company and so is unlikely to occur, unless a trypanocidal drug emerged

from a screening process designed towards more 'profitable diseases' such as cancer, or alternatively by exploiting traditional remedies for the treatment of sleeping sickness [Kigotho, 1997].

The possibility of generating a universal vaccine for the protection against the disease has been hindered by the ability of the parasite to undergo antigenic variation, as over a thousand possible different surface coats can be expressed by a trypanosome [Barry, 1997] and even further variants could be generated by gene conversion between VSG genes during the process of switching. Despite being covered by a dense surface coat, the trypanosome cell must take up nutrients, such as glucose, from the external environment. The invariant receptors responsible for nutrient uptake are buried deep in the cell body at the base of the flagellum in a structure known as the flagellar pocket, which is the site of endocytosis in *T. brucei* [Webster and Russell, 1993]. This structure ensures that, should antibodies bind to proteins in the pocket, they will be unable to stimulate macrophage killing of the parasite due to the small size of the pocket.

One of the most successful control measures for the disease has proved to be the method of vector control using biconical insecticide impregnated traps, which use chemical attractants to lure the tsetse flies. Placed in known foci, such traps have been known to lower significantly the number of seropositive people within a local area [WHO, 1995]. Although cheap and safe, these traps require high levels of maintenance. By using a combination of control measures such as surveillance, early diagnosis, treatment and vector control using traps, the 1980s epidemic in Uganda was brought under control [Molyneux, 1997]. However this required a concerted effort on the part of the authorities and removal of the control methods is likely to result in the re-emergence of the disease. A sterile tsetse release programme on the island of Zanzibar has proved highly successful in eradicating tsetse flies and eliminating the disease [Hide, 1999], although such a programme could only be effective in an isolated region, such as an island, where there is little chance of reinfestation with tsetse flies from the surrounding area. In other countries social and political unrest has resulted in a breakdown of control measures and the movement of people to tsetse infested areas, resulting in a upsurge in the number of infected individuals. The estimated 1995 figures for sleeping sickness cases in the Democratic Republic of Congo and Angola have reached 250,000 and 100,000, respectively, with a prevalence of 20-50% being reported in some villages [WHO, website] and, in 1997, an epidemic in Sudan was reported with 10 million people at risk [Kigotho, 1997].

Genomic Organisation of *T. brucei*

Direct cytogenetic examination of *T. brucei* chromosomes to determine the karyotype and ploidy has not been possible as trypanosome chromosomes do not appear to condense at any stage during mitosis [Vickerman and Preston, 1970]. Therefore, indirect evidence of

ploidy was sought using a number of different techniques: isoenzyme data [Gibson *et al.*, 1980; Tait, 1980], RFLP analysis of single copy genes [Gibson *et al.*, 1985] and the measurement of DNA content and kinetic complexity [Borst, 1982]. The results obtained are all consistent with bloodstream trypanosomes being diploid.

The suggestion that *T. brucei* may be a sexual organism [Tait, 1980] led to the search for a haploid stage (gametes) in its life cycle. Trypanosomes isolated from different life cycle stages, in both *T. brucei* and other trypanosome species, were analysed by cytofluorimetry and their DNA contents compared to bloodstream forms [Shapiro *et al.*, 1984; Kooy *et al.*, 1989]. These results indicated that all life cycle stages of the trypanosome that were investigated are diploid. The findings of these two studies directly contradict one report which, based on microfluorometry, suggested that metacyclic forms of the parasite were haploid and the products of obligatory meiosis during trypanosome development in tsetse flies [Zampetti-Bosseler *et al.*, 1986]. If this suggestion were correct then it would be predicted that cloned trypanosome lines heterozygous for a number of loci, when transmitted through a tsetse fly and sub-cloned from single metacyclics, should be homozygous for all loci. Analysis of a series of metacyclic derived sub-clones all appeared to be identical to the original clone, *i.e.* all heterozygous loci remained heterozygous [Tait *et al.*, 1989], demonstrating that metacyclics were indeed diploid and not the products of obligatory meiosis.

The development of pulse field gel electrophoresis (PFGE) has greatly increased our understanding of the *T. brucei* karyotype, with chromosomes falling into three distinct categories, mini (50-150 kb), intermediate (150-700 kb) and large chromosomes (800 kb-6Mb). It has been estimated that there are 50-150 minichromosomes in the *T. brucei* genome which are believed to act as a reservoir for VSG genes [Van der Ploeg *et al.*, 1984]. The number and size of the intermediate chromosomes, which also contain VSG genes, varies between strains, and both mini and intermediate chromosomes are probably essentially haploid [Gottesdiener *et al.*, 1990]. The large chromosomes have recently been resolved into 11 chromosome pairs ranging in size from 1Mb to approximately 5.2Mb [Melville *et al.*, 1998]. These chromosomes are diploid and appear to contain housekeeping genes. Within a stock, homologous chromosomes can differ widely in size, while non-homologous chromosomes can be of the same size and co-migrate under PFGE [Gottesdiener *et al.*, 1990; Melville *et al.*, 1998]. PFGE has also revealed a remarkable diversity in chromosome patterns between trypanosome species and strains [Gibson and Borst, 1986; Melville *et al.*, 1998], with chromosomes of all classes varying between strains.

A concerted effort to map and sequence the ~70 Mb comprising the *T. brucei* genome has begun and has lead to significant progress in generating both a genetic [M. Turner, D. Masiga and A. Tait, personal communication] and a physical map [Melville *et al.*, 1998], with both bacterial artificial chromosomes (BAC) and P1 libraries covering the entire

genome and cosmid contigs for portions of the genome, now available. The *T. brucei* sequencing project has succeeded in sequencing most of chromosome I and a considerable part of chromosome II, also the sequencing of random cDNA clones is being pursued by some groups as a direct approach to obtaining coding gene sequences [for example, El-Sayed *et al.*, 1995]. Undoubtedly the primary sequence of the genome will in itself prove invaluable as a basic tool for genome analysis. However, in order to understand fully the structure and function of the *T. brucei* genome, the nature of DNA sequence variation between different strains needs to be elucidated as well as the mechanisms which generate such variation. Variation between genomes can be generated by specialist mechanisms such as antigenic variation or by haplotypic mixing, with resultant homologous recombination and crossing over during sexual recombination. The processes of mutation and sexual recombination in *T. brucei* are poorly understood.

Genomic Variation

The *T. brucei* karyotype is known to exhibit high levels of size polymorphism, with homologous chromosomes within a genome differing in size by as much as 20%, while chromosome size differences between stocks can be as large as two-fold [Melville *et al.*, 1998]. As syntenic groups are maintained in all stocks examined to date, the size differences are thought to be due, to a large extent, to expansions and contractions of repetitive sequence throughout chromosomes and at the telomeres, although the insertions of transposable elements and variable location of VSG expression sites may also be responsible for some of the variation observed. However, the nature of the sequences involved in these size differences is largely unknown, as are the mechanisms leading to the generation of such size differences.

Tandemly repeated genes. Many housekeeping genes in the *T. brucei* genome (approximately half of the genes studied to date) are present as multiple tandem copies [Swindle and Tait, 1996] and a number of these genes have been shown to be polymorphic in their copy number between stocks, for example PARP [Mowatt and Clayton, 1988], tubulin [Seebeck *et al.*, 1983] and the glucose transporter genes [Barrett *et al.*, 1996]. These clusters are assumed to have arisen from a single source gene, via duplication and rounds of unequal exchange to produce arrays of genes which are able to diverge and form the basis of multigene families, for example, the VSG genes [Barry, 1997].

Variant Surface Glycoproteins (VSG) expression sites. The dense layer of a single glycoprotein which covers the surface of the bloodstream *T. brucei* cell is encoded by VSG genes which are transcribed from large polycistronic expression sites found at telomeres. There are several such telomeric expression sites (probably about 20) in the genome although only one site is transcriptionally active at any one time. The bloodstream expression sites (believed to be hemizygous) have a unique architecture: at the 5' region of the expression site lies a 50bp tandem repeat, followed by at least 8

expression site associated genes (ESAGs) and another large stretch of tandem repeats with a repeat size of approximately 70bp, next to the VSG gene which is situated adjacent to the telomere. The metacyclic expression sites, which express M-VSG genes, have a simpler organisation consisting of a small number of 70bp repeats next to the VSG gene at the telomere. As well as VSG genes in expression sites, there are estimated to be approximately one thousand other VSG genes within the *T. brucei* genome, residing either at internal sites within the large chromosomes or at the telomeres of mini- and intermediate chromosomes. Antigenic variation, the process of the serial expression of different VSG genes [reviewed in Barry, 1997 and Borst, 1997] can occur by a number of different mechanisms. Three types of mutation processes which cause VSG switching involve genome rearrangements. They are: (1) duplicative transposition, arguably the major mechanism, in which a VSG gene from an internal site (or from a minichromosome) is copied into the active expression site and which can result in the formation of chimeric VSGs ; (2) telomere conversion, where the VSG gene and telomere in the expression site is replaced by a copy of a VSG gene and telomere from another telomeric site leaving the donor site unchanged or (3) reciprocal telomeric exchange in which the existing VSG gene and telomere at the expression site is reciprocally exchanged with the VSG gene and telomere of another telomeric site.

Two other mechanisms which do not involve major chromosomal rearrangements but which may generate antigenic variation are point mutations in the VSG gene (although this is believed to be rare and of little significance [Barry, 1997]) and switching transcription from one expression site to another.

Retrotransposon-like elements. The main dispersed repeat in the *T. brucei* genome (excluding the VSG genes and their associated tandem repeats) is the element termed *ingi*. This is a 5.2kb sequence which has a copy number of ~200 and is located throughout the genome accounting for approximately 1Mb of the genome [Kimmel *et al.*, 1987]. The *ingi* element is flanked on either side by half a RIME element (~250bp). The RIME elements are also widely dispersed and have a copy number of ~ 400, but are usually associated with *ingi* elements. The *ingi* element shows features characteristic of transposable elements (short direct terminal repeats, poly-A tract at the 3' end and sequence homology to a reverse transcriptase) which can be expressed as part of larger transcripts [Kimmel *et al.*, 1987]. Unlike the dispersed nature of *ingi* and RIME, another retrotransposon-like element, SLACS, is confined to one locus, the mini exon cluster, and is found in 20-30 copies [Aksoy *et al.*, 1990]. Although possessing many of the features of retrotransposons, SLACS have not been shown to transpose. Dispersed repeats may act as initiators of non-homologous recombination which could lead to deleterious deletions or duplications, but could also account for some of the chromosome size polymorphisms [Melville *et al.*, in press].

Variable number tandem repeats (VNTRs). The most variable loci in eukaryotic genomes are the so called variable number tandem repeats (VNTRs) which, as their name suggests, show variation in the copy number of tandemly repeated sequences. The presence of repetitive DNA is a ubiquitous feature of higher eukaryotic genomes. The term VNTR can be used to include a range of variable loci including mononucleotide repeats, dinucleotide repeats, or microsatellites, up through minisatellites, midisatellites, telomeric, rDNA and satellite repeats. Such loci often show extreme levels of variation in terms of repeat unit copy number with high levels of heterozygosity. Such high allele length variability is based directly upon rapid mutation to new length alleles by insertion/deletion mechanisms.

Genes containing tandem repeats have been identified in *T. brucei,* and include the gene encoding the cytoskeletal paraflagellar rod protein, with a number of repeats of 11 amino acids [Woodward *et al.*, 1994], the gene encoding the PARP which covers the surface of the insect stage parasites [Roditi *et al.*, 1987], and two other membrane proteins which contain repeat motifs, the cysteine-rich, acidic integral membrane protein (*CRAM*) and *292*, [Lee *et al.*, 1990; Lee *et al.*, 1994; see also Chapter 3]. Also with the *T. brucei* genome sequencing project underway and the sequencing of expressed sequence tags (ESTs) [El-Sayed *et al.*, 1995] many tandemly repeated sequences, such as microsatellites, have been identified in non-coding (and coding) regions of the genome. Size polymorphisms between stocks in the repeated regions of these coding and non-coding sequences, due to variations in the number of repeats, could be a rich source of polymorphic markers; however little work has been carried out in this area, with the exception of the work of Sasse who identified a series of microsatellite markers in *T. brucei* [Sasse, 1998], and the minisatellite identified by Barrett [Barrett *et al.*, 1997; see also Chapter 3].

Genetic Exchange in *T. brucei*

Experimental crosses. The first evidence for genetic exchange in *T. brucei* was provided by Tait in 1980, who examined the isoenzyme patterns of a number of field isolates and concluded that the large number of different genotypes and the apparent random assortment of alleles at different loci was due to sexual recombination, although it was not until 1986 that the first genetic cross between two different *T. brucei* stocks was demonstrated in the laboratory [Jenni *et al.*, 1986]. To date nine different experimental crosses between trypanosome stocks have been conducted, together with one backcross and two experiments in which the products of self-fertilization were detected (summarised in Table 1.2). In all cases the procedure for performing the crosses involved cotransmitting two different *T. brucei* stocks through tsetse flies. Clones of trypanosomes derived from the infected flies were then analysed using a variety of techniques; RFLPs of single copy genes and of kDNA, isoenzymes, DNA content and karyotype. A total of 140 hybrid

progeny clones and 10 selfers have been identified to date (Table 1.2), although not all of these recombinant progeny are the products of different mating events, *e.g.* two of the first three clones identified are identical [Jenni *et al.*, 1986].

The analysis of these crosses has demonstrated that genetic exchange is non-obligatory. For example, a proportion of flies infected with two different trypanosomes develop mature salivary gland infections consisting of only one parental type. Also, although the majority of flies which have mixed-infected salivary glands produce hybrid trypanosomes, a proportion of the metacyclics are still of a parental genotype, with the proportions of parental and hybrid genotypes varying with time; early in infections, parental trypanosomes are produced, but if sampled later during the course of the infection, hybrid trypanosomes are also produced [Schweizer *et al.*, 1988]. Marker analysis of clones from mixed-infected flies has shown that some trypanosomes were equivalent to F1 progeny, in that alleles appeared to segregate in a Mendelian fashion. For example, for loci at which parents were homozygous but different from each other, the hybrids were heterozygous, and for loci heterozygous in the parents, segregation and independent assortment of alleles into the progeny was observed [Wells *et al.*, 1987; Tait *et al.*, 1988; Gibson, 1989; Sternberg *et al.*, 1989; Turner *et al.*, 1990; Gibson and Garside, 1991; Gibson *et al.*, 1992]. No barriers to mating have yet been identified as three stocks have been successfully crossed in all combinations [Sternberg *et al.*, 1989; Turner *et al.*, 1990], although it is conceivable that a mating type system exists and at least one of the stocks was heterozygous for alleles at a putative mating type locus.

Another class of recombinant progeny from a mixed transmission has been identified as the product of self-fertilization [Tait, 1983; Tait, *et al.*, 1996; Gibson *et al.*, 1997]. It is interesting to note that such an event appeared to occur only in the context of cross-fertilization. Single stock transmissions through flies have so far only ever produced clones which were identical to the original stock [Tait *et al.*, 1996] and the double drug selection of progeny from a cross between two drug resistant transformants of the same strain did not generate the products of self-fertilization [Gibson *et al.*, 1997].

Another interesting feature of the sexual process is that a proportion of analysed hybrid clones demonstrated an elevated DNA content of approximately 1.5 times the normal level, [Gibson *et al.*, 1992; Tait *et al.*, 1993]. Further examination of these clones demonstrated trisomy for those chromosomes analysed. It is possible that such hybrid clones are triploid and have arisen as a result of failure of meiosis in one parent, generating a diploid gamete which then fuses with a haploid gamete [Gibson, 1995], or through fusion of a normal diploid cell with a haploid gamete. However, due to the limited numbers of progeny clones analysed to date it is difficult to determine whether the increased DNA content is indicative of a specific aspect of the genetic exchange mechanism or is merely an aberrant phenomenon.

Summary of *T. brucei* Genetic Crosses							
Parental clones for each cross						No. of Hybrids	Reference
Parent 1			Parent 2				
Clone	Origin	Host	Clone	Origin	Host		
Crosses							
STIB 247	Tanzania	Hartebeast	STIB 386	Ivory Coast	Human	3	Jenni *et al.*, 1986
							Paindavoine *et al.*, 1986
							Wells *et al.*, 1987
						1	Sternberg *et al.*, 1988
						8	Sternberg *et al.*, 1989
						5	Turner *et al.*, 1990
STIB 247	Tanzania	Hartebeast	TREU 927/4	Kenya	Tsetse	24	Turner *et al.*, 1990
STIB 386	Ivory Coast	Human	TREU 927/4	Kenya	Tsetse	9	Turner *et al.*, 1990
STIB 247-LF	Tanzania	Hartebeast	STIB 777-A	Uganda	Tsetse	4	Schweizer *et al.*, 1994
058	Zambia	Human	196	Ivory Coast	Pig	10	Gibson W. C., 1989
							Gibson *et al.*, 1992
196	Ivory Coast	Pig	J10	Zambia	Hyena	12	Gibson and Garside 1991
STIB 831-K cl 1	Uganda	Tsetse	STIB-831-K cl 2	Uganda	Tsetse	9	Degen *et al.*, 1995
058H	Zambia	Human	KP2N	Ivory Coast	Tsetse	7	Gibson and Whittington 1993
						29	Gibson and Bailey 1994
TH2N	West Africa	Human	058H	Zambia	Human	6	Gibson *et al.*, 1997
Backcross							
058H	Zambia	Human	P20 (hybrid of 058 x KP2)			13	Gibson *et al.*, 1995
Self-fertilisation							
STIB 247	Tanzania	Hartebeast	(in the presence of STIB 386)			5	Tait *et al.*, 1996
TH2N	West Africa	Human	(in the presence of 058H)			5	Gibson *et al.*, 1997

Table 1.2. A Summary of *T. brucei* genetic crosses. Each hybrid clone is not necessarily the product of a unique mating event. Hybrids with identical genotype patterns are likely to be vegetative derivatives of a single mating event. STIB, Swiss Tropical Institute Basel; TREU, Trypanosome Research Edinburgh University.

Chromosomal inheritance. The three different classes of chromosomes in *T. brucei* are inherited by F1 progeny in different ways. Mini and intermediate chromosomes are inherited in a non-Mendelian fashion. Wells *et al.* (1987) demonstrated that hybrid progeny clones inherited a full complement of minichromosomes from each parent. The hybrid progeny clones examined in these experiments all had elevated DNA contents and the inheritance of all minichromosomes from both parents accounted for only a small proportion of their increased DNA content. However, diploid progeny clones have not been investigated for minichromosome inheritance. Consequently, it is possible that these results are aberrant and do not reflect the true pattern of inheritance of these chromosomes. Analysis of intermediate chromosomes has shown that parental chromosomes appear to be randomly distributed among the progeny, as would be expected if the chromosomes were not segregating as homologues at meiosis [Sternberg *et al.*, 1987; Wells *et al.*, 1987; Le Page *et al.*, 1988; Gibson, 1989]. Large chromosomes however appear to segregate in a Mendelian fashion, *i.e.* progeny inherit one homologue from each parent. It would appear that different sized chromosomes, which are structural homologues (from Southern analysis), behave as genetic homologues [Gibson, 1989; Gibson and Garside, 1991; Melville *et al.*, 1998]. These findings provide further evidence for the occurrence of meiosis. Although some features of chromosomal inheritance appear to oppose the model of a traditional Mendelian genetic system, *i.e.*, as mentioned previously, some hybrid clones appear to be trisomic for a number of chromosomes [Gibson *et al.*, 1992], have raised DNA content [Paindavoine *et al.*, 1986; Gibson *et al.*, 1992; Tait *et al.*, 1993], and novel non-parental size chromosomes which appear to be specifically associated with genetic exchange are often observed in progeny clones [Gibson, 1989; Gibson and Garside, 1991; Tait *et al.*, 1993; Melville *et al.*, 1998; Tait *et al.*, manuscript in preparation].

Kinetoplast inheritance. The inheritance of the kinetoplast organelle was examined by analysing polymorphic regions within maxi- or minicircle DNA. Initial studies indicated that maxicircles were inherited uniparentally from either parent [Gibson, 1989; Sternberg *et al.*, 1989; Gibson and Garside, 1990] and minicircles were inherited biparentally [Gibson and Garside, 1990]. However more recent studies of hybrid clones during their early stages of bloodstream growth appear to indicate that both maxi- and minicircles are inherited biparentally with subsequent stochastic segregation at each mitotic division [Turner *et al.*, 1995]. This model would predict rapid fixation to a uniparental pattern for the maxicircles (but not for the minicircles due to their high copy number), and would explain the results of earlier reports and the observation that maxicircles are homogeneous. The evidence is strengthened by the observation of two hybrid clones, which have identical nuclear genotypes, mixed mini-circle genotypes but differing maxicircle genotypes [Gibson and Garside, 1990]. The clones are believed to be the two daughter cells from the same mating event, suggesting that parental kinetoplasts

fuse followed by the segregation of the maxi-circles in subsequent mitotic divisions [Gibson and Garside, 1990].

Stage of mating. It is clear from the results of the laboratory crosses that mating takes place at some stage during the trypanosome life cycle in the tsetse fly, *i.e.* at the procyclic, proventricular, epimastigote or metacyclic stages. Analysis of trypanosomes vegetatively derived from single metacyclics, obtained from flies in which trypanosome mating is occurring, have been shown to be hybrid. It is reasonable to assume that such metacyclics are the products of mating and that mating occurs at an earlier stage [Kooy *et al.*, 1989; Tait *et al.*, 1989], contrary to a previous report [Zampetti-Bosseler *et al.*, 1986] which suggested that metacyclics were haploid gametes. The crossing of drug resistant stocks of *T. brucei* and the double drug selection of the progeny [Gibson and Whittington, 1993; Gibson *et al.*, 1997], has shown that hybrids can be isolated from the salivary glands but not the midgut of mixed-infected tsetse flies, indicating that the probable site of genetic exchange is the salivary glands. A report [Schweizer and Jenni, 1991] has described the detection of hybrids in the midguts of tsetse flies, using isoenzyme markers but as no clones were isolated and demonstrated to be of the predicted hybrid phenotype, the significance of this report is unclear. On this basis, the life cycle stage at which mating takes place in *T. brucei* is still open to question.

Most flies with a mature mixed infection will eventually produce hybrid trypanosomes [Sternberg *et al.*, 1989; Tait, personal communication]. By analysing the metacyclics produced by these flies over time, it is clear that early in infection the parental stocks are transmitted without mating, but that, after approximately 25-50 days post infection, hybrid trypanosomes (as well as parentals) are observed, and continue to be produced for the life time of the fly [Schweizer *et al.*, 1988; Sternberg *et al.*, 1989; Turner *et al.*, 1990] indicating that mating is non-obligatory.

Models of genetic exchange. The observed segregation and independent assortment of alleles in *T. brucei* provides strong evidence that genetic exchange involves meiosis and syngamy; as the majority of markers used in the analysis of crosses appear to be inherited in a Mendelian fashion. Although the common mechanism of mating for diploid eukaryotes involves the formation of haploid gametes, and such a mechanism has been described for other flagellates [Cleveland, 1956], other models for genetic exchange in *T. brucei* have been proposed.

Paindavoine *et al.*, (1986) proposed a fusion/chromosome loss model whereby diploid nuclei fuse to form a tetraploid nucleus followed by random chromosome loss to restore diploidy. This attempts to explain the elevated DNA content found in some hybrids, but would predict that the DNA levels fall during vegetative growth, a phenomenon Paindavoine *et al.* report. However Wells *et al.* (1987) analysed the same hybrids and found no such loss. This model also predicts non-Mendelian allelic segregation, which has not been observed. There is strong evidence for meiosis and syngamy [Turner *et al.*, 1990],

although, it is still not clear whether meiosis precedes fusion or meiosis follows fusion. For the classical Mendelian model, the limited number of progeny clones identified is insufficient to obtain statistically significant proof (or rejection) of Mendelian segregation ratios. Other data appear not to fit the Mendelian model. The elevated DNA content of some hybrids [Paindavoine *et al.*, 1986; Gibson *et al.*, 1992; Tait *et al.*, 1993] cannot be directly explained by this model, but may be aberrant products of meiosis, perhaps resulting from fusion of a diploid cell with a haploid gamete (or nucleus). The detection of two hybrid clones that are identical for nuclear markers but differ in maxicircle DNA, has led to the suggestion that an intermediate stage in hybrid formation exists, in which both parental kinetoplasts are present [Gibson, 1989; Gibson, 1995]. The model of fusion followed by meiosis involves the fusion of diploid parents to form an intermediate polyploid cell in which meiosis occurs to yield haploid nuclei which then fuse in pairs, resulting in diploid progeny. A failure of meiosis of one nucleus could result in diploid nuclei being produced which fuse with haploid nuclei and generate triploids. This model may also explain the failure to identify a haploid stage in the *T. brucei* life cycle, (although the search for haploid gametes has been limited).

Population Structure of *T. brucei*

***Identification of* T. brucei.** Although trypanosomiasis has probably been prevalent in Africa for centuries, it was the first recorded in the late 19th century by European colonists. For example, Livingstone, in 1857, referred to the 'tsetse disease' which caused wasting in cattle and horses [Vickerman, 1997]. In 1895, Bruce discovered that trypanosomes were the aetiological agent for the disease and were transmitted by the tsetse fly, which was soon followed by the recognition that human sleeping sickness was also caused by trypanosomes [Vickerman, 1997].

There are three morphologically identical subspecies of *T. brucei: T. b. gambiense* and *T. b. rhodesiense,* which cause the chronic and acute forms of human trypanosomiasis, respectively, and *T. b. brucei* which is non-human infective. The distinction between these subspecies has been based on the area of isolation (*T. b. gambiense* in West and Central Africa, *T. b. rhodesiense* in East Africa and *T. b. brucei* throughout Africa) and on the host from which the sample was isolated (*T. b. gambiense* and *T. b. rhodesiense* from humans and *T. b. brucei* from other mammals).

Human Infectivity. Given that tsetse flies feed on both humans and other mammals, the possibility existed that animals could be infected with *T. b. gambiense* or *T. b. rhodesiense, i.e.* that animals could act as a reservoir of human infective trypanosomes. To address this question a number of early studies used human volunteers to assess the ability of trypanosomes isolated from animals to infect humans, demonstrating that wild animals were a reservoir for human infective trypanosomes [Heisch *et al.*, 1958] and that *T. b. gambiense* and *T. b. rhodesiense* could retain their ability to infect humans even after long

periods of cyclical transmission between tsetse flies and sheep [Willett and Fairbairn, 1955]. However, some *T. b. rhodesiense* strains could lose their ability to infect humans after serial passage in laboratory rodents, and some stocks isolated from animals were unable to infect humans [Rifkin *et al.*, 1994].

Early observations revealed that human serum has a lytic effect on some strains of trypanosome isolated from animals while other stocks appear to be resistant to this trypanolytic effect. Exploiting this effect, a number of tests have been designed to differentiate between human infective and non-human infective trypanosomes without recourse to human experimentation. For example, the blood incubation infectivity test (BIIT) involves the incubation of trypanosomes with human serum and then either inoculating laboratory rodents [Rickman and Robson, 1970] or measuring cell lysis *in vitro* [Brun and Jenni, 1987] to determine whether any trypanosomes have survived. This method was widely used to identify a number of wild and domestic animals as reservoirs for human serum resistant trypanosomes [Robson *et al.*, 1972], although the test has been shown to give variable results with some *T. b. rhodesiense* samples (isolated from humans), as the phenotype can vary depending on the passage history [Rifkin *et al.*, 1994].

Mechanisms of human infectivity. The mechanism(s) of resistance to human serum has been the subject of a great deal of research over the past decade, however, it is still not fully clear how these organisms evade lysis. Investigations into the mechanisms of human serum resistance have identified two trypanolytic factors in human serum. Rifkin, (1978) identified the trypanolytic factor, (TLF1) as being a high density lipoprotein and in 1996 a second trypanocidal factor, TLF2, was identified as being a non-high density lipoprotein [Tomlinson and Raper, 1996]. TLF2 was discovered to be the main lytic factor [Raper *et al.*, 1996] which binds to trypanosome receptors, enters the cell by endocytosis, accumulates in the lysosome and triggers cell lysis. Resistance to this process is believed to be due, not to a reduction in TLF binding but to a reduction in internalisation of TLF [Hager and Hajduk, 1997].

In 1989 a possible genetic basis of resistance to human serum was proposed, when a VSG-like gene (the *SRA* gene) was found to be expressed only in human serum resistant trypanosomes [De Greef *et al.*, 1989]. More recently the *SRA* gene has been identified as being an ESAG (expression site associated gene) of the Etat 1.10 expression site [De Greef and Hamers, 1994], which, if transfected into a human serum sensitive stock, can confer resistance, clearly suggesting a link between *SRA* gene expression and resistance [Van Xong *et al.*, 1998]. However several pieces of evidence indicate that this cannot be the complete explanation of resistance. Firstly, following this model, human serum resistant stocks would be limited to only using one expression site, unless another *SRA* gene is present and expressed from another part of the genome. Secondly, this expression site is not active in metacyclics and so would not confer resistance for the first few crucial days in which the parasite is in the mammalian host, before differentiation into bloodstream

forms. Thirdly, at least two human serum resistant stocks do not express the *SRA* gene [Lindergard, 1999].

Sleeping sickness epidemics. Human sleeping sickness is characterized by long periods of endemicity and occasional short term epidemics confined to specific areas or foci which expand or contract with environmental changes. Many foci have existed for decades, for example, the Busoga focus in Uganda was the site of the epidemic of 1900 and is still active today. This grand epidemic was estimated to have infected half a million people and was thought originally to be due to *T. b. gambiense* but is more likely to have been due to the morphologically identical *T. b. rhodesiense* [Koerner *et al.*, 1995].

It is important to be able to distinguish between the different subspecies of *T. brucei* and to be able to track different strains through populations, so that questions regarding the epidemiology of the disease and the population structure of the parasite can be addressed. For example, are the trypanosomes from different foci the same or different? If it was possible to define a marker for human infectivity, then it would allow human infective trypanosomes to be identified from any host or vector and so facilitate detailed epidemiological analysis. To this end much research has been focused on biochemical and molecular approaches for the identification of the strains and subspecies of *T. brucei.*

***Distinguishing* T. b. gambiense *from* T. b. rhodesiense *and* T. b. brucei.** Isoenzyme analysis has been used extensively for several years to distinguish *T. brucei* strains and, although no single isoenzyme marker can define *T. b. gambiense,* the results with several enzymes has succeeded in distinguishing *T. b. gambiense* from *T. brucei brucei* and *T. b. rhodesiense* [Godfrey and Kilgour, 1976; Gibson *et al.*, 1978; Gibson *et al.*, 1980; Mehlitz *et al.*, 1982; Tait *et al.*, 1984; Godfrey *et al.*, 1987]. *T. b. gambiense* stocks have been divided into two groups based on isoenzyme analysis. The majority of human infections from West Africa fall into the first category (group 1), which have a limited isoenzyme variability and are highly resistant to human serum, but a second group of trypanosomes (group 2) are more heterogeneous in their isoenzyme patterns and display variable resistance to human serum [Gibson *et al.*, 1980; Mehlitz *et al.*, 1982; Tait *et al.*, 1984; Gibson, 1986]. Further analysis of *T. b. gambiense* using repetitive DNA probes and RFLPs confirms the existence of two distinct groups of West African trypanosomes, the well defined *T. b. gambiense* and a heterogeneous group of trypanosomes which shows a close relationship to West African *T. b. brucei,* and could be considered as the West African equivalent of *T. b. rhodesiense* [Paindavoine *et al.*, 1989; Hide *et al.*, 1990].

***Distinguishing* T. b. rhodesiense *from* T. b. brucei.** *T. b. rhodesiense* stocks appear to be indistinguishable from *T. b. brucei* stocks by isoenzyme analysis. Although human infective strains have a more restricted repertoire of zymodemes (sets of stocks which share the same isoenzyme profile) than *T. b. brucei* and certain zymodemes predominated in the *T. b. rhodesiense* samples, there was no particular pattern which was exclusively associated with *T. b. rhodesiense* and could distinguish the two subspecies

[Gibson *et al.*, 1980; Gibson and Gashumba, 1983; Gibson and Wellde, 1985]. Polymorphisms in the maxi-circles of kDNA also failed to differentiate between *T. b. rhodesiense* and *T. b. brucei*, which led to the suggestion that they were not different subspecies but that *T. b. rhodesiense* was a host range variant of *T. b. brucei* [Borst *et al.*, 1981], further suggested, on the basis of isoenzyme data, by Tait (1985).

Using RFLPs in repetitive DNA sequences to examine different trypanosome stocks, Hide found that it was possible to distinguish between human serum resistant and human serum sensitive samples on the basis of a complex cluster analysis, although two samples which were human serum sensitive clustered with the human serum resistant types [Hide *et al.*, 1990; Hide *et al.*, 1994]. The fact that human infective field samples cluster together and are a semi-defined group suggests they are distinct, although the fact that expression of human serum resistance can vary and sensitive derivatives of cloned human isolates can be obtained by passage in rodents, suggests at first sight, that it would be unlikely that *T. b. rhodesiense* is genetically distinct from *T. b. brucei*. However, fully human serum sensitive strains occur and so the distinction, at this level, is in the ability to express human serum resistance. Early studies used human volunteers to determine human infectivity [Heisch *et al.*, 1958] and although such methods have been replaced by tests such as the BIIT, one must be cautious when equating human serum resistance to human infectivity, as stocks may have the ability to survive in human serum, but may not be able to establish a human infection.

Animal reservoirs. If *T. b. rhodesiense* stocks are host range variants of *T. b. brucei* it follows that animals could be a reservoir for the human disease. This was demonstrated in 1958 when human volunteers were infected with trypanosomes which had been isolated from wild game. The resulting disease indicated that an animal reservoir for human trypanosomiasis existed [Heisch *et al.*, 1958]. This work was supported by a number of studies which used the blood incubation infectivity test (BIIT) in place of human volunteers and demonstrated that a range of wild and domestic animals could act as reservoirs for human infective trypanosomes [Robson *et al.*, 1972]. More recently, molecular approaches have been used to identify human-infective *T. b. gambiense* and *T. b. rhodesiense* isolates from wild and domestic animals. Isoenzyme and RFLP patterns of some isolates from both man and animals have been shown to be identical, suggesting there is a reservoir of human infective types within non-human hosts [Gibson *et al.*, 1978; Gibson *et al.*, 1980; Scott *et al.*, 1983; Tait *et al.*, 1984; Gibson and Wellde, 1985; Enyaru, 1993; Hide *et al.*, 1994].

Analysis of data from a number of studies in the Lambwe Valley district of Kenya examined by Mihok showed that ~27% of *T. brucei* infected domestic animals possessed trypanosomes which were identified as being human infective by DNA and isoenzyme analysis [Mihok *et al.*, 1990]. In these studies human serum resistance was not tested directly, but the analysis was based on 'identity' of isoenzyme or RFLP patterns between

isolates of human and animal origin as a criteria for identifying human infective strains. In a similar study of the Tororo district of Uganda, Hide *at al.*, (1996) examined (using isoenzyme and RFLP polymorphisms) isolates that had been collected in 1988-90 during an epidemic and directly measured human serum resistance, to show that ~23% of animal isolates were human infective [Hide *et al.*, 1996]. Examination of blood meals from tsetse flies during the epidemic in SE Uganda demonstrated the importance of the cattle reservoir by showing that a tsetse fly was 5 times more likely to have picked up *T. b. rhodesiense* from cattle than from humans [Hide *et al.*, 1996].

Clonality vs. sexual reproduction. Sexual recombination in *T. brucei* has been shown to occur in laboratory experiments when two genotypically distinct trypanosome stocks are cotransmitted though a tsetse fly, resulting in the generation of novel hybrid genotypes Jenni *et al.*, 1986; Turner *et al.*, 1990; Gibson and Bailey, 1994; see also Chapter 5]. Although, *T. brucei* stocks can undergo sexual recombination in the laboratory, it is still not clear to what extent mating takes place in natural trypanosome populations. The importance of determining the role of mating in the field lies in the possibility that traits such as human infectivity and drug resistance can be spread through a population of non-human infective trypanosomes by genetic exchange. A high level of sexual recombination would also increase the diversity within a population thus creating a wide range of genotypes capable of providing the means for the organism to adapt to changes in the environment. In recent years a number of studies have attempted to define the contribution of sexual recombination to the generation of the genetic variation observed in *T. brucei.*

The first evidence for sexual recombination in *T. brucei* populations came in 1980 when Tait analysed 17 trypanosome isolates from Lugala, in Uganda, for isoenzyme polymorphisms [Tait, 1980]. Examination of allele and genotype frequencies indicated that there was a high degree of allele diversity with genotype frequencies exhibiting no significant difference from those expected if the population was randomly mating, *i.e.* the population was in Hardy-Weinberg (HW) equilibrium. Similar analysis of two further populations, Sindo and Kiboko, also indicated agreement with HW expectations, leading to the proposal that sexual recombination was frequent [Tait, 1983]. However, this conclusion was later criticised, for although agreement to HW expectations is consistent with the occurrence of a randomly mating population, the agreement with expectation could occur by chance, especially if the sample size is small [Cibulskis, 1988]. The high level of allelic diversity found in the *T. brucei* populations analysed [Gibson *et al.*, 1980; Godfrey *et al.*, 1990], and the observation of a range of allelic combinations, suggests that it is highly unlikely that mutation alone could account for the observed variation, as the same mutations would have to occur repeatedly. These considerations led Cibulskis to come to the conclusion that some sexual recombination was occurring, but that it was not sufficiently frequent to fully break up associations between loci [Cibulskis, 1988].

Subsequent studies by a number of different groups would also appear to suggest a less important role for sexual reproduction in the field, for, although they found large numbers of different phenotypes and genotypes, the number of different allelic combinations remained lower than that theoretically expected from a randomly mating population, resulting in deviations from HW expectations [Gibson and Wellde, 1985; Tait *et al.*, 1985].

A population in which recombination is frequent, may still display deviation from HW expectations and linkage disequilibrium for a number of reasons. Mating may not be random, for example, self-fertilization may occur which would generate an excess of homozygotes. Self-fertilization has been demonstrated in *T. brucei* [Tait *et al.*, 1996; Gibson *et al.*, 1997; see also Chapter 5]. The samples analysed may not be from a single population, but a mixture of two or more genetically isolated populations. This could occur if two species are wrongly considered as a single panmictic unit or if two genetically isolated populations are combined. Sub-populations could be created by geographical barriers or by host specificities, creating populations each of which may be randomly mating. Combining data from sub-populations (or different species) could result in departure from HW. Another reason for deviation from HW equilibrium, could be sample bias whereby different genotypes may have different likelihoods of being included in the sample, perhaps due to variation in the ability of trypanosomes to grow in the laboratory. All samples analysed to data have been initially grown in rodents, sometimes for many passages. Indeed it has been observed that only a proportion of *T. brucei* isolates from tsetse salivary glands were able to grow in laboratory rodents [Goebloed *et al.*, 1973]. Such sample bias could be overcome by genotyping trypanosomes directly by PCR without recourse to growth in the laboratory. While it has been generally assumed that selection in rodents will be random in relation to the types of markers used in population genetic analysis, this has never been formally tested. Other factors such as natural selection, bottlenecks, and migration may all result in a deviation from HW equilibrium in an organism in which sexual recombination is frequent. It is also possible that a combination of a number of different factors may exert such effects on the analysis of the population structure of *T. brucei.*

When considering the life cycle of *T. brucei,* the degree of sexual recombination is dependent on opportunities for genetic exchange, which in turn is dependent on the feeding behaviour of flies and the number of mixed infections. It is clear that not all tsetse flies harbour mixed infections and of those that do, laboratory studies show that only a proportion produce hybrid trypanosomes together with parental genotypes [Sternberg *et al.*, 1989]. It would therefore appear unlikely that *T. brucei* populations would demonstrate a truly panmictic population structure, although sexual recombination may still play a significant role in generating new genotypes.

In 1990 after examining isoenzyme variation in populations of a number of medically important protozoa, including *Plasmodium, Leishmania* and *Trypanosoma*, Tibayrenc proposed a clonal theory of population structure, whereby sexual recombination plays only a minor role [Tibayrenc *et al.*, 1990]. Tibayrenc's criteria for clonality centre on the absence of segregation of alleles at a single locus and the lack of recombination between alleles at different loci and are presented in Table 1.3.

Tibayrenc's criteria for clonality	
criteria for segregation	fixed heterozygosity absence of segregation genotypes deviation from H-W
criteria for recombination	identical genotypes widespread, over given geographical areas absence of recombinant genotypes linkage disequilibrium correlation between independent sets of genetic markers (*e.g.* kDNA and nuclear DNA)

Table 1.3. Criteria for clonality given by Tibayrenc et al., (1990).

A number of other studies on *T. brucei* have supported the findings of Tibayrenc in that deviations from HW expectations were detected [Mihok *et al.*, 1990; Cibulskis, 1992; Enyaru, 1993]. However, one reason for this could be that the populations sampled were sub-structured by, for example, geographical barriers, which would disrupt a panmictic population structure and result in linkage disequilibrium and deviation from HW equilibrium. Indeed many of the studies, including those of Tibayrenc, have examined *T. brucei* stocks isolated from a wide variety of locations, time points and hosts and so do not take into account the possibility of population sub-structuring when collecting samples. Indeed, when Tibayrenc analysed 'non-*gambiense*' *T. brucei* samples from West and East Africa separately, there was no statistically significant evidence for linkage disequilibrium, but when the West and East African samples were combined, significant linkage disequilibrium was observed [Tibayrenc *et al.*, 1990]. Given that these populations are very likely to be geographically sub-structured, his argument for a clonal population of *T. brucei* is weak.

Tibayrenc argued that if deviation from HW expectations was due to population sub-structuring this would be associated with fewer than expected numbers of heterozygotes, but the data showed a larger proportion of heterozygotes than predicted, suggesting that even if population sub-structuring was present, the effect was not sufficient to counteract the excess of heterozygotes generated by clonal propagation. The persistence and over-representation of heterozygotes over generations suggested a lack of meiotic segregation and self-fertilization (which would generate homozygotes). Stronger evidence for clonality was provided by the detection of identical multilocus genotypes from different widespread geographical areas. The association of alleles at different loci suggests a lack of sexual recombination which would tend to break up these associations and the fact that such

genotypes were detected from samples isolated at different times suggests that clonal propagation has persisted for many years.

In 1993 Maynard-Smith proposed a population structure for *T. brucei* distinct from either random outbreeding or clonality: an epidemic population structure, whereby a randomly mating population structure is obscured by a recent explosive increase in one (or a few) particular genotype(s) [Maynard-Smith *et al.*, 1993]. This was demonstrated by Maynard-Smith using isoenzyme data from Kenyan samples of *T. brucei*. Using a measure of linkage disequilibrium, the Index of Association, I_A, (whereby I_A is zero for randomly mating population and non-zero if recombination is rare or absent), significant linkage disequilibrium (a non-zero value of I_A) was detected in this population, suggesting a non-randomly mating population structure. However when each common multilocus genotype was considered as an individual sample in the analysis, then the I_A returns to zero, suggesting a panmictic population structure. This indicates that one or a few genotypes have expanded within this population, becoming abundant and widespread, and so have concealed the true role of genetic exchange in the underlying population.

The concept of an epidemic population structure for *T. brucei,* proposed by Maynard-Smith, was supported by Hide in 1994 in a study of sympatric *T. brucei* isolates in SE Uganda, collected at the same time (between 1989 and 1990) from two different hosts (man and cattle), thus eliminating the possibility of population sub-structuring due to geographical barriers or time. Analysis of these isolates, (which were divided into human infective and non-human infective types) was carried out using Maynard-Smith's Index of Association on all the samples and again on electrophoretic types only, whereby the common multilocus genotypes were considered as single samples to reduce any population distortion created by an epidemic spread of one (or a few) genotype(s). The results of this analysis revealed that the non-human infective stocks appeared to show evidence of sexual recombination, whereas the underlying sexual population structure for the human stocks was initially obscured by an epidemic explosion of one or a few genotypes. Indeed it would appear that during epidemics such a population structure may be prevalent, with selection for the human infectivity phenotype driving clonal reproduction but only within the context of a 'background' population undergoing frequent recombination [Stevens and Welburn, 1993].

In summary, although it appears that genetic exchange takes place in natural populations the extent of sexual recombination and the question as to whether it occurs between human infective and non-human infective stocks remains to be elucidated. With the development of more informative DNA markers which can analyse genotype rather than phenotype and be interpreted genetically, more discriminatory tests can be carried out to determine the population structure and whether genetic exchange can affect the spread and inheritance of traits of medical and economic importance.

Substructuring in trypanosome populations due to geography. If the population of *T. brucei* were highly substructured due to geography, then each focus would contain genetically distinct trypanosomes, and combining data from different foci would give misleading results. For example, a deviation from Hardy-Weinberg equilibrium would be obtained although each subpopulation was undergoing random mating. The evidence presented by Hide *et al.*, (1994), based on RFLPs in rDNA and a repetitive probe pBE2, using isolates from Kenya, Uganda, and Zambia, suggests that there is clear substructuring, as parasites from different geographical regions are distinct. For example, *T. b. rhodesiense* stocks from Zambia are distinct from the *T. b. rhodesiense* stocks from the Ugandan/Kenyan focus [Hide *et al.*, 1990; Hide *et al.*, 1994]. Also examination of isolates from the Lambwe Valley in Kenya indicate that the frequency of particular zymodemes can vary significantly over quite short geographical distances [Cibulskis, 1992].

Substructuring in trypanosome populations due to host species. Different trypanosome strains can exhibit different levels of host range specificities, for example human infective strains can infect both humans and animals, whereas non-human infective stocks can infect cattle but not humans, probably reflecting an adaptation to specific transmission cycles. This adaptation could result from resistance to the lytic effects of human serum or from variation in the binding of essential transferrin in different hosts, as different transferrin binding affinities have been implicated in the ability of stocks to infect different hosts [Borst *et al.*, 1997; Bitter *et al.*, 1998]. Clearly, if different stocks have different host range specificities this will result in population substructuring. Evidence for such sub-structuring has been obtained by isoenzyme studies from the Lambwe Valley in Kenya [Cibulskis, 1992], where zymodemes appear to be restricted in the range of hosts they infect, with only a limited number of human infective zymodemes being detected. Further evidence for population sub-structuring due to host specificities was obtained by isoenzyme analysis of trypanosome isolates from Lugala in Uganda, where samples from humans appeared to be extremely homogeneous [Gibson and Gashumba, 1983], whereas samples from tsetse flies from the same area were remarkably diverse [Tait, 1980]. Mihok *et al.*, (1990), analysing isolates from different hosts, proposed that the high level of diversity found in tsetse flies could not be explained by the flies feeding solely on cattle or humans, but that wild animals must be the source of the range of different genotypes found [Mihok *et al.*, 1990]. However little work has been done on the analysis of samples isolated from wild game or from tsetse flies largely due to the practical difficulties involved in collecting such samples. For example, only 0.1% to 1% of tsetse flies are infected with *T. brucei*, so many thousands of tsetse flies must be screened to obtain a reasonable sample size. The pioneering work of Tait, in which sexual recombination was proposed as a mechanism for generating diversity in *T. brucei* [Tait, 1980], was carried out on samples isolated from tsetse flies, whereas most of the work supporting the theory of

clonality has been based on the more homogeneous human or cattle isolates [Tibayrenc *et al.*, 1990]. It is possible that substructuring due to host specificity has lead to an incorrect conclusion that *T. brucei* is clonal whereas, in fact, sexual reproduction is common. There is an obvious need for a large scale analysis of field isolates, sampled correctly from the vector and all hosts from the same place and time to address these issues adequately.

Epidemics and 'new' strains. It has been suggested that epidemics may be caused by the introduction of a 'new' particularly virulent strain of parasite into an area. In a small survey of isolates from the South Nyanza focus in early 1980s, Gibson *et al.*, found that 'new' stocks not previously observed in this region were prevalent [Gibson and Wellde, 1985]. The detection of new strains could be due to the spread of virulent strains from other foci or they could have arisen within the area, perhaps due to changes in the environment which have resulted in a higher transmission of human infective stocks that were already circulating in that area in both man and animals, but were previously undetected due to low abundance and inadequate sample size. Alternatively the new strains could have been generated by genetic exchange or mutation. DNA analysis of stocks from the Tororo district of Uganda sampled in 1960 and 1988-90, and the surrounding geographical area of Busoga in 1982, does not support the idea of the occurrence of new strains but suggests that a single strain present before the epidemic was responsible for the outbreak of disease [Hide *et al.*, 1994; Hide *et al.*, 1996]. This idea is supported by the fact that, in general, epidemics appear to arise in the same foci time after time suggesting that they are caused by pre-existing strains becoming prevalent when conditions within the area are favourable for transmission.

Sampling considerations. It is clear that a great deal still needs to be elucidated regarding the epidemiology of *T. brucei*, and further analysis of field samples is desirable. Because of the difficulties of sampling field isolates often only small sample sizes are available for analysis. Although it is tempting to pool isolates from different hosts, locations and time, this is only valid if there is no population substructuring.

The handling of field samples in the laboratory is also an important consideration. The universal practice of growing field isolates in the laboratory in order to obtain enough material for analysis undoubtedly applies selection for particular stocks and the loss of some genotypes, for example Goebloed reported only a proportion of isolates from salivary glands were successfully grown in laboratory rodents [Goebloed *et al.*, 1973]. The development of markers that can be scored directly by the sensitive PCR technique could enable the direct analysis of field isolates without the requirement of growth in the laboratory, although mixed samples could not be scored for genotypes without isolating single cells and either analysing these directly using single cell PCR or by amplifying in mice prior to analysis.

Another consideration is that the field isolates collected are appropriate for the question under investigation and in sufficient numbers to allow statistical evaluation. For example,

if examining substructuring due to host specificity, the samples must come from different hosts but from the same area and isolated at the same time. The same issue applies to differences due to geography, in that samples must come from the same host, isolated at the same time, from different regions. To date these basic requirements have rarely been satisfied.

When investigating the importance of genetic exchange in the field it may not be appropriate to sample during an epidemic, as the available evidence suggests that a few genotypes predominate. On this basis, a full understanding of the population structure may require sampling endemic areas. The final consideration is that the appropriate typing system be used and this, again, will depend on the questions being addressed.

Typing Systems

A sensitive typing system for *T. brucei,* which would allow strain identity to be established unambiguously, would have a number of applications in the study of trypanosome biology. The identification of particular strains could provide evidence for the ongoing debate on whether trypanosome populations are predominantly clonal or panmictic, as clonally expanded lineages would be obvious. The epidemiology of *T. brucei* infections would also be aided by such strain identification, for example, by identifying if new foci are due to novel genotypes or are as a result of a re-emergence of human infective trypanosomes from the local population.

Human infective trypanosome genotypes could potentially also be distinguished from non-human infective lineages, perhaps revealing that human infectivity has arisen independently in different areas or that all human infective stocks share a common ancestry. Alternatively, the question could be addressed; are *T. brucei rhodesiense* stocks distinct from the local *T. brucei brucei* population or should *T. brucei rhodesiense* samples be considered host range variants of a single subspecies? If human infective strains could be identified by some *T. b. rhodesiense*-specific marker, this would allow local trypanosome populations to be assessed in terms of their potential human infectivity.

Isoenzyme analysis. Isoenzyme analysis, the detection of differences in the electrophoretic mobility of particular enzymes, has been used extensively in the population and genetic analysis of a wide range of organisms, including trypanosomes [for example, Gibson *et al.*, 1980; Godfrey *et al.*, 1990]. This method of analysis can distinguish between some stocks, data from some enzymes can be interpreted genetically and allele frequencies for population analysis can be measured. Those strains with identical isoenzyme patterns are called zymodemes and closely related zymodemes belong to strain groups. However there are drawbacks with using this method of analysis. As large amounts of purified trypanosome material are required, only a proportion of the variation is measured (*i.e.* substitutions which affect protein charge), the markers are not very polymorphic and convergent evolution may result in unrelated stocks having the same

zymodeme pattern. Despite these drawbacks isoenzyme analysis has been the method of choice for most studies regarding the identification of *T. brucei* and although no one marker is capable of distinguishing between *T. b. gambiense* and *T. b. brucei* or *T. b. rhodesiense*, by using several markers in combination, the sub-species *T. b. gambiense* can be reliably identified [Tait *et al.*, 1984; Paindavoine *et al.*, 1989]. Isoenzyme analysis has also revealed the presence of *T. b. rhodesiense*-like trypanosomes in West Africa, termed *T. b. gambiense* type 2 by Gibson [Gibson, 1986]. Although isoenzyme analysis has been unable to distinguish between *T. b. rhodesiense* and *T. b. brucei*, as identical patterns can be found in isolates from both humans and animals, a number of studies have indicated that trypanosomes isolated from humans are more homogeneous than those isolated from other hosts, suggesting that *T. b. rhodesiense* represents stocks which are host range variants of *T. b. brucei* [Tait *et al.*, 1985; Mihok *et al.*, 1990; Cibulskis, 1992; Enyaru *et al.*, 1997].

Restriction fragment length polymorphisms (RFLPs). DNA typing has been used in a number of studies; for example, RFLPs in kDNA were used to type a small collection of samples, the results of which suggested that there is no difference between *T. b. brucei*, *T. b. rhodesiense* and *T. b. gambiense* stocks [Borst *et al.*, 1981]. Clearly these RFLP markers do not have the level of discrimination of alloenzyme markers. Maxi-circle kDNA is now believed to be biparentally inherited with stochastic segregation at each mitotic division leading to fixation for one parental type [Turner *et al.*, 1995], therefore maxi-circle inheritance may not reflect nuclear differences between stocks and variable results could be obtained from typing kDNA depending on the number of mitotic divisions undergone prior to analysis. Taken together, it would appear that kDNA markers are unsuitable for population genetic studies.

RFLPs in nuclear DNA have been used to a limited extent in the analysis of *T. brucei* populations, but have been used to provide further evidence for two types of *T. b. gambiense* stocks in West Africa [Paindavoine *et al.*, 1989]. By using RFLPs in repetitive ribosomal genes Hide *et al.* have generated molecular fingerprints of *T. brucei* stocks [Hide *et al.*, 1994]. Pairwise comparisons of the fingerprint banding patterns were used in cluster analysis to produce dendrograms, illustrating the similarity/differences of different stocks. Several stocks with different DNA patterns were found to have the same isoenzyme pattern, indicating that this method is more sensitive at detecting differences than isoenzyme analysis [Hide *et al.*, 1994]. Although this method reveals high levels of variation, genetic interpretation is not possible as the relationship between specific restriction fragments and the different loci detected has not been established.

Random Amplified Polymorphic DNAs (RAPDs). In recent years the PCR based DNA typing system, RAPDs, in which random fragments of DNA are amplified by PCR, has increased in popularity and has been used to complement alloenzyme analysis [Stevens and Tibayrenc, 1995]. The RAPD technique is quick and easy to perform, detects many

variable fragments of DNA, yet requires no sequence information, and can analyse small amounts of trypanosome DNA. However, RAPD data are often not reproducible, cannot readily be interpreted genetically, are unable to identify mixed or contaminated samples and require purified parasite DNA. Analysing samples from West and East Africa, Mathieu-Daude *et al.*, found this technique more discriminating than isoenzymes and were able to distinguish *T. b. gambiense* type 1 stocks from all other samples, although they were unable to differentiate between *T. b. brucei* and *T. b. rhodesiense* [Mathieu-Daude *et al.*, 1995]. Komba *et al.* (1997) in their analysis of Tanzanian samples combined a number of techniques including RAPD analysis. They also concluded that RAPD analysis revealed more diversity than isoenzymes and that the Tanzanian stocks were distinct from the *T. b. rhodesiense* stocks from other East African foci. RAPD analysis has been used in the study of *T. cruzi*, confirming previous results obtained using isoenzymes that this parasite appears to have a clonal population structure made of two clonal lineages [Tibayrenc *et al.*, 1993].

Karyotype. The full karyotype of the *T. brucei* stock 927/4 has been determined [Melville *et al.*, 1998] and comparisons with other stocks has revealed remarkable variability. For this reason, pulse field gel electrophoresis (PFGE) has been used as a tool to differentiate between some strains [Komba *et al.*, 1997], although this is a time consuming technique which requires large amounts of purified parasite material for analysis.

Multilocus DNA fingerprinting. Classic multilocus DNA fingerprinting is achieved by hybridizing restriction digested DNA to a minisatellite probe (at low stringency) to produce a complex pattern of bands which are individual-specific. This technique has been extensively used in the analysis of human DNA samples [Jeffreys *et al.*, 1985]. It can be used for the analysis of parasite samples not only for the identification of progeny clones but also for distinguishing between different stocks. *T. cruzi* and *Leishmania* species have both been examined in this way using the human minisatellite probe 33.15 [Macedo *et al.*, 1992; Macedo *et al.*, 1992]. Although the high level of diversity revealed by DNA fingerprinting is ideal for individual/strain identification, the high level of intra-population variation revealed by this technique would normally preclude the use of this technique for inter-population comparisons. However, in situations where there are a small number of founders, a severe bottleneck or an inbred or clonal population which reduced the intra-population differences, then the multilocus fingerprinting approach can be applied and can reveal intra- and inter- population differences which are not detected by other systems [Gilbert *et al.*, 1990].

Unfortunately this technique requires large amounts of high molecular weight purified parasite DNA, the fingerprint results cannot be interpreted genetically as many loci are detected at once, and the results can be difficult to interpret if they are from samples containing mixtures of genotypes.

Single Locus Minisatellite Probes. Single locus minisatellite analysis has been used extensively in human genetics to complement the multilocus approach. Although each test is less informative than the classic fingerprint, single locus probe (SLP) analysis offers a number of advantages. Less DNA is required for analysis, the results can be interpreted genetically, and mixed samples can be identified and analysed. Also, by sequential analysis using a series of different probes, highly informative profiles can be obtained [Wong *et al.*, 1987].

Although several genes encoding tandem repeats (*i.e.* coding minisatellites) have been identified in *T. brucei*, no work has been carried out to determine if these loci are polymorphic, with the exception of the *PARP* gene [Mowatt and Clayton, 1988] and *MS42* [Barrett *et al.*, 1997; and Chapter 3]. The use of these coding minisatellites as genetic markers may have a number of applications in the analysis of *T. brucei* populations.

Microsatellites. Polymorphic microsatellites, or short tandem repeats (STRs), which are amenable to amplification by PCR [Weber and May, 1989], have proved extremely useful in the analysis of population genetics and for linkage mapping in a variety of species. These markers (and small minisatellite markers) have many advantages over other typing systems as they can be amplified by PCR and so require very little DNA, they are locus specific and so can be interpreted genetically, are able to detect mixed samples and can be amplified from DNA contaminated with DNA from another species. Polymorphic microsatellite markers were used to analyse the population genetic structure of *T. cruzi* and have provided evidence for clonality and a high incidence of mixed *T. cruzi* infections in animals and insects compared to humans [Oliveira *et al.*, 1998]. Despite the advantages of this typing system, no studies have used these markers in the analysis of *T. brucei* populations although they have been used as markers in laboratory crosses [Sasse, 1998].

The choice of markers for the analysis of *T. brucei* populations should reflect the specific question under consideration. Each typing method has its own advantages and disadvantages. Table 1.4 illustrates some of the typing systems available for analysing populations and their different strengths and weaknesses. For example, RAPDs are quick and simple to perform and do not require any sequence information, but they do not provide any allele frequency data and the interpretation can be confused if samples are mixed or contaminated with host material. Isoenzymes on the other hand can generate allele frequency data and can identify mixed samples, but, these markers are less polymorphic than RAPDs or minisatellites. Highly polymorphic markers are required to distinguish between different strains (intra-species analysis), whereas less polymorphic markers are more suitable for examining the differences between species (inter-species analysis). Highly polymorphic markers would therefore be desirable if a population is suspected of being clonal, as individual strains can be traced through populations. For the analysis of *T. brucei* populations a series of highly polymorphic markers, such as minisatellites, which are among the most polymorphic loci identified to data in eukaryotes,

Typing methods		Informat -iveness	Genotype	Identity	Kinship analysis	intra-species analysis	inter-species analysis	mixed samples	Degraded/ small quantities material	Technical simplicity
Probes	multilocus probes (DNA fingerprinting)	++++	-	++++	+++	++	-	-	-	+
	multi-allelic RFLPs	+++	+	+++	++	++	+	+++	-	++
	RFLPs	+	+	+	+/-	+/-	++	-	-	++
PCR	minisatellite PCR	+++	+	+++	++	++	+	+++	++	+++
	microsatellite PCR	++	+	++	+	+	+	++	+++	+++
	RAPDs	+	-	++	+/-	+	+	-	+++	+++
	MVR-PCR	++++	+	+++	++	++	+	+	++	++
Proteins	isoenzymes	++	+	++	++	++	++	+/-	-	++

Table 1.4. Comparison of different typing methods.

would be desirable to address the population genetics issues of clonality and population sub-structuring.

Minisatellites in Eukaryotes

Tandemly repeated loci are highly abundant in most, if not all, eukaryotic genomes. However, most of the knowledge gained about tandem repeats, and minisatellites in particular, has come from the pioneering work on the analysis of these regions in humans. As a result of this, a large part of this section is based on human repeat loci, but relevant work on parasite genomes will be referred to, where available.

Tandemly repeated regions are present in a wide range of sizes, allowing them to be divided into different classes depending on both the size of the repeat unit and/or the size of the tandem array. However the categorisation of repeat blocks is arbitrary with no clear boundaries between the classes. In fact there is probably a continuum of repeat sizes from small simple repeats to very large midisatellites and satellite arrays. Nonetheless, for reasons of simplicity, the following classes are recognised in this discussion. Microsatellites are short mono-, di-, tri- or tetrameric nucleotide repeats, with arrays of up to 1kb in length; minisatellites are arrays of 1-30kb, usually made up of more complex repeats of 8-100bp in size, whereas midisatellites are much larger with repeat arrays of 50 -200kb [Armour *et al.*, 1993].

Minisatellites. Wyman first observed the phenomenon of multi-allelic restriction fragment length polymorphism (RFLP) in 1980 [Wyman and White, 1980]. Other highly variable RFLPs were subsequently identified flanking a number of human genes including the insulin gene [Bell *et al.*, 1982] and the *Harvey ras* oncogene [Capon *et al.*, 1983]. In some cases the variability was shown to be due to variation in the number of repeats within a tandem array [for example, Bell *et al.*, 1982]. Although these variable regions were recognised as useful markers, there was no efficient systematic approach available for isolating more tandemly repeated loci.

Multilocus detection of minisatellites. The breakthrough came in 1985 when Jeffreys demonstrated that it was possible to detect multiple tandem repeat loci of similar repeat sequence simultaneously on Southern blots by hybridization at low stringency, with probes which contained tandem repeats [Jeffreys *et al.*, 1985]. This produced a complex banding pattern for each human DNA sample. Each probe could detect approximately 17 variable DNA fragments ranging in size from 3.5kb to over 20kb (termed minisatellites for the first time) from one individual. The patterns, which were somatically stable, were shown to be so variable that they could uniquely identify an individual, and became widely known as DNA fingerprints [Jeffreys *et al.*, 1985]. Also, the high degree of germline stability of these complex banding patterns allowed them to be used for relationship testing [Jeffreys *et al.*, 1985] as bands present in a child must have been inherited from either the mother or the father. *De novo* mutations which can generate novel-sized bands can be

directly observed in DNA fingerprinting patterns but they are sufficiently infrequent not to compromise the use of these fingerprint patterns in kinship analysis [Jeffreys *et al.*, 1991].

Other potential applications were quickly realised. In forensic medicine, DNA extracted from a wide range of biological samples could be analysed by multilocus fingerprinting allowing genetic material from a crime scene to be compared to that of a suspect [Gill *et al.*, 1985]. The success of bone marrow transplants can be monitored [Thein *et al.*, 1986] and tumour specific variation, such as large scale rearrangements, can be detected [Thein *et al.*, 1987]. Other applications include the identification of cell lines [Thacker *et al.*, 1988] and of monozygotic twins [Hill and Jeffreys, 1985]. Later, this technique was extended to the identification/paternity testing of other species *e.g.* pigs, marmosets and Waldrapp ibises [Signer and Jeffreys, 1993] and has also been used in the study of protozoa, including *T. cruzi* [Macedo *et al.*, 1992] and *Leishmania* species [Macedo *et al.*, 1992].

Cloning of minisatellite loci. DNA fingerprinting provides DNA patterns rather than genotypes so that information on loci and alleles is unavailable. At high stringency, minisatellite probes can detect their cognate single locus (single locus probes, SLP) on Southern blots of human DNA samples [Jeffreys *et al.*, 1990], allowing the generation of information on locus specific genotypes. The ability of repeat probes, *e.g.* the original Jeffreys' 33.6 and 33.15 probes and other multilocus probes *e.g.* M13 [Vassart *et al.*, 1987], to cross hybridise to many related minisatellite sequences allowed the direct cloning of many more minisatellite loci from humans [Jeffreys *et al.*, 1985; Wong *et al.*, 1986; Nakamura *et al.*, 1987; Wong *et al.*, 1987; Armour *et al.*, 1990] and other species *e.g.* pigs [Signer and Jeffreys, 1993].

Many of these minisatellite loci showed extreme allele length variability with heterozygosities in excess of 90% [Wong *et al.*, 1987], while other loci were invariant. Nakamura distinguished between these two types by calling variable minisatellites, variable number tandem repeats (VNTRs) [Nakamura *et al.*, 1987]. Many of the highly variable minisatellites were used as extremely informative genetic markers. In fact, panels of SLPs were used for individual human identification, kinship analysis, forensic analysis [Wong *et al.*, 1987] and the detection of chromosome abnormalities/allele loss [Vogelstein *et al.*, 1989].

Another important use of SLPs is in linkage mapping [Nakamura *et al.*, 1987]. However minisatellites do not appear to be distributed evenly throughout human chromosomes, but are clustered towards the ends of chromosomes [Royle *et al.*, 1987; Amarger *et al.*, 1998], while such sequences in the mouse genome are more evenly distributed [Jeffreys *et al.*, 1987].

Internal variation in minisatellites. Most (if not all) hypervariable minisatellites which have been characterised in detail have not only shown variation in the length of the repeat array but also variation in the sequence of the tandem repeat unit. Frequently repeat

units within an array carry base substitutions or small deletions/insertions relative to other repeats. These variant repeat types do not appear to be clustered together but are distributed over the entire array of repeats. Thus, each minisatellite allele can be composed of an interspersed mixture of two or more minisatellite variant repeat (MVRs) types. The distribution pattern of MVR types along an allele was first mapped by restriction site mapping for the minisatellite locus, Ms32 [Jeffreys *et al.*, 1990]. More recently a simpler PCR based method for MVR mapping alleles was developed, MVR-PCR [Jeffreys *et al.*, 1991] and has been extend to other human minisatellites [Armour *et al.*, 1993; Neil and Jeffreys, 1993; Urquhart and Gill, 1993], mouse minisatellites [Bois *et al.*, 1998] and to the circumsporozoite gene of the malaria parasite, *Plasmodium falciparum* [Arnot *et al.*, 1993].

Coding minisatellites. Most minisatellites identified to date in humans are non-coding. A few examples of coding variable human minisatellites exist and include the gene for the mucin protein, MUC1, which has a length variation due to different numbers of a 60bp repeat (20 amino acids), resulting in a protein which can range in size from 120 to 225 kDa [Gendler *et al.*, 1990]. Proline rich proteins, usually found in saliva, also show variation in the number of their repeats [Azen *et al.*, 1984]. The gene coding for involucrin (an epidermal protein) is also highly variable [Simon *et al.*, 1991; Urquhart and Gill, 1993]. In parasites the *P. falciparum* genes for many surface antigens have been found to contain variable repeats, for example the circumsporozoite gene [Arnot *et al.*, 1993] and coding minisatellites have also been identified in *T. brucei* [Lee *et al.*, 1990; Lee *et al.*, 1994; Barrett *et al.*, 1997].

Microsatellites. Although short simple tandem repeats have been known in humans for some time and variation in repeat copy number has been identified by sequence analysis there was no easy way of assessing variation at these loci, until the advent of PCR [Saiki *et al.*, 1988] which enabled these regions to be analysed simply and rapidly revealing that many of these loci were polymorphic with heterozygosities of up to 90% [Weber and May, 1989]. Because microsatellites are very abundant and widely distributed with no apparent clustering, they make ideal markers for linkage analysis [Wang and Webber, 1992], and have been used extensively in this respect. Use of such markers has been extended to mapping projects for humans [Dib *et al.*, 1996], *Plasmodium falciparum* [Su and Wellems, 1996] and *T. brucei* [Sasse, 1998; see also Chapter 5].

Mutations in tandem repeats. Although most analysis of the mutation processes involved in tandem repeat turnover has been performed on minisatellites because they can be readily cloned, it is likely that the molecular processes operating on one class of tandem repeat are shared between classes, with varying relativities and may be dependent on factors other than the size of repeat array and the sequence of the repeat units.

Some minisatellites have a high mutation rate *e.g.* Ms32 has a 1% mutation rate to new length alleles allowing measurement of *de novo* mutations in pedigrees [Jeffreys *et al.*,

1991] and by single sperm analysis [Jeffreys *et al.*, 1990; Jeffreys *et al.*, 1994]. This high level of germline mutations is in contrast to the very low mutation frequency in somatic cells. Minisatellite variant repeat mapping of germline specific mutant alleles compared with their progenitor alleles, demonstrated that the mutation events involve complex processes whereby blocks of repeats are copied and transferred from one allele to another. These transfers are usually restricted to one end of the repeat array and appear to be influenced by, as yet unidentified, *cis* acting elements [Monckton *et al.*, 1994]. However, somatic mutations do not appear to be generated in the same way, in that they are mainly intra-allelic events and are not restricted to one end of the minisatellite array.

MVR analysis of mutants from three minisatellites isolated from mice demonstrate that, unlike humans, mouse mutation events involve non-polar intra-allelic mutation events, with a frequency of mutations below 10^{-3} per gamete, much lower than for the human minisatellites analysed to date. A correlation between the size of the minisatellite and the number of alleles at the locus has also been described, suggesting instability is directly related to array size, a phenomenon not observed in human minisatellites [Bois *et al.*, 1998].

The human minisatellites analysed so far are among the most variable loci identified in any organism, although no moderately variable minisatellites have been analysed. It is possible that the mouse minisatellites reflect the mutation mechanisms involved in more common but less variable human minisatellite loci. Indeed one human minisatellite has been recently shown to be similar to the mouse minisatellites in that it mutates in a non-polar fashion [Andreassen and Olaisen, 1998].

Applications of minisatellite markers to* T. brucei *genetics. In the analysis of trypanosome genetics in both laboratory crosses and field studies, research has been hampered by the lack of highly informative molecular markers. The identification of hypervariable minisatellites in *T. brucei* would provide the tools necessary investigate these areas of trypanosome biology. It would also be interesting to study minisatellite repeat turnover in an evolutionary ancient organism such as *T. brucei.*

Aims of this Study

There were several aims to this study:

1) To investigate the mechanisms of genetic exchange in laboratory crosses and to determine the level and rate of recombination between homologous chromosomes.

2) To investigate the role of genetic exchange in natural populations of *T. brucei.*

3) To investigate the population structure of *T. b. brucei* and *T. b. rhodesiense* in relation to host and geography and specifically to address whether *T. b. rhodesiense* is clonal and can be defined as a distinct sub-species.

4) To analyse whether *T. b. rhodesiense* strains are stable over time in a focus of sleeping sickness and from this whether epidemics arise as a result of new strains or are due to the same set of strains.
5) To analyse whether strains of *T. b. rhodesiense* from different geographically separate foci have the same origin or have arisen independently.

The approach taken to address these questions was to use genetic markers that could be applied to both the laboratory crosses and to the population analysis. To address the outstanding questions concerning the mechanisms of genetic exchange, the approach taken was to develop a system for genotyping *T. brucei* so that progeny clones could be genotyped, and, if the genotyping system could be extended to single cells, different life cycle stages within the tsetse fly could be examined to determine the stage at which mating takes place and, possibly, identify intermediates in the mating process. To undertake this approach would require the identification of highly polymorphic markers that could be amplified by PCR. On the basis of studies undertaken in human genetic analysis, mini- and microsatellite markers would be ideal in this respect and so the initial phase of the research was directed at isolating and identifying such markers and developing a system for single trypanosome PCR. These highly polymorphic markers were then used to analyse a large number of trypanosomes derived from genetic crosses.
To address the questions concerning the population structure and biology of *T. b. brucei* and *T. b. rhodesiense* it was necessary to consider whether the genetic markers used to date were appropriate and applicable. For the reasons described in the previous sections, the available markers suffered from a range of disadvantages and so could not be used. While it is probably true to say that there is no class of marker that is ideal for addressing the full range of population genetics questions, minisatellite markers offer many advantages, particularily in relation to their high levels of polymorphism and ability to identify different genotypes with a high level of sensitivity. On this basis, minisatellite markers were chosen to address the questions outlined earlier. Thus the results presented here primarily address the mechanisms and frequency of genetic exchange in both laboratory crosses and natural populations but they also have lead to an evaluation of a class of marker not previously used in trypanosome genetics.

Chapter 2

Materials and Methods

Materials

Chemicals, enzymes and other molecular biology reagents. All reagents used were standard, purchased from established suppliers of molecular biology reagents, (Fisons, FMC Bioproducts, Gibco BRL, New England Biolaboratories, Sigma and Pharmacia).

Oligonucleotides. Oligonucleotides for PCR amplification were synthesized to order by Cruachem Ltd., Glasgow. The sequences of the oligonucleotides described in this thesis are presented in Table 2.1.

T. brucei *DNA samples*. Trypanosome DNA samples were supplied by A. Tait and G. Hide and prepared using standard methods as described in Turner *et al.*, [1990].

***Plasmid sample*.** DNA from plasmid p42Sc3, containing the minisatellite *MS42*, was provided by M. Barrett [personal communication].

T. brucei *stocks*. Trypanosome stocks, either cloned or uncloned, were supplied by A. Tait as stabilates (mouse blood infected with bloodstream form trypanosomes frozen in liquid nitrogen) or as pellets of trypanosomes from procyclic culture. The origins of the *T. brucei* field isolates are given in Table 2.2. The parental stocks of laboratory crosses have been described previously [Jenni et al., 1986; Turner *et al.*, 1990] and are listed in Table 2.3. Clones derived from experimental crosses are listed in Table 2.4, some of the hybrid stocks have been described elsewhere [Sternberg *et al.*, 1989; Turner *et al.*, 1990; Tait *et al.*, 1996]. A series of 'new' hybrid bloodstream clones were generated (by A. Tait) during the course of this work (Table 2.5.), by cloning directly from bloodstream stabilates which were derived from mice infected with metacyclic stage trypanosomes from mixed infected tsetse flies.

***Tsetse flies*.** Teneral tsetse flies (*Glossina morsitans*) were provided by I. Maudlin and S. Welburn.

Mice stocks. Irradiated (600 rads) MF1 mice were used to amplify trypanosome stocks.

Primer	5′ - 3′ sequence
CRAM-C	gcaacgaggcaggtccaatctttg
CRAM-D	ctcctccatcataacctcgttgtc
CRAM-G	ctgctgatgccgtacatgatgatttc
CRAM-H	aactccctcccgatcgatcacaac
292-C	acggaagcagtgcgggtagttaag
292-D	atcacttgccgatgatgtatcacc
292-G	acacccccctctccacttcagatac
292-H	gctgaacctgtgggcccctcaattg
MS42-A	cttctccacaatttctgcc
MS42-B	cgcgttcaaaaaatatggcc
MS42-BR	ggccatatttttgaacgcg
MS42-C	ggatttcttcatagcgaaggcattc
MS42-D	aacttcccacttgtgcggtcgtta
MS42-CR	gaatgccttcgctatgaagaaatcc
MS42-F	ttgtgcggtcgttaacgcgcgttcaa
MS42-F8	agttgttgccaggacagcaac
MS42-F8R	acaaggttagttgctgtcctgg
MS42-H2	aaatgcactagccacatgtgactc
MS42-*Hinc*II+	cattccgttgcgtttttgaagtcaac
MS42-K	gtcgaacatctcgctcctcctttg
MS42-KR	gacgacagcattgccgtaaggag
MS42-L	cgaatttggagggttcacaggttc
MS42-LR	gaacctgtgaaccctccaaattcg
MS42-*Mbo*II+	gcgttgactgagatatgcagcgaa
MS42-Q	gaagggcgttcaggcattcgttc
MS42-QR	gaacgaatgcctgaacgcccttc
MS42-R	attggaggacatgcgctctgccc
MS42-RR	gggcagagcgcatgtcctccaat
MS42-S	cttaccgctgctcatggtcagg
MS42-SR	cctgaccatgagcagcggtaag
MS42-T	aaccgctgcacgactaagcttcac
MS42-V	cattattccacggacgcgaagcagc
MS42-W	ggtgattcatcggctcccttacca
MS42-WR	tggtaagggagccgatgaatcacc
TIM-A	ccacgcatcatccgcagctg
TIM-B	aacacccccctattgttccctctcc
TIM-C	caacttactggggacgctgctatc
TIM-D	ctacactctcttttcctctcccag
TIM-E	tgccgttgagtgggtgaagatagc
TIM-F	ctccctgctacctgtctttacatc
JS2-A	gattggcgcaacaactttcacatacg
JS2-B	ccctttcttccttggccattgttttactat
JS2-C	agtaatgggaatgagcgtcaccag
JS2-D	gatcttcgcttacacaagcggtac
D2-A	ggaagtgaggggagacggaagac
D2-B	cggcaggggaagggagaa
E5-A	atgaagcaaagacaccttctctcc
E5-B	tggctgatgtgttagttccgcagc
F3-A	gtccgttcacccacacaactaaag
F3-B	tttcggcagtctagttaggcatcc

Table 2.1. Oligonucleotide primer sequences.

Table 2.2

Origin	Sample identification no.	Other identification numbers	Cloned/ uncloned (no. of passages)	Host	HS	Year
Zambia-Luangwa	Z 210		clone	man	R	1982
Z-L	Z 269		clone	d.a.	ND	1983
Z-L	Z 90		clone	man	R	1981
Z-L	Z 274		clone	man	R	1983
Z-L	Z 194		clone	man	R	1982
Z-L	Z 218		clone	man	R	1982
Z-L	Z 222		clone	man	R	1982
Z-L	Z 208a		clone	man	R	1982
Z-L	Z 199	GUP 2540	clone	man	R	1982
Z-L	Z 220	GUP 2590	clone	man	R	1982
Z-L	Z 273	GUP 2560	clone	d.a.	ND	1982
Z-L	Z 203	GUP 2469	clone	man	R	1982
Z-L	Z 221	GUP 2548	clone	man	R	1982
Z-L	Z 231	GUP 2491	clone	man	R	1982
Z-L	Z 244	GUP 2492	clone	man	R	1982
Z-L	Z 267	GUP 2546	clone	d.a.	ND	1983
Z-L	Z 212		clone	man	R	1982
Z-L	Z 270		clone	tsetse	ND	1983
Z-L	Z 185		clone	man	R	1982
Kenya-Nyanza	N 97	EATRO 97, GUP 1052	clone	man	R	1961
K-N	N 96	EATRO 96, GUP 1051	clone	man	R	1961
K-N	N 2340	EATRO 2340, GUP 2498	clone	man	R	1977
K-N	N 156	EATRO 156, GUP 791	clone	man	R	1961
K-N	N 95	EATRO 95, GUP 1043	clone	man	R	1961
K-N	N 94	EATRO 94, GUP 1042	clone	man	R	1961
K-N	N 106	EATRO 106, GUP 2077	clone	man	R	1961
K-N	N 116	EATRO 116, GUP 2078	clone	man	R	1961
K-N	N 7	EATRO 7, GUP 795	clone	tsetse	ND	1961
K-N	N 110	EATRO 110, GUP 2088*	clone	man	R	1961
K-N	N 98	EATRO 98, GUP 1074	clone	man	R	1961
K-N	N 111	EATRO 111	clone	man	R	1961
K-N	N 112	EATRO 112	clone	man	R	1961
K-N	N 115	EATRO 115	clone	man	R	1961
K-N	N 502	EATRO 502	clone	man	R	1961
K-N	N 113	EATRO 113	clone	man	R	1961
K-N	N 120	EATRO 120	clone	man	R	1961
K-N	N 605	EATRO 605	clone	man	R	1961
K-N	N 609	EATRO 609	clone	man	R	1961
K-N	N 102	EATRO 102, GUP 2088*	clone	man	R	1961
K-N	N 105	EATRO 105, GUP 2079	clone	man	R	1961
K-N	N 149	EATRO 149, GUP 784	clone	man	R	1961
K-N	N 148	EATRO 148, GUP 788	clone	man	R	1961
K-N	N 18	EATRO 18, GUP 793	clone	tsetse	ND	1961
K-N	N 118	EATRO 118, GUP 2089	clone	man	R	1961
K-N	N 103	EATRO 103, GUP 790	clone	man	R	1961
Uganda-Busoga	B EA 174	EATRO 174, GUP 1075	clone	man	R	1959
U-B	B EA 3	EATRO 3, GUP 1301	clone	tsetse	R	1960
U-B	B UTAR 3	GUP 2067	clone	man	R	1981
U-B	B UTAR 4	GUP 2069	clone	man	R	1982
U-B	B EA 2274	EATRO 2274, GUP 1011	clone	man	R	1976
U-B	B papol 60		uncloned	cattle	S	1990
U-B	B B 76		uncloned	cattle	S	1988
U-B	B B 135		uncloned	cattle	S	1988
U-B	B I 155		uncloned	cattle	S	1988

Table 2.2 continued

Origin	Sample identifica -tion no.	Other identification numbers	Cloned/ uncloned (no. of passages)	Host	HS	Year
U-B	B M 12		uncloned	cattle	S	1990
U-B	B UgC90		uncloned	man	ND	1990
U-B	B UgE90		uncloned	man	ND	1990
U-B	B U89/8		uncloned	man	R	1989
U-B	B UgL		uncloned	man	R	1988
U-B	B UgA90		uncloned	man	R	1990
U-B	B EO		uncloned	man	R	1990
U-B	B UgI		uncloned	man	R	1988
U-B	B UgK		uncloned	man	R	1988
U-B	B URI		uncloned	tsetse	R	1988
U-B	B UgJ		uncloned	man	R	1988
U-B	B Fly 48		uncloned	tsetse	ND	1990
U-B	B U89/2		uncloned	man	R	1989
U-B	B UgB90		uncloned	man	ND	1990
U-B	B UgM		uncloned	man	R	1988
U-B	B MAP		uncloned	man	R	1988
U-Busia	B 3194	KETRI 3194	uncloned	man	R	1989
U-Busia	B 3196	KETRI 3196	uncloned	man	R	1989
U-Busia	B 3200	KETRI 3200	uncloned	man	R	1989
U-Busia	B 3202	KETRI 3202	uncloned	man	R	1989
U-Busia	B 3203	KETRI 3203	uncloned	man	R	1989
U-B	B M31		uncloned	cattle	ND	1990
U-B	B M32		uncloned	cattle	S	1988
U-B	B M42		uncloned	cattle	ND	1990
U-B	B M66		uncloned	cattle	R	1988
U-B	B M80		uncloned	cattle	S	1990
U-B	B M85		uncloned	cattle	ND	1990
U-B	B papol 33		uncloned	cattle	S	1990
U-B	Bpapol103		uncloned	cattle	R	1988
U-B	B S14		uncloned	cattle	S	1990
U-Busia	B 3205	KETRI 3205	uncloned	man	R	1989
U-Busia	B 3206	KETRI 3206	uncloned	man	R	1989
U-B	B B 23		uncloned	cattle	S	1988
U-B	B B 25		uncloned	cattle	ND	1988
U-B	B I 147		uncloned	cattle	S	1988
U-B	B Mag 18		uncloned	cattle	S	1988
U-B	B Mag 40		uncloned	cattle	S	1988
U-B	B M3		uncloned	cattle	S	1988
U-B	B S28		uncloned	cattle	S	1990
U-B	B S38		uncloned	cattle	ND	1990
Kenya-Kiboko	K 1337	LUMP 1337, K10	clone	tsetse(*G.pall*)	ND	1969
K-K	K 1008	TREU 1008, K 29	clone (3-11)	tsetse(*G.pall*)	ND	1970
K-K	K 1027	TREU 1027, K2	uncloned(3-11)	tsetse(*G.pall*)	ND	1969
K-K	K 1027	TREU 1027	clone	tsetse(*G.pall*)	ND	1969
K-K	K 927	TREU 927	uncloned(3-11)	tsetse(*G.pall*)	ND	1970
K-K	K 927cl1B	TREU 927cl1B	clone	tsetse(*G.pall*)	ND	1970
K-K	K 927cl4B	TREU 927cl4B	clone	tsetse(*G.pall*)	ND	1970
K-K	K 927cl4	TREU 927cl4	clone	tsetse(*G.pall*)	ND	1970
K-K	K 927cl5	TREU 927cl5	clone	tsetse(*G.pall*)	ND	1970
K-K	K 927cl9	TREU 927cl9	clone	tsetse(*G.pall*)	ND	1970
K-K	K 984	K21	uncloned(3-11)	tsetse(*G.pall*)	ND	1969
K-K	K 984		clone	tsetse(*G.pall*)	ND	1969
K-K	K 854	TREU 854, K4	uncloned(>8)	tsetse(*G.pall*)	ND	1969
K-K	K 854	TREU 854	clone	tsetse(*G.pall*)	ND	1969
K-K	K 975	K18	uncloned(3-11)	tsetse(*G.pall*)	ND	1970
K-K	K 975		clone	tsetse(*G.pall*)	ND	1970

Table 2.2 continued

Origin	Sample identification no.	Other identification numbers	Cloned/ uncloned (no. of passages)	Host	HS	Year
K-K	K 925		uncloned(3-11)	tsetse(*G.pall*)	ND	1969
K-K	K 925		clone	tsetse(*G.pall*)	ND	1969
K-K	K 982	K38	clone(3-11)	tsetse(*G.pall*)	ND	1970
K-K	K 853		clone	tsetse(*G.pall*)	ND	1969
K-K	K 1009	TREU 1009, K31	uncloned(3-11)	tsetse(*G.pall*)	ND	1970
K-K	K 1009	TREU 1009	clone	tsetse(*G.pall*)	ND	1970
K-K	K 851	TREU 851, K2	uncloned(9)	tsetse(*G.pall*)	ND	1969
K-K	K 851	TREU 851	clone	tsetse(*G.pall*)	ND	1969
K-K	K 852	TREU 852, K10	uncloned(6)	tsetse(*G.pall*)	ND	1969
K-K	K 852	TREU 852	clone	tsetse(*G.pall*)	ND	1969
K-K	K 978	K37	uncloned(3-11)	tsetse(*G.pall*)	ND	1970
K-K	K 978		clone	tsetse(*G.pall*)	ND	1970
K-K	K 997	mouse K269	uncloned(3-11)	tsetse(*G.pall*)	ND	1969
K-K	K 997		clone	tsetse(*G.pall*)	ND	1969
K-K	K 258	LUMP 258, K9	uncloned	tsetse(*G.pall*)	ND	1969
K-K	K 258	LUMP 258	clone	tsetse(*G.pall*)	ND	1969
K-K	K 936	TREU 936	uncloned(3-11)	tsetse(*G.pall*)	ND	1969
K-K	K 926	TREU 926	uncloned(3-11)	tsetse(*G.pall*)	ND	1969
K-K	K 869	K2	uncloned	tsetse(*G.pall*)	ND	1969
K-K	K 994	K36	uncloned(3-11)	tsetse(*G.pall*)	ND	1970
K-K	K 994		clone	tsetse(*G.pall*)	ND	1970
K-K	K 981	K4	uncloned(3-11)	tsetse(*G.pall*)	ND	1969
Uganda-Lugala	L 834	TREU 834, EATRO 1253	uncloned(14)	tsetse(*G.pall*)	ND	1969
U-L	L 844	TREU 844, EATRO 1317	uncloned(6)	tsetse(*G.pall*)	ND	1969
U-L	L 929	TREU 929, EATRO 1321	uncloned(5)	tsetse(*G.pall*)	ND	1969
U-L	L 933	TREU 933, EATRO 1644	uncloned(>9)	tsetse	ND	1970
U-L	L 934	TREU 934, EATRO 1497	uncloned(8)	tsetse(*G.pall*)	ND	1970
U-L	L 941	TREU 941, EATRO 1386	uncloned(9)	tsetse(*G.fux*)	ND	1969
U-L	L 832	TREU 832, EATRO 1242	uncloned	tsetse(*G.pall*)	ND	1969
U-L	L 845	TREU 845, EATRO 1448	uncloned(10)	tsetse(*G.pall*)	ND	1970
U-L	L 791	TREU 791, EATRO 1297	uncloned(13)	tsetse(*G.pall*)	ND	1969
U-L	L 836	TREU 836, EATRO 1279	uncloned(7)	tsetse(*G.pall*)	ND	1969
U-L	L 944	TREU 944, EATRO 1581	uncloned(10)	tsetse(*G.pall*)	ND	1970

Table 2.2. List of all stocks used in this study. Column 1 indicates the country and region from which each sample was isolated. Z-L, Zambia-Luangwa; K-N, Kenya-Nyanza; U-B, Uganda-Busoga; K-K, Kenya-Kiboko; U-L, Uganda-Lugala. Column 2 shows the identification number used throughout this study. Column 3 gives additional identification numbers such as stabilate numbers and those used by other research groups. TREU, Trypanosome Research Edinburgh University; EATRO, East African Trypanosomiasis Research Organisation; LUMP, London University Medical Protozoology; KETRI, Kenyan Trypanosomiasis Research Institute; GUP, Glasgow University Parasitology. Column 4 indicates whether an isolate is cloned or uncloned and the number of passages in mice before analysis. Column 5 gives details of the host from which each sample was isolated. *G. pall, Glossina pallidipes; G.fux, Glossina fuxcipes;* d.a., domestic animal. Column 6 indicates human serum sensitivity. R, resistant; S, sensitive; ND, not determined. Two criteria were used to identify human serum resistance/sensitivity. Samples isolated from humans were considered human serum resistant. Parasites isolated from non-human hosts were assayed for human serum resistance using the *in vitro* test of Brun and Jenni (1987) and the results published in Hide *et al* 1994. Column 7 gives the year in which each sample was isolated. The Luangwa stocks were described in [Godfrey *et al.*, 1990], the Nyanza stocks were isolated by the East African Trypanosomiasis Research Organisation, the Busoga stocks by Hide [Hide *et al.*, 1994] and the Kiboko and Lugala stocks were isolated by Goebloed [Goebloed *et al.*, 1973].

Details of parental stocks				
Stock number	**Place of isolation**	**Host**	**Serum Resistance**	**Year of isolation**
STIB 247	Tanzania	hartebeest	S	1971
STIB 386	Ivory Coast	man	R	1978
TREU 927/4	Kenya	tsetse (*G.pall*)	I	1969

Table 2.3. Details of the origin of each parental stock described in Turner *et al.*, [1990]. Human serum (HS) resistance data was provided by G. Lindergard based on an *in vivo* assay [Lindergard, 1999]. R, resistant; S, sensitive and I; intermediate; *G. pall, Glossina pallidipes.*

Table 2.4.

Trypanosome clones derived from genetic crosses		
Clones previously described		
Cross	**Clone**	**GUP number (if available)**
STIB 247 x TREU 927/4	F124/28 bscl A1	
	F124/28 bscl B3	
	F124/28 bscl A6	
	F124/28 bscl C5	
	F532/53 mcl 1	3114
	F532/63 bscl 2	
	F532/63 bscl 3	
	F532/63 bscl 5	
	F532/63 bscl 7	
	F532/63 bscl 8	
	F532/72 mcl 1	3128, 4391
	F532/72 mcl 2	3129
	F532/72 mcl 3	3130
	F532/72 mcl 4	3131
	F532/72 mcl 5	3132, 4392
	F532/72 mcl 6	3133, 4393
	F532/72 mcl 7	3134
	F532/72 mcl 9	3136, 4366
	F532/72 mcl 10	3135, 4360
	F124/28 bscl B1	
	F124/28 bscl C3	
	F124/28 bscl C2	
	F974/70 mcl 1	3083
	F974/70 mcl 2	3084
	F974/70 mcl 3	3085
	F974/70 mcl 4	3086
	F974/70 mcl 6	3088
	F974/70 mcl 7	3089
	F974/70 mcl 8	3090
	F974/78 mcl 1	3092
	F974/78 mcl 2	3093
	F974/78 mcl 3	3094
	F974/78 mcl 4	3095
	F974/78 mcl 6	3096
	F974/78 mcl 7	3097
STIB 386 x TREU 927/4	F296/44 bscl 1	3199
	F296/44 bscl 2	3200, 4261
	F296/44 bscl 3	3201
	F296/44 bscl 4	3204
	F296/44 bscl 5	3203
	F296/44 bscl 6	3210
	F296/44 bscl 7	3202
	F296/44 bscl 8	3205
	F296/44 bscl 9	3206
	F296/44 bscl 11	3208
	F296/44 bscl 12	3211
	F296/44 bscl 13	3209
	F296/46 mcl 1	3196
	F296/46 mcl 12	3197
	F296/56 mcl 1	3218
	F296/56 mcl 2	3219
	F296/56 mcl 3	3212
	F296/56 mcl 4	3213
STIB 386 x TREU 927/4	F296/56 mcl 6	3215
	F296/56 mcl 5	3214
	F296/56 mcl 7	3220
	F296/56 mcl 8	3216
	F296/56 mcl 9	3217

Table 2.4 continued

Trypanosome clones derived from genetic crosses		
	Clones previously described	
Cross	Clone	GUP number (if available)
STIB 247 x STIB 386	723VI-L	
	723CAB	
	F9/28 mcl 3	2846
	F9/34 mcl 1	
	F9/34 mcl 2	
	F9/45 mcl 2	3300
	F9/45 mcl 4	4295
	F9/45 mcl 7	
	F9/45 mcl 9	3290
	F9/45 mcl 10	3291
	F9/45 mcl 11	3287
	F9/45 mcl 12	3288, 3296
	F57/50 mcl 2	
	F18/50 mcl 4	
	F18/50 mcl 8	
	F19/31 mcl 2	
	F19/31 mcl 3	
	F492/50 mcl 1	2834
	F492/50 mcl 4	2838
	F492/50 mcl 5	2839
	F492/50 mcl 6	2840
	F492/50 mcl 7	2841
	F492/50 mcl 8	2847
	F492/50 mcl 9	2852
	F492/50 mcl 10	2853, 4312
	F492/50 mcl 11	2854, 4364
	F492/50 mcl 12	2843
	F492/50 mcl 13	2856
	F492/50 mcl 15	2857, 3365
	F492/50 mcl 16	2858, 3363
	F492/50 mcl 17	2859
	F492/50 mcl 18	2860, 3364
	F492/50 mcl 19	
	F492/50 mcl 20	2861
	F492/50 mcl 21	
	F492/50 mcl 22	
	F492/50 mcl 23	
	F492/50 mcl 24	2848
	F492/50 mcl 25	2855
	F492/50 mcl 26	2862

Table 2.4. Previously isolated trypanosome clones derived from genetic crosses. The crosses from which each clone is derived is given in column 1. Column 2 indicates the hybrid clone identification number used throughout this study. Columns 2 gives the Glasgow University Parasitology (GUP) number for the stabilates. Clone nomenclature follows that described by [Sternberg *et al.*, 1989]. For example F492/50 bscl 25, was the twenty-fifth clone isolated from fly 492, on day 50 post infection.

Table 2.5

Trypanosome clones derived from genetic crosses	
Newly generated clones	
Cross	**Clones**
STIB 386 x TREU 927/4	F296/39 bscl 1
	F296/39 bscl 2
	F296/39 bscl 4
	F296/39 bscl 5
	F296/39 bscl 6
	F296/39 bscl 7
	F296/39 bscl 8
	F296/39 bscl 9
	F296/39 bscl 10
	F296/39 bscl 11
	F296/39 bscl 12
	F296/39 bscl 13
	F296/39 bscl 14
	F296/39 bscl 15/1
	F296/39 bscl 16
	F296/39 bscl 17
	F296/39 bscl 18
	F296/39 bscl 19
	F296/39 bscl 27
	F296/39 bscl 22/1
	F296/39 bscl 22/2
	F296/39 bscl 23/3
	F296/39 bscl 24/1
	F296/39 bscl 26/2
	F296/39 bscl 32
	F296/39 bscl 40
	F296/39 bscl 41
	F296/39 bscl 44
	F296/39 bscl 47
	F296/39 bscl 57
	F296/39 bscl 58
	F296/39 bscl 59
	F296/42bscl 26
	F296/42bscl 29
	F296/42bscl 41
	F296/42bscl 44
	F296/42bscl 47
	F296/42bscl 48
	F296/42bscl 49
STIB 247 x STIB 386	F492/50 bscl 1
	F492/50 bscl 2
	F492/50 bscl 3
	F492/50 bscl 4
	F492/50 bscl 5/1
	F492/50 bscl 6
	F492/50 bscl 7
	F492/50 bscl 8
	F492/50 bscl 9
	F492/50 bscl 11
	F492/50 bscl 12
	F492/50 bscl 13
	F492/50 bscl 14
	F492/50 bscl 15
	F492/50 bscl 16
	F492/50 bscl 17
	F492/50 bscl 18
	F492/50 bscl 19
	F492/50 bscl 20

Table 2.5 continued

Trypanosome clones derived from genetic crosses	
Newly generated clones	
Cross	**Clones**
STIB 247 x STIB 386	F492/50 bscl 21
	F492/50 bscl 22
	F492/50 bscl 23
	F492/50 bscl 25/1
	F492/50 bscl 27

Tables 2.5. Newly generated trypanosome clones derived from genetic crosses. The crosses from which each clone is derived is given in column 1. Column 2 indicates the hybrid clone identification number used throughout this study. Clone nomenclature follows that described by Sternberg *et al.*, [1989]. For example F492/50 bscl 25/1, was the first reclone (/1) of the twenty-fifth clone isolated from fly 492, on 50 days post infection.

Methods

Trypanosome and tsetse fly manipulations

Optical cloning. Clones were made by direct observation of single parasites in a drop of blood from an infected mouse diluted in guinea pig serum in a well of a humidified Terasaki plate. Immediately after the presence of a single cell was detected, the trypanosome was removed in 20µl of 50% PBSG (phosphate buffered saline/1% glucose)/50% guinea pig serum and injected into an irradiated MF1 mouse. The parasites were harvested at peak parasitaemia (10^7-10^8 parasites/ml) for subsequent PCR analysis. Cryopreserved stocks (stabilates) of the first peak parasitaemias were generated by the addition of DMSO to a final concentration of 7% and storing in liquid nitrogen. All optical cloning was performed by A. Tait.

Isolation of single trypanosomes. Each single trypanosome was observed in a single drop of blood from an infected mouse diluted in guinea pig serum in a well of a humidified Terasaki plate. The single trypanosome was then transferred to a PCR micro-tube by the addition of 10µl of 1 x PCR buffer (as described in Chapter 4).

Tsetse flies and cyclical transmission. Teneral tsetse flies were fed through a silicone membrane on thawed stabilates of bloodstream form trypanosomes of parental clones STIB 386 and TREU 927/4 diluted in defibrinated sheep blood to 3 x 10^7 trypanosomes/ml. The two stocks were mixed in a 1:1 or 3:1 ratio immediately before feeding to flies. Fly handling and maintenance were performed as described by Schweizer *et al.*, [1988]. Briefly, infected flies were maintained in individual tubes and membrane fed three times a week on uninfected horse or sheep blood. From day 21 onwards, flies were allowed to probe (spit) onto warm (35°C) microscope slides to enable the detection of salivary gland infections by microscopy. Positive fly probes were then analysed by PCR to identify the trypanosome genotype present in the saliva. Midgut trypanosomes were obtained by fly dissection (performed by M. Hope).

Generation of clones from stabilates. Bloodstream stabilates derived from mice infected with metacyclic trypanosomes from mixed infected flies were thawed and examined for live trypanosomes. Single parasites were optically identified and cloned directly from these stabilates (by A. Tait) using the method described above. The clones are listed in Table 2.5.

Molecular Biology Techniques

Crude lysate and DNA preparation. Crude lysates from infected mouse blood were prepared as follows: 500µl of PBS was added to 500µl of infected mouse blood and centrifuged at 2500g for 5 minutes, after which 750µl of the supernatant was discarded.

This washing process was repeated three times. The final parasite/blood pellet was then resuspended in 50µl lysis buffer (50mM Tris-HCl pH8, 100mM EDTA 0.5% SDS, 0.64mgml^{-1} proteinase K), incubated overnight at 56°C and stored at 4°C. The lysates were then diluted 1/100 in deionised water and the proteinase K was heat inactivated at 95°C for 5 minutes. One µl of lysate was then used as a template for each of the subsequent PCR reactions. DNA of purified parasites from infected mouse blood were prepared as described by Turner *et al.*, [1990] and provided by A. Tait.

DNA handling. General methods for handling DNA, gel electrophoresis, Southern blotting *etc.* were performed as described by [Sambrook *et al.*, 1989].

Enzymatic manipulations. DNA restriction enzymes and other modifying enzymes were used according to the manufacturers' instructions with the supplied buffer systems.

Preparative gel electrophoresis. For the preparation of size fractionated DNA, the DNA was electrophoresed on agarose gels in 0.5 x TBE buffer (45mM Tris-HCl, 45mM Boric acid, 0.125mM EDTA) containing 0.5µgml^{-1} ethidium bromide and the required size fraction or band was excised from the gel under UV illumination. The gel slice was then loaded onto a Spin-X column (Costar) and centrifuged at 15000 rpm for 5 minutes. The solution was then diluted 1/20 or 1/50 before subsequent amplification by PCR.

General PCR [Saiki et al., 1988]. Due to the sensitivity of the polymerase chain reaction, precautions were taken to ensure that reagents and materials used for PCR were kept free of contaminating DNA, *i.e.* PCR dedicated reagents were used and pipette tips and tubes were taken directly from the manufacturers packaging to minimise exposure to the laboratory environment. Preparation of PCR reagents, single genome dilutions and single molecule PCR reactions were performed in a laminar flow hood to prevent aerial contamination. All PCR reactions were performed in conjunction with the appropriate zero DNA controls, which consistently gave no products.

All PCR reactions, unless otherwise stated, were performed in 10µl reaction volumes in 45mM Tris-HCl pH 8.8, 11mM $(NH_4)_2SO_4$, 4.5mM $MgCl_2$, 6.7mM 2-mercaptoethanol, 4.4µM EDTA, 113µgml^{-1} BSA, 1mM each of the four deoxyribonucleotide triphosphates, 1µM of each oligonucleotide primer and one unit of Amplitaq Polymerase (Perkin Elmer, Cetus USA) using a template of either 5ng of genomic DNA or 1µl of diluted crude lysate. Reaction mixtures were overlaid with mineral oil and amplifications were carried out in a Robocycler gradient 96 (Stratagene). The cycling conditions, unless stated otherwise in the text, were as follows: 96^{0}C for 50s, 64^{0}C for 50s and 70^{0}C for 180s, for a total of 28 cycles. PCR products were separated by electrophoresis in a 1% Seakem agarose gel in 0.5 x TBE buffer or a 4% Nusieve gel for

products under 800bp, and visualised by ethidium bromide staining (0.5µg/ml) and UV illumination.

PCR amplification from fly probes. Positive fly probes containing metacyclic stage trypanosomes were identified by microscopic examination. Deionised water (10µl) was then added to the spot of saliva on each microscope slide, mixed with the dried saliva and transferred to a PCR micro-tube where it was heated to 95°C to disrupt the cells. One µl of each sample was amplified using primers specific for the *TIM* as described in Chapter 5.

PCR amplification from single trypanosomes. The nested PCR approach to the amplification of single trypanosomes is described in full in Chapter 4 and MacLeod *et al.*, [1997].

Minisatellite Variant Repeat (MVR) -PCR. The MVR-PCR technique was developed for the *T. brucei* minisatellite, *MS42*, to specifically amplify variant minisatellite repeats and was adapted from the method of [Jeffreys *et al.*, 1991]; the principles of the technique are described in Chapter 8. The gel extracted DNA from each allele was MVR mapped as follows. One µl of a 1/20 or 1/50 dilution of DNA was amplified by PCR in a 10µl reaction using 0.01µM of *MS42*-TAG-A or *MS42*-TAG-G primer, and a high concentration (1µM) of primers TAG and *MS42*-W. The primers are described in Table 2.1. The cycling conditions were; 50s at 95°C, 50s at 65°C and 3 min at 70°C for 18 cycles. PCR products were electrophoresed through a 40 cm long 1% Seakem agarose gel in 0.5 x TBE, 0.5µgml^{-1} ethidium bromide. DNA was denatured, transferred by Southern blotting onto Magna membrane (MSI) and hybridized to the ^{32}P-labelled *MS42* repeat probe overnight. The preparation and labelling of the *MS42* repeat probe is described in the DNA hybridization section. Autoradiography was for 24 hours at -70°C.

Small Pool (SP) -PCR. The SP-PCR method for the detection of novel sized *MS42* alleles was adapted from [Jeffreys *et al.*, 1994] and is described in Chapter 8. Briefly, limited quantities of DNA (serial dilutions ranging from the equivalent of 1 genome to 100 genomes) were amplified under standard PCR conditions using primers *MS42*-W and *MS42*-F, for 24 cycles. The PCR products were electrophoresed though a 40 cm 1% Seakem agarose gel. The DNA was denatured, transferred by Southern blotting onto Magna membrane (MSI) and hybridized to ^{32}P-labelled *MS42* repeat probe overnight. Autoradiography was for 24 hours at -70°C.

Estimation of the number of input amplifiable molecules per SP-PCR reaction. For every SP-PCR experiment, 12 PCR reactions for each dilution of DNA were amplified. For reactions at high dilutions of DNA, the number of successful amplifications was determined. From this number Poisson analysis was used to determine

the number of amplifiable input molecules in each reaction, which was then used to extrapolate the number of input molecules in reactions containing lower dilutions of DNA.

Southern blotting. Standard procedures were used for the preparation of agarose gels for Southern blotting. Briefly, for genomic blots, 5µg of total genomic DNA was digested with the appropriate enzyme and fractionated by agarose gel electrophoresis. The gel was washed in 0.25M HCl for 2 x7 minutes, followed by 1.5M NaCl, 0.5M NaOH for 2 x 15 minutes and 1.5M NaCl, 0.5M Tris-HCl pH 8.0 for 2 x 15 minutes. The gel was transferred to a platform which was covered with a layer of 3MM filter paper, the ends of which were immersed in 20 x SSC (3M NaCl, 0.3M tri-sodium citrate pH 7.0). The gel was overlaid with a nylon membrane (MSI Magna membrane), several layers of dry filter paper and a weight. In this way the DNA was transferred to the nylon membrane by capillary action. After transfer the DNA was cross-linked to the membrane by placing the membrane on a UV transilluminator for 45 seconds.

DNA hybridization. *MS42, 292* and *CRAM* probes were generated by amplifying these loci from genomic DNA by PCR using primers *MS42*-A and *MS42*-B, *292*-C and *292*-D, and *CRAM*-C and *CRAM*-D, respectively (for primer sequences see Table 2.1), under standard PCR conditions. The PCR products were fractionated by electrophoresis on a low melting point 1% agarose gel and gel slices containing the DNA fragments were excised from the gel. The gel slices were melted (60°C for 5 minutes) and diluted to a final concentration of 2ng/µl with deionised water. Probes (20ng) for DNA hybridization were labelled using the random hexamer priming kit, Prime-it (Stratagene), incorporating α-^{32}P-dCTP using the method described by the manufacture. All Southern blots were hybridized to probes in Church and Gilbert hybridization buffer (7% SDS, 0.5M Na_2HPO_4 pH 7.0, 1mM EDTA) [Church and Gilbert, 1984], at 65°C in a rotating bottle hybridization oven (Hybaid). Filters were washed at high stringency (0.1 x SSC, 0.1% SDS) for 30 minutes to 1hour at 65°C. Autoradiography was performed as described [Sambrook *et al.*, 1989] with exposure times of 1 hour to 1 week depending on the band intensity required.

DNA sequencing. DNA sequencing reactions were performed in accordance with the ABI protocol for automatic sequencing. Sequencing electrophoresis was performed by the Molecular Biology Sequencing Unit within the University.

Computing. DNA sequencing analysis was performed using the Genetic Computer Group Sequence Analysis Software Package, developed at the University of Wisconsin, run on a UNIX mainframe computer. Digital MVR data was analysed with software written by J. A. L. Armour run on a Apple Macintosh personal computer [Armour *et al.*, 1996]. Analysis of allele frequencies was performed on a personal computer using the Genetic Data Analysis program written by P. Lewis [Lewis and Zaykin, 1999].

The methods used during the course of this work followed standard molecular biological procedures, which have largely been described elsewhere, except in some specific instances. Thus the materials and methods outlined here do not contain an exhaustive list of standard protocols. Instead only brief descriptions are given of the general methods used, with minor modifications and exact experimental conditions used being described, in context, within each results chapter.

Chapter 3

Polymorphic Loci in *T. brucei*

Introduction

A number of questions remain to be addressed regarding the basic biology, genetics and epidemiology of the parasitic protozoan, *T. brucei,* as outlined in Chapter 1. This is particularly true of the sexual recombination process, regarding both the mechanisms of genetic exchange and the frequency of the sexual process in the field. In order to address these issues an effective genotyping system for *T. brucei* must be available, so that individual isolates of *T. brucei* can be readily identified thus allowing, for example, progeny clones to be distinguished from parental types.

A number of different typing systems have been used to analyse stock variation in *T. brucei.* The most extensively used technique, isoenzyme analysis, has been employed for the past two decades to study the relationships between trypanosomatids [Gibson *et al.*, 1980; Gibson and Gashumba, 1983; Tait *et al.*, 1984; Tait *et al.*, 1985; Godfrey *et al.*, 1990; Tait, 1990; Turner *et al.*, 1990; Tait *et al.*, 1993; Tait *et al.*, 1996]. In recent years the PCR based DNA typing system, RAPDs (randomly amplified polymorphic DNAs), has been increasingly applied to *T. brucei* genetics and has been used to complement isoenzyme analysis [Gibson *et al.*, 1995; Stevens and Tibayrenc, 1995]. Restriction fragment length polymorphisms (RFLPs) have also been used to a limited extent in the analysis of *T. brucei* genomes, for example to distinguish F1 hybrids from parental trypanosomes [Sternberg *et al.*, 1989], however such RFLPs display limited variability. RFLPs in repetitive ribosomal genes can reveal greater variability, and have been used in the analysis of trypanosome populations, generating molecular fingerprints [Hide *et al.*, 1994], although genetic interpretation is not possible with the data generated in this way.

Hypervariable minisatellites or variable number tandem repeat (VNTR) loci in the *T. brucei* genome, which vary in a strain-specific manner, should provide a means of identifying and tracking individual strains as well as allowing the allele and genotype frequencies of *T. brucei* populations to be determined. The use of locus-specific primers to PCR amplify microsatellite and small minisatellite markers should enable the genotyping of trypanosomes even when contaminated with large quantities of host DNA in addition to allowing the analysis of small quantities of DNA by PCR amplification. Isoenzyme, RFLP

or RAPD analysis requires the amplification of parasites in mice to generate enough material for analysis, however with trypanosome specific PCR based methods this can be circumvented. Because of their high level of polymorphism, minisatellite markers are particularly useful in determining variation between populations, analysing progeny from crosses and detecting heterogeneity within a sample (mixed stocks). This high level of variability allows strains to be tracked through populations and should be useful in identifying trypanosomes which are responsible for epidemics. However, the high degree of variability of minisatellite markers limits the use of these markers to intra-species analysis.

The aim of the work reported in this chapter was to identify and characterize highly polymorphic mini- and/or microsatellites which were small enough to be easily amplified by PCR and which were informative for crosses between the three stock STIB 247, STIB 386 and TREU 927/4, so that they could be used for the analysis of genetic crosses (see Chapter 5).

Results

Tandemly repeated regions are present in the *T. brucei* genome, some of which have already been characterised; for example, the genes encoding the procyclic acidic repeat protein (PARP)[Mowatt and Clayton, 1988]. A literature and database search revealed several single copy sequences which had known repetitive regions, although none of these repetitive loci had been analysed for allele length polymorphisms. Those sequences which contained repetitive regions small enough to be readily PCR amplified were chosen for further analysis. At the time of the searches, the database consisted primarily of coding sequence as the *T. brucei* sequencing project had not been initiated, and thus the tandemly repeated sequences found were in genes.

CRAM. The sequence of the single copy gene from *T. brucei*, encoding the cysteine-rich acidic integral membrane protein, *CRAM*, has been described previously from EATRO 427 and shown to contain 66 copies of a 36bp motif [Lee *et al.*, 1990]. Using the published sequence, a pair of primers flanking the repeated region were designed. These primers were used initially to amplify, by PCR, samples of DNA from stocks STIB 247, STIB 386 and TREU 927/4 (Figure 3.1 A and B). Two amplified fragments of different sizes were detected in stock STIB 386 and TREU 927/4 and presumed to represent allelic variation in the number of repeat units, with the two stocks being heterozygous for different sized alleles, (stock STIB 247 produces one band and so is presumed to be homozygous at this locus). In order to demonstrate that these bands were in fact different alleles, F1 progeny clones from a cross between STIB 386 and TREU 927/4 were analysed by PCR analysis for the *CRAM* marker (Figure 3.2A). It is clear that each F1 progeny clone has inherited one band from each parent. The results confirm that the two bands are alleles that segregate in the F1 progeny and are therefore allelic size variants.

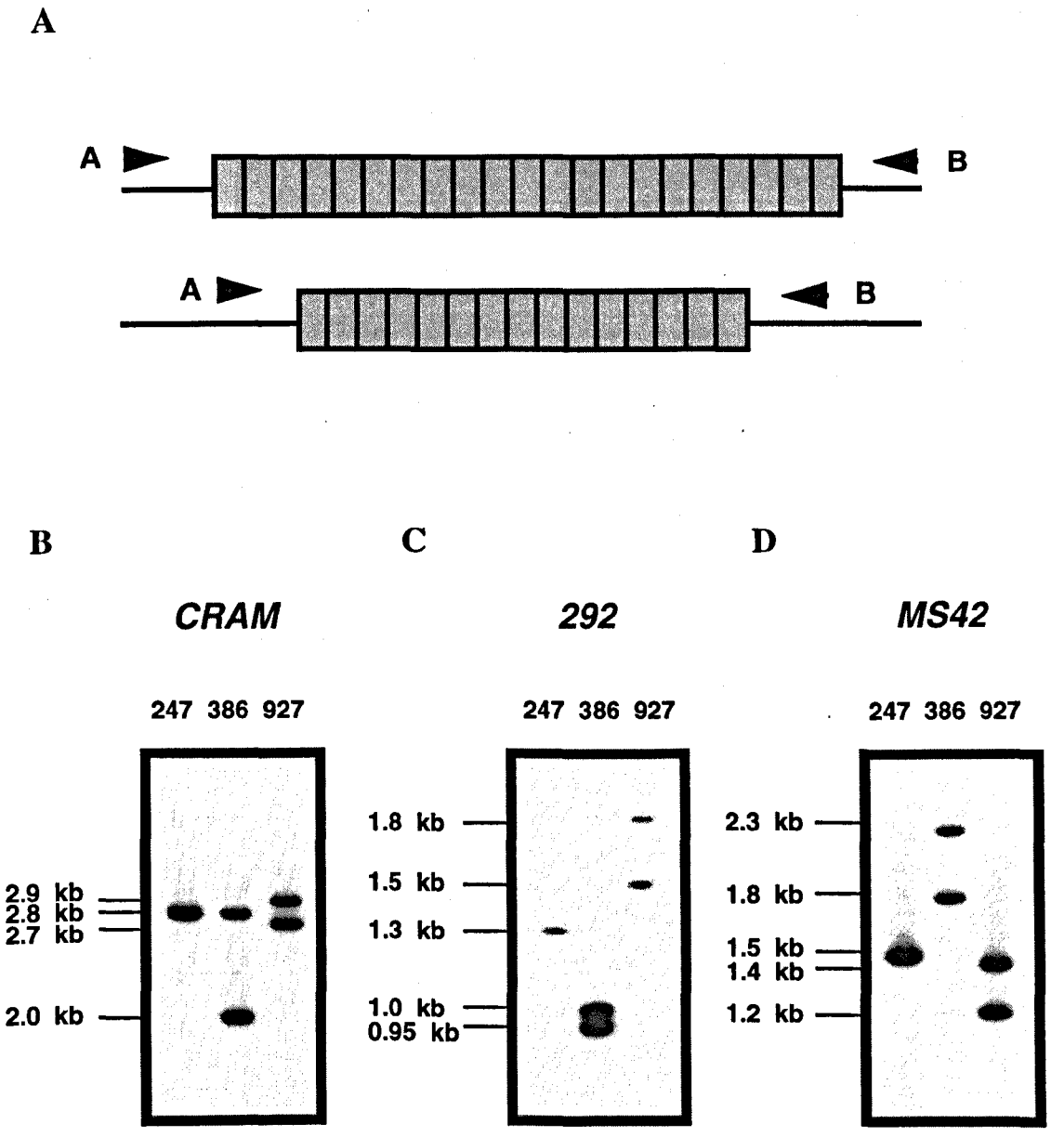


Figure 3.1. Amplification by PCR of minisatellite loci from the stocks STIB 247, STIB 386 and TREU 927/4.

(**A**) Diagrammatic representation of amplification of a minisatellite locus. Two alleles of different sizes are drawn as rectangles. Boxes represent repeat units, PCR primers, A and B, designed to the DNA flanking the repeats, are drawn as arrowheads.

PCR amplification of the minisatellites, (**B**) *CRAM,* (**C**) *292* and (**D**) *MS42* from DNA isolated from the stocks STIB 247, STIB 386 and TREU 927/4, using primer pairs *CRAM*-G/*CRAM*-H, *292*-G/*292*-H, and *MS42*-W/*MS42*-F, respectively. PCR conditions and primer sequences were as described in Materials and Methods. PCR products were separated on a 1% Seakem agarose gel and visualised by ethidium bromide staining. Band sizes estimates were obtained by comparison to λ*Hin*dIII and ϕx *Hae*III markers (not shown).

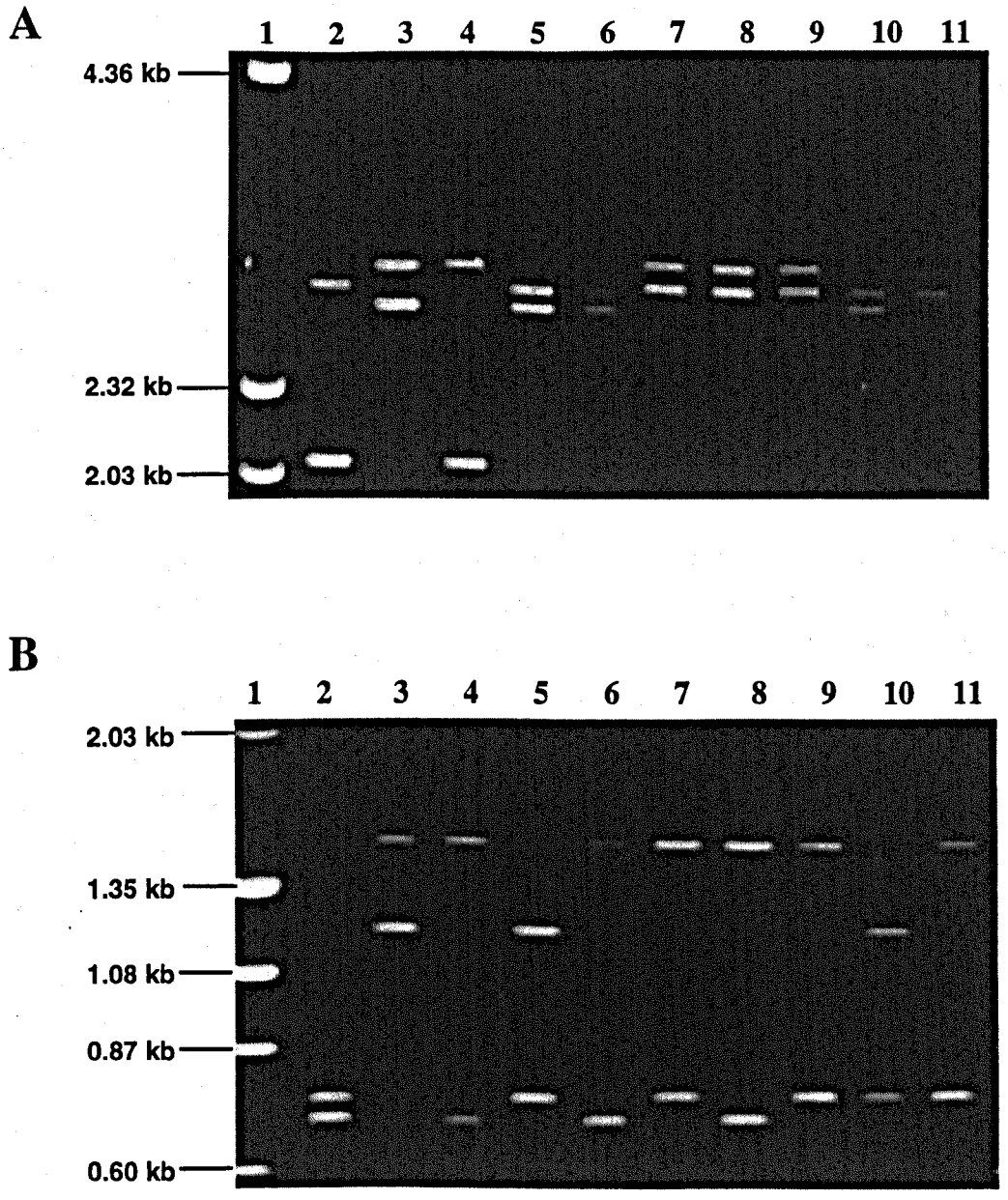


Figure 3.2. Mendelian inheritance of minisatellite markers. Amplification by PCR of the variable regions of the *CRAM* and *292* genes from DNA isolated from stocks STIB 386, TREU 927/4 and hybrid clones derived from a STIB 386 x TREU 927/4 cross. PCR conditions and primer sequences are described in Materials and Methods. PCR products were separated on a 1% Seakem agarose gel and visualised by ethidium bromide staining.
(**A**) Amplification by PCR of the *CRAM* locus, using primers *CRAM*-G and *CRAM*-H, from the parental stocks STIB 386 and TREU 927/4 and 8 progeny from a cross between these stocks.
(**B**) Amplification by PCR of the *292* locus, using primers *292*-G and *292*-H, from the same material as in (**A**).

Lane 1; Lambda *Hind* III and ϕx *Hae* III markers (Advanced Biotechnologies). Lanes 2-11; STIB 386; TREU 924/4; F296/44bscl1; F296/44bscl4; F296/39bscl7; F296/44bscl3; F296/44bscl12; F296/39bscl22/1; F296/44bscl8; F296/39bscl2.

292. A similar approach was taken in the analysis of the gene encoding the membrane-associated protein *292* [Lee *et al.*, 1994]. The cloned gene from *T. brucei*, (EATRO 427) contained a repeated region, consisting of 47 copies of a 24bp repeat unit. By designing primers flanking the repeats, the repeated region of this locus was amplified by PCR from stocks STIB 247, STIB 386 and TREU 927/4 (Figure 3.1C). As with the *CRAM* locus, two stocks, STIB 386 and TREU 927/4 showed two bands differing in size consistent with these stocks being heterozygous for allelic length variants at this locus. PCR amplification of this variable region from the same F1 progeny of a cross between stocks STIB 386 and TREU 927/4 also demonstrates that these variable size PCR products are in fact different alleles which are inherited in a classical Mendelian manner (Figure 3.2B).

STIB 247 appeared to be homozygous for both *CRAM* and *292* which was confirmed by analysis of F1 progeny from STIB 247 and either STIB 386 or TREU 927/4 crosses (see Chapter 5).

MS42. The minisatellite, *MS42*, was originally cloned fortuitously (plasmid pTGR3) by M. Barrett from EATRO 427, in experiments designed to isolate the 6-phosphogluconate dehydrogenase gene, *GNU*, of *T. brucei* [Barrett and Le Page, 1993; Barrett *et al.*, 1997]. The plasmid pTGR3, was found to comprise almost exclusively copies of a 42bp repeat sequence, named *MS42*. This repeated region, unlike most minisatellites isolated in other systems *e.g.* in humans [Wong *et al.*, 1987], was short (~1.6kb) and so was amenable to PCR amplification. Approximately 50bp on either side of the repeat region had been sequenced (Accession number X70187) allowing locus specific PCR primers to be designed to the DNA flanking the repeats [Barrett *et al.*, 1997], which could then be used to amplify the repeated region. This locus, containing tandem repeats, was amplified from genomic DNA of the three stocks, STIB 247, STIB 386 and TREU 927/4, (Figure 3.1D). It is clear from this figure that alleles at this locus vary in size between the different stocks, and that two bands were generated during the PCR reaction in the stocks STIB 386 and TREU 927/4, suggesting these stocks were heterozygous for this locus, while STIB 247, which gave a single PCR product, was homozygous.

Analysis of a genomic Southern of stock EATRO 427, digested with various restriction endonucleases, and of chromosomes separated by pulsed field gel electrophoresis (PFGE), followed by hybridization to the *MS42* repeat sequence [Barrett *et al.*, 1997], revealed that the *MS42* locus is a single copy sequence, hybridizing to the two homologues of chromosome I. A genomic Southern blot of parental stocks STIB 247 and TREU 927/4 and F1 progeny from a STIB 247 and TREU 927/4 cross, revealed different sized bands probably due to allelic variation in the number of repeat units within the tandem array [Barrett *et al.*, 1997]. To further demonstrate that this variation was indeed due to allelic polymorphism, F1 progeny clones a cross between stocks STIB 386 and TREU 927/4 were analysed by a genomic Southern blot of *Pst* 1 digested DNA and hybridized to the *MS42* repeat sequence (Figure 3.3), revealing that each allele was a different size and was

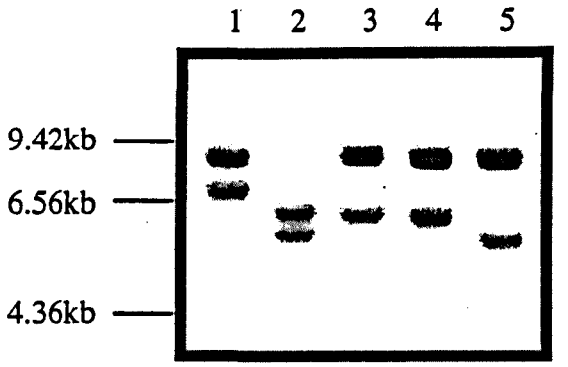


Figure 3.3. Genomic Southern of DNA from STIB 386, TREU 927/4 and 3 F1 progeny clones probed with *MS42*. 5µg of total genomic DNA was digested with *Pst* I, electrophoresed through a 1% Seakem agarose gel, blotted onto MSI Magna membrane and hybridized to a *MS42* repeat probe. The sizes given are from λ *Hin*dIII markers (not shown). Lane 1, STIB 386; lane 2, TREU 927/4; lane 3, F296/44bscl 1; lane 4, F296/44bscl 3; lane 5, F296/44bscl 8.

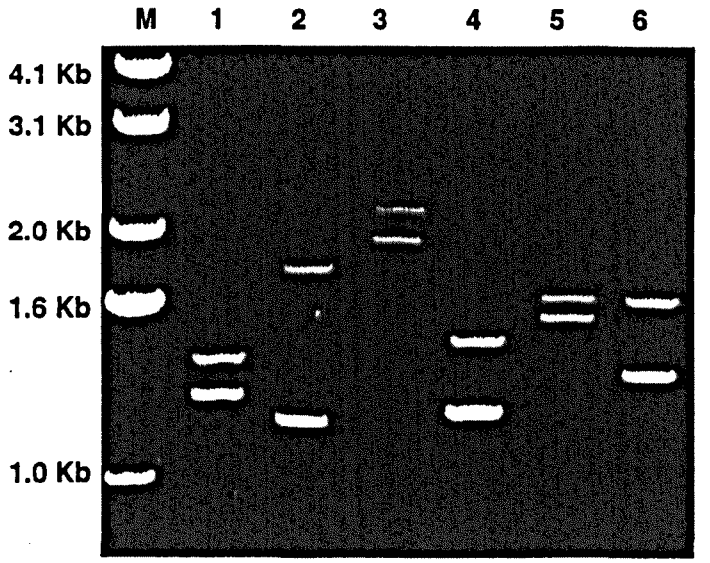


Figure 3.4. Amplification by PCR of a range of *MS42* alleles from different *T. brucei* isolates. The PCR conditions and oligonucleotide sequences (*MS42*-W and *MS42*-F) were as described in Materials and Methods. PCR products were separated on a 1% Seakem agarose gel, stained with ethidium bromide and visualised by ultra violet transillumination. Lane 1, K984; lane 2, K997; lane 3, Z218; lane 4, Z210; lane 5, N96; lane 6, N7. Lane M, 1 Kb ladder (BRL). Prefix K, Kiboko, Kenya; N, Nyanza, Kenya and Z, Luangwa Valley, Zambia.

inherited in a Mendelian fashion. For example, progeny clone F296/44bscl 8 (lane 5, Figure 3.3), appears to have inherited the upper allele of STIB 386 and the lower allele of TREU 927/4.

It is also clear from the patterns of allele inheritance that the alleles of the three minisatellite loci are inherited independently from each other (see Chapter 5). This independent assortment of alleles is expected for these loci, as they are located on different chromosomes: *MS42* on chromosome I; *CRAM* on chromosome X, and *292* on either chromosome III or V [Barrett *et al.*, 1997; Lee *et al.*, 1990; Lee *et al.*, 1994; Hope *et al.*, *in press*]. A similar analysis, for all three loci, was undertaken with the progeny from other crosses, *i.e.* between stocks STIB 247 and TREU 927/4 and stocks STIB 247 and STIB 386. The data (described in full in Chapter 5) are again consistent with Mendelian segregation, further supporting the conclusion that the size differences in the repeat regions of these genes represent allelic variation at each of the two loci.

To investigate if these loci were highly polymorphic and therefore could be useful markers for distinguishing different field isolates, each locus was amplified from a small collection of cloned stocks isolated from Kiboko and Nyanza in Kenya and Luangwa in Zambia. A number of distinct alleles were identified on the basis of differences in size of the PCR products of the *MS42* locus (Figure 3.4) [Barrett *et al.*, 1997], indicating that this locus was extremely polymorphic, varying in a strain specific manner. Similar results were obtained for *CRAM* and *292* minisatellites. Extensive analysis of all three loci in a large collection of field isolates is presented in Chapters 6, 7 and 8.

Minisatellite stability. Minisatellite sequences in other systems have been shown to have a high mutation rate, generating novel sized alleles [Jeffreys *et al.*, 1994]. In order to test the mitotic stability of *MS42* and *CRAM*, 33 lines derived from single cloned stocks were analysed. These lines had been passaged as procyclic forms, bloodstream forms, transmitted through tsetse flies, and selected for resistance to either melarsen or suramin. No alteration in the repeat length was observed in any of the cloned lines (Table 3.1)[Barrett *et al.*, 1997]. Further analysis of mutation processes involved in the generation of *MS42* variants is presented in Chapter 8.

Sequencing of MS42. Sequence analysis of plasmid, pTGR3, containing *MS42*, revealed the 42bp repeat unit [Barrett *et al.*, 1997]. Like many minisatellites, the repeat unit is not precisely reiterated over the tandem array, but contains variant repeat units. Figure 3.5A gives the repeat sequence including the most common variable position, an A-G polymorphic site at one end of the repeat unit.

Single stock fly transmissions					
Clones	GUP no.	Lysate no.	Markers		Comments
			CRAM	*MS42*	
Reference 927/4			3-4	3-4	
Reference 247			1-1	5-5	
Reference 386			1-2	1-2	
927/4 -F222 84/31mcl 2	3305	117	nd	3-4	Parental
927/4 -F222 84/31mcl 3	3306	118	nd	3-4	Parental
927/4 -F222 84/31mcl 4	3307	119	nd	3-4	Parental
927/4 -F222 84/31mcl 5	3308	120	nd	3-4	Parental
927/4 -F222 84/31mcl 7	3309	121	nd	3-4	Parental
927/4 -F222 84/31mcl 1	3310	122	nd	3-4	Parental
927/4 -F222 84/31mcl 6	3311	123	nd	3-4	Parental
247- F125 mcl 1	3104	125	1-1	5-5	Parental
247- F125 mcl 2	3105	127	1-1	5-5	Parental
247- F125 mcl 3	3106	128	1-1	5-5	Parental
247- F125 mcl 4	3107	129	1-1	5-5	Parental
386- F62/47 mcl 1	3266	135	1-2	1-2	Parental
386- F62/56 mcl 1	3270	136	1-2	1-2	Parental
386- F62/56 mcl 2	3271	137	1-2	1-2	Parental
386- F62/56 mcl 3	3272	138	1-2	1-2	Parental
386- F62/56 mcl 4	3273	139	1-2	1-2	Parental
386- F62/56 mcl 6	3275	141	1-2	1-2	Parental
386- F62/56 mcl 7	3276	142	1-2	1-2	Parental
386- F62/56 mcl 8	3277	143	1-2	1-2	Parental
386- F62/56 mcl 10	3294	145	nd	1-2	Parental
386- F62/56 mcl 11	3278	144	nd	1-2	Parental

Long term passaged clones						
Clones	GUP no.	Lysate no.	No. of passages	Markers		Comments
				CRAM	*MS42*	
247	3280	156	30	nd	5-5	Parental
247	3228	130	49	1-1	5-5	Parental
247 (culture)		157	30	nd	5-5	Parental
247 Melarsen R		132	16	1-1	5-5	Parental
247 Melarsen R	3650	131	16+ fly trans.	1-1	5-5	Parental
247 Suramin R		133	9	1-1	5-5	Parental
927/4	3163	124	30	nd	3-4	Parental
386 Suramin R		150	13	nd	1-2	Parental
386	3247	149	52 + fly trans.	nd	1-2	Parental
386 Melarsen R		151	7	nd	1-2	Parental

Table 3.1. Minisatellite analysis of clones from long term passaged and single stock fly transmissions. The clone identification number used throughout this study and additional identification numbers *i.e.* stabilate and lysate numbers, are given. The results of minisatellite analysis is shown, alleles are numbered 1-5 for the minisatellites, *CRAM* and *MS42*. The interpretation of results for each clone is given. For details of drug resistant lines see [Scott *et al.*, 1996]. Fly trans., fly transmitted; GUP, Glasgow University Parasitology; R, resistant; nd, not determined.

Plasmid pTGR3 contained very little DNA flanking the *MS42* repeat units. In order to obtain more sequence surrounding the repeats, plasmid pTGR3 was used as a probe to screen a cosmid library, which was made from a STIB 247 x TREU 927/4 hybrid, clone F532/72 mcl 7 [M. Barrett, personal communication]. This hybridized to cosmid clone, F8, which was digested with *Pst* I and the fragments subcloned into PUC9. One of those subclones hybridized to the pTGR3 plasmid, and so was believed to contain the *MS42* minisatellite. This plasmid clone, p42Sc3, had a larger insert (~4.2kb) which contained the minisatellite sequence *MS42* and more of its flanking DNA [M. Barrett, personal communication].

In an attempt to characterize the *MS42* locus and obtain more flanking DNA sequence, the insert of plasmid clone, p42Sc3, a gift from M. Barrett, was sequenced using an automated ABI sequencing protocol. The sequencing strategy involved sequencing using the forward and reverse primers specific for the vector, PUC9, and designing specific *MS42* primers from the newly generated sequence data, as it was obtained. In this way, the DNA flanking the *MS42* repeats was sequenced [with the technical help of Y. Shafi and the Molecular Biology Service Unit (MBSU)]. Only seven complete *MS42* repeat units within the plasmid were sequenced although, from the size of the insert, it was estimated that 26 repeats were present. The sequence of the repeat units obtained from p42Sc3 was identical except for position 1, which could be either an A or G (see Figure 3.5A), and was similar to the repeat sequence obtained from plasmid pTGR3 although they were isolated from different stocks (F532/72mcl 7 and EATRO 427, respectively). More detailed analysis of repeat sequence variation is given in Chapter 8. Figure 3.5B gives a simple map of the plasmid p42Sc3 clone, with the relative positions of the *MS42* specific primers used to sequence the insert. The primer sequences are given in Material and Methods, (Chapter 2, Table 2.1), and the sequence of plasmid clone p42Sc3, with seven complete repeat units, is given in the Appendix (Figure A1).

The orientation of the sequence was not known and, prior to the sequence of p42Sc3 being completed, 5' and 3' labels were applied arbitrarily to respective ends of the insert. The 5' end was arbitrarily defined as the side of the repeat array where the primer sequence *MS42*-W was situated. This labelling system was continued throughout the following analysis and all subsequent chapters (in particular Chapter 8).

MS42 Open Reading Frame (ORF). The stock EATRO 427 has been shown to carry two different sized alleles for *MS42* [Barrett *et al.*, 1997]. Northern blots of RNA from the bloodstream stage of stock EATRO 427 probed with the *MS42* sequence, have revealed two different sized transcripts (approximately 4.2Kb and 4.9Kb), which could represent transcripts of the different sized alleles [M. Barrett, personal communication].

Analysis of the sequence data of p42Sc3, using the Genetic Computer Group (GCG) sequence analysis programs, developed at the University of Wisconsin, indicated that one large ORF spanned almost the entire sequence (see Figure 3.5 for diagram). The predicted

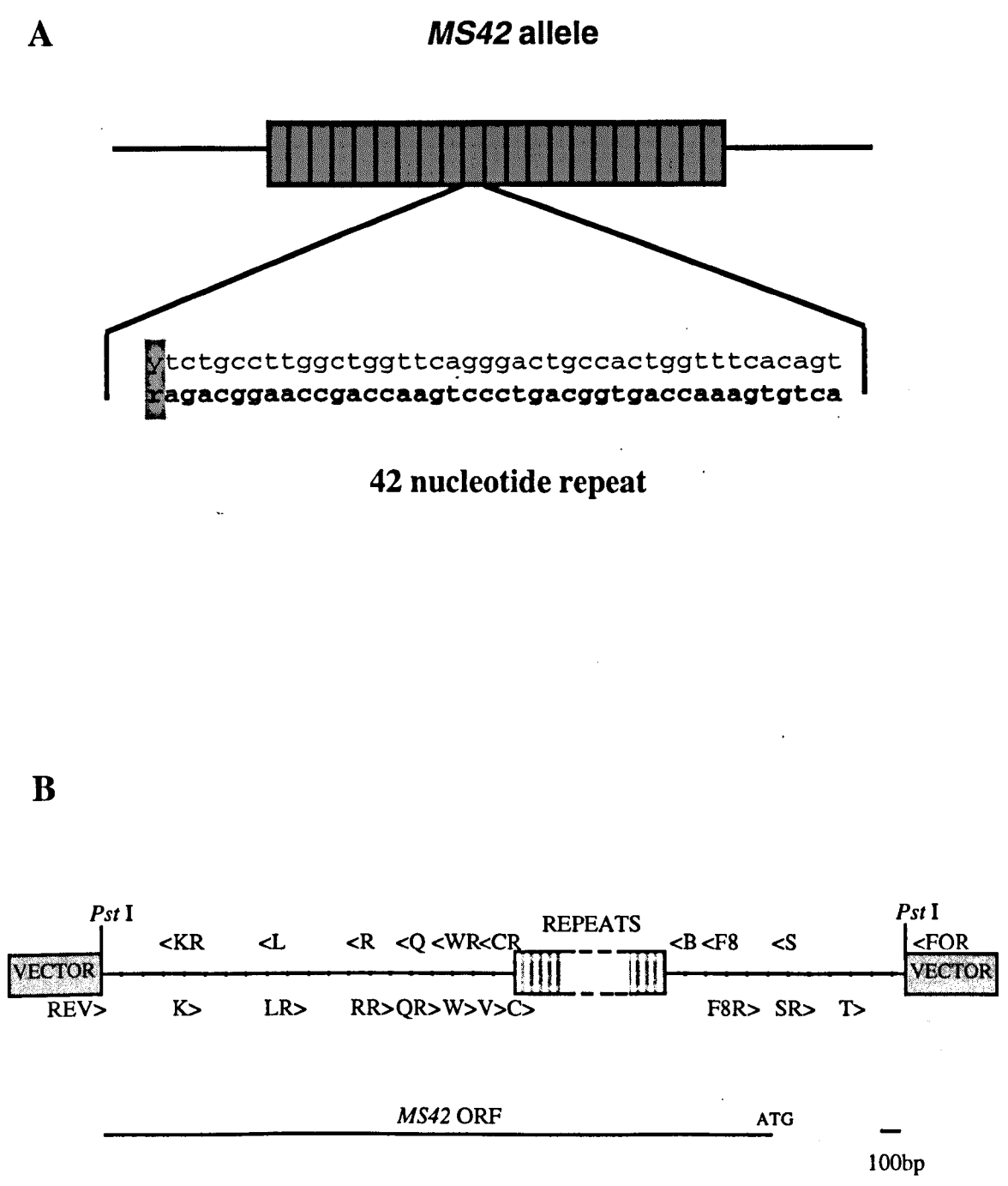


Figure 3.5. Diagrammatic representation of *MS42*.
(A) The repeat array is represented as a large rectangle, with each repeat unit being a shaded box. Both strands of the 42 bp sequence of the repeat unit are given, with the coding strand in bold. The A-G transition site at one end of the repeat unit is highlighted, r indicates either a or g; y indicates t or c.
(B) A 4.2kb fragment containing the minisatellite *MS42* was cloned into the *Pst* I site of PUC9 [Barrett *et al.*, 1997], creating the clone p42Sc3. The diagram illustrates the insert of the clone p42Sc3 and the positions of the *MS42* primers used for sequencing and PCR amplification. Arrowheads indicate the direction of the primers. FOR and REV primers are the universal primers designed to the vector. *MS42* primer sequences are given in Material and Methods. The repeat array is represented by a rectangle. The ORF is drawn below the map.

amino acid sequence of this ORF is given in the Appendix (Figure A2). The ORF consists of 2757bp (excluding the repeats), starts at position 3116 on the p42Sc3 sequence (Appendix, Figure A1) and continues for the length of the insert (to position 1). No stop codon within the insert was detected, thus it is unlikely that the entire ORF has been sequenced. The predicted amino acid sequence of the repeats are proline rich (3 prolines out of 14 amino acids).

The length of the ORF from the upper allele of stock EATRO 427, which was estimated to contain 38 repeats, has a minimum length of 4.35Kb (*i.e.* at least 2757bp of non-repetitive DNA and 38 x 42bp of repeats), not allowing for any 3' untranslated DNA. Since the estimated band size from a Northern blot was 4.9Kb [M. Barrett, personal communication], it was concluded that approximately 500bp of the ORF and untranslated sequence had not been cloned or sequenced.

Similarity searches of the p42Sc3 DNA sequence against the EMBL database did not identify any similar sequences from any organism. Protein searches, however, using the predicted amino acid sequence, found similarity to a number of proteins in the EMBL database. All of the proteins showed similarity to the repeated region and no other part of the sequence. The repeated region has similarity to other genes which contain proline rich repeats. Two proteins with the best matches over the repeat region are, myosin-light-chain kinase (from rabbit) and an unknown protein from *Mycoplasma hyopneumoniae*, which contain repeat units of 12 and 10 amino acids in length, respectively. The repeat units both contain a ET-KPVA motif, but the functional significance of this is unclear.

Discussion

In this chapter, the first polymorphic minisatellites identified in *T. brucei* have been described. These minisatellite markers have many distinct advantages over typing systems previously used for the analysis of trypanosomes; they are highly polymorphic and generate information which can be interpreted genetically. As these minisatellites are in coding regions of the genome, and therefore presumably have a functional significance, they may be under selection. For this reason, it is preferable to use several unlinked loci in any typing system and not to rely on one locus.

This chapter describes how three previously identified genes containing repeats, which demonstrate allelic polymorphism, were exploited to develop a genotyping system. Other genes containing tandem repeats were considered for analysis, but were rejected if they were not single copy, or were too large to be easily amplified by PCR.

A more systematic (but more time consuming) approach to identifying minisatellite loci would be to screen genomic libraries (at low stringency) with minisatellite sequences, for example Jeffreys' probes, 33.6 and 33.15 [Jeffreys *et al.*, 1985]. Alternatively, with more genomic sequence becoming available from the *T. brucei* genome project, computer

programs designed to identify tandem repeats could be used quickly and easily to find more mini- and microsatellite loci [Benson, 1999].

Chapter 4

Genotyping Single Trypanosomes

Introduction

The analysis of experimental crosses of two genotypically different trypanosomes has been limited to only a few F1 hybrids cloned either directly from metacyclics in the salivary glands of a mixed-infected tsetse fly or from bloodstream infections of mice initiated by the bite of an infected tsetse fly. The isolated cloned parasites were then grown in mice to generate enough parasites for examination by isoenzymes, RFLPs and PFGE [Jenni *et al.*, 1986; Sternberg *et al.*, 1987; Wells *et al.*, 1987; Schweizer *et al.*, 1988; Sternberg *et al.*, 1988; Gibson, 1989; Sternberg *et al.*, 1989; Sternberg and Tait, 1990; Turner *et al.*, 1990; Gibson and Garside, 1991; Gibson *et al.*, 1992; Gibson and Whittington, 1993; Gibson and Bailey, 1994; Schweizer *et al.*, 1994; Degan *et al.*, 1995; Gibson *et al.*, 1997]. Although the results of this analysis demonstrated that genetic exchange can occur, it did not determine where or when trypanosome mating takes place or the frequency of cross and self-fertilization.

Direct analysis of individual metacyclic cells present in the salivary glands of tsetse flies sampled over a time course of infection could determine the proportion of trypanosomes present in the salivary glands that are hybrid, parental or products of self-fertilization and so indicate the extent of cross and self-fertilization. Also, examination of single cells from other life cycle stages *e.g.* proventricular and epimastigote forms could help identify the stage at which recombination occurs and whether this involves a haploid trypanosome cell.

The polymerase chain reaction is sensitive enough to allow the analysis of DNA from an individual cell [Jeffreys *et al.*, 1988; Li, 1988]. Single cell PCR technology has been used extensively in human genetics to address a range of genetic and biological questions. For example, analysis of large numbers of single sperm have been used to estimate the recombination frequencies between genetic loci to a higher resolution than had been possible by pedigree analysis [Cui *et al.*, 1989]. Handyside *et al* (1990) applied single cell typing to pre-implantation disease diagnosis by blastomere biopsy and analysis. Such analysis was initially limited to 2-3 loci, but, the development of primer extension PCR (PEP), which randomly amplifies ~80% of the sequences in the genome, extended this technique to allow the amplification of a number of different

from individual cells, including blastomeres [Zhang *et al.*, 1992, Snabes, 1994 #191]. Another development, which relies on the amplification of DNA diluted to the equivalent of a single molecule, has greatly improved linkage mapping for those organisms where the number of progeny from crosses are limited. This technique, 'HAPPY mapping', examines large fragments of DNA for the presence of specific sequences to determine linkage [Piper *et al.*, 1998]. PCR from small amounts of DNA has also proved very useful in the analysis of forensic material [for example, Hopkins *et al.*, 1994] and ancient DNA samples [Hagelberg *et al.*, 1990].

PCR amplification of the multicopy satellite sequence from different species of African trypanosomes has been used to detect these parasites both from infected blood [Moser *et al.*, 1989] and tsetse flies [Masiga *et al.*, 1992] and it has been shown that a PCR signal can be obtained from DNA or trypanosome extracts diluted to less than a single genome equivalent [Moser *et al.*, 1989; Masiga *et al.*, 1992]. While these reports show that amplification of multiple copy sequences can be used to detect trypanosomes to a high degree of sensitivity, PCR based techniques using single copy sequences have not yet been used to detect or genotype single trypanosomes. Studies with *Plasmodium falciparum* have shown that single copy gene sequences can be amplified from single oocysts [Ranford-Cartwright *et al.*, 1991], although this life cycle stage contains many nuclei.

A single cell PCR technique could be applied to the analysis of trypanosomes to address a series of issues. For laboratory genetics, single trypanosomes isolated from the salivary glands of mixed-infected tsetse flies could be individually genotyped to assay the extent of cross and self-fertilization and to address the question of ploidy for each life cycle stage trypanosome. Other possible applications include the genotyping of trypanosomes from wild tsetse flies (or mammals) to study the extent of mixed infections in the field. Single cell analysis could also be used in for the analysis of recombination and other genomic rearrangements, such as VSG switching, in single trypanosomes rather than in populations.

The aim of the work presented here was to develop a method for determining the genotype of single trypanosomes with a high degree of efficiency using target single copy gene sequences. The technique was intended to be used to genotype the different trypanosomes present in the salivary glands of mixed-infected tsetse flies in which trypanosomes are undergoing mating. Parasites would be collected from tsetse flies by allowing them to probe into serum. Trypanosomes could be sampled every few days for the life of the fly, and then be individually genotyped. This would allow an extensive direct assessment of the genotypes of trypanosomes present in the salivary glands of a tsetse fly. The percentage of parental trypanosomes, F1 hybrids and selfers could be determined and any change in their proportions over time could be measured.

This method was applied to the analysis of laboratory crosses of two different *T. brucei* stocks to determine which tsetse flies harboured mixed infections and were producing hybrid trypanosomes. However, although a small number of tsetse flies were mixed-infected, none produced hybrid trypanosomes (see Chapter 5 for details of attempted crosses).

Results

Choice of marker loci for single trypanosome PCR. Three polymorphic markers were examined for their possible use in single cell analysis. The marker, *CRAM*, (described in detail in Chapter 3) was one of the first highly polymorphic single copy minisatellite markers to be identified. This is a highly informative locus, being able to distinguish between all four alleles of the *T. brucei* stocks STIB 386 and TREU 927/4 (the parental stocks in the attempted crosses, see Chapter 5). However the size of the variable region to be amplified was large and so the PCR amplification of this locus could be less efficient. A smaller single copy locus containing a dinucleotide repeat, the triosephosphate isomerase gene (*TIM*), although not as informative as the *CRAM* locus, was considered for single cell analysis. The third marker to be considered for single cell PCR, which was both small enough for efficient amplification in a PCR reaction and polymorphic enough to distinguish all four alleles in stocks STIB 386 and TREU 927/4, was the microsatellite marker, JS2, which was identified by J. Sasse [Sasse, 1998].

PCR from diluted DNA. In order to develop the technique, initial experiments to PCR amplify a single copy sequence from dilutions of STIB 386 DNA equivalent to a single haploid genome (0.04pg) [Borst, 1982] per reaction were performed using the *CRAM* locus. Details of the PCR conditions are given in the legend of Figure 4.1. The PCR products were size separated on a 1% agarose gel, Southern blotted and hybridized to a *CRAM* repeat probe. The resulting autoradiograph (shown in Figure 4.1) demonstrates that it was technically feasible to amplify a single copy sequence from trypanosome DNA diluted to the single cell level. STIB 386 is heterozygous for the *CRAM* locus with alleles of 2 and 2.8 Kb available for amplification. In 5 out of the 23 reactions both alleles amplified and in 8 reactions only one allele amplified. Overall 56% of reactions generated a PCR product from one or other allele. Since smaller fragments amplify more efficiently [Jeffreys *et al.*, 1988], it was believed that by changing to a smaller polymorphic locus *e.g.* a microsatellite, the efficiency of the PCR reactions would improve and would allow amplification of single cells to visible levels on an ethidium stained gel, using a nested PCR approach.

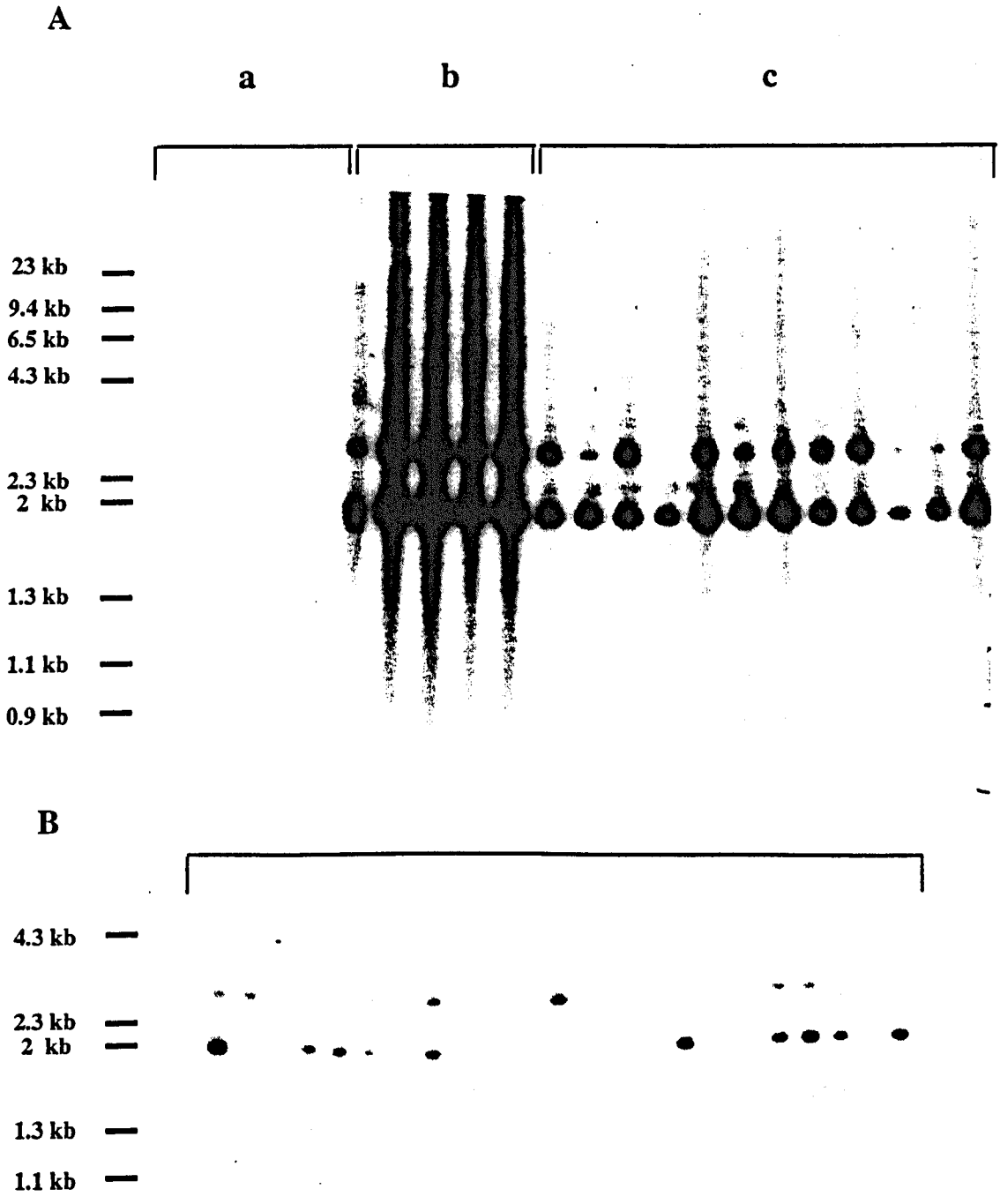


Figure 4.1 Amplification by PCR of the *CRAM* locus from diluted DNA. The *CRAM* locus was amplified using primers *CRAM*-G and *CRAM*-H from STIB 386 DNA (for primer sequences see Materials and Methods). The PCR products were size separated by gel electrophoresis, transferred to a nylon membrane and hybridised to a radiolabelled *CRAM* repeat probe. The autoradiograph was for 24 hrs.
(**A**) Panel a, 5 zero DNA controls; panel b, 5 PCR amplification reactions from 4pg DNA; panel c, 10 PCR reactions from 0.4pg DNA.
(**B**) 23 PCR amplification reactions from 0.04pg of DNA. Size markers based on lambda *Hin*dIII are shown.

TIM polymorphism. The single copy triosephosphate isomerase gene (*TIM*) sequence of *Trypanosoma brucei* contains a microsatellite sequence within the 3' untranslated region, consisting of fourteen repeats of the dinucleotide TA in the sequence obtained from stock EATRO 427 [Swinkels, 1986]. Using the published sequence, several pairs of primers were designed. Two of these, *TIM*-E and *TIM*-F (Figure 4.2A), were used to amplify (by PCR) samples of DNA from stocks STIB 386 and TREU 927/4 of *T. brucei*. The products were run out on a 3% Nusieve agarose gel and visualised with ethidium bromide (Figure 4.2B). The sizes of the fragments differ between the two stocks (STIB 386-100bp; TREU 927/4-80bp) indicating a polymorphism in the length of each repeat array, with each stock being homozygous for a different sized allele. This has been confirmed by sequence analysis (data not shown). The lower band is due to primer dimers and is present in the absence of any DNA template (track 3, Figure 4.2B).

PCR of the TIM locus from single trypanosomes. To determine whether it was possible to amplify a single copy sequence from a single trypanosome, bloodstream trypanosomes were isolated from mice which had been infected with either a mixture of stocks STIB 386 and TREU 927/4 or solely with TREU 927/4. The trypanosome-infected blood was diluted in guinea pig serum, spotted onto Terasaki plates, and drops containing single parasites were identified optically using an inverted microscope. Each well containing a single cell was flooded with 10µl of 1 x PCR buffer, transferred to a thin walled microtube and subjected to PCR amplification using various primer combinations for the *TIM* locus, (details of which are given in the legend of Figure 4.2 and Materials and Methods). Briefly, the primary amplification was conducted using primers *TIM*-C and *TIM*-D, which should amplify a 1390kb region of the gene. One µl of this primary amplification product was used as a template for the second amplification reaction, using internal primers *TIM*-A and *TIM*-B. This generated a 600bp product which could be visualised by UV illumination of the ethidium stained agarose gel.

In the initial experiment 22 single bloodstream trypanosomes of a single stock (TREU 927/4), were subjected to PCR amplification (using primer pairs *TIM*-C/*TIM*-D followed by *TIM*-A/*TIM*-B) and eight gave an ethidium bromide stained product of the predicted size (600bp; see Figure 4.3A for an example, tracks 4-19). The positive controls used in these experiments were drops containing >1 trypanosome and produced an amplified fragment (Figure 4.3A, tracks 1 and 2). In order to provide rigorous negative controls, several (individual) drops from the infected blood which contained no trypanosomes by microscopic inspection, were also subjected to PCR amplification using primers *TIM*-C/*TIM*-D followed by *TIM*-A/*TIM*-B (data not shown). One of these

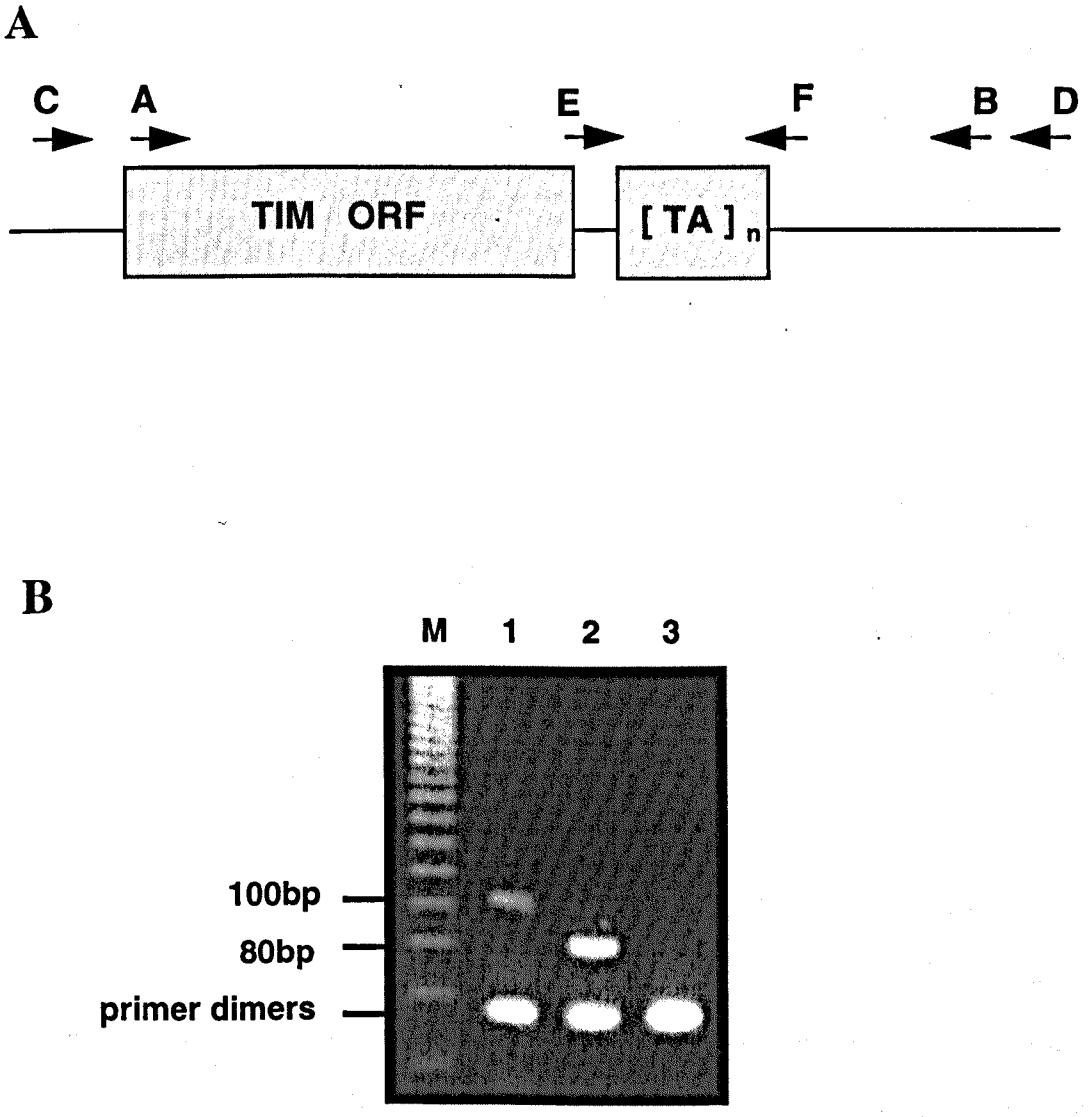


Figure 4.2 Amplification by PCR of the *TIM* locus from stocks STIB 386 and TREU 927/4.
(A) The schematic diagram indicates the *TIM* locus with the approximate positions of the TA dinucleotide repeat and the primers (*TIM*-E and *TIM*-F) used for PCR amplification, under standard PCR conditions (see Materials and Methods).
(B) Ethidium bromide stained 3% Nusieve agarose gel separation of the PCR amplified products obtained from genomic DNA of stocks STIB 386 and TREU 927/4 using the primers *TIM*-E and *TIM*-F. Track 1, STIB 386; track 2, TREU 927/4; track 3, zero DNA control; M, 20bp ladder (Advanced Biotechnologies).
Primer sequences are given in Materials and Methods.

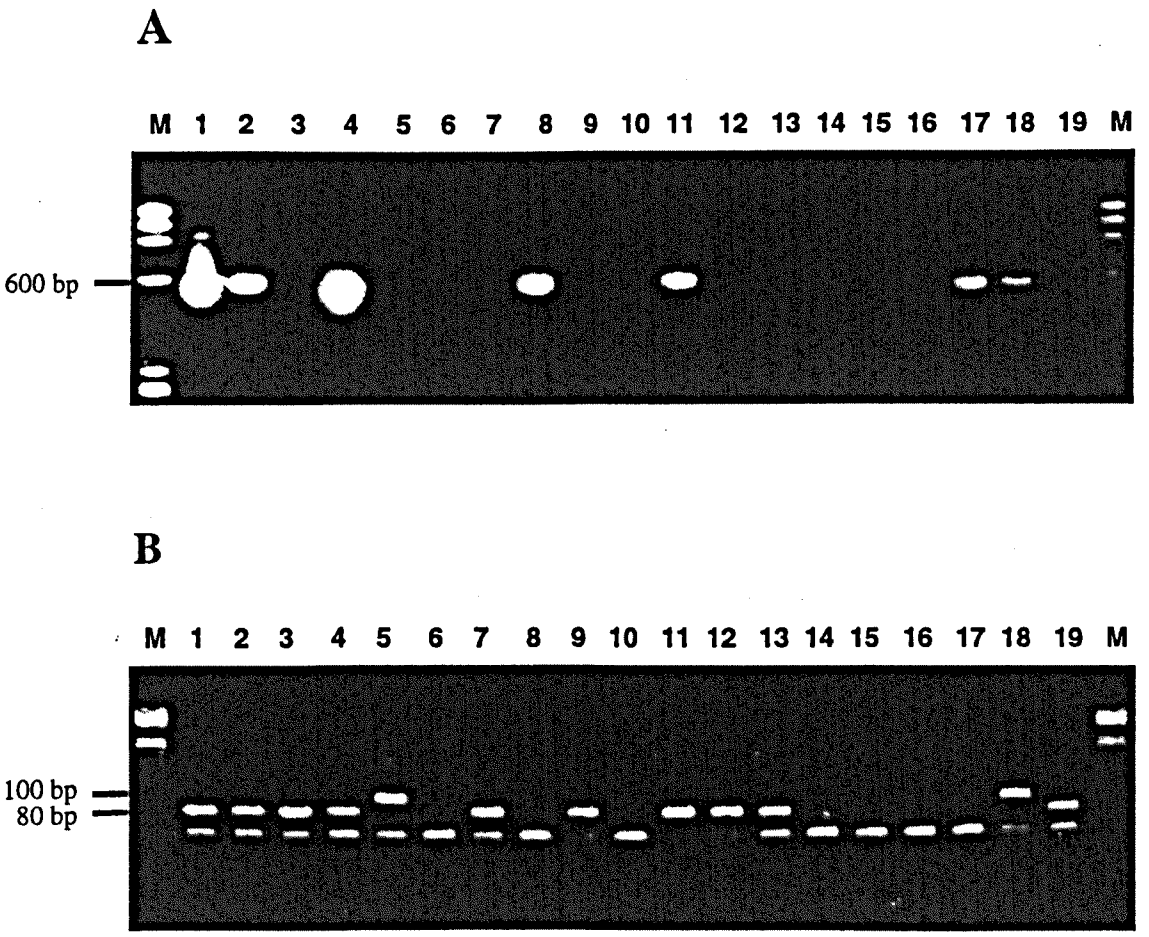


Figure 4.3 PCR amplification of the *TIM* locus from single trypanosomes using a series of nested primers.
(A) Sixteen single bloodstream stage trypanosomes were isolated optically, transferred to a PCR microtube by the addition of 10µl of PCR buffer, described in Material and Methods, overlaid with mineral oil and heated to 96°C for 5 mins (to disrupt the cell). The samples were then subjected to PCR amplification, by the addition of primers *TIM*-C and *TIM*-D to a final concentration of 0.2µM, under the following cycling conditions: 96°C for 1 min., 64°C for 1 min. and 70°C for 90 sec. for a total of 26 cycles. A one µl sample of the product was transferred to a fresh PCR microtube and subjected to a second PCR amplification using primer pair *TIM*-A and *TIM*-B under the same conditions for a further 26 cycles. The final products were separated on 1% Seakem agarose, stained with ethidium bromide and visualised by UV illumination. Lanes 1 and 2, each contained more than one trypanosome; lane 3, zero trypanosome control; lanes 4-19, drops each containing a single trypanosome; M, ϕx174 *Hae*III markers.
(B) Ethidium bromide stained 3% Nusieve agarose separation of the PCR products of single drops containing a single trypanosome isolated from a mixture of stocks TREU 927/4 and STIB 386. The conditions for PCR amplification were identical to those described in (A) except that primer pair *TIM*-A/*TIM*-B was used for the initial amplification and primer pair *TIM*-E/*TIM*-F for the second. Lanes 1-19, single trypanosomes in a single drop; M, ϕx174 *Hae*III markers.

drops in a collection of 10 produced an amplified fragment which is presumed to be due to overlooking a trypanosome in this drop. However, given only 1/10 of the drops lacking trypanosomes gave a positive signal, we conclude that the amplified product is dependent on the presence of trypanosomes rather than extraneous contaminating DNA. These results suggested that it was possible to amplify a single copy gene sequence from a single trypanosome cell, however, the efficiency was low (36%). The inability to amplify a fragment from all samples may be attributed to either a failure to transfer single cells to the PCR tubes or a failure of the primers to amplify successfully from the two target molecules present. The predicted PCR product from the primary amplification was 1390bp and ~600bp for the secondary amplification. In an attempt to increase the efficiency of the nested PCR technique, primers *TIM*-E and *TIM*-F were used which generated smaller products (~100bp for the secondary amplification of TREU 927/4), since smaller products are likely to be amplified more efficiently than larger ones. In four subsequent independent experiments, each using 20 single trypanosomes, the cells were isolated and subjected to the nested PCR amplification, using the *TIM*-E and *TIM*-F primer combination for the secondary amplification. The efficiency of PCR recovery from a single cell rose to an average of 48% (see Figure 4.3B for an example).

One other possible reason for a loss in efficiency of the single cell PCR method was that the single trypanosome failed to be transferred from the Terasaki well to the PCR tube. In order to determine if this was the case, each Terasaki well was washed twice in 10μl of water after the trypanosome was removed. The water from these washings was then amplified as described before. Out of a total of 36 wells, 4 produced a PCR product of the size expected (data not shown), thus showing that some trypanosomes (at least 11%) were not being transferred to the PCR tube in the first instance resulting in a reduced efficiency of PCR. However this only accounts for a small proportion of the PCR failures.

Genotyping single trypanosomes. In order to test whether individual trypanosomes could be identified from mixtures of trypanosome stocks, single cells were isolated (as described above) from infected mouse blood containing two stocks of the parasite, TREU 927/4 and STIB 386. These stocks have different lengths of the *TIM* dinucleotide repeat and so generate different sized PCR products when amplified using primers *TIM*-E and *TIM*-F (Figure 4.2B). In this experiment, a total of 19 drops containing single trypanosomes were PCR amplified and the results obtained are shown in Figure 4.3B. A total of 12 out of 19 drops containing single trypanosomes gave an ethidium bromide stained fragment of either 80bp (TREU 927/4) or 100bp (STIB 386) but none gave two fragments indicating that none of the drops contain two trypanosomes (the lower size band on these gels is due to primer dimers). These data

support the conclusion not only that single copy genes can be amplified from single trypanosomes but that single cells can be genotyped in this way. The results presented here compare well with the success rate for developing clones in mice injected with single trypanosomes, demonstrating that single cell PCR is a reliable technique for defining the genotype of single trypanosomes. To improve the utility of this technique in laboratory genetic analysis, it was necessary to extend this technique to another more informative polymorphic locus *i.e.* a marker which can differentiate between all four alleles present in STIB 386 and TREU 927/4. This led to the search for another microsatellite locus.

JS2 polymorphic locus. The sequence of the single copy *T. brucei* locus, JS2, which contains a polymorphic CA dinucleotide repeat and was obtained from J. Sasse [Sasse, 1998]. Primers flanking the repeat were designed based on this sequence and used to amplify the locus from the stocks STIB 247, STIB 386 and TREU 927/4 (Figure 4.4A). The products were run on a 4% Nusieve agarose gel and visualised with ethidium bromide (Figure 4.4B). Unlike the *TIM* microsatellite, each stock is heterozygous for different sized alleles, allowing all six alleles from the three stocks to be easily distinguished from each other, which makes this marker extremely informative in strain identification and useful in the analysis of the F1 progeny from crosses between these stocks.

PCR of the JS2 locus from single trypanosomes. Using a nested PCR strategy with the oligonucleotides JS2-C and JS2-D followed by JS2-A and JS2-B (Figure 4.4A), the JS2 locus was successfully amplified from single trypanosomes of the stock TREU 927/4, at an efficiency of 71% (*i.e.* 37/52 produced a PCR product: Figure 4.4C). This increased efficiency is probably because the primary PCR amplification product is only about 300bp, compared to the primary amplification product of the *TIM* locus which was 1390bp. As TREU 927/4 is heterozygous for the JS2 locus, two different and easily distinguishable alleles are available for amplification from each individual trypanosome, resulting in two PCR products of ~190bp and ~105bp (Figure 4.4C). In contrast the *TIM* locus is homozygous with two alleles of the same size so that it is impossible to determine whether the PCR product was derived from one or both alleles. With the JS2 locus, 8 single trypanosomes produced only one visible PCR product with the other allele having failed to amplify (for example Figure 4.4C, lanes 7 and 18). The relative proportions of samples which generated two PCR products (56%), a PCR product from the upper allele only (9.6%), the lower allele only (5.8%), or no product at all (29%) are presented in the pie chart in Figure 4.5. The failure of one allele to amplify could be due to sequence variation in the flanking DNA, which affects the efficiency of one or more of the primers, although this is unlikely as no one allele failed to amplify significantly more times than the other (χ^2 test of

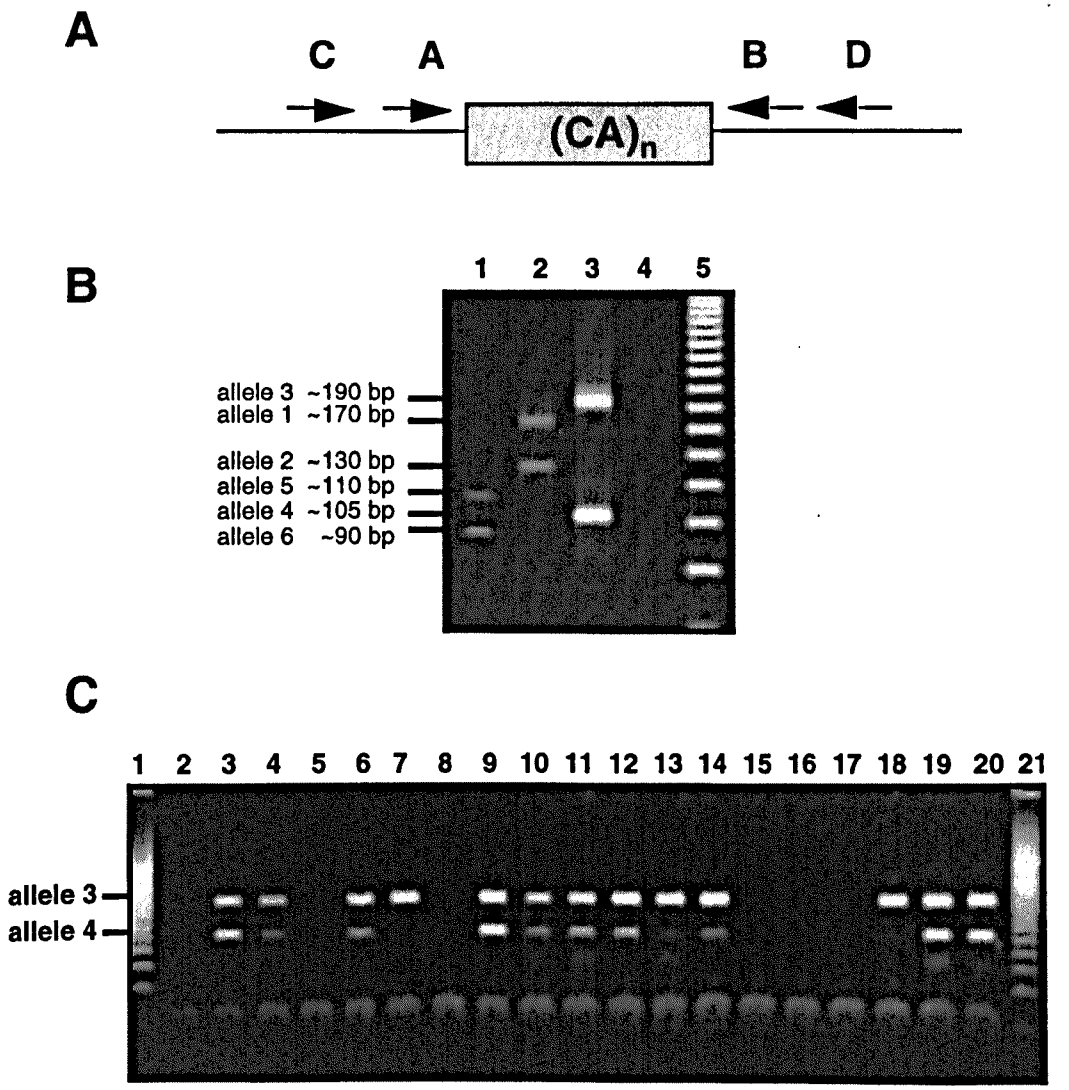


Figure 4.4 PCR amplification of the JS2 locus from DNA and single trypanosomes.
(**A**) A schematic diagram of the JS2 locus indicating the relative positions of the PCR primers. Primer sequences are given in Materials and Methods.
(**B**) Ethidium bromide stained 4% Nusieve agarose gel separation of the PCR amplified products obtained from genomic DNA of stocks STIB 247 (lane 1), STIB 386 (lane 2) and TREU 927/4 (lane 3), using primers JS2-A and JS2-B. Lane 4, zero DNA control. Lane 5, 20bp ladder (Advanced Biotechnologies). Allele numbers and approximate band sizes are shown.
(**C**) An example of an ethidium bromide stained 4% Nusieve agarose gel separation of JS2 nested PCR amplification products from single trypanosomes. Bloodstream stage trypanosomes of the stock TREU 927/4 were isolated optically, transferred to a PCR microtube by the addition of 10µl of PCR buffer, described in Materials and Methods, overlaid with mineral oil and heated to 96°C for 5 mins (to disrupt the cell). The samples were then subjected to PCR amplification, by the addition of primers JS2-C and JS2-D to a final concentration of 0.2µM, under the following cycling conditions: 96°C for 50 sec., 56°C for 50 sec. and 70°C for 50 sec. for a total of 26 cycles. A one µl sample of the product was transferred to a fresh PCR microtube and subjected to a second PCR amplification using primer pair JS2-A and JS2-B under the same conditions for a further 26 cycles. Lanes 2-20, PCR products from single drops each containing a single trypanosome; Lane 21, 20 bp ladder (Advanced Biotechnologies).

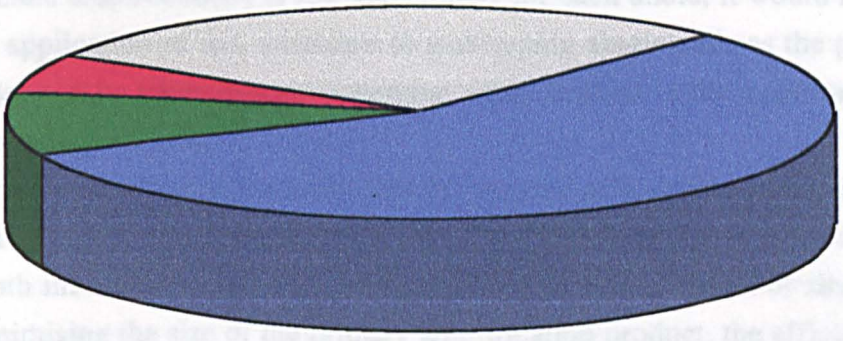


Figure 4.5. A pie chart illustrating the results generated by the PCR amplification of the JS2 locus from single TREU 927/4 trypanosomes. The proportion of samples which generated two products is coloured blue (56%); allele 3 only in green (9.6%); allele 4 only in red (5.8%) and those which failed to amplify are coloured grey (29%). n = 52.

homogeneity = 0.5, d.f.=1, p=0.5, leading to acceptance of the null hypothesis that there is no allele bias). It is possible that the alleles simply did not amplify sufficiently to be visible on an ethidium stained gel. Southern blotting followed by hybridization could possibly detect these alleles. However, the vast majority (78%) of the trypanosomes that generated a PCR signal contained the products of both alleles. Because the percentage of single allele amplifications is low and similar for each allele, it would not appear to hinder the application of this technique to genotyping single cells as the proportion of failed alleles can be taken into consideration when applied to the quantitative analysis of trypanosome genotypes.

A summary of the different properties and the relative efficiencies of the three markers used in this study is presented in Table 4.1. It is clear from these results that the JS2 locus is both informative and highly efficient for the amplification of single cells and that by minimising the size of the primary amplification product, the efficiency of PCR from single trypanosomes can be increased.

Markers	No. of different alleles in stocks 386 and 927/4	Size of PCR product		Efficiency of PCR
		primary	secondary	
CRAM	4	2 - 2.8kb	NA	56% [a]
TIM 1	2	1390bp	600bp	36% [b]
TIM 2	2	1390bp	~100bp	48% [b]
JS2	4	300bp	~200bp	71% [b]

Table 4.1. Summary of experiments with different markers. Column 1 lists the different markers used in these experiments. Two different primer combinations were used to amplify the *TIM* locus (see text). Column 2 indicates the number of different alleles distinguished by each marker for the two stocks STIB 386 and TREU 927/4. Column 3 gives the estimated size of the PCR products from the primary and secondary (nested) reactions. For the *CRAM* locus, a single amplification reaction was performed followed by Southern blotting and hybridization to a specific probe. Column 4 indicated the efficiency of PCR amplification from either (a) DNA diluted to the equivalent of a single cell or (b) from single trypanosomes.

Discussion

The JS2 locus is a highly informative polymorphic locus, as it is heterozygous for the stocks STIB 247, STIB 386 and TREU 927/4 for different sized alleles. This makes it an excellent marker for the analysis of hybrids derived from the genetic crosses of these stocks. Coupled with the powerful technique of nested PCR, it is possible to genotype individual trypanosomes from any life cycle stage and so identify where and when hybrids are produced and to what extent cross and self-fertilization occur.

The results obtained from the analysis of single trypanosomes using JS2 compare well with those obtained from single human sperm [Li, 1988], where 76% of the PCR reactions gave a signal which was consistent with only single cells being amplified. Where two loci were amplified from the same haploid cell, which should generate two PCR products, 61% gave a signal from both loci. This can be compared to 56% of single (diploid) trypanosomes, producing PCR products from both alleles from the same locus. Although it is not expected that PCR amplification from a single cell will be 100% efficient it is obvious that by optimising the conditions of the reactions the efficiency of the technique can be improved. However, because of the sensitive nature of the single cell PCR technique, the data generated by this method may be subject to error. Such errors can arise in a variety of ways: one of the alleles may not be amplified to a visible level, the trypanosome sample may be contaminated with PCR products carried over from other amplification reactions, the sample tube may not contain a cell or may contain more than one cell. However these errors for single trypanosome PCR are minimal. For example, none of the reagent negative controls ever produced a PCR product and results from mixed trypanosome experiments reveal that optical isolation of single cells is reliable. Statistical estimations of these errors are possible [Cui *et al.*, 1989] and so they can be taken into consideration in the analysis of the data.

By extending this technique to incorporate other microsatellites in the same reaction *i.e.* multiplexing [Jeffreys *et al.*, 1988], or by employing the PEP technique [Zhang *et al.*, 1992], it should be possible to analyse single cells for several different markers, although with more markers, the proportion of samples in which at least one allele failed to amplify is increased (assuming the probability of each allele to amplify is independent of the next allele). Using this technique, the independent assortment of alleles in hybrid metacyclic cells isolated from the salivary glands of mixed-infected tsetse flies could then be examined or if the markers were located on the same chromosome, recombination frequencies could be measured. The detection of the products of self-fertilization, however, would be indistinguishable from samples in which an allele has failed to amplify. Although if selfers were present in the population, examination of large numbers of cells should reveal a significantly higher proportion of samples exhibiting a single allele pattern compared to the controls. This approach is limited only by the availability of tsetse flies harbouring hybrid metacyclic trypanosomes.

Chapter 5

Analysis of Genetic Exchange in *T. brucei*

Introduction

The first suggestion that *T. brucei* was a diploid sexual organism came from the analysis of the nature and frequency of isoenzyme variation of different trypanosome stocks [Gibson *et al.*, 1980; Tait, 1980; Tait, 1983]. A system of genetic exchange was confirmed in 1986 when Jenni and co-workers demonstrated the first genetic cross between two genetically and phenotypically distinct *T. brucei* stocks [Jenni *et al.*, 1986]. The cross, which involved the co-transmission of two stocks through tsetse flies, produced trypanosomes with non-parental genotypes and phenotypes, indicating that some form of genetic exchange took place. Since then eight other crosses have been performed in the laboratory, generating hybrid clones which appear to be equivalent to F1 progeny [Gibson, 1989; Sternberg *et al.*, 1989; Turner *et al.*, 1990; Gibson and Garside, 1991; Gibson *et al.*, 1992; Gibson and Whittington, 1993; Gibson and Bailey, 1994; Degan *et al.*, 1995; Tait *et al.*, 1996; Gibson *et al.*, 1997]. One backcross has also been performed [Gibson *et al.*, 1995]. However, the nature of this genetic exchange, *i.e.* the mechanisms involved in the generation of hybrids, still remains to be fully elucidated.

To date 141 hybrid clones from the 10 different crosses have been isolated (see Chapter 1, Table 1.1). Figure 5.1 illustrates the procedure involved in making a genetic cross, whereby individual trypanosomes derived from a mixed-infected tsetse fly are isolated and grown as a clonal population by mitotic replication. There are two distinct methods of analysing the products of a cross: 1) Single non-dividing metacyclic trypanosomes are taken directly from the dissected salivary glands of a mixed-infected tsetse fly, and injected into a mouse where they differentiate to bloodstream forms and replicate. The resulting population is then analysed (for a novel hybrid genotype) using a series of genetic markers that differ between the two parental stocks. Since metacyclic cells are believed to be cell-cycle arrested [Vickerman, 1985], *i.e.* they only replicate after they have differentiated into bloodstream forms in the mammalian host, each hybrid metacyclic is assumed to be the product of a single genetic exchange event, and is therefore considered to be unique. 2) Either a mixed-infected tsetse fly is allowed to bite a mouse or part of the dissected salivary glands of the fly is injected into a mouse, thereby infecting the mouse with several thousand trypanosomes, which differentiate and multiply. Single

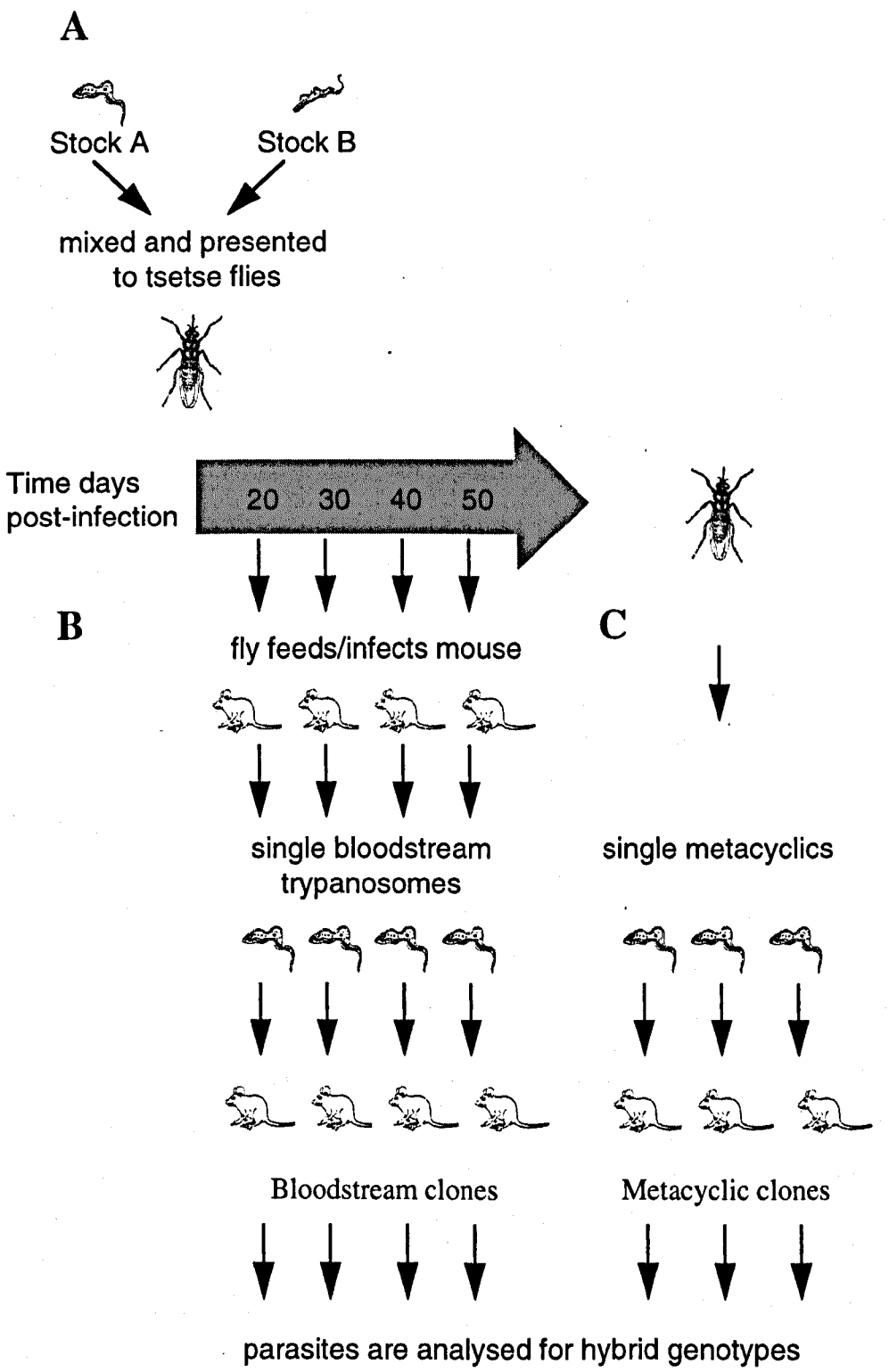


Figure 5.1. A schematic diagram illustrating the procedure involved in the generation of a genetic cross. (**A**) Two genetically distinct *T. brucei* stocks are mixed and fed to tsetse flies. (**B**) Mixed infected tsetse flies are fed on mice at regular intervals. The mice develop infections. Single bloodstream trypanosomes are then isolated and regrown in mice. Trypanosomes isolated in this way are called bloodstream clones. (**C**) The tsetse flies are sacrificed and single metacyclic trypanosomes are removed from the salivary glands. The metacyclics are grown in mice. Trypanosomes isolated in this way are called metacyclic clones.

trypanosomes are then isolated from the infected mouse blood and injected into separate mice where they multiply as clonal populations. Since there is a pre-amplification step in the first mouse before cloning, some of the resulting bloodstream hybrids are identical clones derived from the one genetic exchange event. For the analysis of hybrid clones presented in this chapter, bloodstream clones isolated from the same fly on the same day and which have identical genotype patterns for three minisatellites and one microsatellite marker, were considered to be vegetative derivatives of the same genetic exchange event. The nomenclature used for clone identification throughout this chapter follows that of Sternberg *et al.*, [1989], for example clone F9/45mcl 7, is the seventh metacyclic trypanosome clone derived from fly 9, on day 45 post infection.

Clones derived from mixed-infected tsetse flies have been examined using a range of genotypic and phenotypic markers, in particular by isoenzymes and RFLPs. From these studies a number of the basic features of genetic exchange in *T. brucei* have been defined. It is clear that three distinct genotypes/phenotypes of trypanosomes can be identified from a cross; parentals, indicating that the process of genetic exchange is non-obligatory [Sternberg *et al.*, 1989], hybrids (equivalent to F1s) [Gibson, 1989; Sternberg *et al.*, 1989; Turner *et al.*, 1990] and selfers (the products of self-fertilization) [Tait *et al.*, 1996; Gibson, *et al.*, 1997]. Analysis of markers for loci on the megabase chromosomes demonstrates that the majority of hybrid clones show chromosome segregation in a fashion that is consistent with a Mendelian system of inheritance [Turner *et al.*, 1990].

The ploidy and karyotype of *T. brucei* has been demonstrated by a number of studies to be diploid and stable during mitosis at all life cycle stages examined so far [Shapiro *et al.*, 1984; Wells *et al.*, 1987; Tait *et al.*, 1989; Tait *et al.*, 1996; Melville *et al.*, 1998]. However the karyotype may be altered by genetic exchange, for example, by the generation of novel sized chromosomes [Tait *et al.*, 1993; Melville *et al.*, 1998] and the raised DNA content of some hybrids [Shapiro *et al.*, 1984; Wells *et al.*, 1987; Gibson *et al.*, 1992]. Indeed the first three hybrids isolated all had elevated DNA content (approximately 1.5x) compared to the parental stocks [Sternberg *et al.*, 1989]. Examination of those clones with raised DNA content by RFLP, isoenzyme and karyotype analysis by [Wells *et al.*, 1987; Gibson *et al.*, 1992] led to the suggestion that these clones were triploid.

A number of models of genetic exchange have been proposed for *T. brucei*. The first model proposed a mechanism to explain the detection of triploids and involved the fusion of diploid parasites to form a tetraploid, followed by DNA loss to produce a triploid organism [Paindavoine *et al.*, 1986]. As more hybrid clones were examined from a number of different crosses it became clear that the majority of hybrid progeny were diploid, which led to the suggestion that genetic exchange involved meiosis, with triploid hybrids being an aberrant minority [Le Page *et al.*, 1988; Sternberg *et al.*, 1989]. This hypothesis was supported by the results from a number of other crosses [Gibson, 1989;

Turner *et al.*, 1990; Gibson and Garside, 1991]. However two further crosses, which involved the double drug selection of resistant hybrid stocks [Gibson and Bailey, 1994; Gibson *et al.*, 1995], generated a far higher number of progeny with raised DNA content; in one case 27/32 (11/14 unique clones) had a DNA content of ~1.5x [Gibson and Bailey, 1994]. This has led to the proposal of another model [Gibson, 1995] involving fusion of diploid cells followed by meiosis. The most recent analysis of a large collection of hybrid clones from three different crosses has revealed that the majority of hybrid clones have a diploid DNA content with only 2/30 demonstrating raised DNA content, which suggests that triploidy is rare, at least for those particular crosses, supporting the view that genetic exchange is carried out by a conventional mating system [Hope *et al.*, 1999]. The different models of genetic exchange are examined in more detail in the discussion (Chapter 9).

Despite a number of successful crosses performed in the laboratory there are several specific questions regarding genetic exchange in *T. brucei* which still need to be addressed:

1) At what stage in the trypanosome life cycle does genetic exchange take place? It is clear that mating occurs at some stage in the tsetse fly. Experiments involving the double drug selection of resistant trypanosomes by Gibson and Whittington, (1993) have indicated that the procyclic (midgut) stage is probably not the stage of genetic exchange, contradicting earlier reports of the detection of hybrid procyclics [Schweizer and Jenni, 1991]. Further analysis of the metacyclic (salivary gland) stage indicate that they are the products of mating [Tait *et al.*, 1989]. Taken together these data indicate that either the proventricular or the epimastigote stages may be involved in genetic exchange.

2) Does sexual recombination follow Mendel's laws of allele segregation and random assortment of genes? Evidence from a number of studies, mainly using RFLPs and isoenzymes, suggest Mendelian inheritance, but ratios of different hybrid genotypes in the F1 progeny, which would confirm or refute Mendelian inheritance, have been unavailable because only a very limited number of F1 progeny clones have been generated.

3) Is triploidy or diploidy the usual outcome of genetic exchange? Analysis of hybrid clones by pulse field gel electrophoresis, followed by Southern blotting and hybridization to gene probes, has indicated that hybrid clones inherit one homologue of each chromosome from each parental clone [Tait *et al.*, 1993], suggesting a normal diploid Mendelian genetic system. However it is clear from the data of Gibson and colleagues [Gibson and Garside, 1991; Gibson and Bailey, 1994; Gibson *et al.*, 1995], that many hybrid progeny clones can be trisomic for a number of chromosomes and are probably triploid.

4) Are bloodstream clones, derived from a mixed-infected tsetse fly, true representatives of salivary gland infections, *i.e.* does the pre-amplification step in the generation of bloodstream forms select for particular genotypes of trypanosome?

5) Do the proportions of selfers, hybrids and parentals alter over time? Preliminary evidence suggests that hybrids are produced late in the tsetse fly infection [Schweizer *et al.*, 1988; Sternberg *et al.*, 1989].
6) Does self-fertilization only occur in the presence of cross-fertilization? Two studies have demonstrated self-fertilization in the context of cross-fertilization, but examination of a limited number of clones derived from single stock transmissions has not revealed any products of self-fertilization [Tait *et al.*, 1996; Gibson *et al.*, 1997].
7) What is the mechanism of genetic exchange? Does genetic exchange involve the generation of haploid gametes, which have not yet been identified, followed by fusion, or do diploid cells fuse and then undergo meiosis?
8) To what extent does crossing over occur? How large is a centiMorgan in *T. brucei*?
Examination of these questions has been hindered by the limited number of available clones derived from experiments. The aim of this chapter was to address some of the questions raised, using two separate approaches involving the techniques described in Chapters 3 and 4.
The first approach was to generate new genetic crosses and to analyse directly individual metacyclic cells present in the salivary glands of tsetse flies sampled over a time course of infection, using the method for genotyping single trypanosomes outlined in Chapter 4. This would determine the proportions of the trypanosome population in the salivary glands that were hybrid, parental or products of self-fertilization and so indicate the extent of cross- and self-fertilization. Also, examination of single cells from other life cycle stages *e.g.* proventricular and epimastigote forms were planned, which could help identify the stage at which sexual recombination occurs and determine whether genetic exchange involves a haploid gamete based system.
The second approach was to examine the existing hybrid and parental material already available from previous crosses using the newly identified mini- and microsatellite markers, to determine if hybrid formation follows Mendel's laws of allele segregation and random assortment and if triploidy is a normal outcome of genetic exchange. Also, since the uncloned bloodstream trypanosomes from mice which had been bitten by mixed-infected tsetse flies were available as frozen stabilates, new clones from the old crosses could easily be generated, greatly increasing the number of hybrid clones available for analysis.
Marker analysis of hybrid clones has previously relied on isoenzyme, RFLP and karyotype analysis. This study, which is focused on the inheritance of the megabase chromosomes, uses the highly informative markers described in Chapter 3 to analyse the clones derived from genetic crosses. The four mini- and microsatellite loci are highly informative, genetically interpretable markers, with each marker being able to distinguish between all four alleles in the parental clones STIB 386 and TREU 927/4. This should allow the identification of the 256 potential F1 genotypes. The high discriminating power

of these markers should also be able to identify the majority of selfers. Selfers can be distinguished from parental clones, as in 50% of loci where the parents are heterozygous, the selfers will be homozygous. Therefore one heterozygous marker can identify approximately 50% of selfers, two markers can identify 75%, three markers can identify 87.5% and four markers, 93.75%.

Results and Discussion

Attempted crosses. Attempts were made to generate new genetic crosses. The method for generating the new genetic cross was similar to that outlined in Figure 5.1, whereby two genetically distinct parental stocks were co-transmitted through tsetse flies. However as well as generating bloodstream or metacyclic clones from these parasites, it was intended that single metacyclic and/or epimastigote cells from the salivary glands of the mixed-infected tsetse flies would be analysed directly using the single trypanosome genotyping method described in Chapter 4. In this way large numbers of cells could be analysed directly. The analyses of metacyclics and epimastigotes would identify the life cycle stage at which mating takes place. Also the proportions of trypanosomes with parental and recombinant genotypes present in the salivary glands could be determined and so indicate the extent of cross and self-fertilization. In order to ascertain if these proportions change over time in individual tsetse flies, metacyclic stage trypanosomes could be obtained from the flies without dissection, by allowing the tsetse flies to probe into mouse serum.

The parental stocks STIB 386 and TREU 927/4 were chosen for the crosses because they are both heterozygous for different sized alleles for the minisatellites described in Chapter 3 and so are the most informative combination of stocks. These stocks have previously been successfully crossed so few problems in generating a cross were anticipated.

In the first experiment 200 teneral tsetse flies (provided by I. Maudlin and S.C. Welburn) were fed blood containing a 1:1 mixture of the two parental stocks STIB 386 and TREU 927/4. After three weeks the tsetse flies were allowed to probe onto microscope slides (as described in Materials and Methods) and the saliva was examined by phase contrast microscopy for the presence of trypanosomes. Those probes which contained metacyclic stage trypanosomes were analysed by microsatellite PCR. (Figure 5.2 is an example of the PCR amplification of the *TIM* locus from fly probes, illustrating that it is possible to identify the genotypes present in the saliva of the tsetse flies and to distinguish mixed-infected flies (for example Figure 5.2, lane 6) by this method). In this way the proportion of flies harbouring mixed or single stock metacyclic infections was determined.

Figure 5.2. Amplification by PCR of the *TIM* locus from 17 tsetse fly probes. Tsetse fly probes from flies harbouring salivary gland infections from a STIB 386 x TREU 927/4 cross (3:1 ratio) were obtained as described in Materials and Methods. Deionised water (10µl) was added to each slide, mixed with the dried saliva, transferred to a PCR microtube and heated to 95°C to disrupt any intact cells. One µl of each sample was amplified using primers *TIM*-E and *TIM*-F described in Materials and Methods. PCR conditions were as follows; 95°C for 50sec., 64°C for 50sec. and 70°C for 50sec. for 30 cycles. The products were separated on a 3% Nusieve agarose gel, stained with ethidium bromide and visualised by UV illumination. Lane 1, control STIB 386; lane 2, control TREU 927/4; lanes 3-19 examples of fly probes.

The proportions of metacyclic infections from the first attempted cross are illustrated in Figure 5.3A. It is evident from this pie chart that the majority (87%) of tsetse flies did not appear to harbour metacyclic infections, however the proportion of metacyclic infections (13%) is within the 'normal' range (10-20%) for this species of tsetse fly (*Glossina morsitans*) under these conditions [A. Tait, personal communication]. It is also apparent from Figure 5.3A that only TREU 927/4 genotypes were detected in the tsetse saliva. It was not clear at this stage why no STIB 386 genotypes were detected. Two possibilities existed; STIB 386 had not established midgut infections or the STIB 386 trypanosomes had not progressed from the midgut to the salivary glands. It has been reported previously that STIB 386 has a lower transmission index and so is less efficient at generating salivary gland infections from midgut infections than other stocks [I. Maudlin and S. Welburn, personal communication].

In order to determine if STIB 386 had established midgut infections, ten tsetse flies were dissected by M. Hope and 1µl of each of the midgut contents was analysed by PCR (under the same conditions as described in Figure 5.2). The results are illustrated as a pie chart in Figure 5.3B. From this small sample, it would appear that no flies had midgut infections established by STIB 386 alone, while 60% had been infected solely with TREU 927/4 and 33% contained mixtures. Thus, while mixed infections had been established, there was a bias with TREU 927/4 establishing in the midgut more effectively than STIB 386. It would appear that a combination of a lower midgut infection rates and a lower transmission rate were responsible for the lack of STIB 386 salivary gland infections. To bias the second cross in favour of STIB 386, and thus compensate for initial lower midgut infections and a lower transmission index, 300 tsetse flies were fed on a 3:1 ratio of STIB

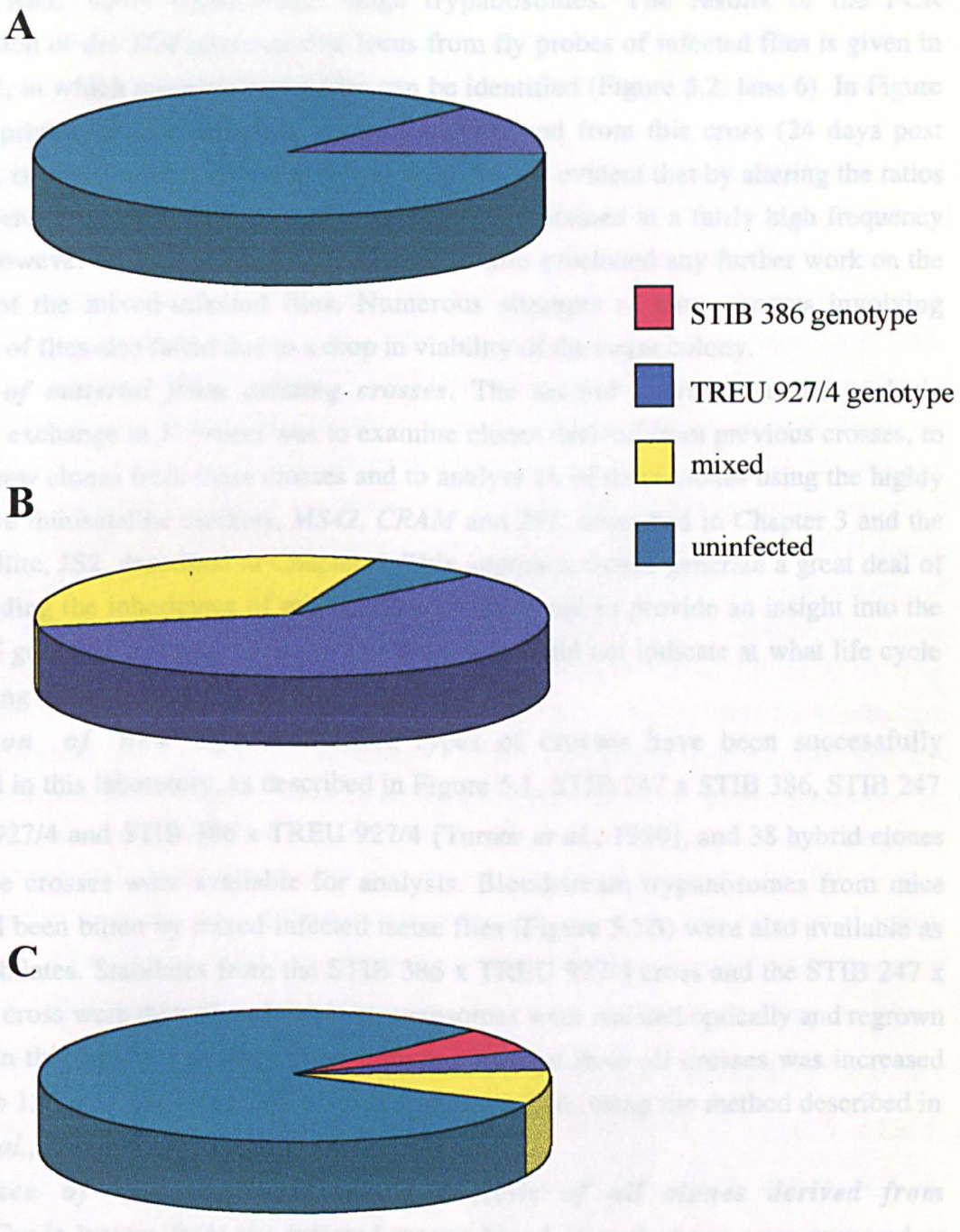


Figure 5.3. Pie charts of tsetse fly infections.
(A) Proportions of tsetse salivary gland infections from flies feed on a 1:1 mixture of STIB 386 and TREU 927/4. STIB 386 genotype 0%, TREU 927/4 genotype 13%, mixed genotype 0% and uninfected 87%. n=100.
(B) Proportions of tsetse midgut infections from flies feed on a mixture (1:1) of STIB 386 and TREU 927/4. STIB 386 genotype 0%, TREU 927/4 genotype 60%, mixed genotype 33% and uninfected 7%. n=15.
(C) Proportions of tsetse salivary gland infections from flies feed on a 3:1 mixture of STIB 386 and TREU 927/4. STIB 386 genotype 5.9%, TREU 927/4 genotype 5.3%, mixed genotype 8.2% and uninfected 80.6%. n=170.

386 to TREU 927/4 bloodstream stage trypanosomes. The results of the PCR amplification of the *TIM* microsatellite locus from fly probes of infected flies is given in Figure 5.2, in which mixed-infected flies can be identified (Figure 5.2, lane 6). In Figure 5.3C the proportions of metacyclic infections obtained from this cross (24 days post infection), is given as determined by fly probe PCR. It is evident that by altering the ratios of the parental stocks, mixed-infected flies could be obtained at a fairly high frequency (8.2%). However the poor viability of these tsetse flies precluded any further work on the analysis of the mixed-infected flies. Numerous attempts at more crosses involving thousands of flies also failed due to a drop in viability of the tsetse colony.

Analysis of material from existing crosses. The second approach to the analysis of genetic exchange in *T. brucei* was to examine clones derived from previous crosses, to generate new clones from these crosses and to analyse all of these clones using the highly informative minisatellite markers, *MS42*, *CRAM* and *292*, described in Chapter 3 and the microsatellite, JS2, described in Chapter 4. This approach would generate a great deal of data regarding the inheritance of minisatellite markers and so provide an insight into the process of genetic exchange in *T. brucei*, although it could not indicate at what life cycle stage mating occurs.

Generation of new hybrids. Three types of crosses have been successfully performed in this laboratory, as described in Figure 5.1, STIB 247 x STIB 386, STIB 247 x TREU 927/4 and STIB 386 x TREU 927/4 [Turner *et al.*, 1990], and 38 hybrid clones from these crosses were available for analysis. Bloodstream trypanosomes from mice which had been bitten by mixed-infected tsetse flies (Figure 5.1B) were also available as frozen stabilates. Stabilates from the STIB 386 x TREU 927/4 cross and the STIB 247 x STIB 386 cross were thawed and single trypanosomes were isolated optically and regrown in mice. In this way the total number of clones derived from all crosses was increased from 91 to 156. All new clones were generated by A. Tait, using the method described in Turner *et al.*, (1990).

Inheritance of minisatellite markers -analysis of all clones derived from crosses. Crude lysates from the infected mouse blood of each clone were prepared as described in Chapter 2 and by MacLeod *et al.*, [1999] and analysed by PCR using the mini- and microsatellites, *MS42*, *CRAM*, *292* and JS2. Two amplified fragments of different sizes were detected in the parental stocks STIB 386 and TREU 927/4 for all four markers and presumed to represent allelic variation in the number of repeat units, with each stock being heterozygous for different sized alleles (see Chapter 3, Figure 3.1). Only one amplified product was detected in the parental stock, STIB 247 for markers *CRAM*, *MS42* and *292*, and so the stock was presumed to be homozygous for these markers. For JS2 two bands were generated for STIB 247 suggesting that this stock is heterozygous for this marker.

Analysis of all clones derived from the three crosses, indicated that the majority of clones fall into one of three categories; parentals, selfers and F1 hybrids. The results for all four markers for every clone derived from the crosses are presented in Table A1 of the appendix.

Parental clones. A large percentage (44%) of clones derived from mixed-infected tsetse flies demonstrate the same allelic pattern as one of the parents for all markers analysed (Table A1 of the appendix). This suggests that sexual recombination is not obligatory.

F1 hybrids. Hybrid clones were identified as having two PCR amplified bands which correspond to one band from each parent (for example, see Chapter 3, Figure 3.2). This demonstrates that the variable size PCR products are different alleles which appear to be inherited in a classical Mendelian manner *i.e.* most combinations of the parental alleles are observed in the F1 progeny. Hybrid clones which possessed three alleles for one or more markers were considered trisomic or triploid and are discussed in the following section. It is also clear from the patterns of allele inheritance shown in Figure 3.2 A and B (Chapter 3) that the alleles for the minisatellite loci *CRAM* and *292* are inherited independently of each other. For example, hybrids F296/44bscl4 and F296/39bscl7 have inherited the same alleles for the *CRAM* locus, allele 1 from STIB 386 and allele 4 from TREU 927/4 (Fig. 3.2 A, lanes 5 and 6) but for the *292* minisatellite, these two hybrids have inherited different parental alleles (Fig. 3.2 B, lanes 5 and 6), indicating that the two loci are inherited independently. Examination of all hybrid clones from all three crosses for the micro- and minisatellite loci (see Table 5.1) also indicate the independent assortment of alleles at these loci. This independent assortment of alleles was expected for all four markers, as they are located on different chromosomes [Lee *et al.*, 1990; Lee *et al.*, 1994; Barrett *et al.*, 1997; Sasse, 1998]. Analysis of the largest collection of hybrids from one cross (STIB 247 x TREU 927/4) for all pairwise combinations of markers using the χ^2 test of independence, did not detect any linkage disequilibrium (*MS42/CRAM*, $\chi^2 = 0.225$, d.f.= 1, P>0.5; *CRAM/292*, $\chi^2 = 0.622$, d.f.= 1, P>0.3; *MS42/292*, $\chi^2 = 0.035$, d.f.= 2, P>0.95; *MS42*/JS2, $\chi^2 = 0.1$, d.f.= 2, P= 0.95; *CRAM*/JS2, $\chi^2 = 0.11$, d.f.=2, P=0.95; *292*/JS2, χ^2=0.389, d.f.= 2, P>0.7; for every comparison, the genotype classes with the lowest expected values were merged to raise the expected value to 5 or more).

Analysis of all 156 clones derived from crosses revealed that 81 were hybrids, 53 of which were the products of unique events. Full marker analysis for all clones derived from the three crosses is presented in Table A1 of the appendix, and for unique hybrid clones in Table 5.1 A-C.

Table 5.1A STIB 247 x TREU 927/4

Clones	GUP no.	Lysate no.	Markers				Comments
			Cram	*292*	*MS42*	JS2	
Reference 247			1-1	5-5	5-5	5-6	
Reference 927/4			3-4	3-4	3-4	3-4	
F124/28 bscl A1		68	1-3	3-5	3-5	3-5	Hybrid
F124/28 bscl A6		72	1-3-4	3-4-5	3-4-5	nd	Mix or triploid
F124/28 bscl B3		73	1-4	3-5	4-5	4-6	Hybrid
F124/28 bscl C5[1]		71	1-4	3-5	4-5	4-6	Hybrid
F532/53 mcl 1	3114	107	1-3	3-5	3-4-6*	4-5	Hybrid-mutant/ trisomic
F532/63 bscl 2		80	1-4	3-5	3-5	3-6	Hybrid
F532/63 bscl 3		78	1-4	4-5	4-5	4-5	Hybrid
F532/63 bscl 5		81	1-3	3-5	4-5	3-6	Hybrid
F532/63 bscl 7		79	1-4	4-5	4-5	3-6	Hybrid
F532/63 bscl 8		82	1-4	4-5	4-5	3-5	Hybrid
F532/72 mcl 1	3128, 4391	84	1-4	4-5	3-5	3-5	Hybrid
F532/72 mcl 2	3129	85	1-4	4-5	3-5	4-5	Hybrid
F532/72 mcl 3	3130	86	1-4	4-5	4-5	4-5	Hybrid
F532/72 mcl 4	3131	87	1-4	4-5	3-5	3-5	Hybrid
F532/72 mcl 5	3132, 4392	88	1-4	4-5	4-5	3-5	Hybrid
F532/72 mcl 6	3133, 4393	89	1-3	3-5	3-5	4-5	Hybrid
F532/72 mcl 7	3134	90	1-4	4-5	4-5	4-6	Hybrid
F532/72 mcl 9	3136, 4366	92	1-4	4-5	4-5	3-6	Hybrid
F532/72 mcl 10	3135, 4360	91	1-4	3-5	4-5	4-5	Hybrid
F974/70 mcl 4	3086	77	1-4	3-5	4-5	3-5	Hybrid

Table 5.1B STIB 386 x TREU 927 /4

Clones	GUP no.	Lysate no.	Markers				Comments
			Cram	*292*	*MS42*	JS2	
Reference 386			1-2	1-2	1-2	1-2	
Reference 927/4			3-4	3-4	3-4	3-4	
F296/44 bscl 1	3199	2	2-3	2-3	1-3	1-3	Hybrid
F296/44 bscl 2	3200, 4261	3	1-2-3	1-2-4	1-2-3	1-2-3	Triploid
F296/44 bscl 3	3201	6	1-3	1-3	1-3	1-4	Hybrid
F296/44 bscl 4	3204	5	1-4	1-4	1-4	1-3	Hybrid
F296/44 bscl 12	3211	9	1-3	2-3	1-2-4	2-3	Hybrid- mix or trisomic
F296/56 mcl 6	3215	1	2-3	2-3	1-3	1-3	Hybrid
F296/39 bscl 2			1-3	1-3	1-3	1-4	Hybrid
F296/39 bscl 6			1-4	1-4	1-4	1-3	Hybrid
F296/39 bscl 7			1-4	2-3	1-3	2-3	Hybrid
F296/39 bscl 17			1-3	2-3	1-2-4	2-3	Trisomic
F296/39 bscl 22/1			1-4	1-3	1-3	1-4	Hybrid
F296/42bscl 26			1-3	1-3	1-3	1-4	Hybrid

Table 5.1C STIB 247 x STIB 386

Clones	GUP no.	Lysate no.	Markers				Comments
			Cram	*292*	*Ms42*	JS2	
Reference 247			1-1	5-5	5-5	5-6	
Reference 386			1-2	1-2	1-2	1-2	
723VI-L[2]		153	nd	nd	1-2-5	nd	Triploid
723CAB[2]		154	nd	nd	1-2-5	nd	Triploid
F9/34 mcl 1		37	1-1	2-5	1-5	2-5	Hybrid
F9/45 mcl 2	3300	30	1-2	1-5	2-5	1-5	Hybrid
F9/45 mcl 7		32	1-2	1-5	1-5	1-5	Hybrid
F9/45 mcl 9	3290	33	1-1	2-5	1-5	1-6	Hybrid
F9/45 mcl 10	3291	34	1-1	2-5	1-5	1-6	Hybrid
F9/45 mcl 11	3287	35	1-2	1-5	1-5	1-5	Hybrid
F9/45 mcl 12	3288, 3296	36	1-1	2-5	1-2-5	1-5	Trisomic
F492/50 mcl 12	2843	39	1-2	1-5	1-5	2-6	Hybrid
F492/50 mcl 13	2856	40	1-1	2-5	1-5	1-6	Hybrid
F492/50 bscl 1			1-1	2-5	1-5	2-6	Hybrid
F492/50 bscl 2			1-1	1-5	1-5	2-6	Hybrid
F492/50 bscl 3			1-2	1-5	1-5	2-6	Hybrid
F492/50 bscl 6			1-2	1-5	1-2-5	1-5	Trisomic
F492/50 bscl 7			1-2	2-5	1-2-5	1-6	Hybrid -mix or trisomic
F492/50 bscl 8			1-1	2-5	1-5	1-6	Hybrid
F492/50 bscl 9			1-2	1-5	2-5	2-5	Hybrid
F492/50 bscl 11			1-2	1-5	1-5	2-5	Hybrid
F492/50 bscl 12			1-1	2-5	2-5	2-5	Hybrid
F492/50 bscl 14			1-1	1-2-5	1-5	1-6	Trisomic
F492/50 bscl 16			1-2	1-2-5	1-2-5	nd	Hybrid -mix or triploid

Tables 5.1 A-C. Mini- and microsatellite analysis of unique hybrid trypanosome clones. (**A**) Cross STIB 247 x TREU 927/4. (**B**)Cross STIB 386 x TREU 924/4. (**C**) Cross STIB 247 x STIB 386. Column 1 indicates the hybrid clone identification number used throughout this study. Columns 2 and 3 give additional identification numbers i.e. stabilate and lysate numbers. Columns 4-7 give the results of mini- and microsatellite analysis. Alleles are numbered 1-5 for the minisatellites, *CRAM*, *292* and *MS42*, and 1-6 for the microsatellite JS2. Column 8 gives the interpretation of results for each clone. Bloodstream clones from the same tsetse fly, sampled on the same day and exhibiting the same pattern for all markers were considered vegetative derivatives of the one sexual recombination event and are represented in this table by one clone. The complete set of results incorporating all clones derived from genetic crosses (including clones of parental genotype) is presented in Table 1 of the appendix. ([1] indicates clone differs from other bloodstream clones by other markers not presented here). Each metacyclic clone was considered to be a product of a unique event. Clones exhibiting three band patterns for one of the markers were considered either a mix or trisomic (lightly shaded rows). Clones which maintained their three band pattern after recloning were considered trisomic (shaded rows). Clones were considered triploid if a three band pattern was obtained for all markers analysed (shaded rows). [2] indicates original hybrid clones identified by Jenni *et al.*, in 1986. Analyses of parental clones are given in rows 1 and 2 of each table, for reference.
GUP, Glasgow University Parasitology; nd, not determined. * indicates novel sized mutant allele.

Assuming a Mendelian system of inheritance, there are 32 possible progeny genotypes which can be detected from the crosses, STIB 247 x STIB 386 and STIB 247 x TREU 927/4, where one parental clone is heterozygous and the other is homozygous for three markers (*CRAM*, *292* and *MS42*) and both are heterozygous for the fourth marker (JS2). For the STIB 247 x TREU 927/4 cross, 13 different genotypes were detected out of 18 hybrids (triploids and trisomics were excluded), with five genotypes being detected twice. No genotypes were detected three times. Similarly for the STIB 247 x STIB 386 cross, 10 different genotypes were detected in 15 hybrids with two genotypes being detected twice and one genotype being detected four times.

In order to test if the ratio of different hybrid genotypes detected was in agreement with those predicted by Mendelian segregation ratios, the two largest data sets of unique hybrids (excluding triploids and trisomics), STIB 247 x TREU 927/4 (18 unique hybrids) and STIB 247 x STIB 386 (15 unique hybrids) were analysed. A comparison of the observed frequency of different hybrid genotypes to the expected genotypes assuming Mendelian inheritance, was made for each marker in turn, using the χ^2 test of homogeneity (Table 5.2). For Table 5.2 the STIB 247 alleles were not included as they were uninformative.

Cross	Marker	STIB 386 or TREU 927/4 alleles inherited	observed (expected)	χ^2	d.f.	Accept/ reject Ho significance of 0.05
STIB 247 x TREU 927/4	*CRAM*	3	3 (9)			reject
		4	15 (9)	8	1	P<0.01
	292	3	8 (9)			accept
		4	10 (9)	0.22	1	0.7>P>0.5
	MS42	3	6 (9)			accept
		4	12 (9)	2	1	0.2>P>0.1
	JS2	3	10 (9)			accept
		4	8 (9)	0.22	1	0.7>P>0.5
STIB 247x STIB 386	*CRAM*	2	7 (7.5)			accept
		1	8 (7.5)	0.07	1	0.9>P>0.7
	292	1	8 (7.5)			accept
		2	7 (7.5)	0.07	1	0.9>P>0.7
	MS42	1	12 (7.5)			reject
		2	3 (7.5)	5.4	1	0.05>P>0.01
	JS2	1	7 (7.5)			accept
		2	8 (7.5)	0.1	1	0.9>P>0.7

Table 5.2. Test for Mendelian inheritance of each marker from the two crosses STIB 247 x STIB 386 and STIB 247 x TREU 927/4. Column 3 indicates the genotypes for each marker examined. Column 4 gives the observed and expected numbers of genotypes. The χ^2 value and degrees of freedom are presented in columns 5 and 6, respectively. Column 7 indicates the acceptance (shaded) or rejection of the null hypothesis at the 0.05 level of significance.

From the data presented in Table 5.2, although from a small sample size, appears to

indicate that the majority of markers do not show deviation from the allele segregation predicted by Mendelian inheritance. The *CRAM* locus for parental stock TREU 927/4 and the *MS42* locus for STIB 386, however, are not in agreement with the null hypothesis of Mendelian allele segregation. This segregation distortion could be because unequal numbers of the different gametes are produced or selection could favour one particular genotype over another.

The third cross, STIB 386 x TREU 927/4, is the smallest data set with only 12 unique hybrids all originating from one fly, including one triploid and two samples which could be either mixed or trisomic (see following sections). Both parental stocks for this cross are heterozygous for each marker, therefore the number of possible distinct progeny genotypes predicted for each marker is 4, and 256 for all four markers combined. Examination of the different genotypes detected from this cross, numbered 1-7, (excluding triploids and bloodstream clones from the same fly and day, which appeared identical), is presented in Table 5.3. It is clear from this table that three genotypes have been detected twice and one genotype three times. Given that there are 256 different genotypes, assuming Mendelian inheritance, then it would be highly unlikely that one genotype would appear three times ($P = 1.5 \times 10^{-4}$) by chance and another two genotypes appear twice ($P = 3.9 \times 10^{-3}$) out of a sample size of 12. This is especially true for the genotype which appears to be trisomic, a usually rare event (see following section), which has been detected twice in this fly.

genotypes	Genotypes of clones from F296 sampled on				
	day 39	day 42	day 44	day 56	Total
1			1	1	2
2	1	1	1		3
3	1		1		2
4	1				1
5	1				1
6	1		1		2
7			1		1
					12

Table 5.3. Different trypanosome genotypes detected from fly 296 sampled at different time points. Column 1 indicates the different hybrid genotypes detected from this fly. Columns 2-5, gives the different genotypes sampled on days 39, 42, 44 and 56 respectively. Column 6 provides the number of times the genotypes were sampled over the time course. Identical genotypes sampled on the same day were removed from this analysis.

There are several possible reasons for these results: 1) Since most of the hybrids from this cross are bloodstream clones, it is possible that the pre-amplification step before cloning, could select for particular genotypes. Although identical bloodstream clones from the same day have been removed from this analysis, it is possible that the same hybrid genotypes are selected for at each time point. 2) It is possible that selection for particular

genotypes occurs in the salivary glands, with certain genotypes not being viable. 3) There is an amplification step in the salivary glands of the tsetse flies. It has been assumed that since metacyclics are non-dividing, there is no amplification in the salivary glands. However it is possible that hybrids are generated as epimastigotes which then produce a number of hybrid metacyclics with the same genotype. 4) The sexual recombination process is not Mendelian. 5) Matings involving STIB 386 and TREU 927/4 may be more problematic than the other crosses, with a high proportion of triploids and trisomics being detected with STIB 386 chromosomes (see following sections). It is therefore conceivable that only a few cells successfully undergo sexual recombination or many of the progeny are selected against.

To summarise, it would appear that the majority of hybrid clones demonstrate allelic segregation for the mini- and microsatellite loci examined. The ratios of the different hybrid genotypes are broadly in agreement with those expected from a Mendelian system for the two crosses, STIB 247 x TREU 927/4 and STIB 386 x STIB 247. The third cross, however, does not appear to follow Mendelian ratios, perhaps due to selection of different hybrid genotypes or limited mating in the salivary glands generating a restricted number of hybrid genotypes which proliferate in the salivary glands. The independent assortment of alleles at different loci has also been demonstrated for the four loci examined. Taken together, the data provide evidence to suggest that the sexual process in *T. brucei* follows a broadly Mendelian system.

Selfers. The third class of clone derived from genetic crosses of *T. brucei* is a result of self-fertilization, which has been demonstrated previously in a fly co-infected with STIB 247 and STIB 386 [Tait *et al.*, 1996]. The products of self-fertilization, 'selfers', would be expected to be homozygous for approximately 50% of the markers for which the parental stock is heterozygous. In the analysis carried out by Tait *et al.*, five products of STIB 247 self-fertilization were identified using a combination of isoenzyme, RFLP and karyotype analysis.

Using four markers for which the parental stock is heterozygous, approximately 93.75% of all selfers should be identifiable. Since the parental stocks TREU 927/4 and STIB 386 are heterozygous for all four mini- and microsatellites, the majority of TREU 927/4 and STIB 386 selfers would be identified using the minisatellites described. However, for STIB 247 only 50% of selfers would be detected because JS2 is the only heterozygous marker. From the analysis of 156 clones derived from the three crosses, 75 clones had a non-hybrid genotype (Table 5.4 A-C). Thirty-nine non-hybrid clones were derived from TREU 927/4, and only one possible TREU 927/4 selfer, F296/46mcl 1, was identified. This clone was homozygous for the 292 minisatellite marker (Figure 5.4, lane 4). Lanes 1-3 of this figure illustrate the normal Mendelian inheritance of this marker for hybrid clones, with lanes 5 and 6 being the parental stocks, TREU 927/4 (genotype 3-4) and

Table 5.4A STIB 247 x TREU 927/4

Clones	GUP no.	Lvsate no.	Markers				Comments
			CRAM	*292*	*MS42*	JS2	
Reference 247			1-1	5-5	5-5	5-6	
Reference 927/4			3-4	3-4	3-4	3-4	
F124/28 bscl B1		69	3-4	3-4	3-4	nd	Parental
F124/28 bscl C3		75	3-4	3-4	3-4	nd	Parental
F124/28 bscl C2		70	3-4	3-4	3-4	nd	Parental
F974/70 mcl 1	3083	93	1-1	5-5	5-5	5-6	Parental
F974/70 mcl 2	3084	94	1-1	5-5	5-5	5-6	Parental
F974/70 mcl 3	3085	95	1-1	5-5	5-5	5-6	Parental
F974/70 mcl 6	3088	96	3-4	3-4	3-4	3-4	Parental
F974/70 mcl 7	3089	97	3-4	3-4	3-4	3-4	Parental
F974/70 mcl 8	3090	98	1-1	5-5	5-5	5-6	Parental
F974/78 mcl 1	3092	99	3-4	3-4	3-4	3-4	Parental
F974/78 mcl 2	3093	100	1-1	5-5	5-5	5-6	Parental
F974/78 mcl 3	3094	101	1-1	5-5	5-5	5-5	Selfer
F974/78 mcl 4	3095	102	3-4	3-4	3-4	3-4	Parental
F974/78 mcl 6	3096	105	1-1	5-5	5-5	5-6	Parental
F974/78 mcl 7	3097	106	1-1	5-5	5-5	5-6	Parental

Table 5.4B STIB 247 x STIB 386

Clones	GUP no.	Lvsate no.	Markers				Comments
			CRAM	*292*	*MS42*	JS2	
Reference 247			1-1	5-5	5-5	5-6	
Reference 386			1-2	1-2	1-2	1-2	
F9/28 mcl 3	2846	43	1-1	5-5	5-5	5-6	Parental
F9/34 mcl 2		159	1-1	5-5	5-5	5-6	Parental
F9/45 mcl 4	4295	31	1-1	5-5	5-5	5-6	Parental
F18/50 mcl 4		165	nd	nd	5-5	nd	Parental
F18/50 mcl 8		166	nd	nd	5-5	nd	Parental
F19/31 mcl 2		160	1-1	5-5	5-5	5-6	Parental
F19/31 mcl 3		161	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 1	2834	46	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 4	2838	49	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 5	2839	50	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 6	2840	51	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 7	2841	52	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 8	2847	56	1-1	5-5	5-5	5-5	Selfer
F492/50 mcl 9	2852	57	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 10	2853, 4312	58	1-1	5-5	5-5	5-5	Selfer*
F492/50 mcl 11	2854, 4364	59	1-1	5-5	5-5	5-6	Selfer*
F492/50 mcl 15	2857, 3365	61	1-1	5-5	5-5	5-5	Selfer*
F492/50 mcl 16	2858, 3363	62	1-1	5-5	5-5	5-6	Selfer*
F492/50 mcl 17	2859	63	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 18	2860, 3364	64	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 19		48	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 20	2861	65	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 21		55	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 22		53	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 23		54	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 24	2848	41	1-1	5-5	5-5	nd	Parental
F492/50 mcl 25	2855	60	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 26	2862	66	1-1	5-5	5-5	6-6	Selfer
F492/50 bscl 22			1-1	5-5	5-5	5-6	Parental

Table 5.4C STIB 386 x TREU 927 /4

Clones	GUP no.	Lysate no.	Markers				Comments
			CRAM	*292*	*MS42*	JS2	
Reference 386			1-2	1-2	1-2	1-2	
Reference 927/4			3-4	3-4	3-4	3-4	
F296/42 bscl 5	3203	22	3-4	3-4	3-4	3-4	Parental
F296/42 bscl 6	3210	23	3-4	3-4	3-4	3-4	Parental
F296/42 bscl 11	3208	24	3-4	3-4	3-4	3-4	Parental
F296/46 mcl 1	3196	12	3-4	4-4	3-4	3-4	Selfer
F296/46 mcl 2	3197	13	3-4	3-4	3-4	3-4	Parental
F296/56 mcl 1	3218	14	3-4	3-4	3-4	3-4	Parental
F296/56 mcl 2	3219	15	3-4	3-4	3-4	3-4	Parental
F296/56 mcl 3	3212	16	3-4	3-4	3-4	3-4	Parental
F296/56 mcl 4	3213	17	3-4	3-4	3-4	3-4	Parental
F296/56 mcl 5	3214	18	3-4	3-4	3-4	3-4	Parental
F296/56 mcl 7	3220	19	nd	3-4	3-4	3-4	Parental
F296/56 mcl 8	3216	20	3-4	3-4	3-4	3-4	Parental
F296/56 mcl 9	3217	21	3-4	3-4	3-4	nd	Parental
F296/39 bscl 1			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 4			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 8			nd	nd	3-4	3-4	Parental
F296/39 bscl 10			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 11			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 13			nd	3-4	3-4	3-4	Parental
F296/39 bscl 14			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 16			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 18			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 32			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 40			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 41			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 44			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 58			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 59			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 22/2			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 23/3			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 26/2			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 57			nd	nd	3-4	3-4	Parental

Table 5.4 A-C. Mini- and microsatellite analysis of selfers and parental clones derived from genetic crosses. (A) Cross STIB 247 x TREU 927/4. (B) Cross STIB 247 x STIB 386. (C) Cross STIB 386 x TREU 924/4. Column 1 indicates the clone identification number used throughout this study. Columns 2 and 3 give additional identification numbers i.e. stabilate and lysate numbers. Columns 4-7 give the results of mini- and microsatellite analysis. Alleles are numbered 1-5 for the minisatellites, *CRAM, 292* and *MS42*, and 1-6 for the microsatellite JS2. Column 8 gives the interpretation of results for each clone. Shaded rows indicate possible selfers i.e. the products of self-fertilisation. Analysis of parental clones is given in rows 1and 2, for reference.
* indicates previously identified selfers [Tait *et al.*, 1996].
GUP, Glasgow University Parasitology. nd, not determined.

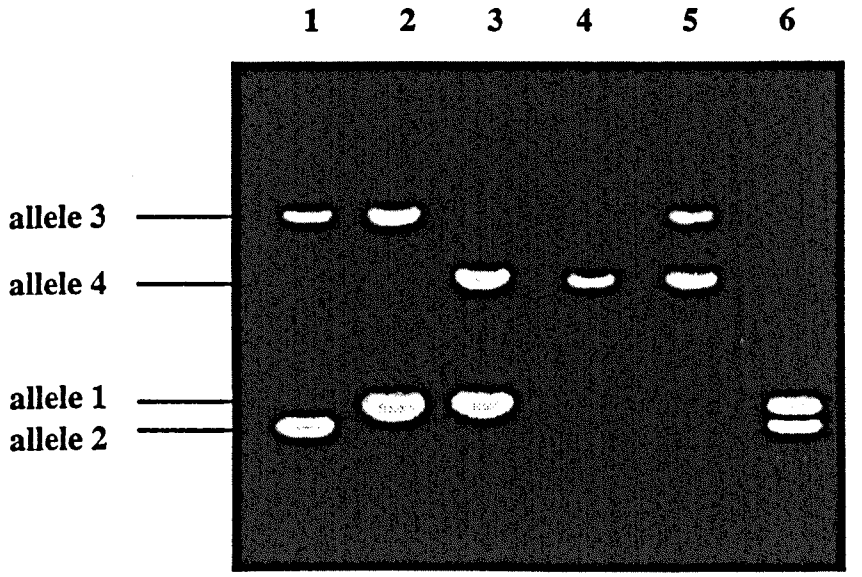


Figure 5.4. Minisatellite marker analysis of hybrid clones, including a putative 927 selfer. PCR amplification of minisatellite marker, *292*, from parental stocks and hybrid clones. PCR conditions were as described in Materials and Methods. The products were size separated on a 1% Seakem agarose gel and visualised by ethidium bromide staining. Lane 1, F296/39bscl 6; lane 2, F296/39bscl 2; lane 3, F296/56mcl 6; lane 4, F296/46mcl 1; lane 5, TREU 927/4; lane 6, STIB 386.

STIB 386 (genotype 1-2), respectively. From this figure it is clear that clone, F296/46mcl 1 has only inherited allele 4 and not allele 3 from the parental stock TREU 927/4. For all other markers examined, this clone was identical to TREU 927/4. To investigate if allele 3 was present in the clone, but amplified poorly in the PCR reaction, this gel was Southern blotted and hybridized to the *292* probe (data not shown). No trace of another amplification product other than allele 4 was detected.

No STIB 386 parental clones or STIB 386 selfers were identified from any of the crosses. The lack of parental STIB 386 genotypes is consistent with the previous observation that STIB 386 has a lower transmission index than the other two stocks. To explain this observation assuming that equal proportions of midgut infections for each stock was achieved in a co-transmission experiment, STIB 386 would not progress to the salivary glands until after the other stock (either STIB 247 or TREU 927/4) had established infections there. It would appear that as soon as STIB 386 infected the salivary glands it began to mate with the stock which was already established there, producing only hybrids.

The parental stock STIB 247 is only heterozygous for one marker, JS2. Therefore only approximately 50% of all STIB 247 selfers could be identified using this marker. In this analysis, of the 37 non-hybrid clones derived from STIB 247, five selfers were identified (two of which had previously been identified by Tait *et al.*, 1996) bringing the total of STIB 274 selfers identified to date to seven (six are described in Table 5.4 A and B, the other in [Tait *et al.*, 1996]). Since only ~50% of selfers can be detected by the microsatellite marker JS2, it is predicted that approximately 10 selfers are present in the 37 non-hybrid clones (27%). This is in marked contrast to the proportion of TREU 927/4 selfers, which constitutes 2.6% of the non-hybrid clones, almost a 10 fold difference. This could indicate that STIB 247 undergoes self-fertilization more readily that TREU 927/4.

In total eight selfers have been identified from the three crosses. The availability of multiple heterozygous markers for each parental stock allows the detection of selfing with a high degree of sensitivity, identifying two further STIB 247 selfers and a TREU 927/4 selfer, which were not identified using isoenzymes or RFLPs [Tait *et al.*, 1996]. The high level of heterozygosity exhibited by mini- and microsatellites makes them particularly useful in this context.

However, one concern regarding the identification of selfers from the PCR analysis of minisatellite markers is the possibility that the homozygosities detected for these markers were not genuine, *i.e.* they were heterozygotes in which one allele has failed to amplify, for this reason all PCR amplifications were performed in duplicate. However, to confirm that each stock was indeed a selfer and homozygous for that particular marker, a genomic Southern of each stock should be performed. Further examination of these clones, such as karyotype analysis would also help to confirm (or refute) the initial interpretation that these clones were the products of self-fertilization. Full analysis of all parental and selfer clones from the three crosses are presented in Table 5.4.

Examination of clones derived from single stock transmissions was also undertaken using one of the minisatellites [Barrett *et al.*, 1997](Chapter 3, Table 3.1), initially to test for minisatellite stability (see Chapter 3), but these experiments would also detect any STIB 386 or TREU 927/4 selfers generated during single stock transmissions. However no selfers were detected from the 19 stocks examined (although this is not surprising if only 2.6% of TREU 927/4 non-hybrid clones are selfers). Similar analysis of clones derived from single stock transmissions by Tait *et al.*, [1989 and by Gibson *et al.*, [1997] have not demonstrated selfing. The fact that self-fertilization has only been detected in the context of cross-fertilization suggests that the sexual process is only induced when two stocks interact with each other, perhaps through the secretion of pheromones [Gibson *et al.*, 1997].

Triploids. It is clear from the analysis presented in Table 5.1, that the vast majority of hybrid clones were equivalent to F1 hybrids, consistent with a Mendelian genetic system and that diploidy is usually maintained, at least for the four markers on different chromosomes analysed here.

DNA content analysis by flow cytometry on 30 of these hybrids, performed by M. Hope [Hope *et al.*, 1999], indicated that the majority of hybrids had DNA contents within the parental range, denoting the maintenance of diploidy. However, two clones, F296/44bscl 2 and F296/44bscl 7, appear to have raised DNA contents of 1.6x and 1.5x those of the parents, respectively.

Analysis of these clones using the three minisatellite and one microsatellite markers indicated that these clones had inherited three alleles for each marker; both alleles from one parent (STIB 386) and one from the other (TREU 927/4). Figure 5.5 illustrates the inheritance patterns for all four markers for two normal diploid hybrids (lanes 3 and 4) and the two hybrids with raised DNA content, the putative triploids, (lanes 5 and 6). The results indicate that clones F296/44bscl 2 and 7 are probably trisomic for each chromosome tested and together with results from the DNA content analysis, suggest that these clones are triploid. Since these clones have identical marker profiles and are bloodstream clones from the same fly, sampled on the same day, it is likely that these clones are vegetative derivatives of the one genetic exchange event.

To rule out the possibility that these extra alleles were due to DNA contamination in the PCR reactions, a Southern blot using *Pst* I digests of genomic DNA of one of the putative triploid clones and several diploid hybrid clones was performed and hybridized to the *MS42* repeat probe (see Chapter 2 for a description of the probe). Figure 5.6 illustrates an autoradiograph obtained from the genomic Southern. The results confirm the PCR data demonstrating that the normal diploid clones have inherited two alleles one from each parental clone (lanes 3-5) and clone F296/44bscl 2 (lane 7) has clearly inherited both alleles from parental clone STIB 386 and allele 3 from TREU 927/4. The three band pattern in lane 6 is discussed in the following section.

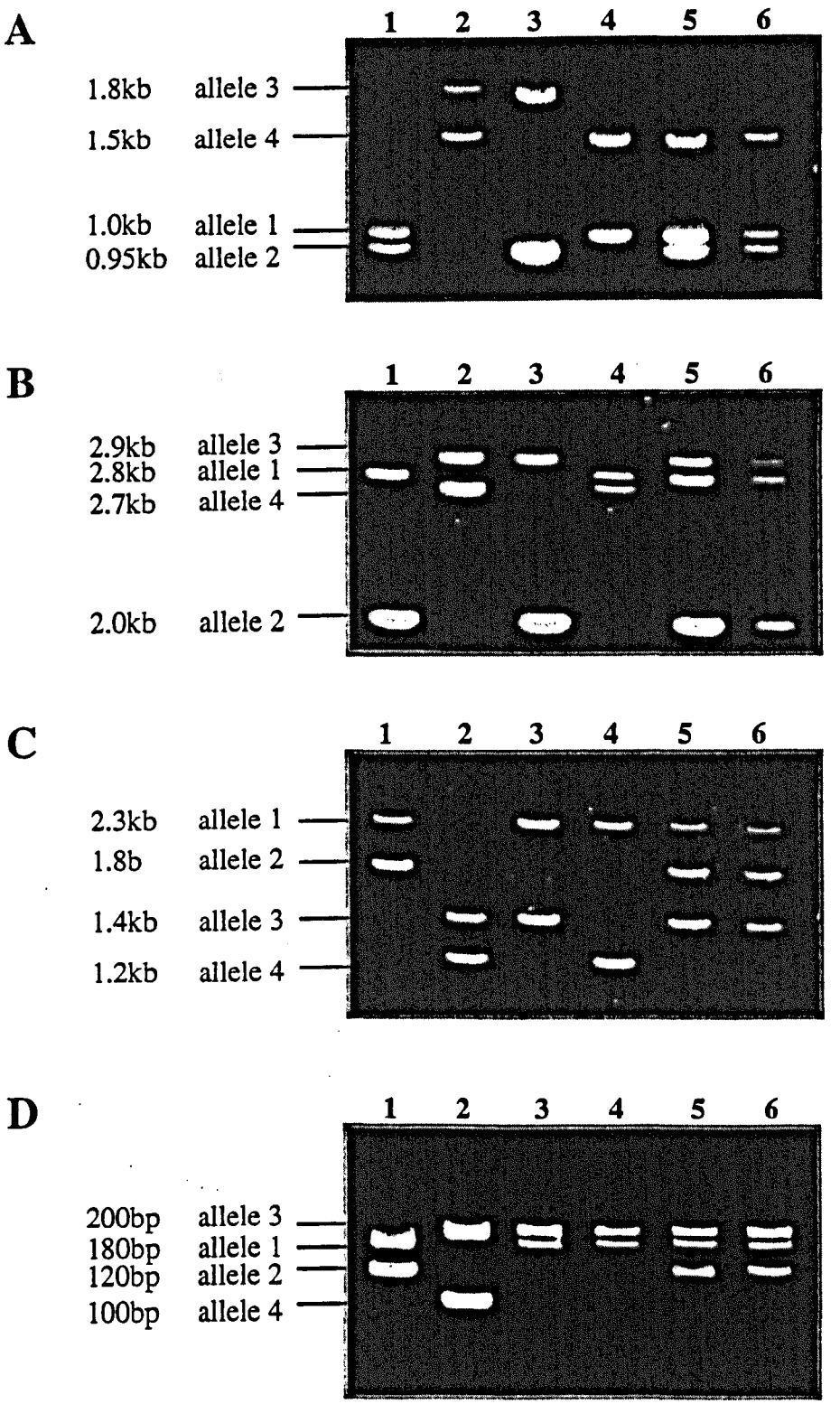


Figure 5.5. PCR amplification of mini- and microsatellite markers from four hybrids two of which are putative triploids. PCR amplification of mini- and microsatellites, (A) *292*, (B), *CRAM*, (C) *MS42* and (D) JS2, from DNA isolated from the parental stocks and hybrid clones. PCR conditions were as described in Materials and Methods. The products for *292*, *CRAM* and *MS42* were size separated on a 1% Seakem agarose gel, for JS2 on a 3% Nusieve gel, and visualised by ethidium bromide staining. Lane 1, STIB 386; 2, TREU 927/4; 3, F296/44bscl 1; 4, F296/44bscl 13; 5, F296/44bscl 2; 6, F296/44bscl 7.

Further examination of the putative triploids was carried out using Southern blots of chromosomes separated by pulse field gel electrophoresis [Hope *et al.*, 1999]. There are 11 megabase chromosomes which contain housekeeping genes [Melville *et al.*, 1998]. Although not all chromosomes were informative by PFGE, the results of analysis of five chromosomes, indicated trisomy for each chromosome. These results, together with the mini- and microsatellite analysis presented above, indicate that a total of 7 chromosome pairs were found to be trisomic using 9 markers, strongly suggesting that these clones were triploid.

From the mini- and microsatellite analysis of all the hybrid clones identified (Table A1, appendix for all clones, and Table 5.1 for unique hybrids only) it is clear that two other clones could be triploid, clone F124/28bscl A6 and clone F492/50bscl 16. It is possible that the sample F124/28bscl A6 is not a triploid clone but a mixture of the parental stocks. In order to prove that it is a genuine triploid the sample must be recloned and re-analysed for the minisatellite markers and for DNA content. Clone F492/50bscl 16 appears to be trisomic for both *MS42* and *292*, which are on different chromosomes but is uninformative for *CRAM*, since the parental stocks STIB 386 and STIB 247 share an allele for this minisatellite. Recloning of this stock followed by re-examination for the markers verified

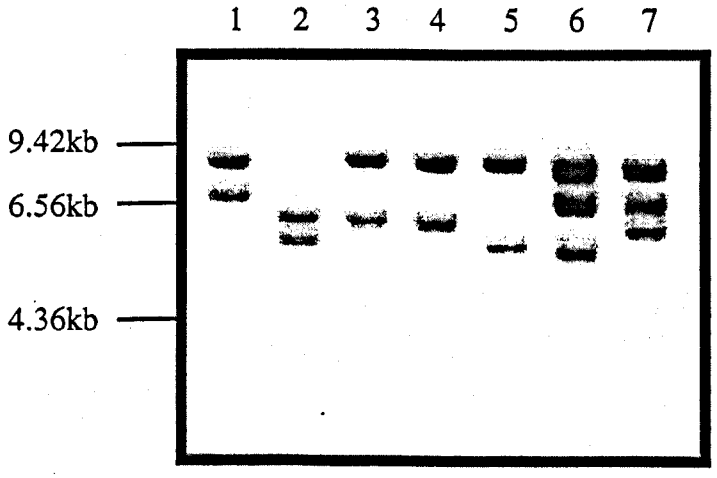


Figure 5.6. Genomic Southern of hybrids, including one triploid and one putative trisomic clone. 5µg of total genomic DNA was digested with *Pst* I, electrophoresed through a 1% Seakem agarose gel, blotted and hybridized to an *MS42* repeat probe. Lane 1, STIB 386; lane 2, TREU 927/4; lane 3, F296/44bscl 1; lane 4, F296/44bscl 3; lane 5, F296/44bscl 8; lane 6, F296/44bscl 12; lane 7, F296/44bscl 2.

that it was not a mixed sample (see Table 5.5). It is possible that this clone is another triploid, however DNA content analysis would have to be undertaken in order to demonstrate this. Also further analysis of this clone for the JS2 microsatellite may provide further evidence for/against triploidy.

The analysis of 53 unique clones, has revealed that one clone is probably triploid [Hope *et al.*, 1999] and another two clones could be triploids, although further analysis would have to be undertaken to verify this. It is evident, therefore, that the proportion of triploids from these crosses is of the order of 2% to 6% of the total number of unique clones. This is in contrast to the finding of Gibson *et al.*, [Gibson, 1995] who observed an increase in DNA content in 12 of 18 (66.7%) genotypically distinct hybrid clones from crosses between *T. b. brucei* and *T. b. rhodesiense* stocks. It is possible that the particular stocks used by Gibson *et al.*, may not be fully genetically compatible, with triploidy being an aberrant product of sexual reproduction that is strain dependent.

Trisomics. From the analysis of the hybrid clones with the minisatellite markers, it is clear that not all hybrids inherit just two alleles from the parental clones. Triploids have been mentioned previously. However there appears to be another class of hybrid which is trisomic for just one locus examined (highlighted in Table 5.1). The genomic Southern of some of the hybrids probed with *MS42* indicate that one clone, F296/44bscl 12, had inherited three *MS42* alleles (Figure 5.6, lane 6). However for the other minisatellite markers this clone appeared to have the normal number of alleles, one from each parental clone (Table 5.1B). There are three possible explanations for this observation. The first is that the sample is actually a mix of two hybrid clones which differ at this one locus. The second is that the clone is trisomic for the chromosome carrying the *MS42* marker (chromosome I). Thirdly a gene duplication event could have generated the extra copy of the locus. There are several clones in the collection which have a similar profile (11 in all), in that they appear to have an extra allele for one marker (see Table 5.5). In order to rule out the possibility that these clones are mixed, some of the samples were recloned and analysed again using the minisatellite markers (the recloning was performed by A. Tait). Three of the samples (F492/50bscl 5, F492/50bscl 25 and F296/39bscl 15) no longer gave a three band pattern after recloning and so the original stock must have been a mixture of two hybrid clones. However five clones (highlighted in Table 5.5) continued to give a three band pattern after recloning, indicating that these clones carried an extra allele at one locus. Clone F492/50bscl 16, is either trisomic for the chromosomes carrying *MS42* and *292*, or triploid and has been discussed in the previous section. Three of the four other possible trisomic clones indicated in Table 5.5 gave a three band pattern for the minisatellite marker, *MS42*, which is located on chromosome I, but a normal two band pattern for the other minisatellite markers. These clones were analysed further using another highly informative marker on chromosome I, D2, which was identified by J. Sasse and is described in detail in the next section (Figure 5.7B). All four clones which had an

Clones	Markers				
	CRAM	*292*	*MS42*	D2	JS2
Reference 247	1-1	5-5	5-5	5-5	5-6
Reference 386	1-2	1-2	1-2	3-4	1-2
Reference 927/4	3-4	3-4	3-4	1-2	3-4
F296/44 bscl 12	1-3	2-3	1-2-4	2-3-4	2-3
F492/50 bscl 7	1-2	2-5	1-2-5	nd	1-6
F532/53 mcl 1	1-3	3-5	3-4-6*	nd	4-5
F9/45 mcl 12	1-1	2-5	1-2-5	3-4-5	1-5
F9/45 mcl 12 reclone 1	1-1	2-5	1-2-5	3-4-5	1-5
F492/50 bscl 5	1-2	1-5	1-5	nd	2-5-6
F492/50 bscl 5 reclone 1	1-2	1-5	1-5	nd	2-6
F492/50 bscl 5 reclone 2	1-2	1-5	1-5	nd	2-6
F492/50 bscl 6	1-2	1-5	1-2-5	3-4-5	1-5
F492/50 bscl 6 reclone 1	1-2	1-5	1-2-5	3-4-5	1-5
F492/50 bscl 6 reclone 2	1-2	1-5	1-2-5	3-4-5	1-5
F492/50 bscl 16	1-2	1-2-5	1-2-5	nd	nd
F492/50 bscl 16 reclone 1	1-2	1-2-5	1-2-5	nd	nd
F492/50 bscl 25	1-1	2-5	1-5	nd	2-5-6
F492/50 bscl 25 reclone 1	1-1	2-5	1-5	nd	2-6
F492/50 bscl 25 reclone 2	1-1	2-5	1-5	nd	2-6
F492/50 bscl 25 reclone 3	1-1	2-5	1-5	nd	2-6
F492/50 bscl 14	1-1	1-2-5	1-5	4-5	1-6
F492/50 bscl 14 reclone 1	1-1	1-2-5	1-5	nd	1-6
F296/39 bscl 15	1-4	1-2-3	1-3	nd	2-3
F296/39 bscl 15 reclone 1	1-4	2-3	1-3	nd	2-3
F296/39 bscl 17	1-3	2-3	1-2-4	2-3-4	2-3
F296/39 bscl 17 reclone 1	1-3	2-3	1-2-4	2-3-4	2-3
F296/39 bscl 17 reclone 2	1-3	2-3	1-2-4	nd	2-3

Table 5.5. Mini- and microsatellite analysis of possible trisomic hybrid clones. Column 1 gives the identification number of each clone and the reclones. Columns 2-6 indicate the results of the analysis with the mini- and microsatellites. Alleles are numbered as in Table 5.1. Shaded rows indicate those clones which gave a three band pattern after recloning. Analysis of parental clones is given in rows 1-3 for reference. Note, *MS42* and D2 markers are both on chromosome I. * indicates novel sized mutant allele. nd, not determined.

extra allele for *MS42* also had an extra allele for the marker D2. These data suggest that the clones are trisomic for chromosome I and have not acquired an extra allele of *MS42* by gene duplication. Only one putative trisomic line has been analysed for DNA content. DNA content analysis was performed on F296/44bscl 12, which appears trisomic for chromosome I, and indicated that the DNA content of this clone was within the normal parental range [Hope *et al.*, 1999], demonstrating that this clone is not triploid.

The majority of the clones identified from previous crosses, for example from Gibson *et al.*, [1992], which have been shown to be trisomic, appear to be trisomic for a number of loci and are probably triploid. It is difficult to prove triploidy as three copies of every megabase chromosome must be demonstrated. However for these clones the raised DNA content to 1.5 x the parental level is highly suggestive of triploidy.

Trisomy appears to be a more common event than triploidy, with up to 8 possible trisomic clones, one of which is likely to be triploid (F492/50bscl 16), bringing the number of trisomics detected to 7, from a total of 53 unique clones. Trisomy is most likely to have arisen from chromosomal non-disjunction during meiosis, and in six of the seven examples identified in this study, the extra allele/chromosome was inherited from the parental stock STIB 386, suggesting that this stock may have difficulty in undergoing meiosis. The majority of possible trisomies (6/7) involve marker *MS42*, on chromosome I, (see Table 5.6), suggesting a bias in the failure of meiosis for this chromosome in stock STIB 386 or an ability for the trisomic trypanosome to survive with an extra chromosome I, where trisomy of other chromosomes may be selected against.

The relatively high incidence of trisomy in the hybrid clones (see Table 5.6), constituting 13% of the total number of unique clones from all crosses is consistent with the occurrence of meiosis in *T. brucei*. Trisomy due to nondisjunction is a common phenomenon in higher eukaryotes, for example, in humans, trisomy occurs in 25% of spontaneous abortions [Koehler *et al.*, 1996].

Cross	Possible trisomics		Marker displaying three band pattern			
	No.	%	*MS42*	*292*	*CRAM*	JS2
STIB 247x TREU 927/4	1/20	5	1	0	0	0
STIB 247 x STIB 386	4/21	19	3	1	0	0
STIB 386 x TREU 927/4	2/12	16.7	2	0	0	0

Table 5.6. Detection of trisomic clones in three crosses.

Bloodstream clones vs. metacyclic clones. The majority of clones derived from crosses have been bloodstream clones, generated as described in Figure 5.1B. However it

is not clear if the pre-amplification step involved in the generation of bloodstream clones selects for particular genotypes of trypanosome. In order to address this issue, 24 bloodstream clones were generated from the available F492/50 uncloned material. Twenty-three metacyclic clones had previously been generated from this fly at this time point [Turner *et al.*, 1990] and so it was possible to make a direct comparison between the metacyclic and bloodstream clones generated from this fly. Analysis of all 47 clones using the three minisatellites and one microsatellite is presented in Table A1 of the appendix and a summary of the results is presented in Table 5.7. It is obvious from these data that there is a startling difference between the bloodstream and metacyclic clones, with the bloodstream clones being mainly hybrids (96%) in direct contrast to the metacyclic clones, which were mainly parental or selfers (8.7% hybrids). It is evident from these results that some form of selection in favour of hybrid trypanosomes has taken place during the pre-amplification step in the generation of the bloodstream clones, despite the use of immunosuppressed mice. This suggests that these hybrid clones are better adapted to growth in mice. STIB 247 is the main genotype of trypanosomes present in the salivary glands of fly 492 (Table 5.7) and multiplies slowly in mice, often failing to reach a high parasitaemia compared to the other parental clone STIB 386 [Turner *et al.*, 1995]. It is therefore not surprising that hybrid clones grow more efficiently in mice than STIB 247. What perhaps is surprising is that this selection procedure has not resulted in the isolation of a STIB 386 genotype, possibly reflecting the fact that very few STIB 386 trypanosomes are present in the salivary glands.

F492/50 clones	Hybrids	Parentals	Selfers
metacyclic clones	2 (2 unique)	17 (all 247 genotype)	4
bloodstream clones	23 (11 unique)	1 (247 genotype)	0

Table 5.7. Comparison between bloodstream and metacyclic clones. Both bloodstream and metacyclic clones were isolated from tsetse fly F492 on day 50 post infection as described in the text and in figure 5.1. This table summarises the analysis of these clones (for complete minisatellite marker results see Table 1 of the appendix).

As mentioned previously, selfing has only been detected in the presence of cross-fertilization, suggesting a possible chemical signal for meiosis to occur only when a different stock is present. It is clear from the metacyclic clones generated from F492 on day 50, (Table 5.7) that the majority of trypanosomes in the salivary glands are parental STIB 247. Therefore if a chemical signal to stimulate meiosis and syngamy were made due to presence of a few STIB 386 cells, then a significant amount of self-fertilization of STIB 247 would occur, due to the abundance of this genotype. Indeed this appears to have been the case with 4 selfers having been detected from the 23 metacyclic clones generated.

Proportions of hybrids/parentals. Preliminary evidence has suggested that hybrid trypanosome genotypes are produced only late in the mixed infection of the salivary glands [Schweizer *et al.*, 1988; Sternberg *et al.*, 1989]. It was possible to examine this phenomenon by analysing 51 clones identified from one fly, F296, which had been infected with the parental stocks STIB 386 and TREU 927/4, sampled over a time course (see Table A1 in the appendix for a full analysis of all clones). It was possible to make a direct comparison between the different types of bloodstream clones sampled at each time point (Table 5.8). It is clear from this table that the proportions of hybrids to parentals (all of TREU 927/4 genotype) has altered over the course of the infection. The proportion of parentals has decreased, and the number of hybrids identified has increased as the infection has progressed. Although bloodstream clones do not give a true representation of the trypanosomes present in the tsetse fly salivary gland (see previous section) and it is unclear if there is a selective (dis)advantage for hybrids over TREU 927/4 parentals, the proportions of hybrids to parentals should reflect the general trend of change within the salivary gland over time.

day sampled	Number of clones derived from fly 296	
	hybrids	Parentals
day 39	13 (41%)	19 (59%)
day 42	7 (70%)	3 (30%)
day 44	9 (100%)	0 (0%)

Table 5.8. Typing of clones derived from one fly over a time course. Column 1 indicates the day on which the trypanosomes were sampled. Columns 3 and 4 give the number of bloodstream clones generated from fly 296, which were hybrids or parentals.

Recombination on chromosome I. Highly polymorphic mini- and microsatellite markers are useful for not only detecting and analysing hybrids, but also to detect recombination or crossing over between homologues of a particular chromosome. Such information can aid the understanding of recombination during the sexual process and also contribute to the construction of a genetic map. To this end a study of recombination on chromosome I was undertaken in collaboration with J. Sasse from the University of Cambridge.

Three polymorphic microsatellites, D2, E5 and F3 have been identified and localised to chromosome I [Sasse, 1998]. Figure 5.7A illustrates the position of these markers and the minisatellite *MS42* on chromosome I, [J. Sasse, personal communication] obtained from mapping these markers to P1 clones that had been ordered on chromosome I.

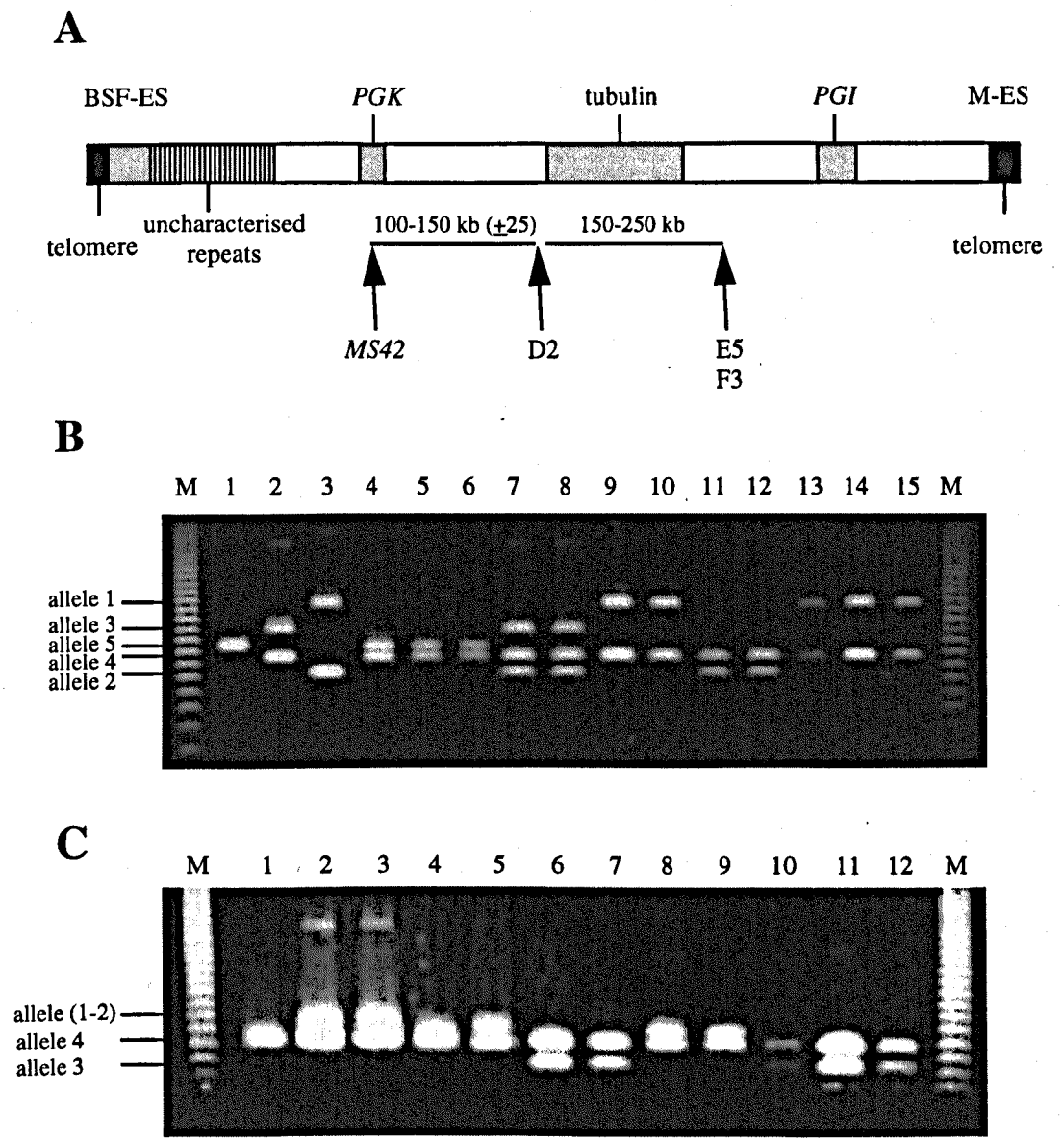


Figure 5.7. Position of mini- and microsatellites on chromosome I and the PCR amplification of markers D2 and E5 from hybrid clones.
(**A**) Diagram of one TREU 927/4 homologue of chromosome I, showing the relative positions of the mini- and microsatellite markers, *MS42*, D2, F3 and E5. The chromosome size is approximately 1115 kb. The diagram is not to scale. BSF-ES, bloodstream from expression site; M-ES metacyclic expression site; *PGK*, phosphoglycerate kinase gene; *PGI* phosphoglucose isomerase gene.
(**B**) PCR amplification of the D2 locus, using primers D2-A 5' and D2-B 5', the sequences of which are presented in Materials and Methods. The products were separated on a 3% Nusieve agarose gel, stained with ethidium bromide and visualised by UV illumination. Lane M, 20 bp ladder (Advanced Biotechnologies); lane 1, STIB 247; lane 2, STIB 386; lane 3, TREU 927/4; lanes 4-6, hybrids from STIB 247 x STIB 386 cross; lanes 7-15, hybrids from STIB 386 x TREU 927/4 cross.
(**C**) PCR amplification of the E5 locus, using primers E5-A, and E5-B, the sequences of which are presented in Materials and Methods. The products were separated on a 3% Nusieve agarose gel, stained with ethidium bromide and visualised by UV illumination. Lane M, 20 bp ladder (Advanced Biotechnologies); lane 1, STIB 386, lanes 2-12, hybrids from STIB 386 x TREU 927/4 cross. Allele 3 from STIB 386 and alleles (1-2) and 3 from TREU 927/4 are indicated.
PCR conditions for both loci were as follows; 95°C for 50 sec., 60°C for 50 sec. and 70°C for 50 sec. for 28 cycles.

Some of the hybrid clones were genotyped for the microsatellites, D2, E5 and F3 by J. Sasse. This analysis was extended for D2 and E5 to incorporate the newly generated hybrids. Since markers F3 and E5 have been located on the same 58 kb fragment of chromosome I, and no crossovers were detected between these markers [Sasse, 1998], microsatellite F3 was not used to analysed the new clones. Figure 5.7B and C gives examples of the data from the PCR assay for the markers, D2 and E5. D2 is a highly informative marker, being able to detect all four alleles of STIB 386 and TREU 927/4. However, as with most markers analysed, STIB 247 is homozygous and so crossovers can only be detected in STIB 386 and TREU 927/4 chromosomes. Marker E5 is less informative as both STIB 386 and STIB 247 are homozygous for the same sized allele, but TREU 927/4 is heterozygous for different sized alleles and so is informative. The allele nomenclature is the same as that followed by J. Sasse, [personal communication]. The complete analysis of all unique hybrid clones for all chromosome I markers is presented in Table 5.9, including the previously obtained data for *MS42*. The shaded rows indicate those clones where a crossover was found. Hybrids were considered vegetative derivatives of one trypanosome if they were bloodstream clones, sampled from the same fly, on the same day and were identical for the following markers; *MS42*, *CRAM*, *292*, JS2, D2 and E5; these clones were removed from the analysis. Hybrid clones from STIB 247 x STIB 386 or STIB 247 x TREU 927/4 crosses which were triploid or trisomic for chromosome I were excluded from the analysis as they were uninformative.

STIB 247 is homozygous for all markers except JS2 and so only TREU 927/4 and STIB 386 chromosomes could be analysed for crossover events. From this analysis eleven unique crossover events were identified and are recorded in Table 5.9 (shaded). The recombinant chromosomes identified are shown schematically in Figure 5.8. Seven recombination events were identified for TREU 927/4 from a total of 30 chromosomes examined, recombination being detected between *MS42* and D2 and between D2 and E5, leading to all four possible recombinant chromosomes being detected. Four recombination events from 26 chromosomes examined were detected for STIB 386, which was only informative for crossovers between *MS42* and D2, but only one of the two possible recombinant chromosomes was detected.

Recombination frequencies for MS42 and D2. Recombination frequencies were calculated from the number of recombinant chromosomes expressed as a proportion of the total number of chromosomes. Four TREU 927/4 chromosomes with a recombination event between the markers *MS42* and D2, were detected out of a total of thirty, resulting in a recombination frequency of 0.133, *i.e.* 13.3 map units or centiMorgans (cM). Four STIB 386 chromosomes with a crossover between the markers *MS42* and D2, were detected out of 26 chromosomes examined, resulting in a recombination frequency of 0.154 (15.4 cM). A comparison between these two figures illustrates there is very little difference between

Recombination	Cross	Chromosome I markers			
		MS42	D2-349	F3-A	E5-349
TREU 927/4	parental	3-4	1-2	1-2	(1-2)-3
STIB 247	parental	5-5	5-5	0-0	4-4
STIB 386	parental	1-2	3-4	0-0	4-4
F532/63blscl 2	247 x 927/4	3-5	1-5	0-2	3-4
F532/63blscl 3		4-5	2-5	0-1	(1-2)-4
F532/63blscl 5		4-5	2-5	0-2	3-4
F532/63blscl 7		4-5	2-5	0-1	(1-2)-4
F532/63blscl 8		4-5	2-5	0-1	(1-2)-4
F532/72 mcl 1		3-5	1-5	0-2	3-4
F532/72mcl 2		4-5	2-5	0-1	(1-2)-4
F532/72mcl 3		4-5	2-5	0-1	(1-2)-4
F532/72mcl 4		3-5	1-5	0-2	3-4
F532/72mcl 5		4-5	2-5	0-1	(1-2)-4
F532/72mcl 6		3-5	1-5	0-2	3-4
F532/72mcl 7		4-5	2-5	0-1	(1-2)-4
F532/72mcl 9		4-5	2-5	0-1	(1-2)-4
F532/72mcl 10		4-5	1-5	0-2	3-4
F124/28blsclA1		3-5	1-5	0-2	3-4
F124/28blsclB3		4-5	2-5	0-2	3-4
F124/28blsclC5		4-5	2-5	0-1	(1-2)-4
F974/70 mcl 4		4-5	2-5	0-1	(1-2)-4
F9/34 mcl 1	247 x 386	1-5	4-5	na	na
F9/45 mcl 2		2-5	3-5	na	na
F9/45 mcl 7		1-5	4-5	na	na
F9/45 mcl 9		1-5	4-5	na	na
F9/45 mcl 10		1-5	3-5	na	na
F9/45 mcl 11		1-5	4-5	na	na
B80C2		2-5	3-5	na	na
F492/50 mcl 12		1-5	4-5	na	na
F492/50 mcl 13		1-5	4-5	na	na
F492/50 bscl 1		1-5	4-5	na	na
F492/50 bscl 2		1-5	4-5	na	na
F492/50 bscl 3		1-5	4-5	na	na
F492/50 bscl 8		1-5	4-5	na	na
F492/50 bscl 9		2-5	3-5	na	na
F492/50 bscl 11		1-5	4-5	na	na
F492/50 bscl 12		1-5	3-5	na	na
F492/50 bscl 14		1-5	4-5	na	na
F296/39 blscl 2	927/4 x 386	1-3	1-4	0-2	3-4
F296/39 blscl 6		1-4	2-3	nd	(1-2)-4
F296/39 blscl 7		1-3	2-4	0-1	(1-2)-4
F296/39 blscl 17/1		1-2-4	2-3-4	nd	(1-2)-4
F296/39 blscl 22/1		1-3	1-4	nd	3-4
F296/42 blscl 26		1-3	1-4	nd	3-4
F296/44 blscl 1		1-3	2-4	0-1	(1-2)-4
F296/44 blscl 7		1-2-3	1-3-4	0-1	(1-2)-4
F296/44 blscl 3		1-3	1-4	0-2	3-4
F296/44 blscl 4		1-4	2-3	0-1	(1-2)-4
F296/44blscl 12		1-2-4	2-3-4	0-1	(1-2)-4
F296/56 mcl 6		1-3	2-4	0-1	(1-2)-4

Table 5.9. Inheritance of mini- and microsatellites from chromosome I. Column 1 gives the unique hybrid clones (as identified in Table 5.1) used in this analysis. Column 2 indicates from which cross the clones were derived. Columns 3-6 gives the results of the analysis for the chromosome I markers, *MS42*, F3-A, D2-349 and E5-349. Analysis of the parental clones are given in rows 1-3 for reference.

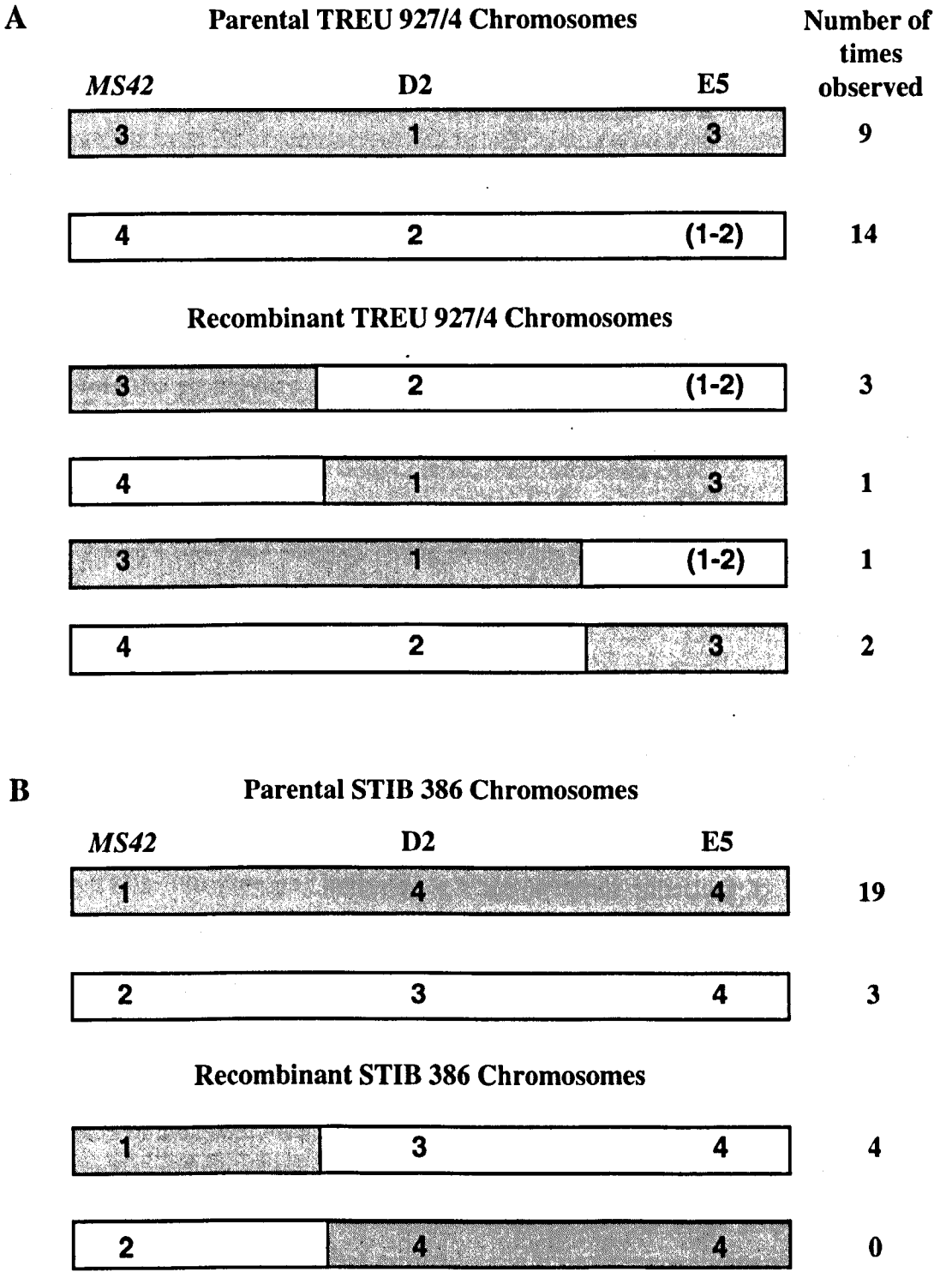


Figure 5.8. Diagram of TREU 924/4 and STIB 386 parental and recombinant chromosomes.
(**A**) The parental and recombinant chromosomes for TREU 927/4, with the alleles for *MS42*, D2 and E5 are drawn. The number of times these chromosomes have been observed is indicated. n=30.
(**B**) The parental and recombinant chromosomes for STIB 386, with the alleles for *MS42*, D2 and E5 are drawn. The number of times these chromosomes have been observed is indicated. n=26.

the frequency of recombination in the two stocks (the average recombination frequency is 0.143).

Recombination frequencies for D2 and E5. Only TREU 927/4 chromosomes were informative for crossovers occurring between D2 and E5. Three recombination events were identified from 30 chromosomes examined, giving a recombination frequency of 0.1 or 10 cM.

Recombination frequencies for MS42 and E5. The genetic map distance between *MS42* and E5 can be calculated by adding the map distances for *MS42*/D2 and D2/E5 together, which results in a distance of 24.3 cM. The probability of a crossover occurring between *MS42* and D2 is 0.143, and the probability of a crossover occurring between D2 and E5 is 0.1, therefore the probability of a double crossover occurring is 0.143 x 0.1 = 0.0143 (or 1.4%), assuming no interference. It is not surprising, therefore, given the small number of chromosomes examined, that no double crossovers were detected.

Relationship between the genetic and physical map distances. At present most, but not all, of chromosome I has been mapped using P1 clones, [Melville, personal communication] and the physical distances between the three markers, *MS42*, D2 and E5 are given as rough estimates in Figure 5.7A. The average physical distance per centiMorgan for *MS42* and D2 is 8.74 kb/cM, and for D2 and E5 is 20 kb/cM, with a minimum and maximum figure ranging between 4.9 kb/cM and 25 kb/cM. The difference in the estimates of a centiMorgan between *MS42* and D2 and between D2 and E5 could possibly reflect a recombination 'hot spot' around *MS42* which is explored in more detail in Chapter 8.

From the data presented in Figure 5.8, there appears to be a significant bias in terms of the particular STIB 386 parental homologue chromosomes inherited by the F1 progeny, compared to the 1:1 ratio predicted for a Mendelian system (χ^2 =11.636, d.f.= 1, P<0.001). The deviation from non-Mendelian ratios is even greater for the inheritance of the STIB 386 *MS42* alleles (see Figure 5.8). Stock TREU 927/4 shows no significant difference from a 1:1 ratio (χ^2 = 1.087, d.f.=1, P= 0.3). There are two possible reasons for this; either there is a bias in the generation of meiotic products (*i.e.* non-Mendelian ratios) or selection for/against particular progeny carrying one of the homologues. However, the markers on other chromosomes appeared to be inherited from STIB 386 in the predicted Mendelian proportions. This is suggestive that normal meiotic products are generated and indicates that selection may account for the chromosome I bias.

Conclusions

The approach of generating new clones from uncloned populations derived from the previous crosses dramatically increased the number of hybrids available for study from 38 to 81 (53 unique hybrids), creating the largest collection of hybrid clones from three crosses. Analysis of the collection of clones derived from these crosses, using the highly

informative mini- and microsatellite markers, has allowed a number of the issues regarding the sexual process in *T. brucei* to be examined. Most important is the examination of allele segregation. The majority of minisatellite markers appear to be inherited in a manner which is consistent with a Mendelian system, although significant deviation from Mendelian ratios has been observed for the *MS42* marker in crosses involving STIB 386. However it must be remembered that the majority of F1 hybrids were cloned from bloodstream trypanosomes and so have undergone a pre-amplification step in mice, which could distort allele segregation ratios. The direct analysis of single trypanosomes in the salivary glands of tsetse flies would avoid this pre-amplification step, although it is possible that the F1 hybrids have multiplied within the salivary glands and only by analysis of gametes (if they exist) can the true segregation ratios be determined.

In this study attempts were made to generate a new genetic cross between STIB 386 and TREU 927/4 which could have been analysed by the single trypanosome PCR method outlined on Chapter 4. Using this technique it should have been possible to genotype metacyclic and epimastigote cells, allowing the extent of cross-fertilization to be examined and perhaps detecting genetic exchange intermediates, (for example, multinucleate cells or gametes) in the salivary glands of mixed-infected tsetse flies, which would reveal at what life cycle stage mating takes place. However no mixed infected tsetse flies were available for analysis.

The analysis of existing hybrids and new hybrids from existing crosses, however, did generate some interesting findings. For example, the incidence of triploidy in F1 hybrids is low (2-6%) compared to 66% detected by Gibson *et al.*, (1994) and the triploids identified here all involved extra copies of STIB 386 chromosomes. It is interesting to note that STIB 386 is a *T. b. gambiense* type 2 (*i.e.* a West African *T. b. rhodesiense*) strain and that all triploids identified by Gibson *et al.*, (1994) also involved extra chromosomes donated by a *T. b. rhodesiense* stock. Such triploids may have been generated by the fusion of a diploid *T. b. rhodesiense* cell (perhaps due to a failure of *T. b. rhodesiense* trypanosomes to undergo meiosis), with a haploid *T. b. brucei* trypanosome. STIB 386 has also been shown to be involved in the generation of trisomics, the new class of hybrid clones identified, which may have been generated by chromosomal non-disjunction, suggesting that this *T. b. rhodesiense* stock is unable to undergo meiosis faithfully.

The availability of hypervariable markers, which have been localised to one chromosome (chromosome I), allowed an investigation into the level of recombination between homologous chromosomes. Although from a small number of cross over events, an estimate of the physical distance per centiMorgan was obtained (4.9 - 25 kb/cM) and is in the same order of magnitude as that for *P. falciparum* (15-30kb/cM)[Walker-Jonah, 1992]. However there are still a number of aspects of genetic exchange in *T. brucei* which remain to be determined. It is still not clear at which life cycle stage genetic exchange takes place.

Do hybrid epimastigotes exist in the salivary glands? Why do hybrids appear late in salivary gland infections? What is the mechanism of genetic exchange? The different models of genetic exchange and how they can be investigated are discussed in detail in Chapter 9.

Chapter 6

A high level of mixed *T. brucei* infections in tsetse flies

Introduction

Although genetic exchange in *T. brucei* has been shown to occur in the laboratory [Jenni *et al.*, 1986], the importance of this sexual process in natural populations of trypanosomes remains controversial. Three models of the population structure of trypanosomes have been proposed, firstly that trypanosomes undergo frequent, if not random mating [Tait, 1980; Tait *et al.*, 1990; Tait *et al.*, 1993], secondly that trypanosomes have a clonal population structure [Tibayrenc *et al.*, 1990; Tibayrenc and Ayala, 1991; Mathieu-Daude *et al.*, 1995] and thirdly that mating occurs at high frequency but is masked by the expansion of a few genotypes in a short time frame - an epidemic population structure [Hide *et al.*, 1994]. An understanding of the extent and importance of genetic exchange in *T. brucei* populations has practical implications, as sexual reproduction would facilitate the spread and inheritance of traits of medical and economic importance as well as generating a population with a high degree of genetic diversity that would make defining common strains causing particular disease patterns difficult.

A difficulty with much of the data on which these models have been based is that they rely on relatively small sample sizes which are diverse in time, host species sampled and geographical location and therefore assume that little or no population sub-structuring occurs. Furthermore, limited attention has been paid to intra-isolate heterogeneity (mixed infections) which is a critical issue if genetic exchange is occurring at any significant level. However, some studies have reported isolates containing more than one genotype of parasite [Letch, 1984; Godfrey *et al.*, 1990; Stevens *et al.*, 1994; Stevens and Tibayrenc, 1995] from both tsetse flies and mammalian hosts. The restraints on examining larger population samples have been the need to expand isolates in laboratory rodents to prepare pure parasites in sufficient quantity for analysis by isoenzyme and RAPD markers while the detection of intra-isolate heterogeneity requires parasite cloning due to the difficulties in applying genetic interpretations to some of the markers used. On this basis there is a need to address these issues using markers that allow parasites to be genotyped without recourse to cloning and growth in the laboratory, facilitating the analysis of large sample sizes and avoiding any selection during vegetative growth. Additionally, genetically

interpretable markers with high levels of polymorphism would be particularly advantageous in examining whether population sub-structuring occurs, whether predominant genotypes are stable in time and place and whether individual isolates are heterogeneous in terms of parasite genotype.

Hypervariable minisatellites or variable number tandem repeat (VNTR) loci, are very useful genetic markers, as they often have a high degree of heterozygosity and many different allelic states based on variation in the number of repeat units in the tandem array. Such minisatellites have been used extensively in human genetics for individual identification, paternity testing [Wong *et al.*, 1987] and linkage mapping [Nakamura *et al.*, 1987] but have only been used to a limited extent in the analysis of parasite genomes. Three minisatellite loci from *T. brucei* have been described in Chapter 3 all of which vary in a strain-specific manner allowing genotypes and allele frequencies to be determined, as well as providing a means of identifying and tracking individual strains. The use of locus-specific primers to PCR amplify minisatellite markers should enable the genotyping of trypanosomes even when contaminated with large quantities of host DNA, as well as allowing the analysis of small quantities of DNA as demonstrated by the detection and genotyping of single trypanosomes [MacLeod *et al.*, 1997; Chapter 4] by PCR amplification of single copy genes. Because of their high level of polymorphism, minisatellite markers are particularly useful in determining variation between geographically distinct populations and detecting heterogeneity within a sample.

This chapter describes the use of the three *T. brucei* minisatellite markers in the analysis of parasite isolate genotype heterogeneity. The degree of genetic diversity in a collection of trypanosomes isolated from wild tsetse flies has been determined in order to examine the extent to which these flies harboured mixed *T. brucei* infections within their salivary glands. The objective was to ascertain to what extent this pre-condition for sexual recombination in *T. brucei* occurs in the field.

Results

T. brucei populations under study. In this chapter two *T. brucei* populations isolated between 1969 and 1970 were examined, Kiboko in Kenya and Lugala in Uganda. Details of the stocks used in this study are presented in Chapter 2, Table 2.1. The stocks were isolated by Goebloed from infected salivary glands of tsetse flies of the subgenus *Glossina* and passaged 3-10 times in mice before transfer to the Glasgow laboratory [Goebloed *et al.*, 1973]. The uncloned stabilates were used to infect irradiated (600 rads) MF1 adult mice and infected blood harvested at peak parasitaemia for subsequent PCR analysis. Single trypanosomes from stabilates of these first peak parasitaemias were used to infect mice for the generation of parasite clones.

Level of polymorphism in Kiboko and Lugala clones. In order to estimate the level of polymorphism shown by the three minisatellites markers, *MS42*, *CRAM* and *292*,

in the two *T. brucei* populations, lysates of single cloned lines from each of 20 isolates from Kiboko were PCR amplified using locus-specific primers. The minisatellite genotypes obtained (see Chapter 8, Table 8.1D) are presented as an estimate of the number of repeats in each allele. In brief, twelve distinct alleles for each of the *CRAM* and *292* loci (estimated from allele size measurements) and sixteen alleles of the *MS42* locus as identified by allele size and MVR maps (described in full in Chapter 9) were detected. The genotypes of each isolate gave heterozygosities of 100% for *MS42* and 95% for *CRAM* and *292*. These results show that the minisatellite loci are highly polymorphic as had been demonstrated previously for the *MS42* locus [Barrett *et al.*, 1997].

Analysis of uncloned* T. brucei *samples from tsetse salivary glands. Because of the hypervariability of the repeat regions within the genes for *MS42*, *CRAM* and *292*, and the ease with which stocks can be genotyped, these markers could be extremely useful in the identification of different genotypes. The data summarised above have been obtained using cloned stocks of the parasite and so do not allow the question of whether isolates contain more than one genotype to be addressed. To resolve this issue the markers were used to determine the number of different genotypes of *T. brucei* present in the collection of uncloned isolates from Kiboko and Lugala derived from the salivary glands of tsetse flies. By using all three minisatellite loci the sensitivity of detecting different genotypes was increased, for example the *CRAM* marker detected a minimum of 36 different genotypes in the 28 uncloned isolates, whereas by using all three markers a minimum of 42 genotypes were identified. Parasite DNA or crude lysates from infected mouse blood were analysed by PCR using all three markers. The results for the Kiboko samples are presented in Figure 6.1, and for all samples in Table 6.1. From the genetic analysis and the genotypes exhibited by cloned trypanosome stocks, one genotype will either contain two different sized alleles (heterozygous) or a single allele (homozygous). A number of the isolates show two alleles at each of the three minisatellite loci, for example tracks 1, 2 and 3 in Figure 6.1A-C, and could represent a single genotype heterozygous at all three loci. Although, formally, such genotypes could represent a mixture of two homozygotes, the high levels of heterozygosity at these loci argue against this. A second group of isolates show more than two alleles (*e.g.* lanes 4, 7, 8, Figure 6.1A-C) up to a maximum of six alleles per locus (lane 17, Figure 6.1C) indicating that they contain a mixture of different trypanosome genotypes. In the isolates from Lugala and Kiboko a significant proportion of the tsetse flies harboured more than one genotype based on the detection of more than two PCR products per locus. Assuming that all the trypanosomes are heterozygous for alleles at each locus and that multiple alleles at more than one locus do not assort independently, the minimum number of genotypes present in each isolate can be determined (Table 6.1). In total, eight out of seventeen (47%) of the Kiboko samples contained more than one genotype and four out of eleven (36%) of the Lugala samples.

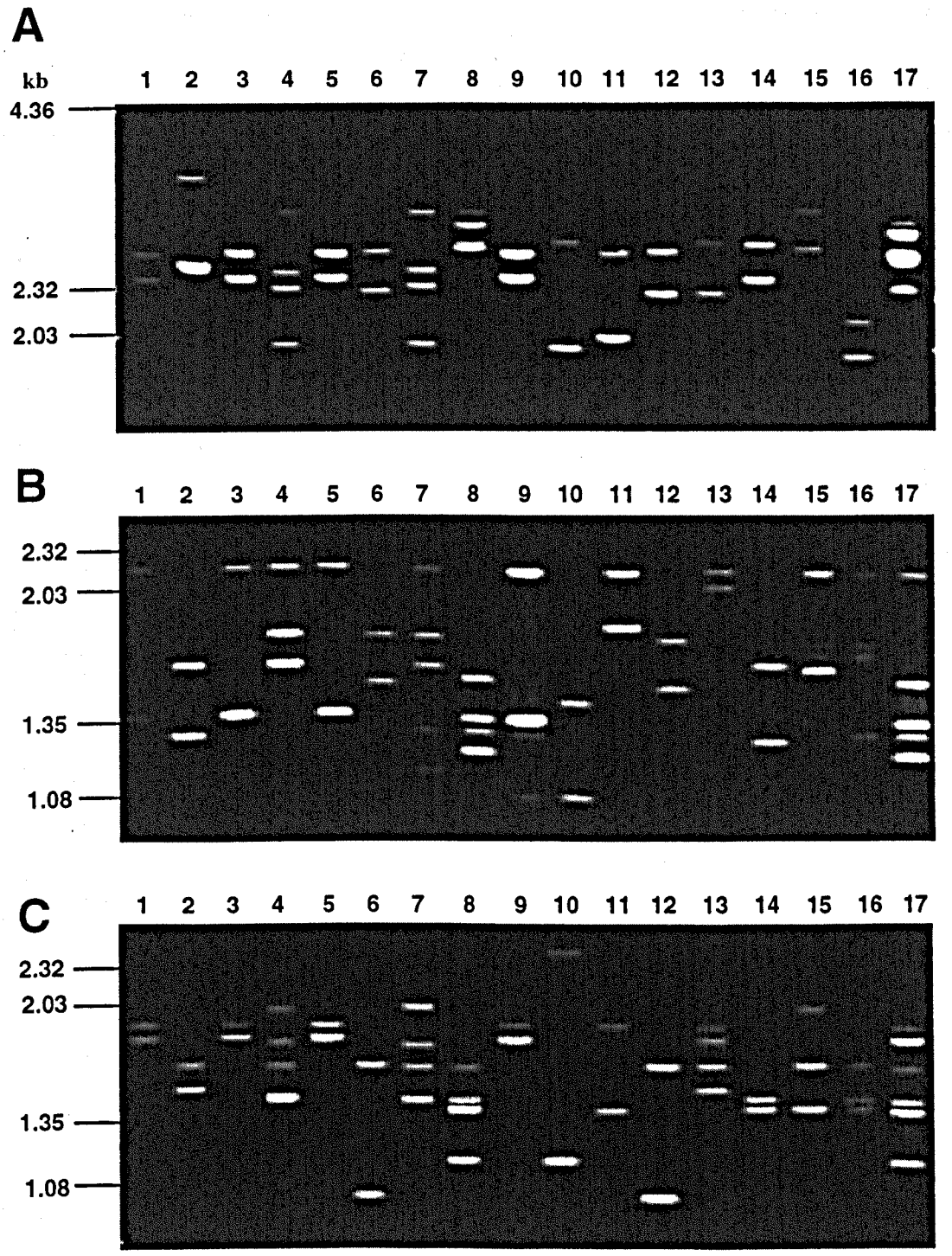


Figure 6.1. Analysis of uncloned tsetse derived trypanosomes using three minisatellite markers. One µl of 1/100 dilution of a crude lysate of infected blood was used in each PCR reaction. All reactions were performed under the same PCR conditions as described in Materials and Methods. PCR products were separated on a 1% Seakem agarose gel and visualised by ethidium bromide staining. Lanes 1-17: uncloned stocks K936; K926; K869; K981; K994; K1027; K984; K854; K975; K925; K1009; K851; K852; K978; K997; K258; K927, using primers specific for the (A) *CRAM* locus; (B) *292* locus; (C) *MS42* locus. The sizes given were determined from lambda *Hin*dIII markers (not shown).

Origin	Sample	No. of different sizes PCR products detected for each minisatellite marker			Minimum no. of genotypes present
		MS42	*CRAM*	*292*	
Kiboko	K936	2	2	2	1
Kiboko	K926	2	2	2	1
Kiboko	K869	2	2	2	1
Kiboko	K981	4	4	3	2
Kiboko	K994	2[a]	2	2	1
Kiboko	K1027	2	2	2	1
Kiboko	K984	4	4	5	3
Kiboko	K854	4	4	4	2
Kiboko	K975	2	2	4	2
Kiboko	K925	2	2	2	1
Kiboko	K1009	2	2	2	1
Kiboko	K851	2	2	2	1
Kiboko	K852	4	2	2	2
Kiboko	K978	2	2	2	1
Kiboko	K997	3	2	4	2
Kiboko	K258	3	2	4	2
Kiboko	K927	6	6*	5	3
Lugala	L836	2	4	2	2
Lugala	L791	2	4	2	2
Lugala	L944	2	2	3	2
Lugala	L834	2	2	2	1
Lugala	L929	2	1	2	1
Lugala	L933	2	2	2	1
Lugala	L941	2	2	2	1
Lugala	L832	2	2	2	1
Lugala	L934	2	2	2	1
Lugala	L844	2	2	2	1
Lugala	L845	4	4	4	2

* Fifth and sixth bands visible after high resolution electrophoresis.

[a] K994 on one occasion gave four *MS42* alleles, but this result was not reproducible.

Table 6.1. Minisatellite analysis of uncloned *T. brucei* isolates. Numbers of different minisatellite alleles detected in uncloned Kiboko and Lugala isolates. The number of different sized PCR products (corresponding to the minimum number of different alleles) detected for three minisatellite markers is indicated in Columns 3-5. The minimum number of genotypes present in the stocks are shown in column 6. Those samples which generated >2 PCR products were considered to harbour mixed trypanosome genotypes.

Most mixed samples probably contained a minimum of two different genotypes, as four alleles at one or more loci were detected. However, two samples must have contained at least three different genotypes, as sample K984 contained five different *292* alleles and sample K927 contained six different *MS42* alleles. It is clear that by using a combination of different highly variable minisatellites rather than relying on one marker, the sensitivity of the analysis was greatly increased. The number of mixed infections is probably an underestimate as the PCR reactions could favour one genotype over another, due to sequence divergence at the primer sites or, if an unequal ratio of two genotypes was present, the most abundant genotype would be preferentially amplified. Also some trypanosome genotypes could have been lost during the amplification in mice and repeated passaging. Analysis of the alleles at the three loci in the Kiboko population of isolates (uncloned and cloned) shows that there are at least nineteen distinct alleles of *MS42*, fifteen distinct alleles of *CRAM* and eighteen alleles of *292 i.e.* a very high level of allelic variation. Analysis of all the uncloned Kiboko isolates using the three loci shows that out of seventeen isolates, sixteen are genotypically distinct (tracks 6 and 12 are identical, Figure 6.1) demonstrating a high level of polymorphism in the trypanosomes within this population.

Cloned parasites from mixed infections. The estimates of the number of genotypes in the isolates showing more than two alleles at two or more loci is a minimum. As the loci are located on different chromosomes and are therefore unlinked [Lee *et al.*, 1990; Lee *et al.*, 1994; Barrett *et al.*, 1997; Chapter 3], the alleles at each locus would potentially assort independently and so many more distinct genotypes could be present in such isolates. As little is known, (in trypanosomes) as to the mechanism by which the repeat length variants are generated, it is possible that they have arisen by mutation during mitosis. To address both these issues, a series of cloned trypanosomes lines were established from two different samples, K927 and L845, both of which harbour multiple trypanosome genotypes. Each cloned line was then analysed by PCR for all three loci. The results from the thirteen K927 clones, analysed for one marker, *292*, are shown in Figure 6.2. Four distinct banding patterns, each containing two bands, clearly demonstrate that more than one genotype of *T. brucei* was present in the original tsetse salivary glands. These genotypes account for four of the alleles observed in the uncloned isolate (lane 2, Figure 6.2) but the fifth allele is not found in any of the cloned lines, implying that further genotypes are present but have not been cloned. The fact that all the cloned lines exhibit only two alleles yet have undergone multiple rounds of mitosis during cloning (from 1 to ~10^8 trypanosomes) suggests that length variants at this locus are not generated at an appreciable frequency during mitosis. The clones obtained were also characterised with respect to the two other minisatellite markers (*CRAM* and *MS42*) and the results are summarised in Table 6.2. There are at least seven different genotypes of trypanosomes present in the tsetse fly's salivary glands represented by isolate K927. Marker analysis of

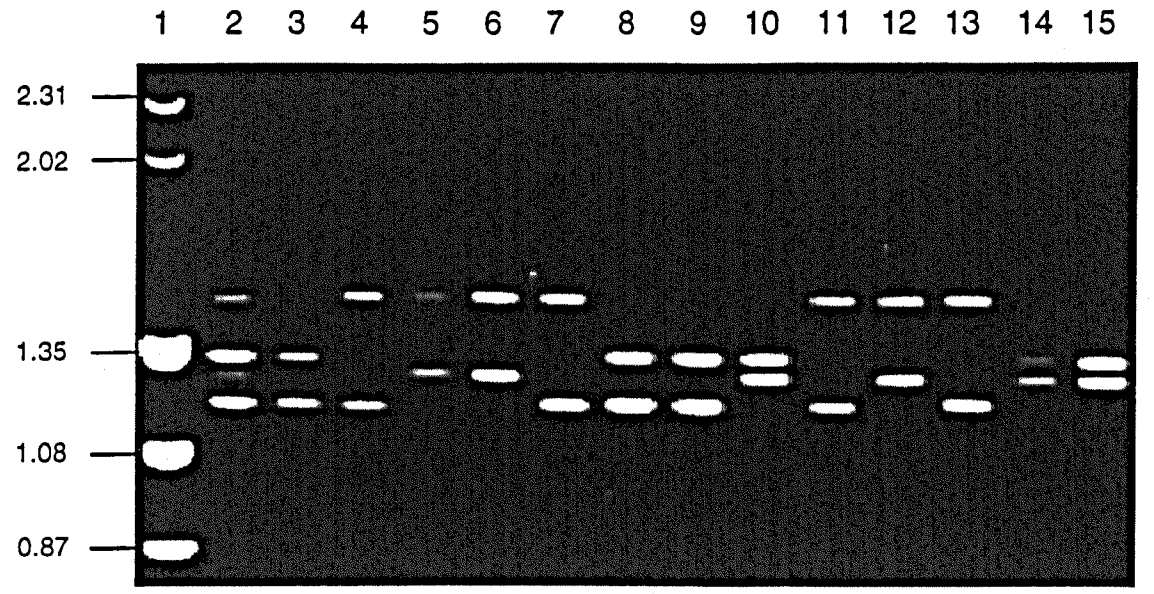


Figure 6.2. PCR amplification of the *292* locus from K927 derived clones. PCR reactions were performed on crude lysates as described in Materials and Methods. PCR products were separated on a 1% Seakem agarose gel and visualised by ethidium bromide staining. Lane 1, Lambda *Hind* III and ϕx *Hae* III markers (Advanced Biotechnologies); lane 2, uncloned tsetse derived sample K927; lanes 3-15, K927 clones, 4B, 4, 5, 9, 10, 12, 13, 14, 18B, 20, 22, 23 and 24.

the eight cloned lines derived from sample 845 (Table 6.2) also indicate that the original sample contained at least two different genotypes of trypanosome. Whether further combinations of alleles are present in this isolate is unknown.

Evidence for meiosis in field sample K927. The five *292* alleles present in the uncloned K927 sample (lane 2, Figure 6.2) have been calculated (by band size estimations) to contain 76, 50, 44, 41 and 38 copies of the 24bp repeat unit. The K927 derived clones (lanes 3-14, Figure 6.2) contain the four smaller alleles in four different combinations out of a possible six combinations. This allele assortment has also been observed for the other markers, *CRAM* and *MS42*, (Table 6.2), where three out of the six possible allele combinations of *CRAM* and four out of six possible *MS42* combinations are found in the clones, excluding those alleles not recovered. Furthermore when the complete multilocus genotypes of each clone are considered seven of the thirteen clones are distinct showing different combinations of alleles at the three loci. It seems likely that the analysis of further clones would identify further allelic combinations and a high level of multiple genotypes within one tsetse fly. These data raise the question of the origins of the trypanosome genotypes identified. These could have arisen by the fly feeding on one mammalian host simultaneously containing at least seven different trypanosome genotypes or these genotypes could represent the products of mating as a result of the

tsetse feeding on a mammal containing two (or more) genotypes. It would be difficult to formally distinguish between these possibilities without analysis of the genotypes present in the midgut of the fly, but the results obtained are highly suggestive of mating having occurred in this tsetse fly.

Samples	Genotypes detected for each minisatellite marker			Multi-locus genotype
	MS42	*CRAM*	*292*	
uncloned K927	35/34/30/26/25/20	70/68/62/61/60/52	76/50/44/41/38	mixed
clone K927/4B	26/25	70/68	44/38	A
clone K927/4	25/20	68/61	50/38	B
clone K927/5	30/20	68/61	50/41	C
clone K927/9	30/20	68/61	50/41	C
clone K927/10	25/20	68/61	50/38	B
clone K927/12	26/25	68/61	44/38	D
clone K927/13	26/25	68/61	44/38	D
clone K927/14	26/25	70/68	44/41	E
clone K927/20	30/20	68/61	50/41	C
clone K927/22	25/20	68/61	50/38	B
clone K927/23	25/20	68/61	44/41	F
clone K927/24	25/20	68/61	44/41	F
clone K927/18B	30/26	62/61	50/38	G
uncloned L845	31/26/21/16	98/73/70/50	75/49/43/12	mixed
clone L845/1	31/16	98/73	49/12	H
clone L845/2	31/16	98/73	49/12	H
clone L845/3	31/16	98/73	49/12	H
clone L845/4	31/16	98/73	49/12	H
clone L845/5	31/16	98/73	49/12	H
clone L845/6	31/16	98/73	49/12	H
clone L845/7	31/16	98/73	49/12	H
clone L845/8	26/21	70/50	75/43	J

Table 6.2. Minisatellite genotypes detected for all clones generated from uncloned samples K927 and L845. Column 1 indicates the isolate number of cloned and uncloned samples. Columns 2, 3 and 4 indicate the *MS42*, *CRAM* and *292* minisatellite genotypes presented as an estimate of the number of repeats in each allele, respectively. Column 5 shows the multilocus genotype, *i.e.* the combined results from the three minisatellites.

Discussion

There are a large variety of methods available to aid the analysis of trypanosome population genetics. The most extensively used technique, isoenzyme analysis, has been used for the past two decades to study the identification, distribution and relationships between trypanosomatids [Gibson *et al.*, 1980; Tait, 1980; Tait *et al.*, 1984; Tait *et al.*, 1985; Godfrey *et al.*, 1990]. In recent years the PCR-based DNA typing system, RAPDs, has increased in popularity and has been used to complement isoenzyme analysis [Stevens and Tibayrenc, 1995]. The RAPD technique is quick and easy to perform, detects many

variable fragments of DNA, yet requires no sequence information, and can analyse very small amounts of trypanosome DNA. However, RAPD data cannot readily be interpreted on a genetic basis and are unable to identify samples of mixed genotypes or contaminated samples.

In order to detect mixed trypanosome samples highly variable single locus markers were required. This chapter has described the use of three highly polymorphic minisatellites loci, present in the coding region of the *MS42*, *CRAM* and *292* genes. By designing specific primers to the DNA flanking the tandem repeats of these genes, it was possible to amplify, by PCR, the hypervariable regions from a range of *T. brucei* samples (even from samples contaminated with host DNA). These single locus markers proved to be extremely informative due to the large number of allelic states at both loci, allowing individual *T. brucei* stocks to be identified and tracked through populations (see Chapter 8). The minisatellite markers were also able to detect those samples which contained more than one genotype of trypanosome.

Using these minisatellites as markers evidence has been obtained for a range of genetically distinct trypanosomes coexisting in salivary glands of a significant proportion of tsetse flies, 47% of Kiboko and 36% of Lugala samples. This is in contrast to the isoenzyme data of Godfrey *et al.*, (1990) where only 9.6% of tsetse samples were shown to be of mixed genotype. This could reflect the fact that isoenzyme analysis is less sensitive for detecting mixtures due to the lower levels of variation compared to minisatellites. Further analysis, (Godfrey, 1990) of a large collection of samples from mammalian hosts from across Africa also detected infections of more than one genotype, in both man (3%) and a range of different animals (2%) as well as two samples containing three genotypes. Stevens *et al.*, (1994) examined the isoenzyme patterns of trypanosomes from the midguts of three tsetse flies and found two flies containing five and nine different *T. brucei* genotypes respectively. Further analysis of cloned lines from the same samples, by the more informative method of RAPD analysis, revealed that even more genotypes were present and that all three tsetse flies harboured mixed infections [Stevens and Tibayrenc, 1995], thus suggesting that previous studies based on isoenzyme analysis alone may under-represent the true degree of trypanosome genetic diversity and number of mixed tsetse infections present in natural populations.

It is possible that all estimates of mixed infections recorded to date, including those presented here, may under-represent the levels of genetic diversity and the frequency of mixed infections, due to the influence of sample bias. All samples analysed, except those from the midgut of tsetse, have been grown in rodents and passaged before analysis. It is likely that the repeated passaging of trypanosome populations in laboratory rodents may serve to filter out less virulent trypanosome types, especially as different genotypes may have different rates of growth in rodents. However the only clear data which suggests that the timing and number of passages in mice or culture can select for different genotypes in

a mixed infection have not been performed on *T. brucei* but on the related species *T. cruzi* [Deane *et al.*, 1984]. In this present study, attempts were made to minimise the loss of less virulent strains by examining trypanosome isolates which had been passaged a limited number of times (about 10). Another factor which may influence the reliability of sampling is selection due to host infectivity, *i.e.* the failure of some trypanosome isolates to grow in laboratory rodents. Host selection has been reported previously, for example approximately 63% of trypanosomes isolated from the salivary glands of tsetse flies in Kiboko were unable to infect rodents and so were lost, similarly, from 23 salivary gland isolates from Lugala only 13 were infective to mice and 3 were lost in passaging [Goebloed *et al.*, 1973]. While the problems associated with selection operating on any sampling procedure involving growth or culture is well recognised, it is assumed that such selection does not bias the analysis of markers which have no obvious phenotype on which selection would operate. Given the nature of the markers used to date, it has been impossible to test this assumption, however the markers described here could readily be used to address this question.

Lysates from infected material, *e.g.* the salivary glands of tsetse flies or mammalian blood, could be analysed directly by PCR amplification of the minisatellite loci. Such an approach could analyse approximately 1000 trypanosomes in each PCR reaction. This would determine far more accurately the number of different *T. brucei* strains present in both tsetse flies and the mammalian host. By comparing these results to those obtained from populations grown in laboratory rodents, it should be possible to ascertain to what extent sample bias due to growth rates in mice and selection due to host infectivity affect the detection of genotype diversity.

It is estimated that 1-0.1% of tsetse flies are infected with *T. brucei*, which would lead to the prediction that 0.01-0.0001% of tsetse flies have mixed infections. Clearly this is not the case for the Kiboko and Lugala isolates, indicating that the probability of acquiring a second infection is not independent of acquiring the first, perhaps by feeding from a mixed infected mammal. Another explanation for the high incidence of mixed infections could be that a proportion of the wild tsetse flies may be more susceptible to *T. brucei* infections than the others, which has been linked to the presence of rickettsia-like organisms (RLOs) in the tsetse midgut [Welburn and Maudlin, 1997].

Tsetse salivary glands are the probable site at which genetic exchange takes place between *T. brucei* stocks and a prerequisite for genetic exchange to occur is that there are at least two different strains of *T. brucei* present in the salivary glands at the same time [Jenni *et al.*, 1986; Turner *et al.*, 1990; Tait *et al.*, 1996]. Results from this chapter have demonstrated that a significant proportion of tsetse flies fulfil this prerequisite in that they harbour mixed *T. brucei* infections in their salivary glands. This suggests that genetic exchange could be occurring in the field. In addition, the detection of seven distinct yet highly related genotypes in one fly (K927) and the allele assortment which was

demonstrated for three polymorphic loci, cannot easily be explained without genetic exchange being involved. The most direct way of testing this would have been to analyse the genotypes of trypanosomes present in this fly's midgut to determine what genotypes had infected the fly and from this analysis deduce whether the genotypes in the salivary gland were recombinant. Unfortunately midgut samples were not collected. An alternative, but less direct approach would be to analyse the frequencies and nature of the different genotypes in the population of fly isolates to estimate the probability that this fly had ingested the seven genotypes detected. Similar numbers of genotypes have been isolated from fly midguts [Stevens *et al.*, 1994] presumably reflecting genotypes ingested when the flies feed on infected mammals and so the observation of multiple genotypes in the salivary glands does not *per se* indicate that these are generated by genetic exchange.
The high levels of mixed infections detected in this study and the recombinant genotypes detected in one fly, together provide evidence that genetic exchange does occur. However the data presented in this chapter do not address how frequently genetic exchange occurs and to what extent it is involved in generating diversity. It is likely that population diversity has been underestimated due to sampling selection and the use of markers with low heterozygosity. This can now be assessed by analysing trypanosomes, directly from their source, without the additional step of growth in laboratory rodents, using the highly polymorphic and informative minisatellite markers described here.

Chapter 7

Analysis of *T. brucei* populations using three hypervariable minisatellites

Introduction

The main goals of molecular epidemiological studies of trypanosomes are to identify the parasite(s) responsible for particular outbreaks of trypanosomiasis and to determine their source and route of transmission. The extent to which such goals can be achieved is very dependent on the genetic variability detected by the markers used in any particular study. For example, *T. b. gambiense,* the aetiological agent of chronic sleeping sickness, has been identified as a distinct population genetics entity, using a number of biochemical and molecular techniques, in particular isoenzymes [Godfrey and Kilgour, 1976; Gibson *et al.*, 1978; Gibson *et al.*, 1980; Tait *et al.*, 1984; Godfrey *et al.*, 1987]. However, such techniques have been unsuccessful in distinguishing *T. b. rhodesiense*, which causes acute human sleeping sickness, from *T. b. brucei*, which is non-human infective, as no particular pattern of isoenzyme (or kDNA) polymorphism is exclusively associated with *T. b. rhodesiense* has been identified [Gibson *et al.*, 1980; Borst *et al.*, 1981; Gibson and Wellde, 1985; Tait *et al.*, 1985]. This does not necessarily imply that differences do not exist, but demonstrates that the techniques used to date may be inadequate for the identification of possible differences. Most of the studies carried out so far on parasite populations have been based on isoenzyme analysis, with the description of stocks sharing particular enzyme patterns as zymodemes. However isoenzymes are not particularly variable, so that stocks sharing identical isoenzyme profiles should not be considered as identical but only as related samples. Other typing systems, for example, restriction fragment length polymorphisms (RFLPs) [Hide *et al.*, 1994] and random amplified polymorphic DNAs (RAPDs) [Mathieu-Daude *et al.*, 1995; Stevens and Tibayrenc, 1995] have been used in conjunction with isoenzymes, but are either cumbersome to apply in the analysis of populations or difficult to interpret genetically. There is a clear need for better marker systems that may allow the more sensitive detection of differences between

isolates as well as being genetically interpretable. Microsatellites would be an obvious improvement on isoenzymes or RAPDs markers for determining whether linkage disequilibrium occurs in populations, as they are often highly polymorphic, detect genotypes rather than phenotypes, can be readily interpreted genetically and different alleles can be sensitively detected by PCR and accurately measured on polyacrylamide gels. Also standard population genetic tests can easily be applied to microsatellites with a limited number of alleles. Unfortunately microsatellites have not been used to examine trypanosomes, as sequence information is required to develop a panel of markers. Furthermore, the level of variation is lower than for minisatellites and so less valuable for tracking individual strains, while scoring microsatellites is often difficult due to PCR artefacts. Minisatellite markers or VNTRs, which have been used extensively in forensic and legal medicine, may overcome some of these problems as many have large numbers of alleles [Wong *et al.*, 1987]. Although VNTR markers are useful in the analysis of populations, they have not been used extensively for this purpose, as most minisatellites are large and so cannot be easily amplified by PCR, requiring restriction digests of genomic DNA and Southern blotting techniques to be employed followed by hybridization with the minisatellite probe. Therefore microsatellites, although less variable, have become more popular in the analysis of populations as they can be quickly, easily and sensitively detected by PCR.

In the case of *T. brucei* neither micro- or minisatellites have been employed in the analysis of populations as only a few have been described [Barrett *et al.*, 1997; Sasse, 1998]. The three *T. brucei* minisatellites described in Chapter 3 are highly polymorphic and all are small enough to be amplified by PCR and so combine the convenience of PCR markers with the extreme variability of a minisatellite for the analysis of populations.

The aim of this chapter is to address questions relating to the population genetics of *T. brucei*. Specifically, the aims were to test for recombination in the field, to examine the population structure (panmictic, clonal or epidemic as proposed by Tait [Tait, 1980], Tibayrenc [Tibayrenc *et al.*, 1990] and Maynard-Smith [Maynard-Smith *et al.*, 1993], respectively) and to test for population sub-structuring due to geography or host specificities. If there is population sub-structuring then this needs to be taken into consideration when analysing populations, as many previous studies have been based on samples from different geographical areas, times and host species [Tibayrenc *et al.*, 1990]. The approach taken in this study to address these issues was to use the highly variable

minisatellite typing method, which has been described in Chapter 3, to analyse field isolates collected from five different populations.

The details of the five populations chosen for this analysis are summarised in Table 7.1 and contain three populations sampled mainly from humans during sleeping sickness epidemics (Busoga, Nyanza and Luangwa) and two other populations from tsetse flies, one from a sleeping sickness endemic area (Lugala) and the other from a region where no sleeping sickness has been reported (Kiboko). Although the sample size for each population is small and therefore far from ideal, these populations were chosen as they were the largest collection of samples available which could provide an insight into sexual recombination and sub-structuring in *T. brucei* populations. By analysing these populations, comparisons can be made as to the extent of recombination in different populations, to determine if there is an association between clonality/panmixia and specific hosts, *e.g.* cattle, human and tsetse. Population sub-structuring due to geography can be examined by comparing the allele distributions observed between geographically distinct populations. Similarly, population divisions due to host specificities can be examined by comparing human infective stocks with non-human infective isolates, which were collected from the same geographical area (Busoga) at the same time (see Table 7.1) Such analysis could also indicate if cattle are an important reservoir for human infective trypanosomes.

Population name	Specific geographical area	Number of samples	Host	Year
Luangwa, Z	Chilbale and Kasyasya in Luangwa Valley in Zambia	19 clones	15 man 3 d.a. 1 tsetse	1981-83
Kiboko, K	Kiboko in Kenya	17 clones 17 uncloned	all tsetse	1969-70
Lugala, U	Lugala in Busoga (Uganda)	11 uncloned	all tsetse	1969-70
Nyanza, K	Central Nyanza in Kenya	26 cloned	24 man 2 tsetse	1961 (one from 1977)
Busoga, U	Busoga (Uganda) and Busia (Kenya) at the Ugandan/Kenyan border	49 uncloned	24 man 22 cattle 3 tsetse	1988-90 (5 from 1959-82)

Table 7.1 Description of samples. Samples from all hosts in each area were included in the analysis except where stated otherwise; d.a., domestic animal.

A brief description of the five populations under study (Table 7.1) follows and a map of the different regions in Kenya and Uganda is provided (Figure 7.1).

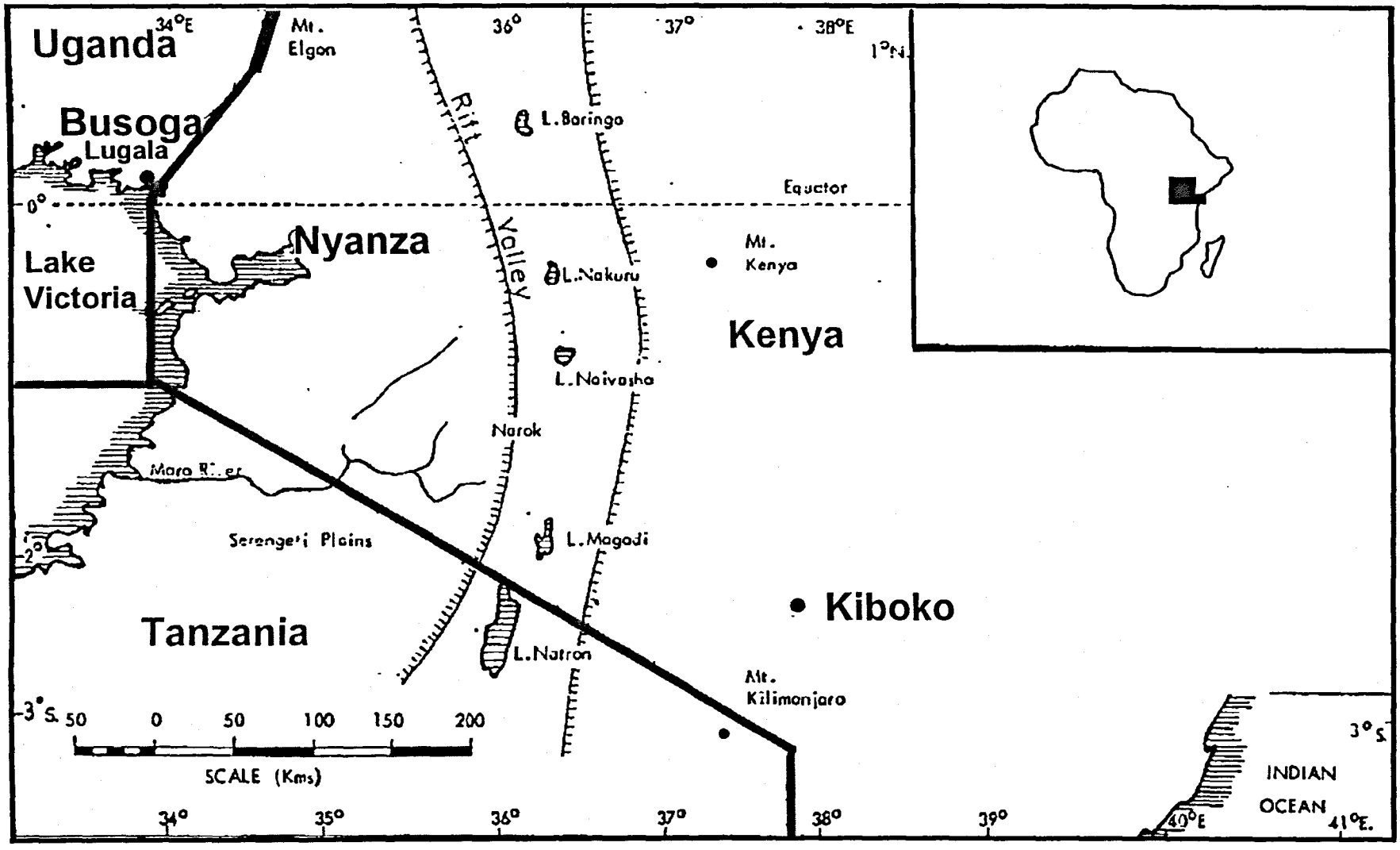


Figure 7.1. Partial map of East Africa. The north east region of Lake Victoria is illustrated, with the sample regions of Lugala and Busoga in Uganda, and Nyanza in Kenya, shown in bold. The Kiboko area in Kenya is approimately 400 Km from Lugala in Uganda.

Kiboko is in the southern region of Kenya, approximately 340Km from Lake Victoria (see Figure 7.1). At the time of sampling (1969-70) the grass savannah area was rich in wild game, with some cattle and few humans. No cases of human trypanosomiasis were reported from this region. The samples collected from this area were isolated from tsetse salivary glands [Goebloed *et al.*, 1973].

Samples from Lugala, in the Busoga region of Uganda, on the north eastern shore of Lake Victoria (Figure 7.1) were also isolated from tsetse salivary glands at the same time (1969-70)[Goebloed *et al.*, 1973]. At the time of sampling this wet savannah area was sparsely populated with few domestic animals but abundant game. Acute human trypanosomiasis was first recorded in the region in the 1940s and was believed to have spread southwards across the Kenyan border to Central Nyanza during the 1950s, where it persisted at a low endemic level, with sporadic outbreaks until an explosive epidemic arose in Alego in Central Nyanza in the 1960s [Baldry, 1972]. The samples from the Nyanza population were collected from humans in 1961 during this epidemic. The recent outbreak of human trypanosomiasis around Lake Victoria in SE Uganda began in 1976, [Abaru, 1985] spreading north to the Tororo district and peaking at over 300 cases in 1990 [Hide *et al.*, 1994]. The trypanosome samples from SE Uganda, taken in 1990, were from both humans and cattle, mainly from Tororo with six samples taken from the same focus from Busia on the Kenyan border.

The Luangwa valley in Zambia was the source of many outbreaks of human trypanosomiasis in the 1960s and with a major outbreak occurring in 1971-74 [Gibson *et al.*, 1980]. The Zambian samples in this study were collected in 1981-3 from two villages in the Luangwa Valley, Chilbale and Kasyasya. This area is home to a large tsetse (*G. morsitans)* population and contains a large wild animal reserve. The incidence of human infections correlates with the movement of animals whereby humans are infected when the wild animals seasonally migrate [Gibson *et al.*, 1980].

Results

In this chapter the hypervariable minisatellite loci (described in Chapter 3) were used to analyse stocks from the five populations (Table 7.1). The alleles at each locus were examined in every stock by PCR amplification using locus specific primers (see Materials and Methods). Because of the large number of alleles at these loci and the fact that the alleles vary in a quasicontinuous manner, unequivocal allele identification is often difficult, this is especially true when the difference between allele bands on an agarose gel

is small. For *CRAM* and *292*, the 'fixed bin' method of allele identification [Devlin and Risch, 1992] was employed, whereby the band (allele) size was determined on the basis of mobilities relative to a reference standard lane (using the Kodak 1D Image Analysis Software) and a window is created around it. Repeated measurements of the same allele were used to calculate the mean and the standard deviation around the mean; two standard deviations corresponded to 2.3% of the estimated allele size. This measurement was then taken as the window size for every band measurement (±2.3% band size) and alleles were considered identical if they are within this window. However there was the possibility that alleles of the same size can differ in repeat structure and so size may not always indicate identity. To circumvent these problems for one minisatellite (*MS42*), alleles were identified by assaying sequence variation in the tandem array as well as size differences by the process of minisatellite variant repeat (MVR) mapping. A full description of this technique, the allele maps derived from it and an analysis of the relationship between alleles is presented in Chapter 8. In this current chapter, this technique is used only to identify unequivocally *MS42* alleles.

Three RFLPs flanking the minisatellite *MS42* have been described and used to type all field samples (see Chapter 8). These data (Table A3, appendix) have been included in the analysis to compare the results obtained for agreement with Hardy-Weinberg (HW) proportions and linkage disequilibrium from the hypervariable minisatellite markers with those from the two allele system defined by RFLP analysis. These markers are within a few kilobases of each other and so are closely linked; therefore the data obtained for these RFLPs can be considered as one locus.

The results of the minisatellite genotype analysis for every stock in each population are presented in Table A2 of the appendix as an estimate of the number of repeat units in each allele for minisatellites *CRAM* and *292*, and as unequivocal defined alleles based on size and MVR analysis for minisatellite *MS42*. Due to the imprecise method of allele identification for *CRAM* and *292*, which is conservative, *i.e.* there is the possibility that different alleles are grouped together, more weight is given to the analysis of the data obtained from *MS42* than for the other two minisatellite markers. Many of the samples in this study were not cloned isolates and could contain more than one genotype of trypanosome, therefore any sample which generated more than two PCR products for any minisatellite marker was considered mixed and so could not be included in the multilocus genotype analysis. However the alleles from mixed samples were included in the allele

frequency analysis. For some Kiboko samples both uncloned and cloned material was available. Where an uncloned sample was identical for all markers to its cloned counterpart, only the cloned sample was included in the analysis but where the genotypes differed between cloned and uncloned material both samples were included.

Variation within populations. Forty-eight different *MS42* alleles were identified using the unequivocal allele identification system described in Chapter 8, compared to 24 alleles which were be identified by band size measurements alone. The nomenclature system for *MS42* alleles is based on the number of repeat units, designated by a number, and the minisatellite variant repeat code, designated by a letter, for example, allele 35c. Band size measurements were used to identify alleles for the other two minisatellite loci, *292* and *CRAM*, which revealed, 27 and 21 alleles, respectively. The frequencies of each allele, for each marker, within each population is given in Tables A4-A6 of the appendix and are presented as summary histograms in Figures 7.2 -7.4. The multilocus genotypes, *i.e.* the combined results for the three minisatellites are also presented in Table A2 of the appendix and in Figure 7.5. It is immediately obvious from the allele/genotype distributions displayed in the figures, that some populations are quite homogeneous, consisting of one or a few genotypes, whereas other populations have a high level of variation, despite the small sample size. For example, for all three minisatellite markers the Kiboko and Lugala populations have a large number of alleles and genotypes, *i.e.* they are diverse, whereas the other three populations are far more homogeneous with fewer alleles and genotypes. The levels of diversity in each population can be measured by examining the number of different alleles and multilocus genotypes as a proportion of the total number of alleles or multilocus genotypes detected in each population (Table 7.2). It is clear from these values that the two populations isolated from tsetse flies (Kiboko and Lugala) have a higher level of diversity than the populations isolated from humans. The isolates from Busoga were sampled at the same time but from humans, cattle and tsetse. The cattle isolates appear to have a level of diversity similar to those of tsetse flies from Kiboko and Lugala or perhaps slightly lower. Examination of the distribution of multilocus genotypes within different hosts (Figure 7.6) from the Busoga population shows that the human isolates consist mainly of the single genotype 21 and are, therefore, far more homogeneous than the cattle samples.

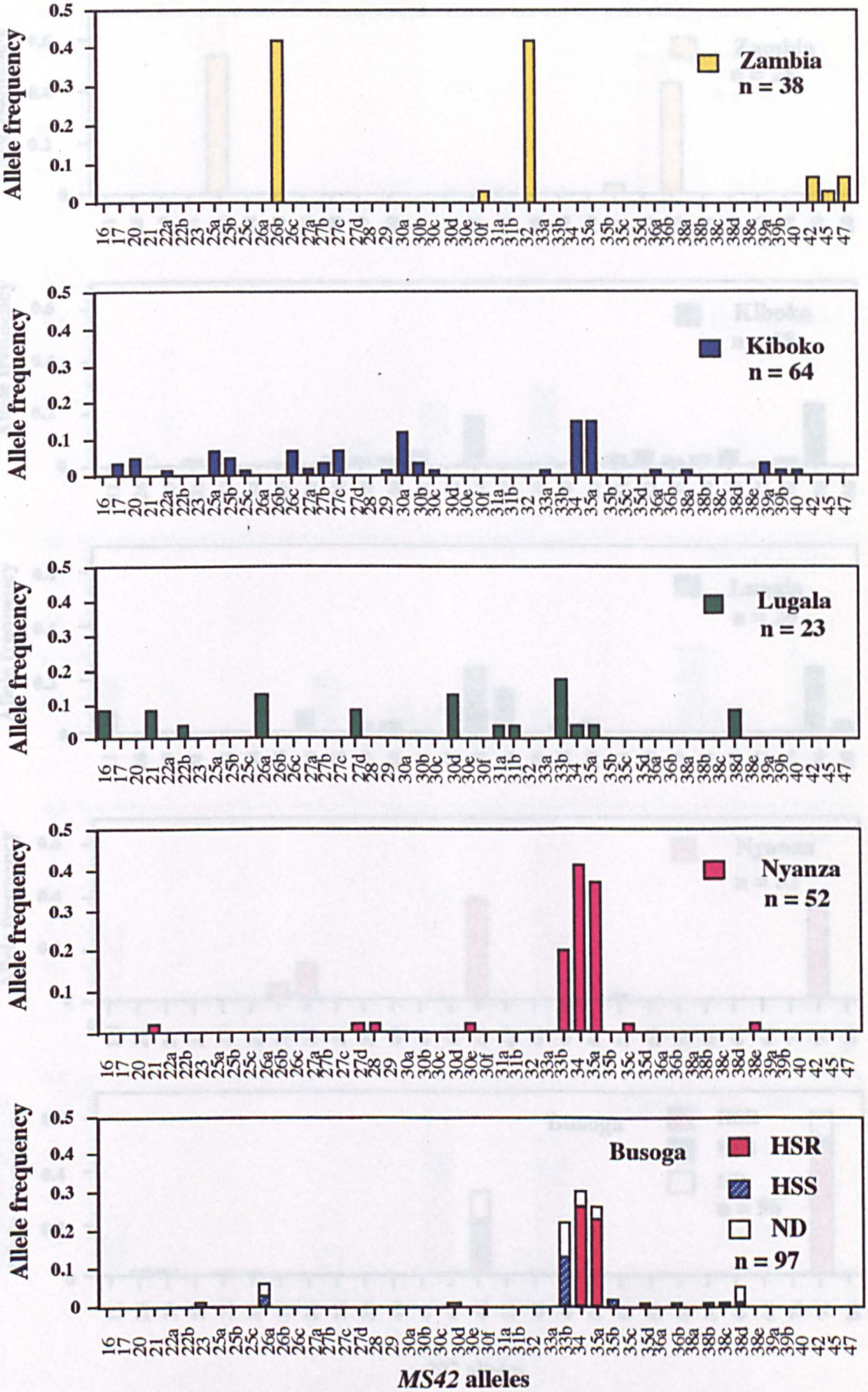


Figure 7.2 . The frequency of *MS42* alleles in each population. HSR, human serum resistant; HSS, human serum sensitive; ND, not determined. n = number of alleles.

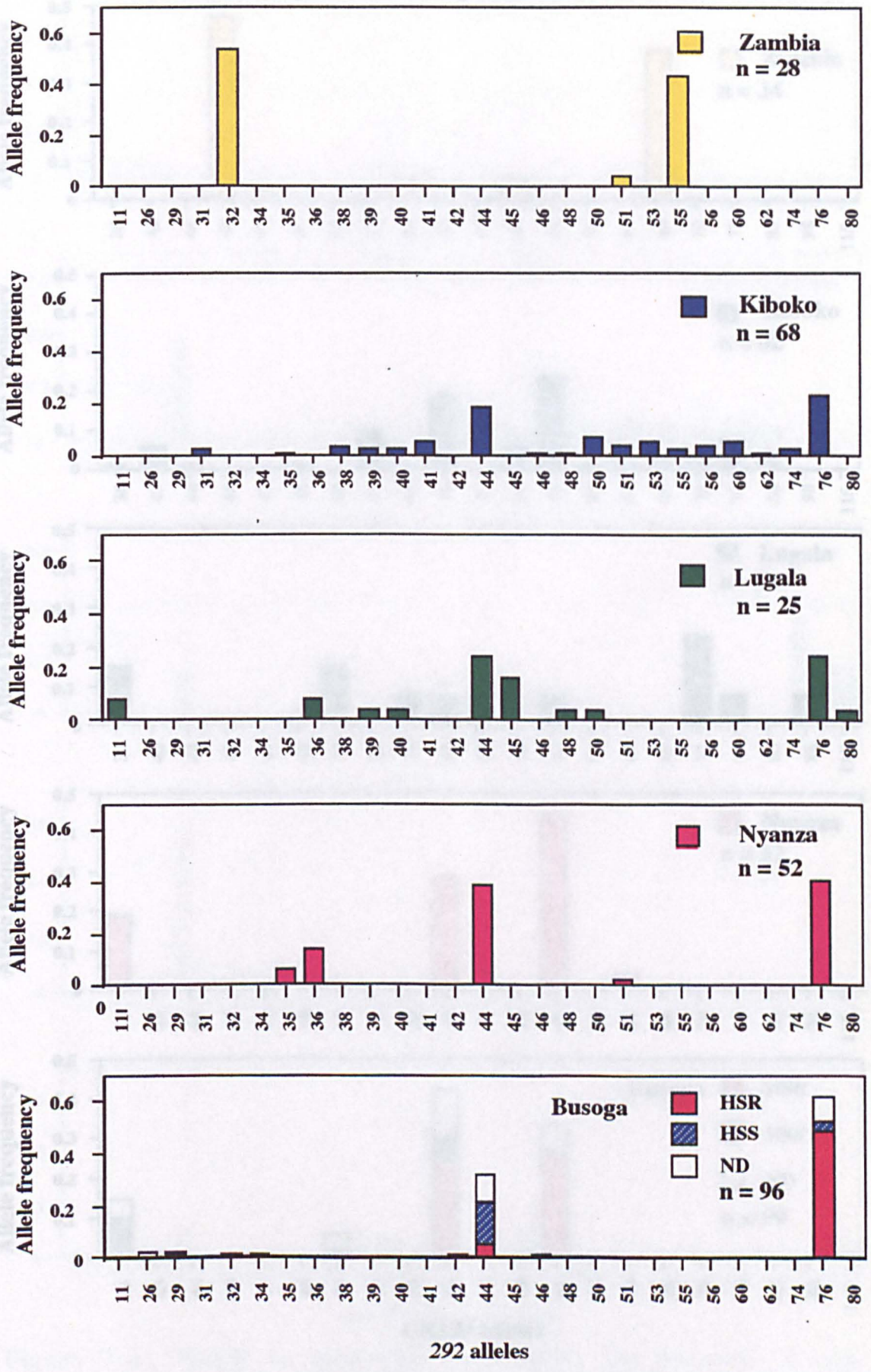


Figure 7.3. The frequency of *292* alleles in each population. HSR, human serum resistant; HSS, human serum sensitive; ND, not determined. n = number of alleles.

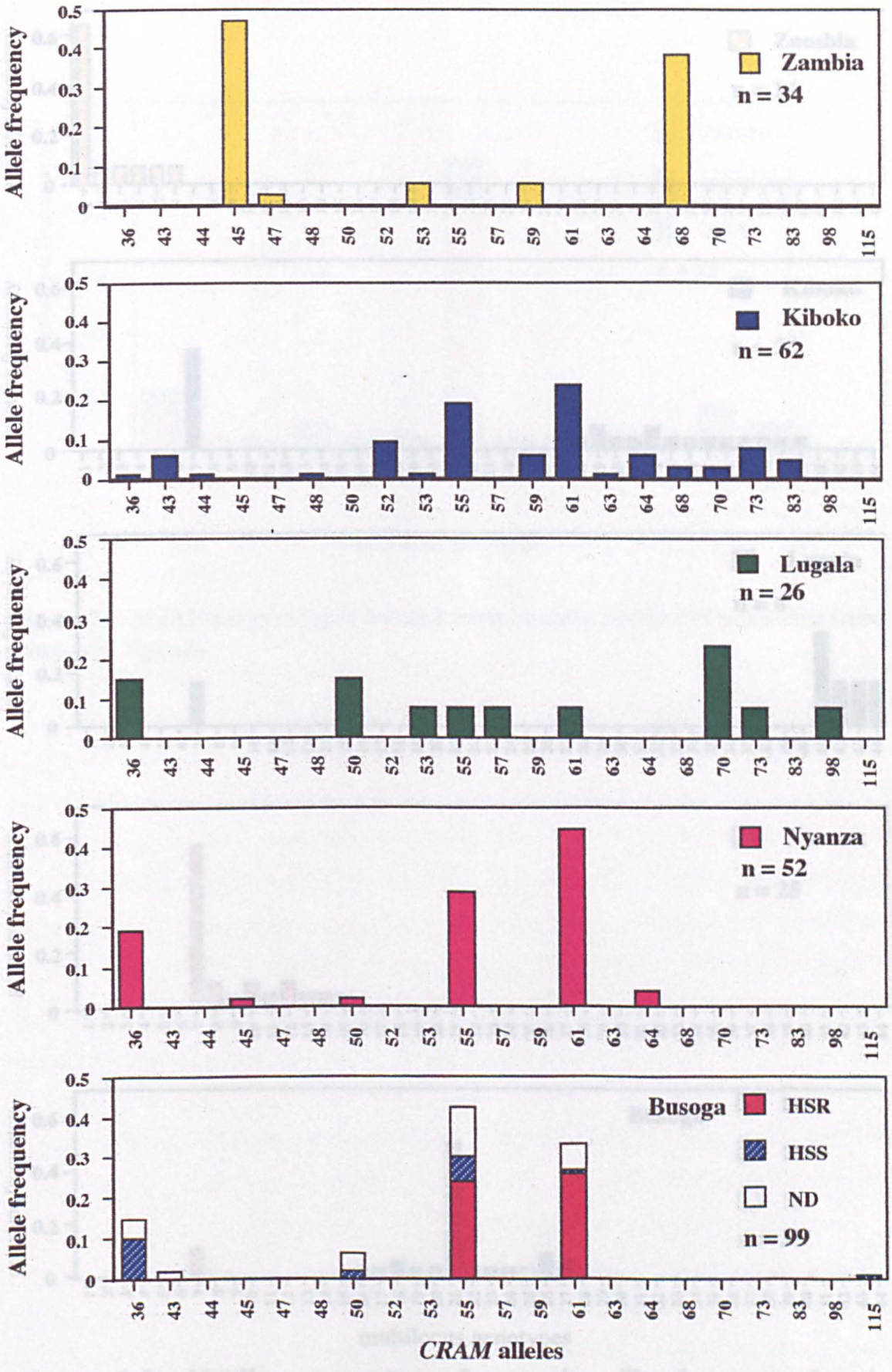


Figure 7.4. The frequency of *CRAM* alleles in each population. HSR, human serum resistant; HSS, human serum sensitive; ND, not determined. n = number of alleles.

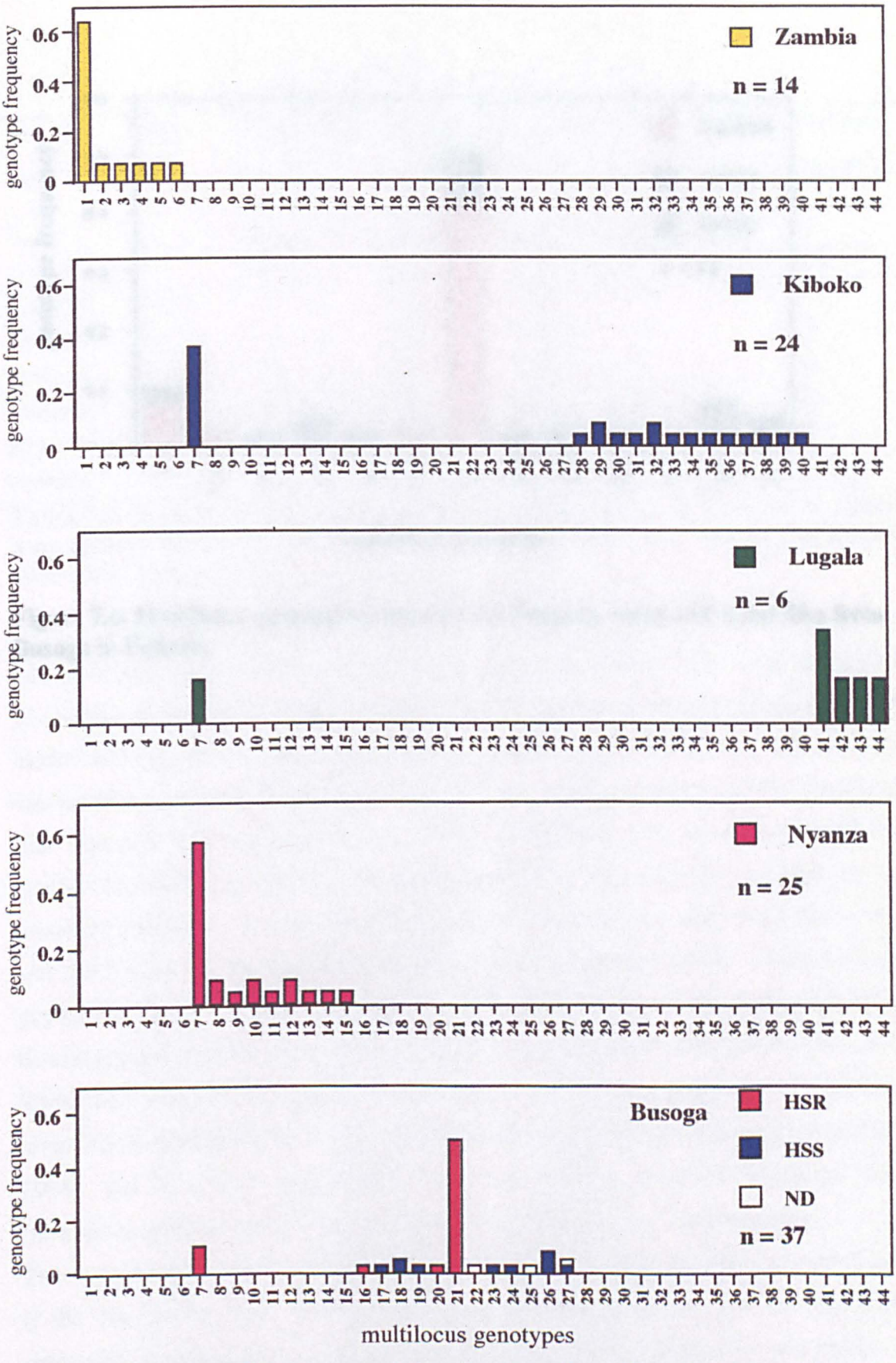


Figure 7.5. Multilocus genotype frequencies. The frequency of each multilocus genotype within each population is illustrated. Multilocus genotypes are described Table A.2, Appendix. HSR, human serum resistant; HSS, human serum sensitive; ND, not determined. n = number of samples.

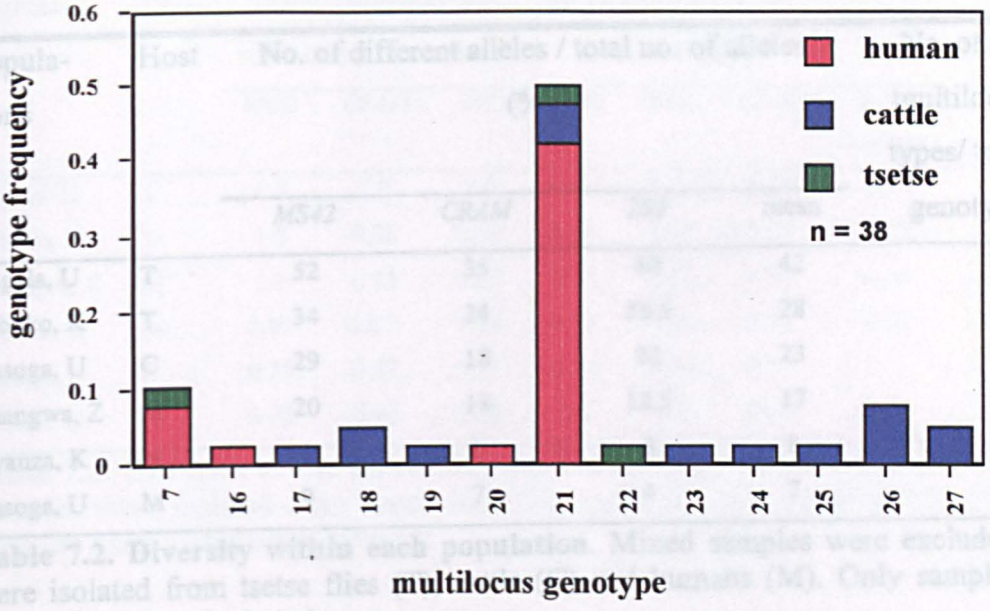


Figure 7.6. Multilocus genotypes isolated from humans, cattle and tsetse flies from Busoga in Uganda.

Popula-tions	Host	No. of different alleles / total no. of alleles (%)				No. of different multilocus geno-types/ total no. of genotypes (%)
		MS42	*CRAM*	*292*	mean	
Lugala, U	T	52	35	40	42	83
Kiboko, K	T	34	24	26.5	28	58
Busoga, U	C	29	18	22	23	69
Luangwa, Z	M	20	18	12.5	17	50
Nyanza, K	M	8	8	8	8	29
Busoga, U	M	9	7	4	7	19

Table 7.2. Diversity within each population. Mixed samples were excluded. Samples were isolated from tsetse flies (T), cattle (C) and humans (M). Only samples from the hosts stated were included.

Hardy-Weinberg and linkage disequilibrium. Although *T. brucei* has been demonstrated to undergo selfing under laboratory conditions [Tait *et al.*, 1996], it is not clear to what extent selfing occurs in natural populations. The fixation index, F_{is}, is a measure of the amount of reduction in heterozygosity observed due to inbreeding (or selfing) compared to that expected if the population was in HW proportions. The heterozygosity for each marker (minisatellites and RFLPs) and the fixation index were calculated for each population using the Genetic Data Analysis (GDA) program [Lewis and Zaykin, 1999], and the results are presented in Table 7.3. Combining the data for all markers, each population gave a negative fixation index value, indicating that there was an excess of heterozygotes rather than homozygotes. The F_{is} values for the RFLP markers were slightly lower than those for the minisatellite markers, probably reflecting the high number of minisatellite alleles relative to sample size, a phenomenon described by [Scribner *et al.*, 1994]. The F_{is} values varied between populations, being closest to 0 for the Kiboko (Kenyan) population and as low as -0.54 for the Luangwa (Zambian) population.

The negative values of F_{is} indicate a deviation from the number of heterozygotes predicted if the populations were in HW equilibrium. To investigate this for each population separately to avoid any deviation from HW expectations caused by combining sub-populations, the number of stocks with the observed genotypes was compared to the

Popula-tions	Host	Heterozygosity for minisatellite markers			Heterozygosity for RFLP markers			F_{is} based on data from		
		MS42	*CRAM*	*292*	*Bgl*II	*Hinc*II	*Mbo*II	mini-satellites	RFLPs	Over-all
Lugala, U	T	1.0	1.0	1.0	0.67	0.37	0.57	-0.16	-0.43	-0.25
Kiboko, K	T	1.0	0.96	0.96	0.36	0.32	0.27	-0.12	-0.17	-0.13
Luangwa, Z	M	1.0	0.93	0.83	0.75	0.83	0.50	-0.54	-0.54	-0.54
Nyanza, K	M	0.91	0.875	0.96	0.78	0.86	0.56	-0.33	-0.58	-0.43
Busoga, U	C	0.75	0.57	0.25	0.33	0.73	0.77			
Busoga, U	M	0.95	0.95	0.19	1.0	1.0	0.89	0.02	-0.70	-0.29

Table 7.3. Estimate of the inbreeding coefficient (F_{IS}). Mixed samples were excluded. Samples were isolated from tsetse flies (T), cattle (C) and humans (M). Only samples from the hosts stated were included.

expected number of genotypes if the population was in HW equilibrium, for each minisatellite marker and for the RFLP markers and the results obtained are presented in Tables 7.4 and 7.5. Given the large number of different alleles in each population, the number of expected genotypes is extremely large, and so many genotypes are not represented in the population samples. Therefore only those genotypes which were observed in the population are given in Table 7.4. Inspection of Table 7.4 reveals that in each population one genotype tends to occur far more frequently than was predicted, for example, in the Luangwa population the *MS42* 32/26b genotype has been observed 15 times compared to the expected 6.7 times. As some genotypes are over-represented in each population, (with the exception of Lugala, although this is the smallest population sample), other genotypes are under-represented or not present at all, which suggests deviation from HW expectations in most populations. The samples from the Busoga collection were separated into two different groups for this analysis depending on their resistance to human serum (samples which have not been tested for human serum resistance but were isolated from cattle were included in the human serum sensitive group). From the data presented in Table 7.4 the Busoga human serum resistant (HSR) population appears to show deviation from the expected whereas the human serum sensitive (HSS) population is in agreement with the expected values. However by dividing the Busoga samples into two groups, the size of the HSS population is small (10 isolates) and so may not give a true reflection of the HSS population. In general for the RFLP data,

Population	Number of times each genotype was observed					
	MS42		*CRAM*		*292*	
	genotype	obs (exp)	genotype	obs (exp)	genotype	obs (exp)
Luangwa, Z	32/26b	15 (6.7)	68/45	13 (6.12)	55/32	11 (6.43)
	45/32	1 (0.54)	59/59	1 (0.06)	32/32	2 (4.02)
	47/42	2 (0.169)	53/45	1 (0.94)	55/51	1 (0.43)
	30f/26b	1 (0.54)	47/45	1 (0.47)		
Nyanza, K	35a/34	16 (8.9)	61/55	15 (6.64)	76/44	18 (8.08)
	35c/28	1 (0.02)	61/61	2 (5.09)	51/44	1 (0.38)
	34/33b	3 (3.5)	50/36	1 (0.19)	76/36	2 (2.83)
	33b/33b	2 (0.92)	61/36	4 (4.42)	36/36	1 (0.47)
	35a/33b	1 (3.1)	64/36	2 (0.38)	36/35	2 (0.40)
	33b/21	1 (0.2)	36/36	1 (0.96)	44/36	1 (2.69)
	33b/27d	1 (0.2)	45/36	1 (0.19)	76/35	1 (1.21)
Busoga, U	35a/34	23 (6.13)	61/55	25 (10.75)	76/44	5 (14.72)
Total	35d/34	1 (0.23)	61/61	1 (4.22)	76/76	23 (14.23)
	38b/26a	1 (0.05)	115/55	1 (0.33)	44/44	8 (3.81)
	33b/33b	5 (1.7)	55/50	2 (1.95)	44/29	2 (0.51)
	36b/33b	1 (0.17)	36/36	5 (0.87)	46/42	1 (0.01)
	34/34	1 (3.53)	55/36	2 (4.88)		
	38d/26a	1 (0.23)	61/36	1 (3.84)		
	38c/26a	1 (0.23)	55/55	1 (6.84)		
	38d/33b	1 (0.81)				
	35b/33b	2 (0.33)				
Busoga, HSR	35a/34	23 (11.96)	61/55	24 (12.48)	76/44	5 (4.5)
	35d/34	1 (0.52)	61/61	1 (6.76)	76/76	20 (20.25)
	34/34	1 (6.76)				
Busoga, HSS	38b/26a	1 (0.15)	115/55	1 (0.3)	76/76	2 (0.36)
	33b/33b	4 (3.03)	55/50	2 (0.6)	44/44	7 (5.11)
	36b/33b	1 (0.55)	36/36	4 (2.5)	44/29	1 (0.68)
	38d/26a	1 (0.15)	55/36	1 (3)	46/42	1 (0.05)
	38c/26a	1 (0.15)	55/55	1 (0.9)		
	35b/33b	2 (0.55)	61/36	1 (0.5)		
Kiboko, K	40/25a	1 (0.05)	83/83	1 (0.06)	48/44	1 (0.12)
	35a/34	9 (1.17)	61/55	9 (2.25)	76/44	9 (1.98)
	30a/17	2 (0.19)	61/52	2 (0.96)	60/50	2 (0.19)
	30c/20	1 (0.04)	68/61	1 (0.32)	41/50	3 (0.19)
	26c/25a	1 (0.19)	70/68	1 (0.04)	44/38	1 (0.25)
	30a/27c	4 (0.38)	73/55	3 (0.64)	76/74	1 (0.31)
	29/27b	1 (0.02)	61/53	1 (0.16)	46/31	1 (0.19)
	38a/20	1 (0.04)	64/43	2 (0.17)	76/62	1 (0.15)
	36a/25b	1 (0.04)	61/44	1 (0.16)	51/39	2 (0.06)
	26c/25b	1 (0.14)	63/55	1 (0.13)	53/40	1 (0.08)
	39a/25b	1 (0.07)	68/36	1 (0.02)	76/56	1 (0.46)
	30a/25c	1 (0.94)	83/59	1 (0.13)	56/56	1 (0.04)
Lugala, U	38d/27d	1 (0.1)	70/36	2 (0.43)	45/36	2 (0.18)
	35a/34	1 (0.02)	61/55	1 (0.07)	76/44	1 (0.81)
	33b/30d	1 (0.27)	50/36	1 (0.28)	45/44	1 (0.54)
	26a/22b	1 (0.07)	61/36	1 (0.14)	76/45	1 (0.54)
	38d/27d	1 (0.09)	70/55	1 (0.21)	80/44	1 (0.13)
	33b/26a	1 (0.27)			76/39	1 (0.13)

Table 7.4. Number of times each genotype that was observed compared to the expected value. The genotypes for a few samples were not determined (see Table A2 in appendix), therefore the number of genotypes analysed in each population for each marker may vary.

Genotype	Number of times each genotype was observed (and expected) in each population						
	Kiboko	Lugala	Busoga Total	Busoga (HSR)	Busoga (HSS)	Nyanza	Zambia
***Hin*cII**							
++	0 (0.56)	1 (0.78)	0 (6.82)	0 (5.75)	0 (1.23)	0 (3.77)	0 (2.16)
+-	7 (5.89)	3 (3.4)	30 (16.36)	23 (11.5)	7 (4.55)	19 (11.47)	11 (6.68)
--	15 (15.56)	4 (3.78)	3 (9.82)	0 (5.75)	3 (4.23)	5 (8.76)	3 (5.15)
***Bgl*II**							
++	14 (14.73)	0 (0.667)	0 (4.13)	0 (5)	0 (0.27)	0 (4)	3 (5.15)
+-	8 (6.545)	4 (2.66)	24 (15.79)	20 (10)	4 (3.46)	20 (12)	11 (6.68)
--	0 (0.726)	2 (2.67)	11 (15.15)	0 (5)	11 (11.28)	5 (9)	0 (2.16)
***Mbo*II**							
++	0 (0.41)	1 (1.29)	0 (5.93)	0 (4.05)	0 (1.93)	0 (2)	0 (0.57)
+-	6 (5.18)	4 (3.43)	28 (16.12)	18 (9.9)	10 (6.16)	14 (10)	4 (2.86)
--	16 (16.4)	2 (2.28)	5 (10.96)	2 (6.05)	3 (4.92)	11 (13)	3 (3.57)

Table 7.5. Number of times each genotype that was observed compared to the expected value. The genotypes for a few samples were not determined (see Table A3 in appendix), therefore the number of genotypes analysed in each population for each marker may vary.

which are presented in Table 7.5, there is better agreement between the number of observed and expected genotypes, especially for the Kiboko and Lugala populations, with the Busoga, Nyanza and Zambian populations showing greater deviation due to an excess of heterozygotes and a deficit of ++ homozygotes for each RFLP marker. As with the minisatellite markers, the RFLP data of all the Busoga samples show deviation from the expected frequencies as does the HSR samples, but the observed genotype frequencies from the HSS samples show agreement with the expected values. This suggests that the Busoga population is sub-structured due to host specificities.

Because the collection of 116 samples is divided into 6 populations (based on geography and host specificities), each population is small making the statistical evaluation of the population difficult. For this reason the Lugala population has been omitted from the Hardy-Weinberg analysis, as this population is the smallest collection of samples (n=7) and therefore agreement with HW expectations could occur by chance due to the small sample size [Cibulskis, 1988]. For the analysis several approaches were taken to examine deviation from HW predictions, although no method can fully compensate for the small

sample sizes. Deviation from HW expectations and linkage disequilibrium were calculated using the shuffling tests of Weir, and the results are presented as probabilities of the exact significance levels in Tables 7.6 and 7.7 for the minisatellites and RFLPs, respectively. A probability of less than 0.05 indicates a statistically significant deviation from HW expectations and linkage equilibrium. With small sample sizes, such as these, the RFLP data are more suitable for HW analysis than the minisatellite data as there are fewer alleles and therefore fewer possible genotypes. The probability of linkage equilibrium measured using the three loci analysed by RFLP, however, is not valid for this population analysis as the three loci are closely linked on Chromosome I (see Chapter 8), therefore only the HW analysis based on RFLP data is presented in Table 7.7. The results with the minisatellite data are consistent with the F_{is} values, indicating that each population (with the exception of Busoga HSS population) is not in agreement with HW predictions, or in linkage equilibrium. The RFLP data for HW should give slightly more statistically robust results as only two alleles are observed. From these results the Busoga (HSS) and Kiboko populations appear to show agreement with the results expected if the populations were in HW equilibrium, with little or no agreement with HW expectations for the Nyanza, Luangwa and Busoga (HSR) populations. Comparison of these results (Table 7.7) with those from the minisatellites (Table 7.6), suggests that the Busoga (HSS) population is in agreement with HW expectations. It is interesting to note that the Busoga populations have fewer minisatellite alleles than the other populations which may explain the agreement between the minisatellite and RFLP results. The conflicting results obtained for the Kiboko population (n=24) for the minisatellite analysis compared to the RFLP data is probably due to the very large number of minisatellite alleles detected in these samples, which generate a large number of different possible genotypes, but the sample size is small so that they are unlikely to be detected. With the two allele system (RFLPs) the small sample size is less crucial and so agreement is obtained. The human serum resistant samples from Busoga appeared not to be in HW equilibrium, whereas the human serum sensitive samples may be in agreement with HW expectations. When the human serum resistant and sensitive samples are combined, the population appears to show a deviation from HW expectations and linkage equilibrium. These results are consistent with those expected if the Busoga population was a mixture of two distinct sub-populations *i.e.* if there was sub-structuring due to host specificities. However it must be remembered that by dividing the samples into two groups the sample size is reduced further (to 25 and 15) respectively and this may affect the results.

Populations	N	Probability of agreement with HW			Probability of agreement with Linkage Equilibrium		
		MS42	*CRAM*	*292*	*MS42/ CRAM*	*MS42/ 292*	*CRAM /292*
Luangwa, Z	19	0.000	0.000	0.025	0.000	0.005	0.000
Nyanza, K	26	0.000	0.000	0.000	0.000	0.000	0.000
Kiboko, K	24	0.000	0.000	0.000	0.000	0.000	0.000
Busoga, U-HSR	25	0.000	0.000	1.000	0.000	0.000	0.000
Busoga, U-HSS	15	0.165	0.227	0.000	0.018	0.001	0.000
Busoga, U-Total	40	0.000	0.000	0.000	0.000	0.000	0.000

Table 7.6. Hardy-Weinberg and linkage equilibrium analysis for each population, calculated for the minisatellite data, using the GDA program. Shaded cells indicate agreement with HW equilibrium or linkage equilibrium *i.e.* probability of >0.05. The Lugala population was omitted due to its small sample size (N).

Populations	N	Probability of agreement with HW		
		*Bgl*II	*Hinc*II	*Mbo* II
Luangwa, Z	19	0.034	0.034	1.000
Nyanza, K	26	0.001	0.003	0.061
Kiboko, K	24	1.000	1.000	1.000
Busoga, U-HSR	25	0.000	0.000	0.000
Busoga, U-HSS	15	1.000	0.219	0.091
Busoga, U-Total	40	0.003	0.000	0.000

Table 7.7. Hardy-Weinberg analysis for each population for the RFLP data, using the GDA program. Shaded cells indicate agreement with HW expectations *i.e.* probability of >0.05. The Lugala population was omitted due to its small sample size (n).

Populations	N	Probability of agreement with HW by			χ^2
		MS42	*CRAM*	*292*	
Luangwa, Z	19	<0.001	<0.001	0.05>P>0.01	
Nyanza, K	26	<0.001	<0.01	<0.001	
Kiboko, K	24	<0.01	0.5	0.05>P>0.01	
Busoga, U-HSR	25	<0.001	<0.001	0.9>P>0.7	
Busoga, U-HSS	15	0.3>P>0.2	=0.05	>0.05	
Busoga, U-Total	40	<0.001	<0.001	<0.01	

Table 7.8. Hardy-Weinberg analysis for each population, calculated using minisatellite data for the two most common alleles and combining all other allele frequencies. Shaded cells indicate agreement with HW *i.e.* probability of >0.05. The Lugala population was omitted due to small sample size (n).

Two alternative methods for calculating deviation from HW, involving binning or grouping alleles together, were also applied to the minisatellite data. All alleles from each population except the two most frequently sampled alleles were combined and treated as a third allele. The observed and expected frequencies were compared and the significance was then calculated using χ^2 test. The probability of agreement to HW expectations for each population is presented in Table 7. 8 and agrees with that obtained from the RFLP analysis, in that the Busoga HSS population (and perhaps the Kiboko population) shows agreement with HW while the other populations appear to deviate from HW expectations. In a similar way, binning alleles according to similarities in their internal structure as defined by MVR mapping (see Chapter 8) gave similar results (data not shown).

A possible explanation for the deviation from HW equilibrium shown in the *T. brucei* populations under study could be an epidemic population structure. From the allele and multilocus genotype frequencies depicted in Figures 7.2-7.5, is clear that one or two genotypes dominate each population, particularly in the populations isolated from humans. Given the large number of alleles present at all three minisatellite loci, the expected frequency of any particular multilocus genotype in the population would be extremely small, assuming HW and linkage equilibrium. Therefore observing a genotype several times in one population would be wholly unexpected, a probability which can be calculated using the multiplication rule (*i.e.* the expected genotype frequency for one locus is multiplied by the expected genotype frequency at another locus). This is illustrated for the most common genotype (multilocus genotype 7) where the expected frequency (assuming HW equilibrium) is far smaller than the observed frequency for each population in which it was observed (see Table 7.9). The same is true for genotype 21, which is the most common genotype present in the Busoga population (data not shown).

Populations	**Frequency of multilocus genotype 7**	
	observed	expected
Luangwa, Z	0	0
Nyanza, K	0.56	0.01894
Busoga, U	0.105	0.0177
Kiboko, K	0.375	0.000313
Lugala, U	0.167	0.0000051

Table 7.9. Observed and expected frequency of genotype 7 in each population.

If deviation from HW equilibrium and linkage disequilibrium were due an epidemic population structure [Maynard-Smith *et al.*, 1993; Hide, 1999], whereby a few common genotypes obscure a randomly mating population structure, the removal of the common genotypes may reveal the true population structure. The frequency of each multilocus genotype detected (see also the genotype frequencies in Figure 7.5) is presented in Table 7.10. Most genotypes were unique, being detected only once, however a few genotypes were detected several times, with genotype 7 being detected 28 times. Following the type of analysis described by Maynard Smith, each multilocus genotype (defined by minisatellite analysis) which was detected more than once, was considered as one individual or electrophoretic type (ET). Re-analysis, using Weir's shuffling tests of the minisatellite data for HW expectations using ETs only revealed that most of the populations were in agreement with HW predictions (see Table 7.11), which would be predicted if an epidemic population structure were prevalent. The same data set of only ETs was re-analysed for the RFLP markers (Table 7.12), and revealed that nearly all populations are in agreement with HW expectations. However, caution must be taken in drawing conclusions as, by removing many stocks from the analysis, the population sizes were reduced further, increasing the probability of obtaining agreement with HW expectations by chance. For example, the Busoga (HSR) population was reduced to 4 genotypes, which is too small a sample size from which to draw any conclusions. The HSS samples (n=14) appears to show agreement with HW predictions. However, when the HSR and HSS samples are combined, deviation from HW expectations is detected (for both minisatellite and RFLP data), which is indicative of sub-structuring due to host specificities. The Nyanza population also shows agreement with HW expectations. However the Kiboko and Luangwa populations do not show agreement with HW predictions for the minisatellite data but show strong agreement for the RFLP data, which may be a reflection of the fact that these populations have a large number of minisatellite alleles.

No. of times each genotype sampled	No. of different genotypes
1	32
2	8
3	1 (genotype 26)
9	1 (genotype 1)
19	1 (genotype 21)
28	1 (genotype 7)

Table 7.10. Common genotypes.

Populations	n	Probability of agreement with HW (ETs)			Probability of agreement with Linkage Equilibrium (ETs)		
		MS42	*CRAM*	*292*	*MS42/ CRAM*	*MS42/ 292*	*CRAM/ 292*
Luangwa, Z	11	0.002	0.019	0.641	0.262	0.889	0.586
Nyanza, K	10	0.806	0.833	0.420	0.619	0.834	0.832
Kiboko, K	16	0.001	0.044	0.009	0.054	0.035	0.104
Busoga, U-HSR	4	1.000	1.000	1.000	0.839	0.839	1.000
Busoga, U-HSS	14	0.145	0.398	0.000	0.127	0.001	0.000
Busoga, U-Total	18	0.006	0.120	0.001	0.007	0.001	0.000

Table 7.11. Hardy-Weinberg and linkage equilibrium analysis for each population, calculated for the minisatellite data for ETs only, using the GDA program. Shaded cells indicate agreement with HW or linkage equilibrium *i.e.* probability of >0.05. The Lugala population was omitted due to its small sample size. The Busoga HSR samples were included from comparison.

Populations	n	Probability of agreement with HW (ETs)		
		*Bgl*II	*Hinc*II	*Mbo* II
Luangwa, Z	11	1.000	1.000	1.000
Nyanza, K	10	0.468	1.000	1.000
Kiboko, K	16	1.000	1.000	1.000
Busoga, U-HSR	4	1.000	0.322	1.000
Busoga, U-HSS	14	1.000	0.173	0.070
Busoga, U-Total	18	0.528	0.022	0.030

Table 7.12. Hardy-Weinberg and linkage equilibrium analysis for each population, calculated for the RFLP data for ETs only, using theGDA program. Shaded cells indicate agreement with HW or linkage equilibrium *i.e.* probability of >0.05. The Lugala population was omitted due to its small sample size. The Busoga HSR samples were included from comparison.

To investigate further the possibility of an epidemic population structure, the Index of Association (I_A) was calculated for each population as described by Maynard-Smith *et al.* (1993). The I_A measures the association of alleles (or genotypes for diploid organisms) at different loci, and has a predicted value of zero (or a negative value) for populations which are randomly mating, while if recombination is rare or absent the I_A is larger with a value significantly different from zero. The I_A values for each population, using the total data set, (*i.e.* all isolates) are presented in Table 7.13, with the Nyanza, Luangwa, Kiboko and Busoga (HSR) populations all having large I_A values, indicating non-random mating populations, whereas the Busoga (HSS) and the Lugala samples have small I_A values, suggesting little linkage disequilibrium and therefore indicating that genetic exchange could be occurring in these populations. The data were also analysed using only ETs to reduce population distortion which may be caused by the epidemic spread of particular strains [Maynard-Smith *et al.*, 1993]. The analysis shows that when the common genotypes are removed the I_A values are greatly reduced all becoming negative, thus indicating that genetic exchange and recombination could be occurring in all populations, but that the population structure was distorted by an epidemic spread of one or two genotypes for the Nyanza, Luangwa, Busoga (HSR) and Kiboko populations. The difference in I_A for all isolates for the HSR and HSS samples from Busoga suggests that there is a different population structure for human infective and non-human infective trypanosomes, which has previously been suggested by Hide *et al.* [Hide *et al.*, 1994]. By inspection of the data, in terms of the numbers of distinct alleles and their frequencies, this is clearly due to a predominance in the HSR population of one major genotype. However it must be remembered that small sample sizes (particularly for the Lugala population) and the few loci analysed in this study may result in misleading conclusions.

Similar re-calculation of the F_{IS} value using only ETs was carried out (Table 7.14), resulting in a increase in the F_{IS} values for most populations, bringing them closer to zero (the predicted value for a panmictic population). Again the lowest values were obtained for Kiboko and the Busoga (HSS) populations.

Variation between populations. From the allele frequency distributions illustrated in Figures 7.2 - 7.4, it is clear that a great deal of variation exists not only within but also between the different populations, with limited allele sharing between populations. Many alleles are specific to the population from which they were derived. Table 7.15 indicates the number of private alleles (alleles not present in another population) per locus,

Population	Number of isolates	I_A all isolates	Number of isolates (ETs only)	I_A (ETs only)
Luangwa (Z)	14	1.447	6	-0.489
Nyanza (K)	25	1.138	9	-0.002
Busoga HSR (U)	25	1.539	4	-0.391
Busoga HSS (U)	13	0.116	9	-0.192
Kiboko (K)	24	1.407	14	-0.488
Lugala (K)	6	0.08	5	-1

Table 7.13. Comparison of Index of Association, I_A, for each population. The I_A was calculated saparately for all isolates and ETs only, for each population as described in Maynard-Smith *et al.*, 1993.

Populations	F_{is} all data	F_{is} common genotypes removed
Luangwa Z	-0.54	-0.24
Nyanza K	-0.43	-0.20
Busoga, U -HSR	-0.84	-0.50
Busoga, U -HSS	-0.05	-0.07
Kiboko K	-0.13	-0.03
Lugala U	-0.25	-0.20

Table 7.14. Comparison of inbreeding coefficient (F_{IS}) estimates using all samples and electrophoretic types only. Mixed samples were excluded.

identified from each population, ranging from 0 for the Nyanza population for the *CRAM* locus, to a striking 15 *MS42* alleles for the Kiboko population. The number of alleles shared between populations is presented in Tables 7.16- 7.18. From these data, the Zambian population appears to be the most genetically distinct with limited allele sharing with any of the other populations, whereas Busoga, Lugala and Nyanza share a considerable number of alleles. This can also be clearly seen in the allele frequency graphs of Figures 7.2-7.4.

The amount of inter-population genetic variation can be assessed by estimating the fixation index, F_{ST} which determines the amount of genetic difference between sub-populations, with a theoretical minimum of 0 indicating no genetic divergence, and a maximum of 1 indicating fixation for each sub-population for different alleles. The F_{IT} is a measure of the amount of inbreeding, taking into account both the effect of non-random mating within the sub-population and the effect of population subdivision. Estimates of F_{ST} and F_{IT} were calculated, using all markers, for the *T. brucei* samples (using the GDA program [Lewis and Zaykin, 1999]), and the results are presented in Table 7.19. From these results the overall estimate of the F_{ST} value was 0.141, with 95% confidence intervals of 0.168 and 0.105 (measured by bootstrapping). F_{ST} values of this magnitude indicate moderate to great genetic difference between populations [Hartl and Clark, 1997], indicating that the populations are sub-divided. Re-analysis using only ETs increased the F_{ST} value marginally to 0.168, reflecting the fact that the common genotype (genotype 7) is the only genotype which is present in more than one population and so by removing it from the analysis the difference between populations is greater. The overall F_{IS} value (the inbreeding coefficient) was -0.318, indicating that there is an excess of heterozygotes in the populations. After removing the common genotypes the F_{IS} value became closer to 0 at -0.1485. The overall estimate of inbreeding measured by F_{IT} which was -0.132, also increased to 0.0447 when the common genotypes were removed.

A pairwise measure of genetic distance (Nei 1978) was calculated using the minisatellite data and the results are presented in Table 7.20. From this table it is clear that the populations with the smallest genetic distance between then are the Nyanza and Lugala populations, followed by Nyanza and the HSR samples from Busoga. Interestingly the HSR and HSS samples from Busoga display a large genetic distance of 0.471, similar to that of Nyanza and Zambia, which provides strong evidence for sub-structuring due to host specificities. However caution must be taken not to over-interpret these results, as the sample sizes are small and so may not be truly representative of the population from

Populations	Host	Number of private alleles/population		
		MS42	*CRAM*	*292*
Lugala, U	T	2	0	2
Kiboko, K	T	15	6	11
Luangwa, Z	M	6	1	2
Nyanza, K	M	2	0	1
Busoga, U	M, C	5	1	2

Table 7.15. Private alleles/population. Samples were isolated from tsetse flies (T), cattle (C) and humans (M). Only samples from the hosts stated were included.

Populations	Number of *MS42* alleles shared between populations				
	Zambia (6)	Nyanza (9)	Busoga (12)	Kiboko(22)	Lugala(12)
Zambia (6)	-	0	0	0	0
Nyanza (9)	-	-	3	2	5
Busoga (12)	-	-	-	2	6
Kiboko (22)	-	-	-	-	2
Lugala (12)	-	-	-	-	-

Table 7.16. Number of distinct *MS42* alleles shared by each pair of populations. Numbers in parentheses show the number of distinct alleles found in each population.

Populations	Number of *292* alleles shared between populations				
	Zambia (3)	Nyanza (5)	Busoga (8)	Kiboko(18)	Lugala(10)
Zambia (3)	-	1	1	2	0
Nyanza (5)	-	-	2	4	3
Busoga (8)	-	-	-	3	2
Kiboko (18)	-	-	-	-	6
Lugala (10)	-	-	-	-	-

Table 7.17. Number of distinct *292* alleles shared by each pair of populations. Numbers in parentheses show the number of distinct alleles found in each population.

Populations	Number of *CRAM* alleles shared between populations				
	Zambia (5)	Nyanza (6)	Busoga (6)	Kiboko(15)	Lugala(9)
Zambia (5)	-	1	0	3	1
Nyanza (6)	-	-	4	4	4
Busoga (6)	-	-	-	4	4
Kiboko (15)	-	-	-	-	6
Lugala (9)	-	-	-	-	-

Table 7.18. Number of distinct *CRAM* alleles shared by each pair of populations. Numbers in parentheses show the number of distinct alleles found in each population.

Genetic variability within and between populations			
	F_{IS}	F_{IT}	F_{ST}
over all markers	**-0.318**	**-0.132**	**0.141**
95% CI upper	-0.174	0.016	0.168
95% CI lower	-0.498	-0.349	0.105
over all markers with common genotypes removed	**-0.1485**	**0.0447**	**0.1682**
95% CI upper	-0.025	0.125	0.235
95% CI lower	-0.330	-0.086	0.121

Table 7.19. Genetic variability within and between populations.

Pairwise estimates of genetic distance (Nei 1978) among populations						
Populations	Kiboko	Lugala	Busoga (HSS)	Busoga (HSR)	Nyanza	Zambia
Kiboko		0.125	0.453	0.256	0.145	0.369
Lugala			0.086	0.250	0.053	0.370
Busoga (HSS)				0.471	0.167	0.651
Busoga (HSR)					0.074	0.975
Nyanza						0.493

Table 7.20. Pairwise estimates of genetic distance (Nei 1978) among populations, calculated from minisatellite alleles.

which they were derived. The largest genetic distance (0.975) was found to be between the HSR samples of Busoga and the Zambian population. These results illustrate what is immediately obvious from the allele frequency histograms, *i.e.* that the Zambian population is the most divergent from the other populations examined and that there is a close relationship between the Nyanza, Busoga and Lugala populations. It is interesting to note that the Kiboko population appears to be closer to the Busoga HSR population than the Busoga HSS population. This is due to the detection of genotype 7 in the Kiboko population, which is a common genotype in the Busoga HSR population and is closely related to major genotype 21. The implication of this result is that there are human infective trypanosomes within the Kiboko population.

Discussion

Choice of markers. As mentioned in Chapter 1 the markers for any analysis should be chosen for the particular question that is being addressed. For evolutionary studies and comparisons between species, markers which are not particularly variable would be desirable, and for more recent population changes markers which are evolving at faster rates are required. To trace particular genotypes through populations and to address the question of clonality, markers which can distinguish between particular genotypes (such as minisatellites) are required. In this chapter, five *T. brucei* populations have been examined, using the three hypervariable minisatellite markers described in Chapter 3, to address the question of clonality and population sub-structuring. However, because the sample sizes are small and there are many alleles at each locus, these markers are probably less than ideal for the analysis of deviations from HW expectation and less variable markers are more appropriate. It was for this reason that three RFLP markers (located on Chromosome 1) have been employed to detect deviations from HW equilibrium.

Clonality. It is without doubt that sexual recombination does occur in natural populations of *T. brucei,* as the high level of diversity would be difficult to explain by mutations alone [Cibulskis, 1988]. In this study the assortment of alleles into different genotypes has been observed many times, indicating that some sexual recombination does occur. However the key question regarding sexual recombination is the level at which it occurs in the field.

To address the issue of clonality in *T. brucei* populations a marker system which would uniquely identify specific clones is required. Previous analysis of populations using markers with limited variability could possibly have grouped different clones together and so incorrectly given the impression of clonality. In this chapter unequivocal allele

identification for one minisatellite locus (*MS42*) combined with allele size measurements for another two minisatellites were used to identify stocks, with most multilocus genotypes being unique. It is clear from the data presented here (Figure 7.5), however, that one or two genotypes in each population are over-represented, (especially pronounced in the populations isolated from humans- Luangwa, Nyanza and Busoga (HSR)), which would initially suggest a clonal population structure as proposed by [Tibayrenc *et al.*, 1990]. Indeed these data would fit Tibayrenc's criteria for clonality (see Chapter 1, Table 1.3), in that there is an excess of heterozygotes, deviation from HW expectations and linkage disequilibrium. However, the common genotypes are not detected over a wide geographical area and the Zambian (Luangwa) population does not share the genotypes which are common to the other populations but has a different common genotype. Further analysis of these data, examining only the ETs (for both HW and I_A), indicated that the underlying population may be in agreement with HW expectations and the population structure may be sexual, which was initially obscured by the over-representation of one or two genotypes. However it must be emphasised that by removing the common genotypes the sample size for each population was considerably reduced, thus decreasing the ability to detect deviation from HW equilibrium. In order to verify an epidemic population structure a far larger number of samples would be required; unfortunately only a small number of samples collected at the same time and from the same place were available for analysis here. The results presented in Chapter 6 in which uncloned mixed isolates were analysed, gives a sample of the larger population and indicates that not only are there more alleles in the population than in the genotypes presented in this chapter, but also that different combinations of alleles exist, suggesting that if a large number of isolates were examined more recombinant genotypes would be identified. It is also possible that an epidemic population structure was observed because the samples were collected during an outbreak of the disease in humans or cattle and that, by sampling from other hosts (or the vector) during periods of endemicity, a different population structure would have been apparent.

Population sub-structuring due to geography -comparison of human isolates. In order to examine the possibility of population sub-structuring due to geography, a comparison can be made between the human isolates from different geographical regions, Luangwa, (Zambian), Nyanza (Kenyan) and Busoga (Ugandan). Examining the allele distribution for each of these populations, for the minisatellite *MS42*, (Figure 7.2), it is immediately clear that the Luangwa (Zambian) stocks do not share any *MS42* alleles with those of any of the other populations. This is highlighted in Table 7.16,

which indicated the number of *MS42* alleles shared between populations. For the *292* locus, two alleles have been found to be shared between the Luangwa stocks and those isolated from Nyanza and Busoga, and for the *CRAM* marker, one allele has been identified that is common in the Luangwa population and has been observed in one Nyanza stock (see Tables 7.17 and 7.18). Interestingly the three alleles shared between Nyanza and Busoga were from the few samples in these populations isolated from non-human isolates, *i.e.* they were isolated from tsetse flies or cattle. Therefore a direct comparison of human isolates from Luangwa with human isolates from Nyanza and Busoga does not identify any alleles in common. These data suggests that the human isolates from the Zambian population are distinct from the human isolates from other populations, supporting the findings of Hide *et al* in 1994. However it must be remembered that the Luangwa samples were isolated at a different time (1981-3) from those of the other populations and so the genetic distance observed here may be due to time differences rather than geographical barriers. However, this is unlikely as the populations from the shores of Lake Victoria (Lugala, Nyanza and Busoga) show remarkable similarity over time, with the same common alleles being present in all populations between 1969 and 1990.

The samples from Luangwa in Zambia were isolated from two villages (Kasyasya and Chilbale) in Luangwa approximately 40 Km apart. Analysis of the genotypes detected (Table 7.21) has revealed that the samples from Kasyasya appear to be clonal with only two highly related genotypes being present in the 15 samples from this village, the genotypes differing by two repeat units in one *MS42* allele (see Chapter 8). The four samples isolated from the Chilbale village, however, are all different. Although from a small sample set, these results indicate that differences in genotypes can be detected between villages which are only a short distance apart, suggesting that sub-structuring can occur over very small distances. A much larger collection of isolates exists from this region [Godfrey *et al.*, 1990] and could be analysed to substantiate fully these findings.

Village	Genotype of Luangwa HSR samples			
	MS42	*CRAM*	*292*	Multilocus genotype
Kasyasya	32/26b	68/45	55/32	1
Kasyasya	30f/26b	68/45	55/32	5
Chilbale	32/26b	59/59	32/32	2
Chilbale	45/32	53/45	55/32	3
Chilbale	47/42	53/45	55/51	4
Chilbale	47/42	47/45	32/32	6

Table 7.21. Genotypes of HSR Luangwa samples.

Population sub-structuring due to geography -comparison of tsetse isolates. To examine further the possibility of population sub-structuring due to geography, two populations, which were both isolated from tsetse flies at the same time (1969-70) but from different locations, Kiboko (Kenya) and Lugala (Uganda), approximately 400Km apart, were compared. From Figure 7.2 it is clear that there are only 2 shared alleles, which constitute the common genotype, out of 34 different *MS42* alleles detected in the two populations. Six *292* and six *CRAM* alleles are also shared between the two populations. It is clear from these data that the two populations are genetically different, each containing several private alleles (for a summary see Table 7.22).

	Markers		
	MS42	*CRAM*	*292*
No. of different alleles (total no. of alleles) -Lugala	12 (23)	9 (26)	10 (25)
No. of different alleles (total no. of alleles) - Kiboko	22 (64)	15 (62)	18 (68)
Shared alleles between Lugala and Kiboko	2	6	6
Private alleles - Lugala	2	0	2
Private alleles - Kiboko	15	6	11

Table 7.22. Comparison between Lugala and Kiboko populations. The total number of alleles in each population are given in brackets.

The differences between the 5 populations in this study are reflected in the F_{ST} values, which suggests that there is an appreciable level of genetic difference between populations. The data presented here, from both tsetse and human isolates, suggests that because significant population sub-structuring exists due to geographical barriers, which has previously been suggested by [Cibulskis, 1992], previous studies in which samples from widely different geographical areas were combined [Tibayrenc *et al.*, 1990] would lead to flawed conclusions.

Population sub-structuring due to host specificity and* T. b. rhodesiense*-specific markers. The possibility of population sub-structuring due to host specificity has rarely been considered in the analysis of *T. brucei* populations, with the exception of Cibulskis *et al.*, in 1992 who provided evidence for a difference in zymodemes isolated from different hosts, suggesting population sub-structuring. In this present study one population was ideally suited for the analysis of sub-structuring due to host specificities. Samples from Busoga in Uganda were collected from both humans and cattle at the same time point, and many of those samples were analysed for human serum sensitivity [Hide *et*

al., 1994]. Examining the difference between HSR and HSS stocks, it is clear that no *MS42* alleles are shared between them (Figure 7.2), but the less discriminatory markers *CRAM* and *292* do show common alleles between the populations, although there are substantial allele frequency differences between the two groups of stocks (Figures 7.3 and 7.4). Comparing the multilocus genotypes, only 4 genotypes have been identified in 25 HSR stocks compared to 8 genotypes in 13 HSS or stocks where human serum resistance was not tested, suggesting that the HSR stocks are far more homogeneous than the HSS. Interestingly the 4 different HSR genotypes are all highly related (see Table 7.23), in that they share the same basic complement of alleles. The relationship between these genotypes suggests that they could have originated by selfing, for example, genotype 7 (heterozygous for alleles at *MS42* and *292*) could have self-fertilised to produce genotype 20, which is homozygous at both loci. This could also be the case for the derivation of genotype 21 and 16 as the self-fertilization of genotype 7 could produce genotype 21, which is homozygous for *292*, and genotype 16, which is homozygous for *CRAM* (the *MS42* allele 35d is a single repeat variant of allele 35a, described in full in Chapter 8). It would appear that for the Busoga focus at least, HSR is associated with one genotype and its self-fertilization products, implying little or no recombination between the HSR and HSS stocks; this is further supported by the high genetic distance value (0.471) between these groups of isolates. This view supports the findings of [Hide *et al.*, 1994] who examined many of the same stocks from the Busoga focus. The fact that several self-fertilization products have been detected in the HSR stocks may not be surprising as during an epidemic (from which these samples were collected) the predominant genotype may only come in contact with trypanosomes of the same genotype, and so any mating would essentially constitute self-fertilization (a phenomenon which has been observed in laboratory crosses). The results presented here suggest that HSR stocks may be clonal in origin, with stocks from just one lineage being able to infect humans, perhaps suggesting that human infectivity is a fairly recently acquired attribute which has not as yet, or cannot, spread through the rest of the *T. brucei* population by sexual recombination.

The majority of the samples were isolated from Busoga between 1988 and 1990, with the exception of five samples which were isolated from humans at earlier time points; 1959, 1960, 1976, 1981 and 1982. Two multilocus genotypes were identified from these samples (genotype 7 and 16, see Table 7.23), neither of which were detected in the large collection of samples from 1988-90. The relationship between the samples suggests that genotype 7, which has been found in all populations except the Zambian population, is the ancestral HSR genotype from which the other genotypes were derived, giving rise to genotypes 20

and 21 (probably by self-fertilization) which are the only HSR genotypes detected in the 1988-1990 Busoga samples.

Year of isolation	Genotype of Busoga HSR samples			
	MS42	*CRAM*	*292*	Multilocus genotype
1959-82	35a/34	61/55	76/44	7
1976	35d/34	61/61	76/44	16
1989	35a/34	61/55	76/76	21
1988-90	34/34	61/55	76/76	20

Table 7.23. Genotypes of HSR Busoga samples.

It is interesting to note that the possible epidemic population structure due to the over-representation of common genotypes from the five populations (as depicted in Figure 7.5), may be related to the ability of the different trypanosomes to infect humans, in that each of the common multilocus genotypes (1, 7, and 21) is associated with human serum resistance. Evidence from the Busoga population suggests that human infective stocks may be a sub-population, defined by a unique set of alleles, with limited genetic exchange with human serum sensitive stocks. One of the multilocus genotypes associated with human infectivity from the Busoga population (genotype 7) has also been detected in the Kiboko samples, which predicts that the Kiboko population may contain trypanosomes which are human infective. This hypothesis could be easily tested using the blood incubation infectivity test (BIIT) to identify any human infective trypanosomes in the Kiboko population.

Analysis of the Busoga HSR samples has revealed one lineage which is associated with human infectivity and could be considered a *T. b. rhodesiense*-specific marker. However analysis of the Nyanza population does not reveal such a tightly defined set of genotypes associated with HSR (see Table 7.24), although the predominant genotype (genotype 7) is common to both populations, and the Nyanza genotype 8 and the Busoga genotype 16 are highly related, with a single *MS42* repeat unit change between them (see Chapter 8). The presence of other alleles in genotypes, 10, 11, 12 and 14 suggests that a pure clonal lineage of HSR genotypes does not exist in this population and that some measure of cross-fertilization has occurred, although there is still a restricted repertoire of alleles and the genotypes are related (for example, it is possible that the Nyanza genotype 11 is the product of self-fertilization of genotype 10).

Genotype of Nyanza HSR samples			
MS42	*CRAM*	*292*	**Multilocus genotype**
35a/34	61/55	76/44	7
35a/34	61/61	76/44	8 *
34/33b	61/36	76/36	10
34/33b	61/36	36/36	11
33b/33b	64/36	36/35	12
35a/33b	61/55	76/44	13
33b/21	36/36	44/36	14

Table 7.24. Genotypes of HSR Nyanza samples. Genotype 8 is highly related to the Busoga HSR genotype 16, having one *MS42* repeat unit difference (see Chapter 8).

Interestingly, the human infective genotype isolated from the Zambian population is very different from the human infective genotype isolated from Busoga focus, with no alleles in common. This indicates that *T. b. rhodesiense* samples from different areas may not be closely related, as their subspecies classification would suggest. Indeed, HSR Busoga (*T. b. rhodesiense*) samples appear to be more closely related to (although still distinct from) HSS Busoga samples (*T. b. brucei)* than to the Zambian *T. b. rhodesiense* samples. Further evidence from the similarity of MVR maps of *MS42* alleles which supports this conclusion is provided in the following chapter.

Most studies of *T. brucei* populations [Stevens and Welburn, 1993; Truc and Tibayrenc, 1993; Mathieu-Daude and Tibayrenc, 1994] have examined samples which were derived largely from human and cattle infections, sampled during an epidemic, and have concluded that *T. brucei* populations are homogeneous with little or no sexual recombination. Based on the data presented here, it is likely that a wider range of genotypes would have been observed if the samples were collected from other mammalian hosts or vectors from an endemic area. Indeed the apparent clonality and genetic isolation of *T. b. rhodesiense* from *T. b. brucei* would mean that previously analysed populations would be a mixture of two sub-species and so would not be in HW equilibrium and would show linkage disequilibrium.

Cattle are a reservoir for human infective trypanosomes. Having identified multilocus genotypes which are associated with human infectivity in the Busoga area of Uganda, examination of the host species they were isolated from (Figure 7.6) revealed that these genotypes were present in the cattle population, indicating that cattle could be a reservoir for human infective trypanosomes, supporting the view of Hide [Hide *et al.*, 1996]. These data also highlight the importance of sub-structuring in the trypanosomes within the cattle host.

Conclusions

The main findings from this study indicate that *T. brucei* populations have an epidemic population structure, whereby a sexual population structure has been obscured by the over-representation of mainly one genotype, which is associated with human infectivity. This is likely to have been accentuated by the non-random sampling programme focused on epidemics of disease in humans. For the Busoga focus the human infective trypanosomes appear to be clonal in origin having been derived from genotype 7, and so for the Busoga population a *T. b. rhodesiense* marker has been identified. The main human infective genotype in the Nyanza population can also be identified using this marker system (and is highly related to the Busoga human infective genotype), although one or two human infective samples may have a variant genotype, suggesting sexual recombination has occurred. Human infective stocks from the Zambian population, however, are unrelated to the human infective samples from Nyanza or Busoga and may have originated separately from the human infective samples of the Busoga/Nyanza focus. Indeed the nature of the disease they cause appears to be different in the two foci, with a very acute disease in the Busoga/Nyanza focus and a low virulence disease in Zambia, with a high incidence of asymptomatic carriers [Hide, 1999].

The results from this chapter also suggest that considerable population sub-structuring can be observed due to geography, with little allele sharing between different populations. Analysis of stocks isolated from different hosts also suggests that sub-structuring exists due to host specificities with human infective trypanosomes having a more restricted repertoire of alleles and genotypes. Taken together, these data suggest that previous studies based on samples from different areas, hosts and time points may give misleading results. For example, deviation from HW could arise by mixing samples from different populations. Although it was tempting to combine the data obtained from the 116 samples examined in this study to measure deviation from HW, the results from the sub-population analysis indicates that the basis for doing this would be seriously flawed.

The analysis presented in this chapter outlines a new approach to the analysis of *T. brucei* populations, by using hypervariable minisatellite markers which can distinguish between the majority of stocks. This approach is particularity useful for tracing genotypes through the population and to examine the theory of clonality. However the question arises as to how appropriate the statistical tests commonly employed are to the data generated by hypervariable markers. A direct comparison of isoenzyme and single-locus minisatellite markers for the analysis of toad populations using traditional statistical tests found a high degree of concordance with the different marker systems and suggests that minisatellites

may be highly informative for population-level analysis addressing questions concerning population sub-division [Scribner *et al.*, 1994]. However in order to derive meaningful results from these markers a large number of samples must be analysed. The sample sizes in this study were too small to yield unequivocal conclusions, although the general trends they reveal, which support the findings of [Cibulskis, 1992; Maynard-Smith *et al.*, 1993; Hide *et al.*, 1994], may be valid.

Chapter 8

Development of a digital DNA typing system to analyse *T. brucei* populations

Introduction

In Chapter 7 hypervariable minisatellite markers were used to analyse a series of *T. brucei* population samples from East Africa, in order to address the issues of population substructuring (due to geography and host species), epidemic population structures and clonality. However, despite the obvious advantages of minisatellite markers over traditional typing systems, there are some technical problems involved in the use of these markers. For example, minisatellite alleles tend to vary in a quasicontinuous fashion making unambiguous allele identification difficult. Thus, when size differences between alleles are small, allele length estimates are not very accurate and variation between gel runs can lead to a failure in identifying matching samples, or the false matching of different samples, thus weakening the statistical power of population databases based on allele length. Also alleles of the same size may differ in repeat structure and so size may not indicate relatedness. The development of minisatellite variant repeat (MVR) mapping by the polymerase chain reaction (MVR-PCR) as a digital approach to DNA typing [Jeffreys *et al.*, 1991] has overcome many of the drawbacks of VNTR length analysis. The system assays the dispersion patterns of MVRs within minisatellite arrays producing an easily interpretable code for each allele. The simple and rapid MVR-PCR technique increases the level of information about each allele so that the ability to define differences is increased. Furthermore the information is generated in a digital format ideal for computer based analysis and for the production of population databases.

MVR mapping has been applied successfully to a number of minisatellites in humans [Jeffreys *et al.*, 1991; Neil and Jeffreys, 1993; Armour *et al.*, 1993; Buard and Vergnaud, 1994], mice [Bois *et al.*, 1998] and the parasite, *Plasmodium falciparum* [Arnot *et al.*, 1993]. The study of variant repeats within minisatellite arrays has proved useful for two main areas of research; individual/stock identification for the analysis of populations and

analysis of the mutation processes involved in the generation of new length alleles at those loci which are highly unstable. For example, MVR-PCR analysis of the human minisatellite, Ms205, has revealed differences in allele diversity between African and non-African populations with a restricted set of allele groups being identified in the non-African populations [Armour *et al.*, 1996]. The non-African alleles appeared to be a subset of the much greater diversity found in the African population, supporting the 'out of Africa' theory of human evolution. Allele differences between British African-Caribbean and Caucasian populations have also been investigated using the MVR-PCR technique on the minisatellite in the involucrin gene, indicting how this gene evolved in human populations and revealing a potential race specific marker, which may have forensic applications [Urquhart and Gill, 1993]. Another human minisatellite found on the non-pseudoautosomal segment of the Y chromosome which is therefore male specific, has been identified by Jobling and co-workers, and internal mapping of this locus by MVR-PCR has been used to trace male lineages [Jobling *et al.*, 1998; Bouzekri *et al.*, 1998].

MVR-PCR has been applied to the study of the mutation processes involved in the generation of new length alleles for three human minisatellites, (Ms32, Ms31 and Ms205) revealing that variation in allelic structures has a marked polarity with mutations occurring at or near one end of the repeat array, although polarity for a fourth minisatellite has not been confirmed [Buard and Vergnaud, 1994] and has not been observed for another human minisatellite [Andreassen and Olaisen, 1998]. Evidence from the analysis of these mutations has also suggested that inter-allelic as well as intra-allelic recombination may play an important role in the creation of new length alleles [Jeffreys *et al.*, 1991; Armour *et al.*, 1993; Neil and Jeffreys, 1993; Buard and Vergnaud, 1994], with the frequency of the mutation process apparently being modulated by *cis* acting elements [Monckton *et al.*, 1994]. The discovery of minisatellite mutations which are associated with human disease, has renewed interest in minisatellite mutation processes [Buard and Jeffreys, 1997].

The MVR-PCR method is a powerful tool in the analysis of minisatellite loci, but not all minisatellites are amenable to analysis by this technique. For the methodology to be successfully applied to a locus, the minisatellite structure must meet a number of criteria. At least one variant repeat must be common among alleles and other variant repeats must be rare. As MVR specific primers are designed to the common variant, the presence of further polymorphic sites within the repeat unit could prevent the primers from amplifying from every repeat. The repeat unit size must be uniform throughout the locus (for diploid mapping) and for single allele mapping, alleles must be small enough to amplify in their

entirety for gel purification or alternatively flanking polymorphisms must exist which can be exploited to generate allele specific primers for the typing of heterozygous samples [Monckton *et al.*, 1993].

This chapter describes the application of the MVR-PCR technique to the *T. brucei* minisatellite, *MS42*, with the aim of obtaining large numbers of single allele maps from the five different populations described in Chapter 7. Using this method, a greater number of alleles were identified at this locus than by size estimates alone. The allelic structures identified also generated cladistic information as well as providing evidence for both intra- and inter-allelic recombination in the generation of new length alleles at this highly variable locus. In order to relate allelic structures to the local haplotypic context, several polymorphisms identified in the DNA flanking the *MS42* repeats were examined. This information could aid the grouping of similar alleles and also indicate the extent of recombination events involving the exchange of flanking markers.

By studying the five different populations of *T. brucei* (see Chapter 7) in this way, a number of questions regarding the population structure and mutation processes can be addressed. Can *T. brucei rhodesiense* stocks be distinguished from the local *T. brucei brucei* population as suggested by Hide *et al.*, (1994)? Can MVR maps provide a rational method for binning or grouping alleles to provide data for population analysis? Did human infectivity arise from the *T. b. brucei* population and did this occur independently in several different regions? How frequently do new length alleles arise, *i.e.* are the repeats in *MS42* mitotically stable or do they represent a recombination hot spot? Do new length alleles arise by intra-allelic events during mitosis or by inter-allelic crossing over during meiosis?

Results and Discussion

The principles of MVR-PCR of MS42. Sequence analysis of the *T. brucei* minisatellite *MS42* (Chapter 3) and partial sequence of several alleles [see following sections and Barrett *et al.*, 1997] shows the presence of variant repeats, the most common being an A/G transition which does not affect the predicted amino acid sequence of the gene product. The presence of one common variant repeat, the fact that the repeat unit size is constant (42bp) and that all alleles are small and so can be easily purified suggests that the minisatellite *MS42* satisfies the strict criteria necessary for the development of a digital DNA typing system.

The MVR-PCR method accesses the dispersion pattern of variant repeats within the repeat array of a minisatellite. In the case of *MS42*, two repeat types exist, which differ by a single base substitution, (A/G transition) at the beginning of the repeat unit; thus, every repeat can have either an A or G base at the first position of the repeat unit. The two classes of repeat unit were designated **a**-type and **g**-type repeats (Figure 8.1). To access this variation two PCR primers were generated which are specific for each repeat type. Using these MVR specific primers and a specific primer located in the flanking DNA, a ladder of PCR products corresponding to the position of each **a**-type and each **g**-type repeat could be generated (Figure 8.1C). Applied to single isolated alleles, a binary code of the distribution of **a**- and **g**-type repeats within the repeat array could be constructed, while if applied to total genomic DNA, (by including both alleles), a ternary code could be derived from the superimposed maps of the two alleles.

Unlike the human minisatellites Ms32 and Ms31 [Jeffreys *et al.*, 1991; Neil and Jeffreys, 1993], which have many alleles >10kb long, all the alleles of the *MS42* variable region are small (less than 2.4Kb), and so every allele can be completely and readily amplified by PCR. This allows each allele to be analysed in full, in a similar fashion to the human minisatellite MS205 [Armour *et al.*, 1993; Armour *et al.*, 1996].

MVR-PCR methodology. Initially the variable region of the *MS42* locus in each isolate was amplified to visible levels on an ethidium stained gel, using the universal flanking primers *MS42*-W and *MS42*-F (as in Chapter 3; Figure 3.2; details of PCR reactions including oligonucleotide sequences are given in Materials and Methods). Each PCR-generated band, corresponding to the different alleles in the sample, was then excised from the agarose gel and used as a template for the subsequent PCR reactions using MVR specific primers.

The two different MVR specific primers prime off either **a**- or **g**-type repeats in conjunction with the universal flanking primer to generate two sets of products, one recording the products of repeats starting with **a** and the other repeat starting with **g**. Because the MVR specific primers can prime internally from the initial PCR products this can lead to progressive shortening of the PCR products with each cycle of amplification. To prevent this occurring each MVR specific primer contains a 20 nucleotide extension, TAG. Thus the MVR-tagged primers provide the specificity of the PCR reactions and are used at low concentrations while the TAG primer is used at higher concentrations to amplify these products. In this way the MVR detection is separated from the subsequent amplification (Figure 8.1). After a limited number of cycles of amplification the products were electrophoresed through a 40 cm 1% agarose gel to obtain the maximum resolution

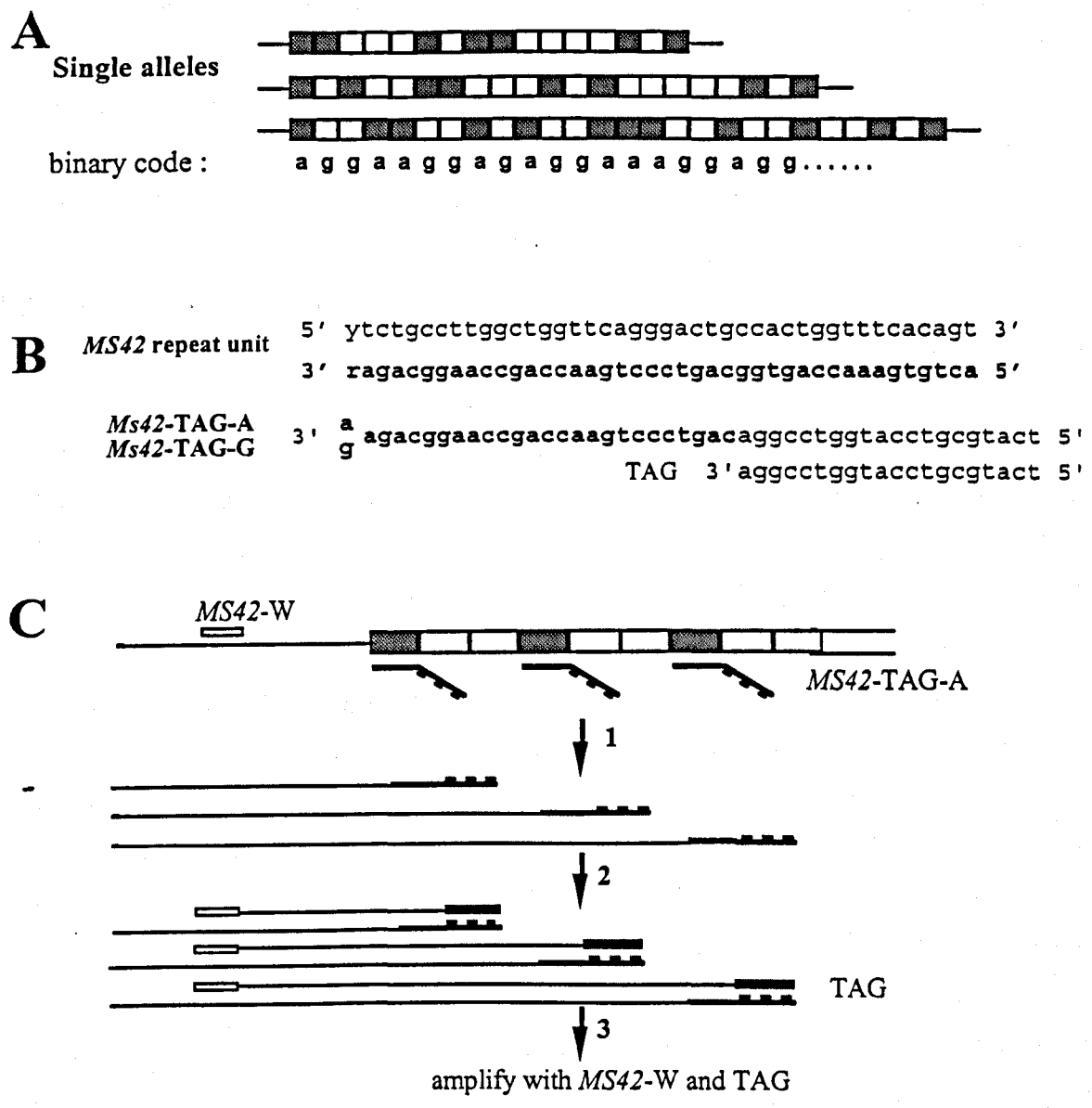


Figure 8.1 The principles of minisatellite repeat coding.
(A) Principle of digital coding. Minisatellite alleles consisting of interspersed arrays of two variant repeat units termed **a**-type (shaded boxes) and **g**-type (open boxes). Individual alleles can be encoded as a binary code extending from the first repeat unit.
(B) Ms42 repeat unit and MVR-PCR primers. Both strands of the consensus 42bp repeat unit sequence of *T. brucei* showing the common polymorphic site, with y indicating C or T and r indicating A or G. The coding strand is in bold. Primers *MS42*-TAG-A and *MS42*-TAG-G are variant repeat specific oligonucleotides terminating at the polymorphic site. Each primer consists of 20 nucleotides of the minisatellite repeat unit (bold) and 20 nucleotides of a non-minisatellite extension identical to the TAG primer.
(C) The principle of MVR-PCR. Illustrated for a single allele amplified using *MS42*-TAG-A. 1. At low concentrations of *MS42*-TAG-A primer, the primer will anneal to approximately one **a**-type repeat unit per target minisatellite molecule and extend into the flanking DNA. 2. Primer *MS42*-W then primes from the flanking DNA, creating a sequence complementary to the TAG sequence. 3. The newly synthesised fragments terminate in W and the complement of TAG. They can now be amplified using high concentrations of *MS42*-W and *MS42*-TAG primers, to create a set of PCR products extending from the flanking *MS42*-W site to each **a**-type repeat unit. By using primer *MS42*-TAG-G in place of *MS42*-TAG-A a complementary set of products terminating at each **g**-type repeat unit can be generated. This figure was adapted from Jeffreys *et al.*, 1991.

of the PCR products. Southern blotting and hybridization to an *MS42* repeat probe revealed a continuous ladder of complementary products extending for the length of the alleles. The number of repeats in each allele, estimated by allele size measurements was compared to the number of repeats indicated by MVR-PCR analysis, to confirm that the entire allele had been mapped. The largest allele detected (47 repeats) generated clearly separated products ranging in size from 319bp to 2377bp. It is possible to amplify the MVR-PCR products to visible levels on an ethidium bromide stained gel, but because smaller products can out-amplify larger ones and the PCR products can collapse [Jeffreys *et al.*, 1988] only partial maps are often obtained.

An example of the variant repeat dispersion patterns of one allele from the stock TREU 927/4 (lower allele) is presented in Figure 8.2A, with grey and red boxes representing **a** and **g** type repeat units, respectively, and arrowheads indicating the primer sites. The binary code for this allele is given below the diagram. The MVR mapping autoradiograph for this allele, illustrating the complementary ladder of PCR products generated by amplification using **a** and **g**-type repeat specific primers, is presented in Figure 8.2B, beside the deduced MVR code generated from it. Further examples of MVR maps generated from several different purified alleles are presented in Figure 8.3.

In this study alleles from the population samples previously used in Chapter 7 were analysed by MVR-PCR, generating a total of 274 MVR codes, and identifying 48 different alleles; the data are summarised in Table 8.1. A forty-ninth allele was detected in stock STIB 247, which does not belong to any of the five populations and so is excluded from subsequent analysis. The proportion of distinct alleles detected varies markedly between populations, with maximum single allele frequencies in the populations ranging from 0.145 to 0.4. The analysis of the distribution of the different alleles in the five populations is presented and discussed in Chapter 7.

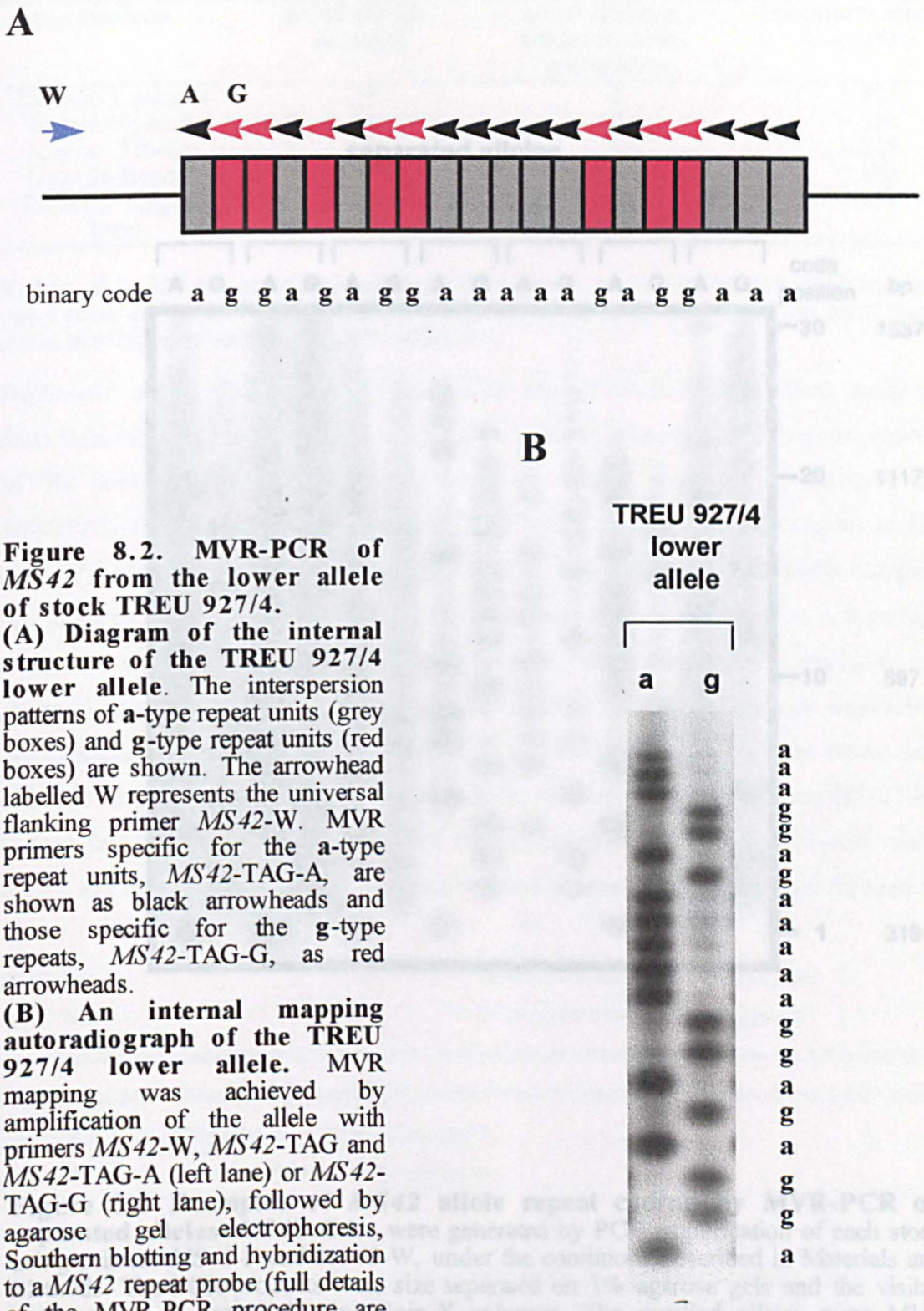


Figure 8.2. MVR-PCR of *MS42* from the lower allele of stock TREU 927/4.
(A) Diagram of the internal structure of the TREU 927/4 lower allele. The interspersion patterns of **a**-type repeat units (grey boxes) and **g**-type repeat units (red boxes) are shown. The arrowhead labelled W represents the universal flanking primer *MS42*-W. MVR primers specific for the **a**-type repeat units, *MS42*-TAG-A, are shown as black arrowheads and those specific for the **g**-type repeats, *MS42*-TAG-G, as red arrowheads.
(B) An internal mapping autoradiograph of the TREU 927/4 lower allele. MVR mapping was achieved by amplification of the allele with primers *MS42*-W, *MS42*-TAG and *MS42*-TAG-A (left lane) or *MS42*-TAG-G (right lane), followed by agarose gel electrophoresis, Southern blotting and hybridization to a *MS42* repeat probe (full details of the MVR-PCR procedure are given in Materials and Methods). The derived internal map of **a**- and **g**-type repeats is shown.

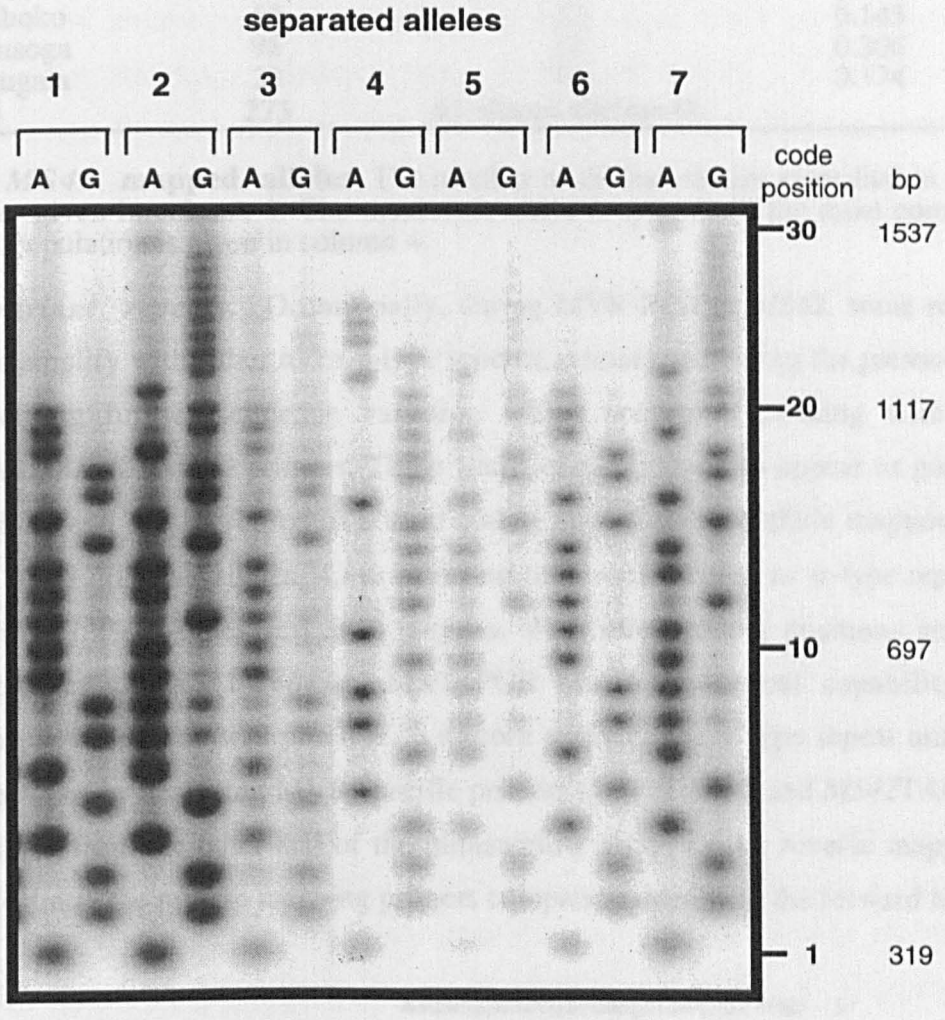


Figure 8.3 Examples of *MS42* allele repeat coding by MVR-PCR on separated alleles. *MS42* alleles were generated by PCR amplification of each stock using primers *MS42*-F and *MS42*-W, under the conditions described in Materials and Methods. The PCR products were size separated on 1% agarose gels and the visible bands were gel extracted using Spin-X columns. The purified alleles were MVR mapped as described in Materials and Methods.
Lane 1, K927cl 1B lower allele (l); lane 2, K927cl 1B upper allele (u); lane 3, K854 allele 4; lane 4, K927cl 4B l; lane 5, K927 uncloned allele 4; lane 6, K925 l; lane 7, K927cl 1B u.

populations	no. of alleles mapped	no. of distinct alleles in each population	maximum allele frequency
Zambia- Luangwa	38	6	0.421
Kenya-Nyanza	52	9	0.365
Kenya- Kiboko	62	22	0.145
Uganda-Busoga	98	12	0.306
Uganda- Lugala	23	12	0.174
Total	273	61-shared alleles=48	

Table 8.1. *MS42* mapped alleles. The number of distinct alleles identified in each population are given in column 3. The maximum allele frequency of the most common allele in each population is given in column 4.

Different variant repeats. Occasionally, during MVR-PCR of *MS42*, some repeat units failed to amplify with either **a**- or **g**-type specific primers, indicating the presence of MVRs containing further sequence variation which prevented priming with the *MS42*TAG-A or *MS42*TAG-G primers. These 'null' or **n**-type repeats appear as gaps in the MVR ladder, and can be scored as a third coding state for single allele mapping. In this analysis of 48 different alleles, 4.8% of repeat units were scored as **n**-type repeats, compared with 57.2% **a**-types and 38% **g**-types. With three coding positions and an average allele length of 31 repeats, MVR-PCR has a theoretical capability of distinguishing 3^{31} alleles. It was possible to rescore many of the **n**-type repeat units as either **a**- or **g**-types by designing MVR specific primers (*MS42*TAG-T and *MS42*TAG-C) to map the allele from the other end of the minisatellite array, *i.e.* by reverse mapping. Figure 8.4 illustrates the reverse mapping primers comparing them with the forward MVR

```
MS42-TAG-A                                    aagacggaaccgaccaagtccctgac+TAG  5'
MS42-TAG-G                                    gagacggaaccgaccaagtccctgac+TAG  5'
ragacggaaccgaccaagtccctgacggtgaccaaagtgtcaragacggaaccgaccaagtccctgacggtgaccaaagtgtca
ytctgccttggctggttcagggactgccactggtttcacagtytctgccttggctggttcagggactgccactggtttcacagt
MS42-TAG-T  5' TAG+gggactgccactggtttcacagtt
MS42-TAG-C  5' TAG+gggactgccactggtttcacagtc
```

Figure 8.4. Forward and reverse MVR-PCR primers. Both strands of two 42bp repeat units of *MS42* with the common polymorphic site (in bold) are shown. The primers *MS42*-TAG-T and *MS42*-TAG-C are variant repeat specific oligonucleotides terminating at the polymorphic site, in the opposite orientation to the forward MVR-PCR primers, *MS42*-TAG-A and *MS42*-TAG-G. Each primer consists of 26 nucleotides of minisatellite repeat unit and an extension identical to the TAG primer. PCR conditions were identical to those used for forward mapping.

specific primers, while Figure 8.5 illustrates MVR codes of alleles containing null repeats,obtained using both the forward and reverse mapping process. From the data it is clear that many **n**-type repeat units can be scored by reverse mapping as either **a**- or **g**-types, however some repeat units which were scored using forward mapping became **n**-type repeats by the reverse process and some null repeat units could not be scored by forward or reverse mapping, indicating that a number of variant sites exist in the minisatellite repeats. By combining the data for both forward and reverse mapping some, but not all, of the alleles could be rescored only in terms of **a**- and **g**-type repeats.

To investigate further the sequence variation within the repeated region of *MS42*, two *MS42* alleles were sequenced. All 26 repeats of allele 26d from the stock STIB 247 and 21 of the 34 repeats from allele 34 of stock BU89/8 were sequenced and the results are presented in Figure 8.6. Allele 26d repeats are the same as the consensus sequence containing either G or A at the polymorphic site used for MVR mapping, (which does not affect the predicted amino acid sequence) and no other variant repeats were present. The sequence of seven repeats from alleles 20 and 25a was determined and revealed no other variants repeats (data not shown). The repeat units in allele 34 are quite different from those in the other alleles, with the normal polymorphic site used in MVR-PCR, being almost homogeneous, consisting mainly of As. Adjacent to this site is another polymorphic position, with an A-G transition which affects the predicted amino acid sequence of *MS42*. Two other variant repeat units were identified which also alter the amino acid sequence (Figure 8.6). Sequence analysis of three null repeats from allele 34 (repeats 30-32) has revealed four variant positions (Figure 8.6). Two of these variants lie within the region of the MVR primers, and so are clearly responsible for preventing the *MS42*-TAG-A and *MS42*-TAG-G primers from amplifying these repeats. It is likely that **n**-type repeats identified by MVR-PCR from other alleles are different from those sequenced here, so that null repeats probably constitute a heterogeneous group of variant repeats. Similar analysis of null repeats for the human minisatellite, ms32, has revealed one major variant and a number of other different sequence variants which generate null repeats [Tamaki *et al.*, 1992]. Further sequence variations in repeat units probably occur at *MS42* which, instead of preventing *MS42*-TAG-A or *MS42*-TAG-G primers from amplifying completely, merely reduce the efficiency of the amplification reaction for certain repeat units, resulting in variations in band intensity observed on mapping autoradiographs. This phenomenon has been noted for other minisatellites including Ms205 [Armour *et al.*, 1993].

N 7 upper allele	forward	aaaaagannnggagaaganaannnaaagaggggggag
N 7 upper allele	reverse	aaaagngagggagaagngaaaaaaaagaggggggag
N 7 upper allele	complete	aaaagagagggagaagagaaaaaaaagaggggggag
L 941 upper allele	forward	aaaaaaaaaaaaaaaaaaannaanaaaaaaaaannnag
L 941 upper allele	reverse	nanaaaanananan?an?nanaaaaaaaaa?aaganng
L 941 upper allele	composite	aaaaaaaaaaaaaaaaaaaaaaagaaaaaaaaaganag
L 836 upper allele	forward	aaaaaanaagannngagaagannngaggng
L 836 upper allele	reverse	aaaaaaaaagngaggagaagngaggnggag
L 836 upper allele	complete	aaaaaaaaagagaggagaagagaggaggag
L 845 allele 4	forward	aagaagnananggggg
L 845 allele 4	reverse	aagang?agaggnggg
L 845 allele 4	composite	aagaagnagaggggggg
L 845 allele 1	forward	aagaagnananaagnaaannaagagnggggg
L 845 allele 1	reverse	aagaaggagagaagnaanaaangnggggggg
L 845 allele 1	composite	aagaaggagagaagnaaaaaaagaggggggg

Figure 8.5. Forward and reverse MVR codes. Five *MS42* alleles containing null repeats (n) were analysed by both the standard (forward) MVR mapping procedure and the reverse mapping method. Reverse mapping was achieved using primers *MS42*TAG-T and *MS42*TAG-C, under the same conditions as described for forward mapping in Materials and Methods. Primer sequences are given in Figure 8.4. Note that the mapping of two alleles was completed by combining the results from both codes. Alleles from the uncloned sample L 845 which contained four alleles were numbered 1-4. Ambiguous positions were indicated (?).

repeat no.	MVR code	allele 26d sequence
consensus	3′	ragacggaaccgaccaagtccctgacggtgaccaaagtgtca 5′
1	a	**a**
2	g	**g**
3	a	**a**
4	a	**a**
5	g	**g**
6	g	**g**
7	a	**a**
8	g	**g**
9	a	**a**
10	g	**g**
11	g	**g**
12	a	**a**
13	a	**a**
14	g	**g**
15	a	**a**
16	a	**a**
17	a	**a**
18	g	**g**
19	g	**g**
20	g	**g**
21	a	**a**
22	g	**g**
23	g	**g**
24	a	**a**
25	a	**a**
26	a	**a**

repeat no.	MVR code	allele 34 sequence
consensus	3′	ragacggaaccgaccaagtccctgacggtgaccaaagtgtca 5′
1	a	**a** **g**[1]
2	a	**a** **t** **g**[1]
3	a	**a** **g**[1]
4	a	**a** **g**[1]
5	a	**a** **g**[1]
6	a	**a** **g**[1]
7	a	**a** **t** **g**[1]
8	a	**a** **g**[1]
9	a	**a** **t** **g**[1]
10	a	**a** **t**[2]
11	a	**a** **g**[1]
23	a	**a** **g**[1]
24	a	**a** **g**[1]
25	a	**a** **g**[1]
26	a	**a** **g**[1]
27	a	**a** **t** **g**[1]
28	a	**a** **t**[2]
29	a	**a** **t**[2]
30	n	**a** **g** **t**
31	n	**a** **g** **t**
32	n	**a** **g** **t**[3]

AA sequence of the *MS42* repeat unit	T	V	K	P	V	A	V	P	E	P	A	K	A	E
variant[1]	**A**													
variant[2]	**I**													
variant[3]						**S**								

Figure 8.6. Sequence of *MS42* repeats of alleles 26d and 34. The consensus sequence of *MS42* repeat unit is given, with the variant repeat used in MVR mapping in bold and the region of the forward MVR primers underlined. The complete sequence of 26 repeats for allele 26d is shown. Positions of variation only are indicated in bold. Sequence of 21 of the 34 repeats of allele 34 is given with positions of variation shown in bold. The MVR codes for each repeat unit is indicated. Note repeats 30-32 of allele 34 which are null repeats by MVR mapping, contain variant positions in the region of the forward MVR specific primers. Three variant positions [1-3] alter the predicted amino acid of the repeat unit. The predicted amino acid (AA) sequence of the consensus *MS42* repeat unit is shown with the three possible variants in bold. Sequencing of each allele was performed in duplicate directly from the PCR-amplified allele.

MVR-PCR of STIB 386 alleles proved to be problematic probably due to a high number of variant repeat types present in the array. Sequencing of some of the repeats of this West African *T. b. rhodesiense* stock revealed some repeat variant positions in common with allele 34, suggesting that these alleles may be related (Figure 8.7).

consensus repeat unit		actgtgaaaccagtggcagtccctgaaccagccaaggcagar
allele 34	variant positions	gt tt g g r
386	variant positions	gt t t r
T. evansi	variant positions	t t r

Figure 8.7. Variant repeats. *MS42* repeat unit consensus sequence is shown. Positions which vary from this in allele 34, STIB 386 alleles, and *T. evansi* stock 2187 are indicated in bold. r, either a or g at that position.

Null type repeats are restricted to a limited number of alleles, with 2-9 **n**-type repeats occurring in each of 11 (out of 48) different alleles (see Table 8.2), indicating a non-random distribution of **n**-type repeats. The percentage of null type repeats within an allele can range from 0% to 27% as illustrated by the analysis presented in Figure 8.8.

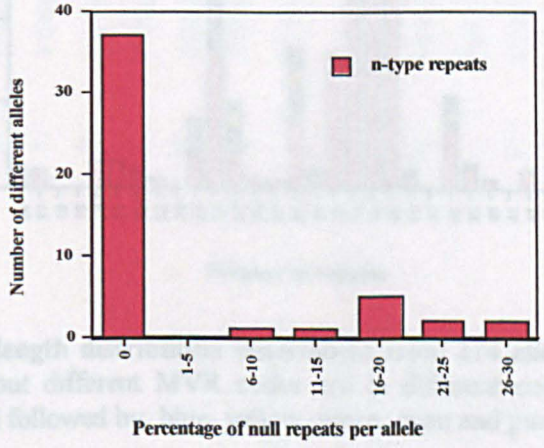


Figure 8.8. Variation in the proportion of null repeats within *MS42* alleles.

MVR maps of the 48 different alleles identified can be grouped in terms of regions of homology, presumably indicating recent common ancestry (see following sections). The distribution of null repeats within these groups is again clustered, with only four homology groups (groups 5, 6, 13 and 16) containing alleles with null repeats (see Figure 8.13). It is likely, given the close relationship of alleles within groups, that the variants causing null repeats in one allele are the same variants in most null repeats within each allele within the group, although the variants are probably not different between groups.

Allelic variation demonstrated by MVR mapping. A total of 48 different alleles,

(containing between 16 and 47 repeats with a mean of 31) were identified from the 274 alleles mapped by MVR-PCR and the data are presented in Table 8.2. Compared to band size estimates, MVR-PCR doubled the number of different alleles detected (assuming 100% accuracy of band size measurements). This is further illustrated by the histogram in Figure 8.9, which shows that the same sized alleles can have different and unrelated allele maps. For example, there are six different alleles with an allele length of 30 repeat units (see Table 8.2, alleles 30a-30f, and Figure 8.9). It is clear from the data in Figure 8.9 that three alleles occurred with high frequency, having been sampled 37, 53 and 59 times while twenty four alleles were sampled only once.

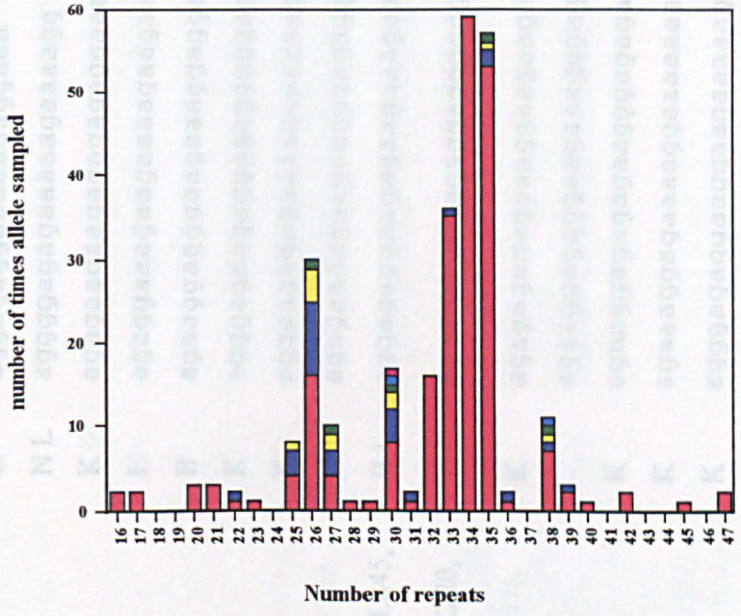


Figure 8.9. *MS42* allele length distribution determined from 274 alleles. Alleles with the same number of repeat units, but different MVR codes are in different colours, the most frequently sampled allele being in red followed by, blue, yellow, green, cyan and purple.

Homozygotes. Because some alleles are the same size but have different internal structures, it is possible that stocks which appear to be homozygous by allele length are in fact heterozygous (pseudo-homozygotes). MVR-PCR of these pseudo-homozygotes would generate a diploid MVR map (see following section), where one allele map is superimposed on the other. In this analysis no pseudo-homozygotes were identified and eight out of a total of 132 stocks were identified as homozygotes on the basis of their unambiguous single allele maps when analysed by MVR-PCR. To confirm that these homozygotes were not heterozygotes in which one allele has failed to amplify at all due to sequence variation in the primer region, each stock was amplified using a different combination of *MS42* primers but still only produced one band (data not shown). However to prove conclusively that these stocks were genuine homozygotes, Southern blots of

Table 8.2

Allele	Isolates with each allele	Pop	MVR map
16	L845, L944	L	aagaagnanangggggg
17	K1027cl, K851un	K	agaaggaggggggaaag
20	K927cl1B, K854un, K925un	K	aggagaggaaaaagaggaaa
21	N149, L845un, L929	N L	aggggagagaaaaagaaaagg
22a	K994un	K	agagaagaagagagaggaaa
22b	L934	L	agaggaaagaagaaaagaagaa
23	BM32	B	agaaggagggaagaaaggaggaa
25a	K927cl4B, K1337cl, K854un, K997un	K	agggagaagaggggaggggagaaa
25b	K997cl, K1009cl, K978un	K	agaaagagagaaaaaaaaaagaga
25c	K258cl	K	agagaaaagaagaggggagagggga
26a	Bpapol60, BM31, BM42, Bpapol 33, BB23, BS38, L845, L934, L832	B L	aggagaggaagagaagaaagaagagg
26b	Z199, Z273, Z210, Z269, Z90, Z274, Z222, Z208, Z220, Z203, Z221, Z231, Z244, Z267, Z212, Z270	Z	agaaggggagagaggagggggggggg
26c	K927cl4B, K981un, K854un, K978un	K	agagagagagaaaggaagaagaggaa
26d	STIB 247		agaaggagaggaagaaagggaggaaa
27a	K994un	K	agnnggagagagaagggagagaaggaa
27b	K925cl, K852un	K	agaaaggagaaaaggaaaaaaaagaga
27c	K984, K854cl, K978cl, K926un	K	agggagaggaaggaagaaaaagggaaa
27d	N18, L834, L941	N L	aggagaggaagagaagaaaagaagagg
28	N7	N	aggagaggaagagaagagggagaagagg
29	K925cl	K	agaaagagagaagagaaaggaagagagaa

Table 8.2 continued

Allele	Isolates with each allele	Pop	MVR map
30a	K984cl, K258cl, K854cl, K978cl, K1027cl, K926un, K851un, K997un	K	aggagaaagagaagaggaggaggggggga
30b	K852un, K981un	K	agaaagagagaagagaaaggaagagagaaa
30c	K927cl1B	K	aggagaaaaagaagaggaggaggggggga
30d	L791, L836, L933, BM32	L B	aaaaaanaagannngagaagannngaggng
30e	N2340	N	aggagaggaagagaaggagaaaagaagagg
30f	Z222	Z	aagaaagagagaaggaaaaggaaaaaagga
31a	L944	L	aagaagnanagaagnaaannaananngggggg
31b	845un	L	aagaagnananaagnaaannaagagnggggg
32	Z210, Z269, Z90, Z274, Z194, Z208, Z199, Z220, Z203, Z221, Z231, Z244, Z267, Z212, Z270, Z273	Z	aagaaagagagaaggaaaaggaaaaaaggaga
33a	K981un	K	agaaaggggaaaggaaaaaaaaaagaaagaga
33b	N111, N112, N115, N502h, N102, N149, N148h, N18, BUgC90, BUgE90, BI147, Bmag18, Bmag40, BM3, BS28, BB76h, BB135h, BI155, BM12h, BS14h, BM85, BS38, BB25, BT168*, L791, L933, Bfly48h, BM80, L836, L832	N B L	agaggaaaaaaaaagaggagaagaaaagaagaa
34	N97, N96, N156, N95, N94, N106, N116, N2340, N110, N98, N111, N112, N115, N113, N120, N605, N609, N105, N103, BEA174, BEA3, BUTAR3, BUTAR4, BEA2274, BUgC90*, BUgA90*, BUgE90, BUg89/8h, BUgL, BUgI, BUgK, BUgJ, BURI*, Bpapol103, BU89/2, BUgM, BMAP, B3194, B3196, B3200, B3202, B3203, B3205, BM66, B3206, BM42, BM31, K1008cl, K851cl, K975cl, K982cl, K853cl, K852cl, K936un, K869un, K994cl, L844, BEO	N B K L	aaaaaaaaaaaaaannaaanaaaaaaaaannnag

Table 8.2 continued

35a	N96, N97, N156, N95, N94, N106, N116, N2340, N110, N98, N113, N120, N605, N609, N102, N105, N103, BEA174, BEA3, BUTAR3, BUTAR4, BUgC90, BUgE90, BUgL, BUgA90, BEO, BUgI, BUgK, BUgJ, BU89/2, BUgB90, BURI*, BUgM*, BMAP, BM66*, B3194, B3196, B3200, B3202, B3203, Bpapol103, B3205, B3206, K1008cl, K851cl, K975cl, K982cl, K853cl, K852cl, K936un, K869un, K994cl, L844	N B K L	aggagaggaagagaaggagaaaagaagaggagagg
35b	BI 147, Bmag18	B	aaaaaaaaaaaaaaannaaanaaaaaaaannnag
35c	N7	N	aaaagagagggagaagagaaaaaaagaggggag
35d	BEA2274	B	aggagaggaagagaaggagaaaaaaagaggagagg
36a	K1009cl	K	agaggggagagagaaagagggaaagaagaagaggaa
36b	BI155	B	aaaaaaaaaaaaaaannaaanaannaaaaannnnag
38a	K925un	K	aaaaagaagaggggaggagaggggggaagaggggggaaga
38b	Bpapol 60	B	aaaaaaaaaaaaaaaaaannaaaaaaaaaaaannnag
38c	BB23	B	aaaaaaaaaaaaaaaanaannaaaaaaaaaaaannnag
38d	BB25, BM3, BS28, Bpapol33, BS38, L834, L941	B L	aaaaaaaaaaaaaaaaaannaaanaaaaaaaannnag
38e	N2340	N	aaaaaaganaaaannaaanaaaaaaaaaannnaggagg
39a	K997cl, K981un	K	agaaggaaggaggggaaggaaggggggaaagagggagaag
39b	K854un	K	agagggagggaggggaaggaaggggggaaagagggagaag
40	K1337cl	K	agaaggaaggaggggaaggggggaaggggagaagggagaag
42	Z218, Z185	Z	agaggaaggaaagagaaagaagaaggagagagaaggaagagg
45	Z194	Z	agaggaaggaaagagaaagaagaaggagagagaaggaaaagaagg
47	Z218, Z185	Z	agaggaaggaaagagaaagaagaaggagagagaaggaaaagaagagg

Table 8.2 Ms42 alleles. Column 1 indicates the name of each different ms42 allele, based on size (number of repeats) and on the different MVR maps obtained for that size of allele. Homozygous stocks are marked (h). Column 2 indicates which isolates have that particular allele. Column 3 gives the populations in which the different alleles have been observed; Z, Zambia; N, Nyanza; B, Busoga; K, Kiboko; L, Lugala. The different MVR maps are presented in column 4. * indicates 1-5 positions in the MVR map could not be scored, however the rest of the positions were identical to the assigned allele.

multiple restriction digests of the homozygous stocks followed by hybridization to the *MS42* repeat probe would need to be undertaken.

Diploid mapping. If MVR-PCR is applied to total genomic DNA a ternary code derived from the superimposed maps of the two alleles could be generated, as for the human minisatellite MS32 [Jeffreys *et al.*, 1991]. This ternary code can be scored as 1 (both alleles are **a**-type at that position, **aa**), 2 (both alleles are **g**-type at that position, **gg**) or 3 (heterozygous for **a** and **g** types at that position, **a g**). With the presence of null or **n**-type repeats three additional codes can be generated; 4 (**a n**), 5 (**g n**) or 6 (**nn**). Identification of the coding states 4-6 requires correct interpretation of MVR-PCR band intensities. However further variant repeat types can result in variations in band intensity observed on mapping autoradiographs, making the detection of these states difficult. Since mapping single alleles generates a simpler, cleaner map and each allele can be easily purified (with the exception of homozygotes), it was decided to only use purified alleles for the MVR-PCR analysis.

Identification of flanking polymorphisms. Alleles which share a recent common ancestor would presumably also share the same sequence of DNA flanking the repeated region of *MS42*. Knowledge of flanking haplotypes would therefore aid in the derivation of allele lineages. Therefore in order to determine the haplotypic context of the alleles mapped, polymorphisms 5' and 3' to the repeats were identified. The ORF of *MS42* has been partially sequenced (Chapter 3). To search for further polymorphisms in the sequenced region of the *MS42* ORF and adjacent sequence, a series of primers were designed to PCR amplify the regions 5' and 3' to the repeats as illustrated in Figure 8.10A. Using primers pairs *MS42*-H2/*MS42*-CR and *MS42*-BR/*MS42*-F8, the regions flanking the repeats were amplified from five different stocks. The resulting PCR products were analysed for restriction fragment length polymorphisms (RFLPs) by restriction digest with a variety of enzymes, revealing polymorphisms for three restriction sites, *Bgl*II, *Hinc*II and *Mbo*II. No polymorphisms were detected for the following enzymes; *Acc*I, *Alu*I, *Ava*I, *Bam*HI, *Cla*I, *Dra*I, *Eco*RI, *Hae*III, *Hin*dIII *Hin*fI, *Kpn*I, *Mlu*I, *Nar*I, *Nco*I, *Nde*I, *Nsi*I, *Pst*I, *Pvu*II, *Rsa*I, *Sal*I, *Sma*I, *Spe*I, *Sst*I, *Stu*I, *Sty*I, *Taq*I, *Xba*I and *Xho*I.

Assays for the polymorphisms and heterozygosity analysis. Figure 8.10B illustrates the polymorphic *Bgl*II restriction site for several different stocks, giving the three possible genotypes, homozygous for the absence of the *Bgl*II site (--) and so the PCR product of 1820bp is undigested, homozygous for the presence of a *Bgl*II site (++), producing bands of 1200bp and 620bp, and heterozygous (+ -), giving all three bands. Similar assays for the *Hinc*II and *Mbo*II polymorphisms are illustrated in Figure 8.10C and D. It was therefore a simple procedure to type each stock for these polymorphisms by PCR amplifying the region containing the restriction site (using primer combinations *MS42*-H2/*MS42*-CR for both *Bgl*II and *Hinc*II sites and *MS42*-BR/*MS42*-F8 for the *Mbo*II site), followed by a restriction digest with the appropriate enzyme (Figure 8.10). Table

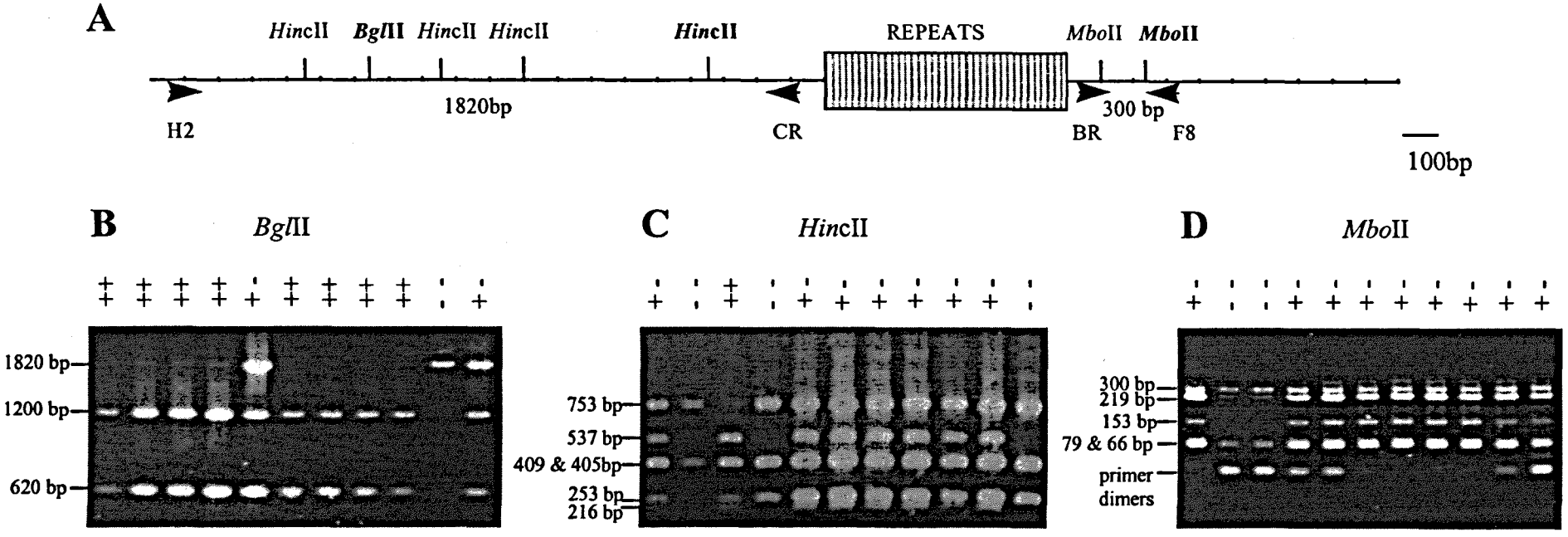


Figure 8.10. Base substitutions in the DNA flanking *MS42* repeats.

(A) Diagrammatic representation of the *MS42* 5' and 3' flanking regions, showing polymorphic restriction sites (in bold) and PCR primers (arrowheads). PCR primer sequences are given in Materials and Methods.

(B) An example of the PCR assay for the *Bgl*II polymorphic site performed on 11 different stocks. 1820bp of immediate 5' flanking DNA was amplified from either 5ng of genomic DNA or 1/200 dilution of a crude lysate (see Chapter 2) using primers *MS42*-H2 and *MS42*-CR. The PCR conditions were; 95 °C for 50s, 50 °C for 50s, 70 °C for 2mins for 30 cycles. 2-5µl of each PCR product was digested with *Bgl*II and resolved by electrophoresis through a 1% agarose gel. Stocks homozygous for the *Bgl*II- allele (–) produce a product of 1820bp. In stocks homozygous for the *Bgl*II+ allele (++) the 1820bp band is digested to generate bands of 1200bp and 620bp. Heterozygous stocks (+-) give all three bands. **(C)** An example of the PCR assay for the *Hinc*II polymorphic site for 11 different stocks. 1820bp of 5' flanking DNA was amplified as above. 2-5µl of the PCR product was digested by *Hinc*II and resolved by electrophoresis as above. Those stocks homozygous for the *Hinc*II- allele (–) produce four bands of 753bp, 409bp, 405bp and 253bp. In this example the 253bp band was too faint to be visible. In stocks homozygous for the HincII+ allele (++) the 753bp fragment is digested further to produce bands of size 537bp and 216bp. Heterozygous stocks (+-) produce all six bands. **(D)** An example of the PCR assay for the *Mbo*II polymorphic site for 11 different stocks. 300bp of immediate 3' flanking DNA was amplified from either 5ng of genomic DNA or 1/200 dilution of a crude lysate using primers *MS42*-BR and *MS42*-F8. The PCR conditions were: 95 °C for 50s, 60°C for 50s, 70°C for 50s, for 30 cycles. 2-5µl of each PCR product was digested with *Mbo*II and resolved by electrophoresis through a 3% NuSieve GTG agarose gel. Stocks homozygous for the *Mbo*II- allele (–) generate products of 219bp and 79bp. The 219bp fragment is digested further in the single stock homozygous for the *Mbo*II+ allele (++), (stock L944, not shown), giving products of 153bp and 66bp. Heterozygous stocks (+-) product all four bands.

2A-E in the appendix records the genotypes of the flanking polymorphisms for all stocks analysed.

The five different populations of *T. brucei* (Chapter 7) constituting the 127 different stocks previously analysed by MVR mapping, were typed for these polymorphisms and revealed heterozygosities ranging from 78.6% to 36% for *Bgl*II, 89.7% to 32% for *Hin*cII and 83% to 27% for *Mbo*II in the different populations (determined from the data in Table 2, appendix). The allele frequencies for each population are given in Table 8.3, demonstrating the wide range of allele frequencies obtained from the different populations, *e.g. Bgl*II+ allele frequency in Busoga is 0.34 compared to 0.82 in Kiboko. These markers were used in the analysis of populations described in Chapter 7.

Locus	Allele	Zambia Luangwa		Kenya Nyanza		Uganda Busoga		Kenya Kiboko		Uganda Lugala	
		no.	freq.	no.	freq.	no.	freq.	no.	freq.	no.	freq.
*Bgl*II	+	17	0.61	20	0.4	30	0.34	36	0.82	4	0.33
	-	11	0.39	30	0.6	58	0.66	8	0.18	8	0.67
*Hin*cII	+	11	0.39	19	0.4	35	0.45	7	0.16	5	0.31
	-	17	0.61	29	0.6	43	0.55	37	0.84	11	0.69
*Mbo*II	+	4	0.29	14	0.28	35	0.42	6	0.14	6	0.43
	-	10	0.71	36	0.72	49	0.58	38	0.86	8	0.57

Table 8.3. *MS42* flanking polymorphism allele frequencies. The number of alleles observed and frequencies of each allele for each population is given for the three flanking RFLPs.

Haplotype analysis of flanking DNA polymorphisms. Direct sequence analysis of the regions containing the polymorphisms from two stocks which were homozygous but different for each polymorphism (++ and --) revealed the *Bgl*II C/T transition at position 1175 (see Chapter 3 for *MS42* flanking sequence), the *Hin*cII C/T transition at position 2167 and the *Mbo*II G/A transition at position 2991(Figure 8.11A and B). All three polymorphic sites do not affect the predicted amino acid sequence.

In order to identify which homologue the flanking *Hin*cII alleles are assigned to in relation to the minisatellite alleles, haplotypic analysis was performed using the following procedure. An allele specific primer for *Hin*cII+ was designed as shown in Figure 8.11A. PCR amplification of the entire minisatellite for each heterozygous stock was performed in duplicate using the allele specific primer and a universal primer *i.e. Hin*cII+ primer with *MS42*-F. One aliquot was amplified at a low annealing temperature (58°C), the other identical aliquot was amplified at a higher temperature (68°C). At the lower annealing temperature both minisatellite alleles amplified, however at the higher temperature only the allele with a perfect match to the primer *Hin*cII+ could amplify. In this way the haplotype of the *Hin*cII polymorphism with respect to the minisatellite array could be determined (Figure 8.11C). Similarly, the haplotype for the *Mbo*II polymorphism was obtained using primers *Mbo*II+ and *MS42*-V (Figure 8.11B and D). The *Bgl*II polymorphism was not analysed further due to time constraints. Therefore only the

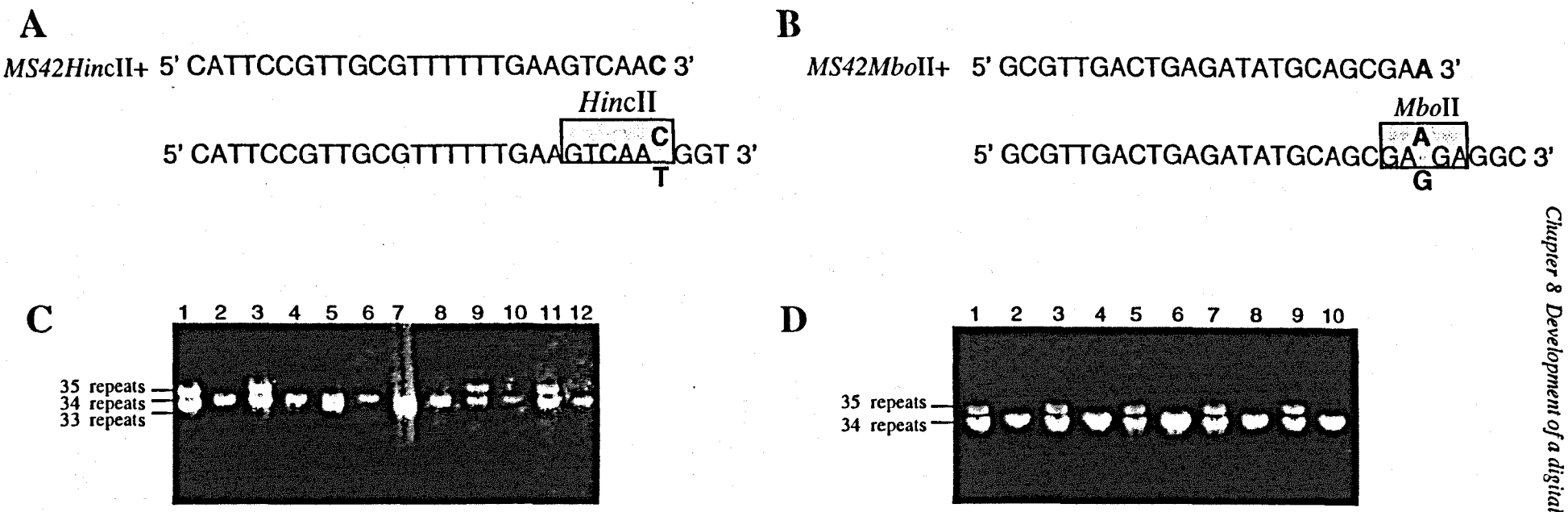


Figure 8.11. The construction of 5' and 3' haplotypes.

(A) The *Hinc*II polymorphism. The genomic sequence is shown, with the position of the polymorphic C/T transition, shaded box represents *Hinc*II recognition sequence. Allele specific primer *MS42Hinc*II+ is shown.

(B) The *Mbo*II polymorphism. The genomic sequence is shown, with the position of the polymorphic A/G transition, shaded box represents the *Mbo*II recognition sequence. The allele specific primers *MS42Mbo*II+ is shown. All primer sequences are given in Materials and Methods.

(C) An example of the PCR *Hinc*II haplotyping assay. The entire *MS42* minisatellite was amplified from 5ng of genomic DNA or 1/200 dilution of crude lysate using the universal primer *MS42*-F and the allele specific primers *MS42Hinc*II+. Each reaction was set up in duplicate. One aliquot was amplified under the following conditions; 95°C for 50s, 58°C for 50s, 70°C for 3mins for 28 cycles. The other aliquot was amplified using an annealing temperature of 68°C. The PCR reactions were loaded next to each other on a 1% agarose gel and electrophoresed in 1xTBE. Odd lanes at 58°C; even lanes at 68°C; lanes 1 and 2, N110; lanes 3 and 4, N98; lanes, N111; lanes 5 and 6, N112; lanes 7 and 8, N113; lanes 9 and 10, N120; lanes 11 and 12, N605.

(D) An example of the PCR *Mbo*II haplotyping assay. The entire *MS42* minisatellite was amplified from 5ng of genomic DNA or 1/200 dilution of crude lysate using the universal primer *MS42*-V and the allele specific primers *MS42Mbo*II+. Each reaction was set up in duplicate. One aliquot was amplified under the following conditions; 95°C for 50s, 58°C for 50s, 70°C for 3mins for 28 cycles. The other aliquot was amplified using an annealing temperature of 68°C. The PCR reactions were loaded next to each other on a 1% agarose gel and electrophoresed in 1xTBE. Odd lanes at 58°C; even lanes at 68°C; lanes 1 and 2, BMAP; lanes 3 and 4, B3194; lanes, B3196; lanes 5 and 6, B3200; lanes 7 and 8, B3202; lanes 9 and 10, B3203.

obvious haplotype of those stocks which were homozygous either for *Bgl*II ++ or -- are presented. All of the *Mbo*II haplotyping analysis and some of the *Hinc*II analysis was performed by K. Brown (including the gels shown in Figure 8.11). Haplotypes for all alleles analysed, compared to the expected values assuming linkage equilibrium, are summarised in Table 8.4. Significant (but not absolute) linkage disequilibrium exists between the two polymorphic sites, with *Hinc*II- *Mbo*II- and *Hinc*II+ *Mbo*II+ over-represented in the population and *Hinc*II+ *Mbo*II- and *Hinc*II- *Mbo*II+ under-represented. The complete data set of minisatellite structures and corresponding flanking haplotypes is presented in Figure 8.13, indicating a clear association between allele groups, based on MVR code similarity and flanking haplotypes (data discussed in following section).

Haplotype *Hinc*II *Mbo*II		Zambia Luangwa		Kenya Nyanza		Uganda Busoga		Kenya Kiboko		Uganda Lugala	
		obs	exp	obs	exp	obs	exp	obs	exp	obs	exp
-	-	9	6.06	28	20.7	30	17.9	39	31.8	8	6.29
+	-	1	3.88[a]	6	13.8	2	14.6	0	6.05	0	2.83[a]
+	+	4	1.58[b]	13	5.38	23	10.6	5	0.99[a]	5	2.13[b]
-	+	0	2.48[c]	1	8.06	1	12.9	0	5.17[b]	3	4.75[c]
χ^2		2.51		23.96		44.5		7.899		0.766	
d.f.		1		3		3		2		1	
P		0.1		<0.001		<0.001		0.05>P>0.01		0.5>P>0.3	

Table 8.4. Haplotype frequencies for *Hinc*II and *Mbo*II polymorphisms. In the calculation of χ^2 some cells were merged to raise the expected numbers above 5. For the Luangwa population cells a, b and c were merged, for the Kiboko population cells a and b were merged and for the Lugala population cells a, b and c were merged. There is significant deviation from the null hypothesis of random association, for the Nyanza, Busoga and Kiboko populations.

Allele alignments and polarity. Variation in MVR patterns between alleles at the three human minisatellites (ms32, ms31 and ms205) are not randomly distributed throughout the allele, but tend to be concentrated at one end of the array *i.e.* they demonstrate a marked polarity [Jeffreys *et al.*, 1991; Armour *et al.*, 1993; Neil and Jeffreys, 1993]. Analysis of new length alleles has demonstrated that mutation events are often inter-allelic occurring at one end of the minisatellite array [Jeffreys *et al.*, 1994; May *et al.*, 1996]. Polarity has also been observed in the repeats of the circumsporozoite gene of *Plasmodium falciparum* [Arnot *et al.*, 1993]. However polarity at minisatellites is not universal, for example the minisatellites in mice have shown that non-polar, intra-allelic events appear to be responsible for the generation of most new alleles [Bois *et al.*, 1998]. Visual inspection of the *MS42* alleles revealed no obvious polarity, however these alleles are small in length by comparison to the human minisatellites analysed by MVR-PCR, and so it may be difficult to recognise regions of similarity, especially as new mutations could obliterate the changes made by older mutation events.

In order to test if the differences between *MS42* alleles are polar, the computer program 'switchbyk', which has been used for human minisatellites [Armour *et al.*, 1996] was applied to the collection *MS42* alleles. The program aligns all the alleles at one end (5')

and performs pairwise comparisons of all allelic structures giving each pair a similarity score. Alleles sharing a minimum similarity score were then assembled into groups. Comparisons were made of the number of different groups formed over a range of different similarity values. This analysis was repeated by aligning all the alleles at the opposite end (3') of the array, as well as aligning 'mock' alleles. To generate mock alleles the order of the repeat units of each allele in the data set was randomised to produce a new map using the computer program 'twister' (the frequency of pairs and triplets of repeats was approximately preserved) [Armour *et al.*, 1996]. The results are represented by plotting the number of groups of alleles obtained at the different levels of similarity used to define groups, using alignments from the 5' or 3' ends and mock alleles (Figure 8.12). From the graph it appears that there are fewer groups generated for the 5' aligned alleles compared to either the 3' aligned and the control of mock alleles. This would suggest that there is a limited amount of polarised variability at this locus, with greater similarity at the 5' end. However, the trend is limited and could be explained by the fact that all the *MS42* alleles are small, ranging from 16-47 repeats and so genuine regions of similarity are difficult to identify or have been 'overwritten' by recent mutation events. Only 48 different allelic structures have been identified, and perhaps, with more allelic structures a more obvious pattern of variability would emerge. It is also possible that some groups of alleles mutate in a polar fashion whereas others do not, and by combining the data for all the alleles the overall trend is towards limited polarity.

The limited polarity of *MS42* is in contrast to the extreme polarity observed in the circumsporozoite gene of *P. falciparum* [Arnot *et al.*, 1993]. Although the circumsporozoite gene has small alleles with a similar repeat number to that of *MS42*, and a maximum of 46 repeat units, MVR-PCR analysis of 18 different alleles has revealed strong polarity with only 6 variants being detected in the first 13 repeat units. In order to generate allele groups for *MS42*, comparisons of all 274 *MS42* alleles aligned at both the 5' and 3' ends were generated using the computer program 'switchsorter' provided by John Armour, as well as pairwise dot matrix analysis on some alleles searching for a match of 10 perfect repeats. Matches of 10 repeats were chosen as it is expected that such matches would only appear (approximately) once by chance in a dataset of this size (1500 repeat units) based on two types of repeat units randomly distributed (probability of a match is 1 in 2^{10} = 1 in 1024). The authenticities of selected matches and the final alignment of allele groups were made by eye. The criteria for a match is reasonably stringent to reveal only those alleles which are clearly related. Relaxing the stringency would create larger groups but would also increase the risk of grouping unrelated alleles together. Figure 8.13 shows the full haplotypes including both repeat structures and flanking genotypes of all 274 alleles grouped in this way. Polarity at the 5' end of the repeat array is clearly demonstrated in group 3. However, in other groups no polarity is obvious, for example group 4 (although these alleles are approximately 10 repeats shorter than those in group 3).

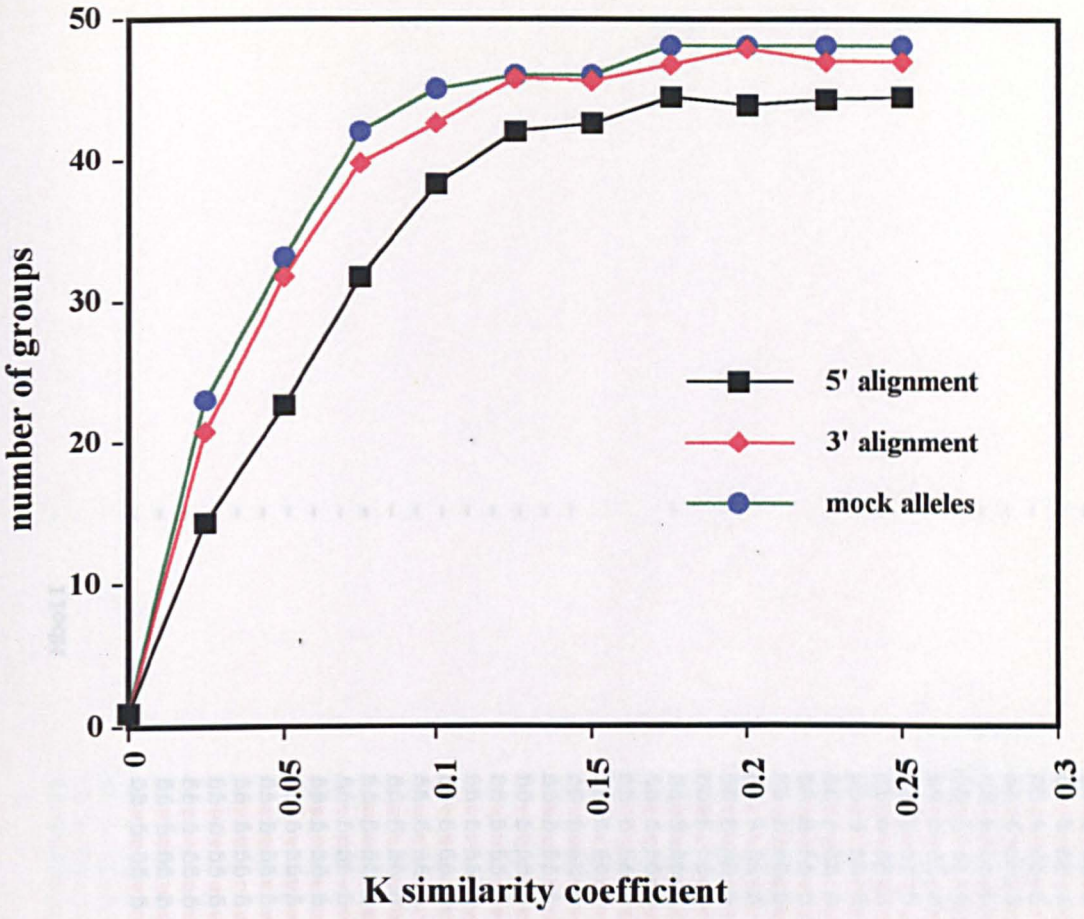


Figure 8.12. Comparison of 5' and 3' *MS42* allele alignments. The number of different groups generated at different similarity coefficients are compared between (i) alleles aligned at the 5' end (black line), (ii) alleles aligned at the 3' end (red line) and (iii) 5' aligned mock alleles (blue line).
Mock alleles were generated by randomly permutating the repeat units of each allele.
Computer programs were kindly provided by John Armour and have been described in Armour *et al.*, 1996.

MS42 Figure 8.13

Stock	HS	allele name	*Bgl*II	*Hinc*II		*Mbo*II
group 1						
N96u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N97u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N156u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N95u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N94u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N106u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N116u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N2340-1	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N110u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N98u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N113u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N120u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N605u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N609u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N102u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N105u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
N103u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
BUgC90u	ND	35a		aggagaggaagagaaggagaaaagaagaggagagg		-
BEA174u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		
BEA3u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		
BUTAR3u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		
BUTAR4u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
BUgE90u	ND	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		
BUgLu	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		
BUgA90u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		
BEOu	R	35a		aggagaggaagagaaggagaaaagaagaggagagg		
BUgIu	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		
BUgKu	R	35a		aggagaggaagagaaggagaaaagaagaggagagg		
BUgJu	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		
BU89/2u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
BUgB90u	ND	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
BURIu	R	35a		agg??aggaagagaaggagaaaagaagaggagagg		
BUgMu	R	35a	-	aggagagga?gagaaggagaaaagaagaggagagg		-
BMAPu	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
BM66u	R	35a	-	aggagaggaaga?aaggag????gaagaggagagg		-
B3194u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
B3196u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
B3200u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-
B3202u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg		-

Figure 8.13

group 1 continued					
B3203u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg	-
Bpapol103u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg	-
B3205u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg	-
B3206u	R	35a	-	aggagaggaagagaaggagaaaagaagaggagagg	-
K1008clu	ND	35a	-	aggagaggaagagaaggagaaaagaagaggagagg	-
K851clu	ND	35a	-	aggagaggaagagaaggagaaaagaagaggagagg	-
K975clu	ND	35a	-	aggagaggaagagaaggagaaaagaagaggagagg	-
K982clu	ND	35a	-	aggagaggaagagaaggagaaaagaagaggagagg	-
K853clu	ND	35a	-	aggagaggaagagaaggagaaaagaagaggagagg	-
K936unu	ND	35a		aggagaggaagagaaggagaaaagaagaggagagg	-
K869unu	ND	35a	-	aggagaggaagagaaggagaaaagaagaggagagg	-
K994clu	ND	35a	-	aggagaggaagagaaggagaaaagaagaggagagg	
L844u	ND	35a	-	aggagaggaagagaaggagaaaagaagaggagagg	-
K852clu	ND	35a	-	aggagaggaagagaaggagaaaagaagaggagagg	
B2274u	R	35d		aggagaggaagagaaggagaaaaaaagaggagagg	-
N2340-4	ND	30e		aggagaggaagagaaggagaaaagaagagg	
L8341	ND	27d	-	aggagaggaagagaaga---aaagaagagg	-
N181	ND	27d	-	aggagaggaagagaaga---aaagaagagg	-
L9411	ND	27d	-	aggagaggaagagaaga---aaagaagagg	-
L934u	ND	26a	-	aggagaggaagagaaga----aagaagagg	-
Bpapol601	S	26a	-	aggagaggaagagaaga----aagaagagg	
BM31-3	ND	26a	-	aggagaggaagagaaga----aagaagagg	-
BM42-3	ND	26a	-	aggagaggaagagaaga----aagaagagg	-
Bpapol331	S	26a	-	aggagaggaagagaaga----aagaagagg	-
BB231	S	26a	-	aggagaggaagagaaga----aagaagagg	-
BS38-3	ND	26a	-	aggagaggaagagaaga----aagaagagg	
L8451	ND	26a		aggagaggaagagaaga----aagaagagg	-
L8321	ND	26a		aggagaggaagagaaga----aagaagagg	-
N71	ND	28	-	aggagaggaagagaagaggga--gaagagg	-
group 2					
K1337clu	ND	40	+ -	agaaggaaggaggggaaggggaa ggggagaagggagaag	-
K997clu	ND	39a	+ -	agaaggaaggaggggaagg--aaggggaaagagggagaag	-
K981unu	ND	39a		agaaggaaggaggggaagg--aaggggaaagagggagaag	
K854unu	ND	39b		agagggagggaggggaagg--aaggggaaagagggagaag	

Figure 8.13

group 3						
Z218l	R	42	+	-	agaggaaggaaagagaaagaagaaggagagagaaggaa-----gagg	-
Z185l	R	42	+	-	agaggaaggaaagagaaagaagaaggagagagaaggaa-----gagg	
Z218u	R	47	+	-	agaggaaggaaagagaaagaagaaggagagagaaggaaaagaagagg	-
Z185u	R	47	+	-	agaggaaggaaagagaaagaagaaggagagagaaggaaaagaagagg	
Z194u	R	45	+	-	agaggaaggaaagagaaagaagaaggagagagaaggaaaaga--agg	-
group 4						
Z210u	R	32		-	aagaaagagagaaggaaaaggaaaaaaggaga	-
Z269u	ND	32		-	aagaaagagagaaggaaaaggaaaaaaggaga	-
Z90u	R	32		-	aagaaagagagaaggaaaaggaaaaaaggaga	-
Z274u	R	32		-	aagaaagagagaaggaaaaggaaaaaaggaga	-
Z194l	R	32	+	-	aagaaagagagaaggaaaaggaaaaaaggaga	-
Z208u	R	32			aagaaagagagaaggaaaaggaaaaaaggaga	
Z199u	R	32			aagaaagagagaaggaaaaggaaaaaaggaga	
Z220u	R	32		-	aagaaagagagaaggaaaaggaaaaaaggaga	-
Z203u	R	32			aagaaagagagaaggaaaaggaaaaaaggaga	
Z221u	R	32		-	aagaaagagagaaggaaaaggaaaaaaggaga	
Z231u	R	32			aagaaagagagaaggaaaaggaaaaaaggaga	
Z244u	R	32		-	aagaaagagagaaggaaaaggaaaaaaggaga	
Z267u	ND	32			aagaaagagagaaggaaaaggaaaaaaggaga	
Z212u	R	32		-	aagaaagagagaaggaaaaggaaaaaaggaga	
Z270u	ND	32			aagaaagagagaaggaaaaggaaaaaaggaga	
Z273u	ND	32			aagaaagagagaaggaaaaggaaaaaaggaga	
Z222u	R	30f		-	aagaaagagagaaggaaaaggaaaaaagga	
K925clu	ND	29	+	-	a-gaaagagagaagagaaaggaagagagaa	-
K981un3	ND	30b			a-gaaagagagaagagaaaggaagagagaaa	
K852un3	ND	30b			a-gaaagagagaagagaaaggaagagagaaa	
K978unl	ND	25b	+	-	a-gaaagagagaa--aaaaaaaaagaga	-
K997cll	ND	25b	+	-	a-gaaagagagaa--aaaaaaaaagaga	-
K1009cll	ND	25b	+	-	a-gaaagagagaa--aaaaaaaaagaga	-
K925cll	ND	27b	+	-	a-gaaaggagaaaaggaaaaaaaagaga	-
K852un4	ND	27b			a-gaaaggagaaaaggaaaaaaaagaga	
group 5						
L845un-4	ND	16			aagaagnana----------ngggggg	+
L944l	ND	16	-	+	aagaagnana----------ngggggg	+
L845unu	ND	31b			aagaagnananaagnaaannaagagngggggg	+
L944unu	ND	31a	-	+	aagaagnanagaagnaaannaanannggggg	+

Figure 8.13

group 6					
BB23u	S	38c		+	aaaaaaaaaaaaaaanaannaaaaaa-aaaaaaannnag +
Bpapol 60u	S	38b		+	aaaaaaaaaaaaaaaaaannaaaaaa-aaaaaaannnag
BB25 u	ND	38d	-	+	aaaaaaaaaaaaaaaaaannaaanaa-aaaaaaannnag +
BM3 u	S	38d	-	+	aaaaaaaaaaaaaaaaaannaaanaa-aaaaaaannnag +
BS28 u	S	38d	-	+	aaaaaaaaaaaaaaaaaannaaanaa-aaaaaaannnag +
Bpapol33 u	S	38d		+	aaaaaaaaaaaaaaaaaannaaanaa-aaaaaaannnag +
BS38 u	ND	38d		+	aaaaaaaaaaaaaaaaaannaaanaa-aaaaaaannnag +
L834 u	ND	38d		+	aaaaaaaaaaaaaaaaaannaaanaa-aaaaaaannnag +
L941 u	ND	38d		+	aaaaaaaaaaaaaaaaaannaaanaa-aaaaaaannnag +
BI147 u	S	35b	-	+	a---aaaaaaaaaaaaaannaaanaa-aaaaaaannnag +
Bmag18u	S	35b	-	+	a---aaaaaaaaaaaaaannaaanaa-aaaaaaannnag +
N2340-3	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
N971	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
N961	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
N156u	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
N951	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
N941	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
N1061	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
N1161	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
N1131	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
N6051	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
N6091	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
N1051	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
BUg89/8h	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
BUgL1	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag
BUgA901	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannn?g
BEO	R	34			a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag
BUgI1	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag
BUgK1	R	34			a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag
BURI1	R	34			a----aaaaaaaaaaaaannaaanaa-aaaaaaannn?g
BUgJ1	R	34			a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag
BUgM1	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
BMAP1	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
B31941	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
B31961	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
B32001	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
B32021	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
B32031	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
BM311	ND	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
BM421	ND	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +
BM661	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag +

Figure 8.13

group 6 continued						
Bpapol1031	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	+
B32051	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	+
B3206	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	+
BUTAR41	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	+
K1008c11	ND	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	+
K975c11	ND	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	+
K982c11	ND	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	+
K853c11	ND	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	
K851c11	ND	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	+
K869unl	ND	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	+
L8441	ND	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	+
BU89/21	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	-*
N1031	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	-*
N1201	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	-*
BEA1741	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	
BEA31	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	
BUTAR31	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	
N1101	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	-*
N981	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	-*
N111u	R	34	-	+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	-*
N112u	R	34	-	+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	-*
N115u	R	34	-		a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	-*
B22741	R	34			a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	-*
BUgC901	ND	34			a----aaaaaaaaaaaaannaaanaa-aaaaaa???nag	-*
BUgE901	ND	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	
BUg89/8h	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	-*
K852c11	ND	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	
K936unl	ND	34			a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	-*
K994unl	ND	34			a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	
BI155u	S	36b	-		a---aaaaaaaaaaaaannaaanaannaaaaannnnag	-*
N2340u	ND	38e			a-----aaaaaganaaaannaaanaaaaaaaaannnaggagg	
group 7						
K1009u	ND	36a	+	-	agaggggagagagaaagagggaaagaagaagaggaa	-
K927c14bu	ND	26c	+	-	agag---agagagaaag-------gaagaagaggaa	-
K981un4	ND	26c			agag---agagagaaag-------gaagaagaggaa	
K854unl	ND	26c			agag---agagagaaag-------gaagaagaggaa	
K978unu	ND	26c	+	-	agag---agagagaaag-------gaagaagaggaa	-

Figure 8.13

group 8						
K984l	ND	27c	+	-	agggagaggaaggaagaaaaagggaaa	-
K854cll	ND	27c	+	-	agggagaggaaggaagaaaaagggaaa	-
K978cll	ND	27c	+	-	agggagaggaaggaagaaaaagggaaa	-
K926unl	ND	27c	+	-	agggagaggaaggaagaaaaagggaaa	-
group 9						
K1027cll	ND	17	+	-	agaaggaggggggaaag	-
K851unl	ND	17	+	-	agaaggaggggggaaag	-
group 10						
K927cl1bu	ND	30c	+	-	aggagaaaaagaagaggaggaggggggga	-
K984clu	ND	30a	+	-	aggagaaagagaagaggaggaggggggga	-
K 258cl	ND	30a		-	aggagaaagagaagaggaggaggggggga	-
K 854clu	ND	30a	+	-	aggagaaagagaagaggaggaggggggga	-
K 978clu	ND	30a	+	-	aggagaaagagaagaggaggaggggggga	-
K 1027clu	ND	30a	+	-	aggagaaagagaagaggaggaggggggga	-
K 926un	ND	30a	+	-	aggagaaagagaagaggaggaggggggga	-
K 851un	ND	30a	+	-	aggagaaagagaagaggaggaggggggga	-
K 997unu	ND	30a	+	-	aggagaaagagaagaggaggaggggggga	-
group 11						
Z1991	R	26b			agaaggggagagaggaggggggggg	
Z2731	ND	26b			agaaggggagagaggaggggggggg	
Z2101	R	26b		+	agaaggggagagaggaggggggggg	+
Z2691	ND	26b		+	agaaggggagagaggaggggggggg	+
Z901	R	26b		+	agaaggggagagaggaggggggggg	+
Z2741	R	26b		+	agaaggggagagaggaggggggggg	+
Z2221	R	26b		+	agaaggggagagaggaggggggggg	
Z2081	R	26b			agaaggggagagaggaggggggggg	
Z2201	R	26b		+	agaaggggagagaggaggggggggg	-*
Z2031	R	26b			agaaggggagagaggaggggggggg	
Z2211	R	26b		+	agaaggggagagaggaggggggggg	
Z2311	R	26b			agaaggggagagaggaggggggggg	
Z2441	R	26b		+	agaaggggagagaggaggggggggg	
Z2671	ND	26b			agaaggggagagaggaggggggggg	
Z2121	R	26b		+	agaaggggagagaggaggggggggg	
Z2701	ND	26b			agaaggggagagaggaggggggggg	

Figure 8.13

group 12						
L934l	ND	22b		-	agaggaaa----------gaagaaaagaagaa	-
N111l	R	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
N112l	R	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
N115l	R	33b	-		agaggaaaaaaaaagaggagaagaaaagaagaa	-
N502h	R	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
N502h	R	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
N102l	R	33b		+*	agaggaaaaaaaaagaggagaagaaaagaagaa	+*
N149u	R	33b		-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
N148h	R	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
N148h	R	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
N18u	ND	33b		-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
BUgC90-3	ND	33b			agaggaaaaaaaaagaggagaagaaaagaagaa	-
BUgE90-3	ND	33b		-	agaggaaaaaaaaagaggagaagaaaagaagaa	
BI147l	S	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
Bmag18l	S	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
Bmag40l	S	33b		-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
BM3l	ND	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
BS28l	S	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
BB76h	S	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	
BB76h	S	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	
BB135h	S	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	
BB135h	S	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	
BI155l	S	33b	-		agaggaaaaaaaaagaggagaagaaaagaagaa	-
BM12h	S	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
BM12h	S	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
BS14h	S	33b			agaggaaaaaaaaagaggagaagaaaagaagaa	
BS14h	S	33b			agaggaaaaaaaaagaggagaagaaaagaagaa	
BM85l	ND	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	
BS38l	ND	33b		-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
BB25l	ND	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
L791u	ND	33b		-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
L933u	ND	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
Bfly48h	ND	33b			agaggaaaaaaaaagaggagaagaaaagaagaa	
Bfly48h	ND	33b			agaggaaaaaaaaagaggagaagaaaagaagaa	
BM80l	S	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
L836u	ND	33b		-	agaggaaaaaaaaagaggagaagaaaagaagaa	-
L832u	ND	33b			agaggaaaaaaaaagaggagaagaaaagaagaa	-
BT168l	ND	33b	-	-	agaggaaaaaaaaagaggagaagaaaagaag?a	-

Figure 8.13

group 13						
L791 u	ND	30d		-	aaaaaanaagannngagaagannngaggng	+
L836 1	ND	30d		-	aaaaaanaagannngagaagannngaggng	+
L933 1	ND	30d	-	-	aaaaaanaagannngagaagannngaggng	+
BM32 1	S	30d	-	-	aaaaaanaagannngagaagannngaggng	+
group 14						
N1491	R	21		-	aggggagagaaaaagaaaagg	-
L845un3	ND	21			aggggagagaaaaagaaaagg	-
L9291	ND	21		-	aggggagagaaaaagaaaagg	
group 15						
K925un u	ND	38a	+	-	a---aaaagaagaggggaggagaggggggaagaggggggaaga	-
K258c11	ND	25c		-	agagaaaagaagaggggagagggga	-
group 16						
K994un-3	ND	27a			agnnggagagagaagggagagaaggaa	
group 17						
K981un1	ND	33a			agaaaggggggaaaggaaaaaaaaaaagaaagaga	
group 18						
N7u	ND	35c		-	aaaagagagggagaagagaaaaaaaagaggggggag	+
group 19						
K927c11b1	ND	20	+	-	aggagaggaaaaagaggaaa	-
K854un4	ND	20			aggagaggaaaaagaggaaa	
K925un1	ND	20	+	-	aggagaggaaaaagaggaaa	-
group 20						
K927c14b1	ND	25a	+	-	agggagaagaggggaggggagaaa	-
K1337c11	ND	25a	+	-	agggagaagaggggaggggagaaa	-
K854un3	ND	25a			agggagaagaggggaggggagaaa	
K997un1	ND	25a	+	-	agggagaagaggggaggggagaaa	-

Figure 8.13

```
group  21
K994un4        ND    22a                    agagaagaagagagaggaaa

group  22
K258c11        ND    25c          -         agagaaaagaagaggggagagggga                -

group  23
BM32-3         S     23     -     -         agaaggagggaagaaaggaggaa                  -
```

Figure 8.13. Groups of aligned *MS42* alleles. For each allele its place of isolation is given as a prefix to its stock number; Z, Zambia; N, Nyanza; B, Busoga; K, Kiboko and L, Lugala. The *MS42* flanking haplotypes, alleles + or -, for each RFLP, *Bgl*II, *Hinc*II and *Mbo*II are given. The absence of a symbol indicates that the flanking haplotype has not been determined in this instance.
The derived MVR code and allele name is shown (**a**-type repeat units are in red, **g**-type repeat units in blue, **n**-type repeat units are in black and ambiguous positions are ?). Homozygous stocks are marked (h) and their alleles have been entered twice. Human serum resistance (R) and sensitivity (S) is given, ND, not determined, data from Hide *et al.*, 1994. Gaps (-) have been introduced to improve alignments. Arrows indicate regions of duplication. Identical alleles showing switching of flanking haplotypes are marked (*). The allele groups were generated based on the criteria outlined in the text, with the authenticity of selected matches and the final alignment of allele groups being made by eye.

Allele alignments and population analysis. Of the 48 different alleles detected, 37 were aligned into 10 groups containing 2-7 different alleles. The remaining alleles failed to align obviously with any other allele using the methods outlined. The authenticity of the allele grouping was supported by the flanking haplotypes, in that nearly every member of a group shares the same flanking haplotype. Only three examples of haplotype switching within a group were observed. Allele N102l in group 12 has changed both *Hinc*II and *Mbo*II sites compared to identical alleles, whereas allele Z20l in group 11 has a switch at the 3' *Mbo*II site. The third example of haplotype switching again involves the *Mbo*II site in group 6. This group comprises 59 examples of allele 34 and 13 other highly related alleles. Of the 56 alleles which have been typed for *Mbo*II in this group, the majority (43) are *Mbo*II+ and 13 are *Mbo*II-. The fact that four different allelic structures in the group all have an *Mbo*II+ haplotype may suggest that *Mbo*II+ was the ancestral haplotype for this allele lineage and the switch to an *Mbo*II- haplotype was a more recent event.

It is interesting to note that all the haplotype switching events involve the 3' *Mbo*II site which would be consistent with the view that the 3' end of the minisatellite array is more variable and that this variability extends into the flanking DNA. However, it appears that haplotype switching is a rare event, suggesting that recombination involving the exchange of flanking markers may not be a major mechanism in the generation of new alleles.

Alleles containing null repeats (**n**-type) appear to have a strong association with the *Mbo*II+ (and to a lesser extent *Hinc*II+) haplotypes. The fact that null containing repeats appear in a limited number of allele groups (groups 5, 6, 13 and 16) and have a restricted haplotype, may suggest that these alleles are distinct lineages having evolved separately from the other alleles. Certainly, alleles in group 6 which consist predominantly of **a**-type repeats with 6-9 **n**-type repeats, are unlike any of the other alleles identified so far. As previously mentioned, the sequence of the repeats from one of the alleles in group 6, allele 34, is quite distinct from those obtained from allele 26d (Figure 8.6) and from the partial sequence of alleles 20 and 25 (data not shown). This suggests a limited (if any) amount of inter-allelic recombination events involving these alleles, as inter-allelic events would result in the transfer of **g**-type repeats into the predominantly **a**-type allele, and the spread of **n**-type repeats into other allele groups.

A significant proportion of alleles fell into one of two groups (groups 1 and 6), containing 6 and 7 different alleles, respectively, suggesting the existence of two relatively ancient lineages. The fact that these groups contain alleles from four of the five trypanosome populations studied supports the view that these groups are evolutionary ancient.

Groups 2, 3, 5, 7-11, 15-23, are all population specific. Twelve of those groups are specific for the Kiboko population, supporting the conclusion from Chapter 7 that the Kiboko population is very heterogeneous. If more alleles were examined, it is probable that some groups would expand to incorporate alleles from different origins. In larger groups, population specific subgroups could be identified (*e.g.* the Kiboko stocks in group 4) sharing greater homology with each other than the rest of the group, presumably reflecting more closely related alleles.

Allele groupings and population structure. The question of how the allele groups relate to population structure can be addressed by analysing the frequency of each group of alleles in the different populations. The data have been plotted for each of the five populations of trypanosomes analysed and are presented in Figure 8.14.

Substructuring due to geography- comparison of human isolates. Three of the populations under study have been isolated from humans or are human serum resistant cattle isolates collected from Zambia, Uganda and Kenya, with the latter two locations being geographically close to one another on the north-west shores of Lake Victoria. The most recent outbreak of human trypanosomiasis near Lake Victoria in SE Uganda began in 1976, spreading North to the Tororo district and peaking to over 300 cases in 1990. It was from this focus that the Busoga stocks used in this study were isolated [Hide *et al.*, 1994]. The Kenyan samples from Central Nyanza were collected during an outbreak of human trypanosomiasis in 1961. From the analysis of allele groups, it is clear (graphs D and E, Figure 8.14) that the Uganda (Busoga) samples and the Kenyan (Nyanza) collection have almost identical patterns of allele group frequencies. This would strongly suggest that the Busoga and Nyanza foci are related, with the same human infective strains prevalent in both outbreaks.

Previous analysis of allele frequencies (Chapter 7) of the Zambian population and the human isolates from the Ugandan (Busoga) and Kenyan (Nyanza) populations indicate that the Zambian population was distinct, with no alleles in common. Similar analysis using allele groupings revealed that no Zambian alleles are in groups containing Busoga Nyanza alleles, indicating that Zambian alleles do not have regions of homology with other alleles from Busoga or Nyanza (Figure 8.14 graphs A, D and E). This is further evidence that the Zambian focus is distinct from that around the shores of Lake Victoria, indicating population substructuring due to geography. However it must be remembered that the Zambian samples were isolated in 1981-3, and so are not directly comparable to those isolated in Busoga (1988-90) and Nyanza (1961), despite the evidence presented in

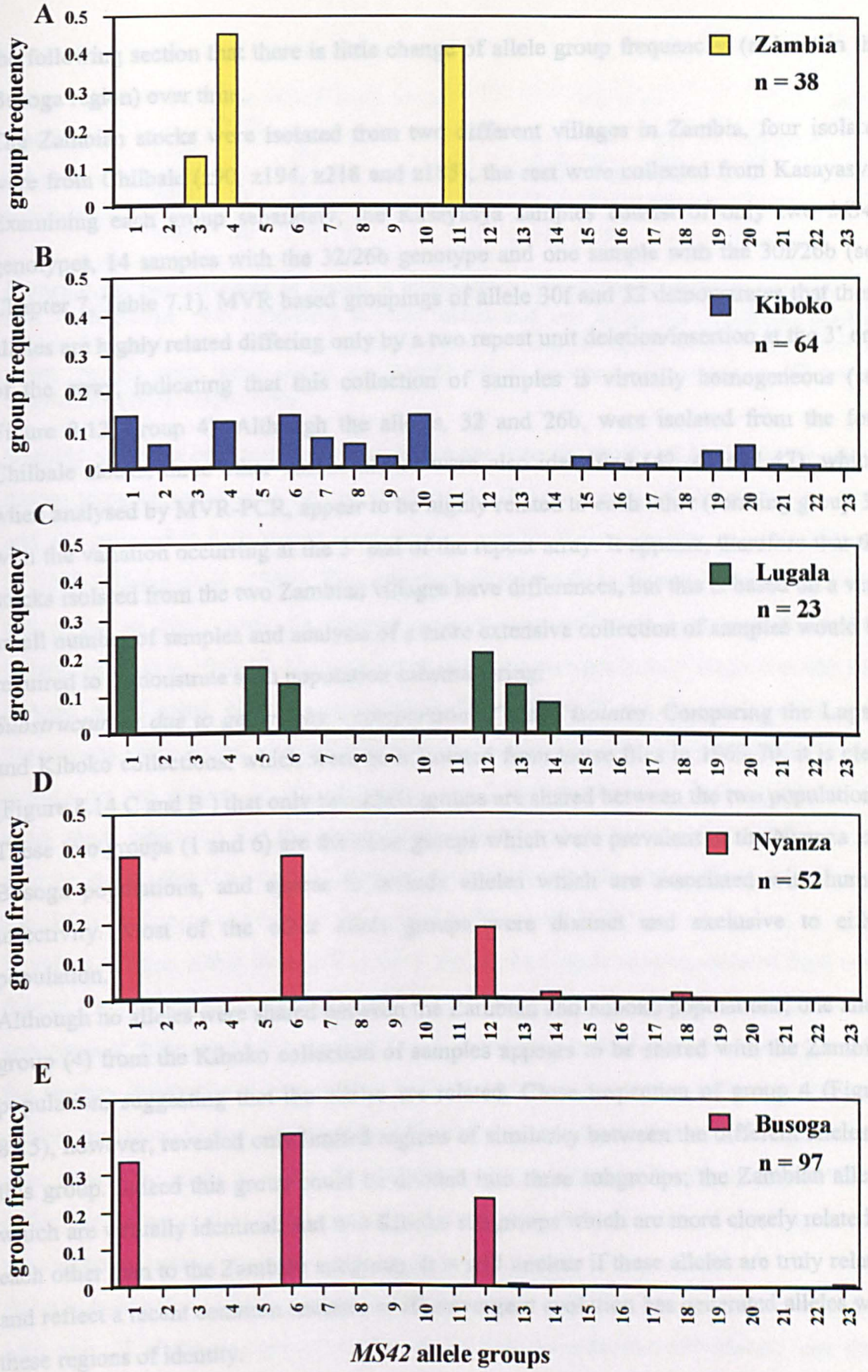


Figure 8.14. Frequency of allele groups in each population. Alleles were grouped according to the criteria indicated in the text. The number of alleles in each population is indicated (n).

the following section that there is little change of allele group frequencies (at least in the Busoga region) over time.

The Zambian stocks were isolated from two different villages in Zambia, four isolates were from Chilbale (z90, z194, z218 and z185), the rest were collected from Kasayasya. Examining each group separately, the Kasayasya samples consist of only two *MS42* genotypes, 14 samples with the 32/26b genotype and one sample with the 30f/26b (see Chapter 7, Table 7.1). MVR based groupings of allele 30f and 32 demonstrates that these alleles are highly related differing only by a two repeat unit deletion/insertion at the 3' end of the array, indicating that this collection of samples is virtually homogeneous (see Figure 8.12, group 4). Although the alleles, 32 and 26b, were isolated from the four Chilbale stocks, three other unique alleles were also identified (42, 45 and 47), which, when analysed by MVR-PCR, appear to be highly related to each other (forming group 3), with the variation occurring at the 3' end of the repeat array. It appears, therefore that the stocks isolated from the two Zambian villages have differences, but this is based on a very small number of samples and analysis of a more extensive collection of samples would be required to demonstrate such population substructuring.

Substructuring due to geography - comparison of tsetse isolates. Comparing the Lugala and Kiboko collections, which were both isolated from tsetse flies in 1969-70, it is clear (Figure 8.14 C and B) that only two allele groups are shared between the two populations. These two groups (1 and 6) are the same groups which were prevalent in the Nyanza and Busoga populations, and appear to include alleles which are associated with human infectivity. Most of the other allele groups were distinct and exclusive to either population.

Although no alleles were shared between the Zambian and Kiboko populations, one allele group (4) from the Kiboko collection of samples appears to be shared with the Zambian population, suggesting that the alleles are related. Close inspection of group 4 (Figure 8.15), however, revealed only limited regions of similarity between the different alleles in this group. Indeed this group could be divided into three subgroups; the Zambian alleles which are virtually identical, and two Kiboko subgroups which are more closely related to each other than to the Zambian subgroup. It is still unclear if these alleles are truly related and reflect a recent common ancestor or if convergent evolution has generated alleles with these regions of identity.

group 4

	stock	human serum	allele	*Bgl*II	*Hin*cII	MVR codes	*Mbo*I
subgroup 1	Z210	R	32		-	aagaaagagagaaggaaaaggaaaaaaggaga	-
	Z222	R	30f		-	aagaaagagagaaggaaaaggaaaaaagga	
subgroup 2	K925	ND	29	+	-	a-gaaagagagaagagaaaggaagagagaa	-
	K981	ND	30b			a-gaaagagagaagagaaaggaagagagaaa	
subgroup 3	K978	ND	25b	+	-	a-gaaagagagaa--aaaaaaaaagaga	-
	K925	ND	27b	+	-	a-gaaaggagaaaaggaaaaaaaagaga	-

Figure 8.15. Subdivisions in group 4. Examples of the different alleles in group 4 split into 3 subgroups, two containing Kiboko alleles and one subgroup containing Zambian alleles. For each allele the flanking haplotypes for RFLPs, *Bgl*II, *Hin*cII and *Mbo*II are given. Stocks with these alleles are human serum resistant (R), sensitive (S) or not determined (ND), respectively.

Stability of alleles over time - analysis of isolates from the shores of Lake Victoria. Acute human trypanosomiasis was first recorded in the Busoga region of Uganda in the 1940s and was believed to have spread southwards across the Kenyan border to Central Nyanza by 1955 [Gibson *et al.*, 1980]. The collection of stocks from Nyanza were isolated from this region in 1961, while the Busoga samples were collected during an outbreak of trypanosomiasis in 1988-90. Comparison of allele groups from human infective stocks from Busoga and Central Nyanza, presented in Figure 8.14 D and E, indicate that these populations share almost identical allele group frequencies. Although geographically these regions are not distant, (~ 100 kilometres apart) the samples were isolated about thirty years apart, indicating that the same human infective strains have been circulating in that area for at least that time and may have been responsible for the earlier Busoga epidemic of the 1940s confirming a previous suggestion by Hide *et al.*, (1997).

Lugala is a region within Busoga (Uganda), and so the Lugala samples, isolated from tsetse flies in 1969-70 can be directly compared to the Busoga stocks. Comparison of the allele groupings of these stocks (Figure 8.14 C and E) shows that four of the five allele groups in the Busoga isolates were also present in the Lugala tsetse samples, collected ~20 years earlier, providing further evidence that the alleles circulating in this area are quite stable over time. One allele in group 23 (Figure 8.14 E), is present in the Busoga 1988-90 collection but not in the Lugala tsetse population, probably as a result of sampling error as the number of isolates from the Lugala population is small.

Substructuring due to host specificity. Although the Lugala collection of samples is small with only 23 alleles, compared to that of the Busoga collection (98 alleles), one allele group (group 5), which consists of 3 different alleles, was identified in the Lugala (tsetse) collection, which was not present in the human isolates from Nyanza and Busoga.

Presumably this is because the tsetse isolates include both human infective and non-human infective trypanosomes and so a proportion of the latter would not appear in the collection of isolates from humans. The tsetse samples appear to have quite different alleles from those detected in the human infective stocks, supporting the suggestion that trypanosome populations are substructured due to host specificities.

A high level of allele diversity was observed in the Kiboko population sampled from tsetse flies (Figure 8.14, graph B), with 15 different allele groups being identified by MVR analysis from 64 alleles. It therefore appears that analysis of trypanosomes from tsetse flies provides a full representation of the *T. brucei* alleles circulating in an area, as they will contain trypanosomes from diverse hosts such as, wild animals, cattle and humans.

Evolution of human infectivity. Evidence based on isoenzyme and RFLP analysis of stocks from Busoga, isolated in 1988-90 from both humans and cattle, has indicated that human infective stocks (*T. b. rhodesiense*) form a fairly homogeneous group which can be distinguished from non-human infective isolates (*T. b. brucei*), with the human infective samples forming a subset of the total *T. brucei* population [Hide *et al.*, 1994]. Minisatellite analysis of many of the same stocks used by Hide *et al.*, have been presented in Chapter 7, and the *MS42* alleles identified are illustrated in Figure 8.16A. Results from this analysis support the conclusions of Hide *et al.*, that the human infective samples were quite homogeneous, with 24 out of 26 stocks having the genotype 35a/34 (see Chapter 7, Table 7.1). One of the remaining human infective stocks (BU89/8) was homozygous for the 34 allele and the other (B E 2274) had the genotype 35d/34, the 35d allele being only one repeat unit different from the 35a allele. It is clear from these data that there is an association between alleles 35a and 34 and human infectivity, since no non-human infective stocks had these alleles. This suggests that the human infective Busoga samples are distinct (genetically isolated) from the non-human infective stocks, and can possibly be defined by the presence of these alleles. The 34 and 35a alleles are also present and associated with human infectivity in the Nyanza population, although one other allele, 33b, is also associated with human infectivity (see Chapter 7, Table 7.1). Information as to how these human infective stocks evolved could be gained by examining the MVR codes derived from their *MS42* alleles. The 12 different *MS42* alleles from Busoga isolates fall into 5 different allele groups, the two most frequent groups containing both human infective and non-human infective stocks (Figure 8.16B; and Figure 8.17). This suggests

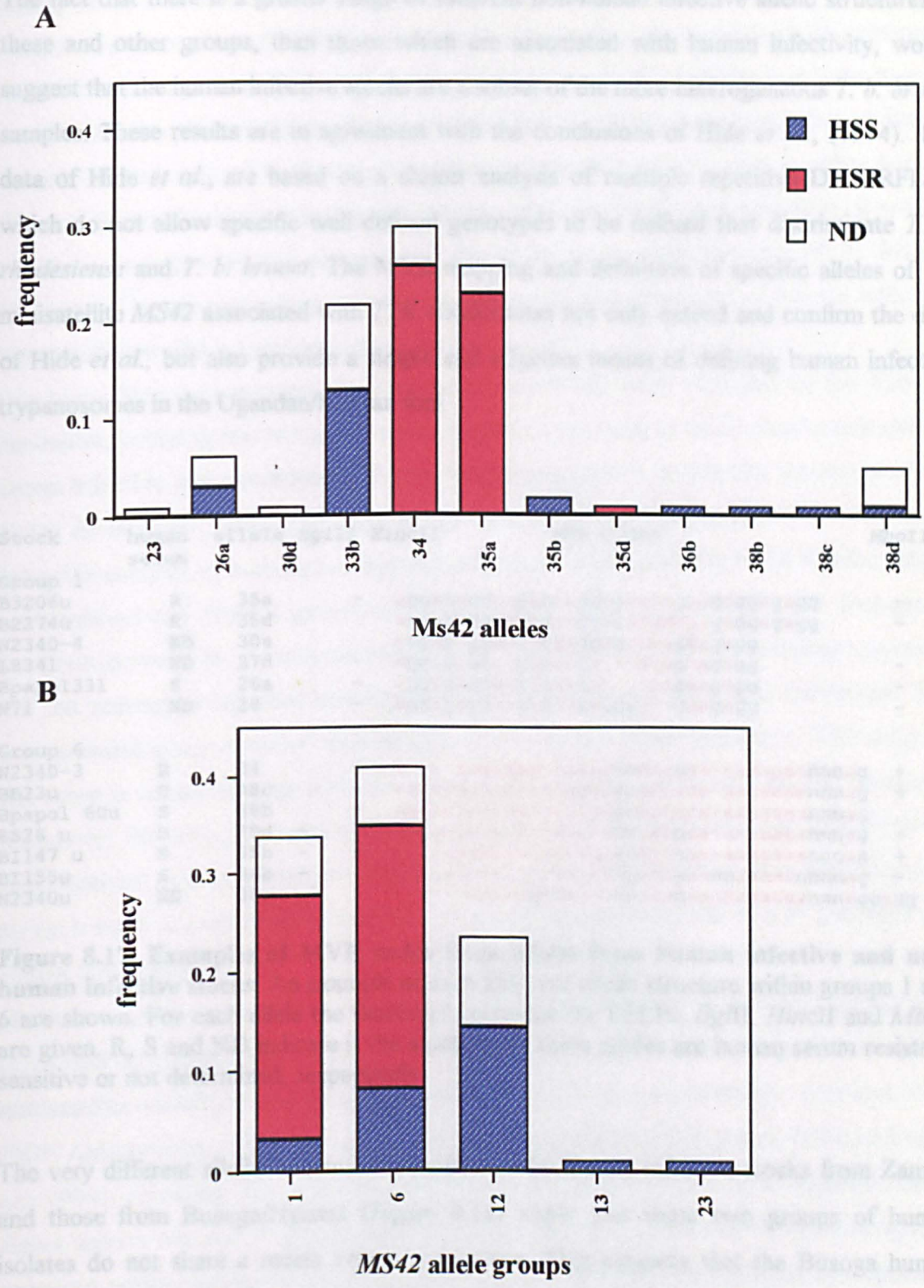


Figure 8.16. Comparison of *MS42* alleles from human serum resistant (HSR) and sensitive (HSS) stocks from Busoga.
(A) *MS42* allele frequencies for HSR (red) and HSS (blue stripes) stocks. ND indicates human serum sensitivity was not determined.
(B) Frequencies of *MS42* allele groups for HSR and HSS stocks.

that the human infective and non-human infective stocks share recent common ancestors. The fact that there is a greater range of different non-human infective allelic structures in these and other groups, than those which are associated with human infectivity, would suggest that the human infective stocks are a subset of the more heterogeneous *T. b. brucei* samples. These results are in agreement with the conclusions of Hide *et al.*, (1994). The data of Hide *et al.*, are based on a cluster analysis of multiple repetitive DNA RFLPs, which do not allow specific well defined genotypes to be defined that discriminate *T. b. rhodesiense* and *T. b. brucei*. The MVR mapping and definition of specific alleles of the minisatellite *MS42* associated with *T. b. rhodesiense* not only extend and confirm the data of Hide *et al.*, but also provide a simple and effective means of defining human infective trypanosomes in the Ugandan/Kenyan foci.

Stock	human serum	allele	*Bgl*II	*Hinc*II	MVR codes	*Mbo*II
Group 1						
B3206u	R	35a	-	+	aggagaggaagagaaggagaaaagaagaggagagg	-
B2274u	R	35d			aggagaggaagagaaggagaaaaaaagaggagagg	-
N2340-4	ND	30e			aggagaggaagagaaggagaaaagaagagg	
L8341	ND	27d	-		aggagaggaagagaaga---aaagaagagg	-
Bpapol331	S	26a	-		aggagaggaagagaaga----aagaagagg	-
N71	ND	28	-		aggagaggaagagaagagga--gaagagg	-
Group 6						
N2340-3	R	34		+	a----aaaaaaaaaaaaannaaanaa-aaaaaaannnag	+
BB23u	S	38c		+	aaaaaaaaaaaaaaannaannaaaaa-aaaaaaannnag	+
Bpapol 60u	S	38b		+	aaaaaaaaaaaaaaaaaannaaaaaa-aaaaaaannnag	
BS28 u	S	38d	-	+	aaaaaaaaaaaaaaaaaannaaanaa-aaaaaaannnag	+
BI147 u	S	35b	-	+	a---aaaaaaaaaaaaaannaaanaa-aaaaaaannnag	+
BI155u	S	36b	-		a---aaaaaaaaaaaaaannaaanaannaaaaannnnag	-
N2340u	ND	38e			a-----aaaaaganaaaannaaanaaaaaaaaaannnaggagg	

Figure 8.17. Examples of MVR codes from alleles from human infective and non-human infective stocks. An example of each different allelic structure within groups 1 and 6 are shown. For each allele the flanking haplotypes for RFLPs, *Bgl*II, *Hinc*II and *Mbo*II are given. R, S and ND indicate if the stocks with these alleles are human serum resistant, sensitive or not determined, respectively.

The very different allelic structures identified in the human infective stocks from Zambia and those from Busoga/Nyanza (Figure 8.14) imply that these two groups of human isolates do not share a recent common ancestor. This suggests that the Busoga human infective isolates are more closely related to the local non-human infective isolates, than they are to other human infective stocks from different regions. Following this argument, human infectivity appears to have originated independently in these different geographical regions, a product of convergent evolution. This would infer that the separate classification

of all human infective stocks from East Africa into the subspecies, *T. brucei rhodesiense* [Hoare, 1972] is incorrect and would be better considered to be host range variants of *T. brucei brucei* or a series of different sub-species. To confirm and extend this conclusion it would be necessary to compare non-human infective stocks from Zambia with the human infective isolates presented here. If this revealed that the Zambian human infective stocks are more closely related to the local non-human infective isolates than to other human infective stocks from other regions (as in the case of the Busoga samples), it would provide a strong case for abandoning the subspecies classification of *T. brucei rhodesiense*, or alternatively geographical subspecies classifications of human infective trypanosomes would have to be introduced.

It is interesting to note that the genotype 35a/34 has been detected in the Kiboko population, implying that human infective trypanosomes may be circulating in this area. If human infective and non-human infective trypanosomes are genetically distinct, with no alleles in common, then two separate groups of trypanosomes should be detected. Extensive analysis by isoenzymes and pulse field gel electrophoresis of 18 Kiboko clones has identified two distinct groups of trypanosomes [Tait, unpublished]. The first group was homogeneous for isoenzyme patterns, with very similar but not identical karyotype and the second group was heterogeneous for both isoenzymes and karyotype. The homogeneous group contains only those stocks with the 35a/34 genotype. This suggests that there is substructuring with the Kiboko population, which may be due to human infectivity. However, no cases of human sleeping sickness have been reported from this area. Analysis of a much larger collection of samples, with human serum resistance data for each stock are required to confirm the association between the 35a/34 genotype and human infectivity.

Mutation processes. *MS42* shows a high level of variability in *T. brucei* populations, as a result of a high mutation rate to new length alleles. The mutation processes driving minisatellite variability has been examined for four human minisatellites [Armour *et al.*, 1993; Armour *et al.*, 1993; Neil and Jeffreys, 1993; Buard and Vergnaud, 1994; Jeffreys *et al.*, 1994], revealing that most mutation events involve the gain of a small number of repeat units at one end of the repeat array, suggesting the existence of localised hotspots of mutation within minisatellites. Some new length alleles at these human loci appear to contain MVR map segments from both parental alleles indicating that these mutant alleles may have arisen through inter-allelic exchange events, either gene conversion or

recombination, although intra-allelic events such as sister chromatid exchange and slippage have also been implicated in the generation of some new alleles.

In order to analyse the mutation processes which generate the variability at *MS42*, two approaches were adopted. The first was to analyse how *MS42* alleles from the different populations are related to each other, which can provide evidence for the different mutation processes (both intra- and inter- allelic events) involved in the generation of new alleles. The second approach was to examine a length mutant generated during a genetic cross and to isolate new length mutants by the process of small pool PCR (SP-PCR) [Jeffreys *et al.*, 1994].

Examination of allele groups- intra-allelic events. Examination of the different allele groups has revealed that many of the alleles identified here appear to have been generated by deletions/duplications, substitutions and micro-gene conversions, which need not involve another allele. For example, small regions of duplications or deletions can be observed in the alleles in group 1 (Figure 8.13), with the terminal 5 repeat units of allele 35a perhaps being generated by a duplication, while alleles 26a and 27d could be generated by internal deletions. Further examples of variation involving duplications/deletions can been seen in group 3, where highly polar events have generated new alleles. Groups 7 and 15 also contain examples of possible deletions. Deletions/duplications and possible substitutions could also generate the different alleles in group 6. Only one **g**-type repeat unit is present in most of the alleles in group 6 and, since **g**-type repeat units constitute 44.8% of all non-group 6 alleles, it is highly unlikely that alleles from group 6 have been generated by inter-allelic events involving alleles from other groups. Another possible example of a simple intra-allelic event generating a new allele can be see in group 5 (Figure 8.13). Here alleles L845un-4 and L845unu (isolated from the same fly) differ by an internal deletion (or insertion). Substitutions or micro-gene conversion events may also play a role in the generation of new alleles, for example in group 1 alleles 35d to 35a only differ by one repeat unit, suggesting such an event. One repeat unit difference has also been observed between alleles 30c and 30a in group 10 which is likely to have been caused by a substitution or micro-conversion event. It is clear that examination of allele groups indicates that intra-allelic events such as duplications, deletions, substitutions and micro-gene converions can all be involved in the generation of new alleles. The mechanism by which these events can occur could either be crossovers between sister chromatids or replication slippage and may occur during mitotic replication.

Examination of allele groups- inter-allelic events. Other alleles could possibly have been generated by inter-allelic events, *i.e.* involving the allele on the homologous chromosome, perhaps by introducing a small patch of repeats by gene conversion or by larger recombination events, presumably occurring at meiosis. For example (Figure 8.18A) allele 39b may have been generated by a crossover between alleles 40 and 39a. Unfortunately the flanking haplotypes for these alleles are not informative, so it is unclear if this crossover event (if real) involved the exchange of flanking markers. Although these alleles have been isolated from different stocks, they were all from the Kiboko population, implying that mixing of alleles takes place in this population. The second example (Figure 8.18B) involves another possible recombination event, where two of the alleles, 30b and 27b, have been isolated from the same stock, indicating that mating may have occurred to generate allele 25b.

Examples C and D in Figure 8.18 are possible gene conversion events whereby a small part of one allele is inserted into another to create a larger hybrid allele, without exchange of flanking markers, which is the predominant mechanism of mutation in human minisatellites [Armour *et al.*, 1993]. In these examples the alleles are all either from within the same population (mainly Kiboko) or are from populations which have been shown to be closely related (*i.e.* Busoga, Nyanza and Lugala), suggesting that these alleles could have been present in a single genome. However, the regions of homology between alleles in these examples are small and so need not necessarily suggest a recent ancestral link but may be the product of convergent evolution.

Another example of a possible mutation event comes from the clone N2340. PCR amplification of the *MS42* locus from this stock revealed the presence of four *MS42* alleles. This stock was a **cloned** isolated from a human in Nyanza in 1977 and so it is unlikely (although not impossible) that it was a mixed sample. Two of the alleles from this stock were the common alleles 34 and 35a, generating strong bands on an ethidium bromide stained gel, the other two alleles (30e and 38e) were fainter and MVR mapping revealed that these alleles were unique. On closer inspection of these alleles it is clear that these alleles were highly related and unlikely to be the result of contamination of the DNA sample. It seems likely that during growth in the laboratory a mutant trypanosome arose resulting in a mixed population of progenitor and mutant cells. Examination of the MVR codes from these alleles indicates that two of the alleles may have been the products of a reciprocal crossover event (Figure 8.19). Whether this involved a small patch of gene conversion or extended into the flanking DNA and involved exchange of the flanking

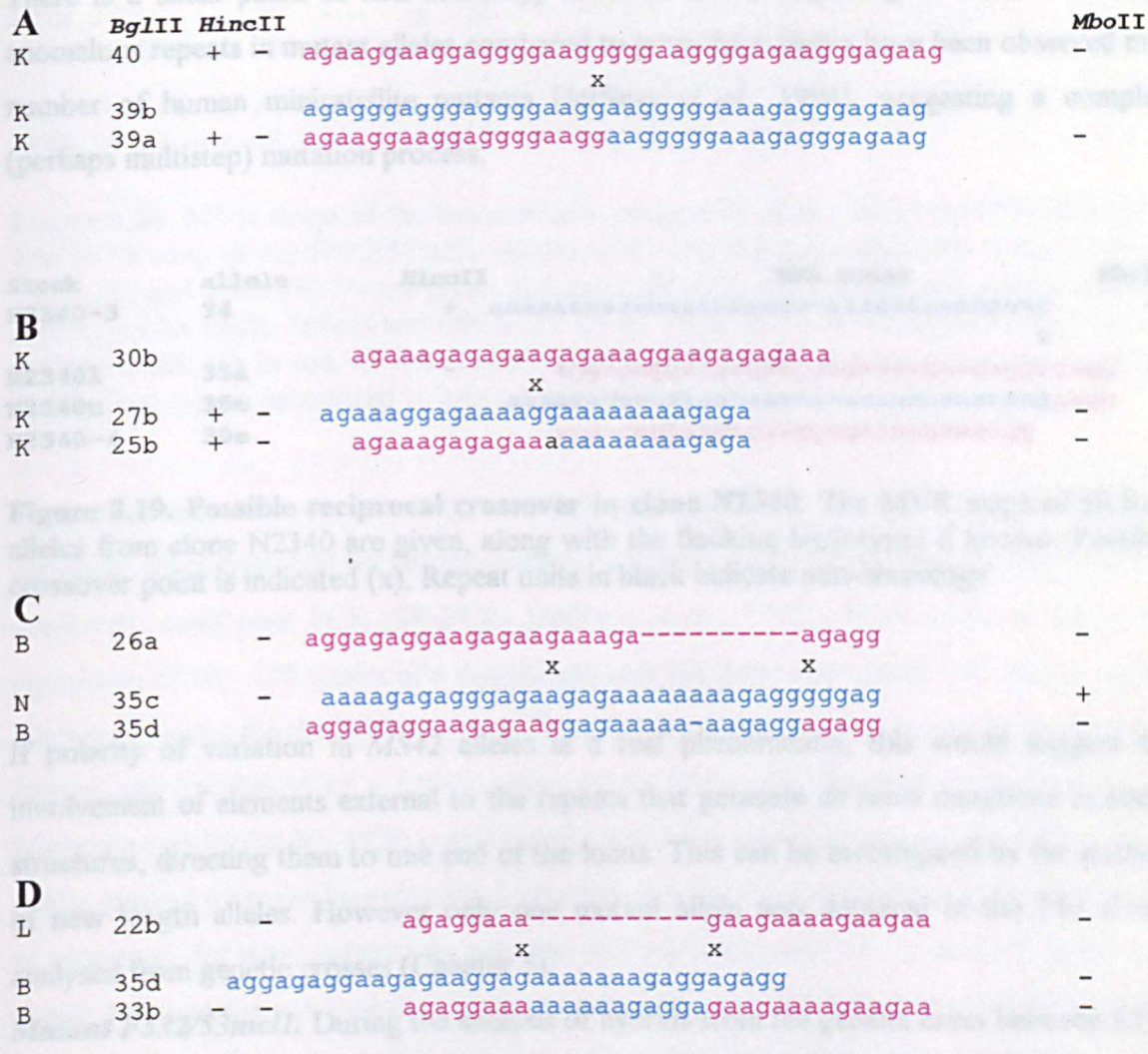


Figure 8.18. Possible inter-allelic events in the generation of new alleles. The MVR maps of the two possible progenitor alleles for each example are shown in pink and light blue with the allele name and the flanking haplotypes for *Bgl*II, *Hin*cII and *Mbo*II. The possible hybrid alleles are shown. Possible points of crossovers are indicted (x).

markers is not known as the flanking haplotypes for alleles 30e and 38e were not obtained. There is a small patch of non-homology detected at the beginning of allele 38e. Such anomalous repeats in mutant alleles compared to progenitor alleles have been observed in a number of human minisatellite mutants [Jeffreys *et al.*, 1994], suggesting a complex (perhaps multistep) mutation process.

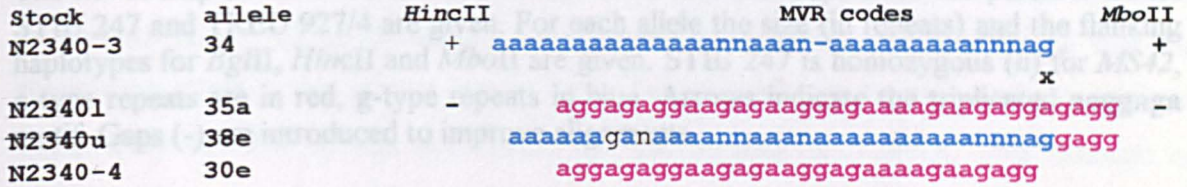


Figure 8.19. Possible reciprocal crossover in clone N2340. The MVR maps of all four alleles from clone N2340 are given, along with the flanking haplotypes if known. Possible crossover point is indicated (x). Repeat units in black indicate non-homology.

If polarity of variation in *MS42* alleles is a real phenomenon, this would suggest the involvement of elements external to the repeats that generate *de novo* mutations in allelic structures, directing them to one end of the locus. This can be investigated by the analysis of new length alleles. However only one mutant allele was detected in the 156 clones analysed from genetic crosses (Chapter 5).

Mutant F532/53mcl1. During the analysis of hybrids from the genetic cross between STIB 247 and TREU 927/4 (Chapter 5), a novel sized *MS42* allele was observed in clone F532/53mcl1, giving a mutation rate of 1/109 meiosis (putative). This clone appears to be trisomic possessing three *MS42* alleles of 20, 25 and 28 repeats respectively. It appeared that this clone has inherited both TREU 927/4 alleles and a mutant STIB 247 allele. By comparing the MVR map of the mutant allele with the parental allele, the mutation process involved in the generation of this allele can be deduced (Figure 8.20). It appears that the mutant allele was generated from the STIB 247 allele by an internal triplication of a **aaggaga** motif and a truncation.

```
Stock          allele  BglII HincII                                    MboII
STIB 247h        26    - -   agaaggaga--------------ggaagaaagggaggaaa
mutant allele    28          agaaggagaaaggagaaaggagaggaag
                               ------>------>------>
TREU 927/4cl4l   20    + -   aggagaggaaaaagaggaaa                      -
TREU 927/4cl4u   25    + -   agggagaagaggggaggggggagaaa                -
```

Figure 8.20. MVR maps of the mutant and progenitor alleles for clone F532/53mcl1. The MVR map of the F532/53mcl1 mutant allele and the maps from the parental clones STIB 247 and TREU 927/4 are given. For each allele the size (in repeats) and the flanking haplotypes for *Bgl*II, *Hin*cII and *Mbo*II are given. STIB 247 is homozygous (h) for *MS42*, **a**-type repeats are in red, **g**-type repeats in blue. Arrows indicate the triplicated **aaggaga** motif. Gaps (-) are introduced to improve alignments.

Frequency of mutations. Since there are a limited number of progeny clones available for the study of minisatellite mutations, another approach to detecting mutants by PCR was employed, small pool PCR (SP-PCR) [Jeffreys *et al.*, 1994]. Small pools of DNA, the equivalent of 10 - 100 copies of a diploid genome, per pool, were amplified using a limited number of cycles for the PCR reaction. The resulting products were run on an agarose gel, blotted and hybridised to a *MS42* repeat probe (for a full description of SP-PCR see Materials and Methods). PCR is sufficiently faithful to allow the detection of new length alleles derived from single molecules. Using this approach it is possible to screen thousands of molecules for new mutants. In this way new mutants could be isolated for analysis to reveal the mutation mechanisms generating variability and the frequency of these mutation events could be determined. The frequency of mutation events is particularly relevant to the use of minisatellite loci as markers, as a high mitotic mutation rate would prevent the ability to track individual alleles, for example, in pedigree analysis. Two stocks, the *T. b. rhodesiense,* STIB 386 and the *T. b. brucei,* TREU 927/4, were analysed by SP-PCR for new length mutants generated during mitotic growth. No mutants were detected in 6500 genomes worth of DNA, a maximum frequency of 0.00015, indicating that this locus is extremely stable mitotically (roughly equivalent to the mutation frequency in somatic cells for the human minisatellite ms32, [Jeffreys *et al.*, 1994]) and so confirming its suitability as a molecular marker.

As indicated by the possible inter-allelic events detected in the population analysis, and the detection of one mutant allele in 53 F1 hybrid clones, mutation events during meiosis may be important in the generation of new alleles. This was investigated using the following approach. Since pure trypanosome meiotic material *i.e.* gametes, are obviously unavailable

for analysis, the closest available material to this was bloodstream (unpassaged) uncloned trypanosomes derived from the salivary glands of a tsetse fly from a genetic cross (STIB 386 x TREU 927/4), which was known to contain both parental types and hybrid progeny. SP-PCR analysis of DNA made from the uncloned material detected three presumptive mutant alleles from 15000 genomes worth of DNA, a frequency of ~0.0002 (Figure 8.21). The frequency of detection of mutant alleles from this material is not significantly different from the values obtained for mitotic mutations (0.00015), although this is a minimum value. All three mutants were increases in size (if it is assumed that the progenitor allele is the closest in size to the mutant allele). Further analysis of these mutants by MVR-PCR was not carried out due to time constraints. This estimate of mutation frequency from this meiotic-enriched material by SP-PCR is approximately 50 fold lower than the estimate based on progeny analysis (one new length allele in 53 progeny clones; mutation frequency of 0.0094), however both are crude estimates.

Conclusions

MVR analysis has revealed a far larger number of alleles at the *MS42* locus than could be detected by band size estimates even assuming 100% accuracy of the band size measurements. Therefore, using this system, unrelated alleles which are of the same size are no longer scored as matching alleles. The distribution patterns of variant repeats within alleles also contains cladistic information as alleles with very similar patterns presumably share a recent common ancestor. This provides a rational method for grouping (or binning) alleles for population genetic analysis.

In this study, comparisons of allele groups has revealed that trypanosomes isolated from tsetse flies have a greater level of diversity compared to those isolated from humans. Examination of allele groupings for human isolates have revealed that the Zambian stocks are quite unrelated to those from Kenya and Uganda, with the Kenyan and Ugandan foci being highly related, demonstrating that population substructuring due to geographical barriers exists. Also, examination of allele groups for human and non-human infective samples from the same area (Busoga), isolated at the same time, indicated that human infective samples could be defined by the presence of specific *MS42* alleles and are more closely related to the local non-human infective samples than to human infective samples from Zambia, suggesting that the subspecies classification of *T. brucei rhodesiense* for all human infective stocks from East Africa is flawed. This analysis could be extended to examine trypanosomes from different regions of Africa; in particular, examination of non-

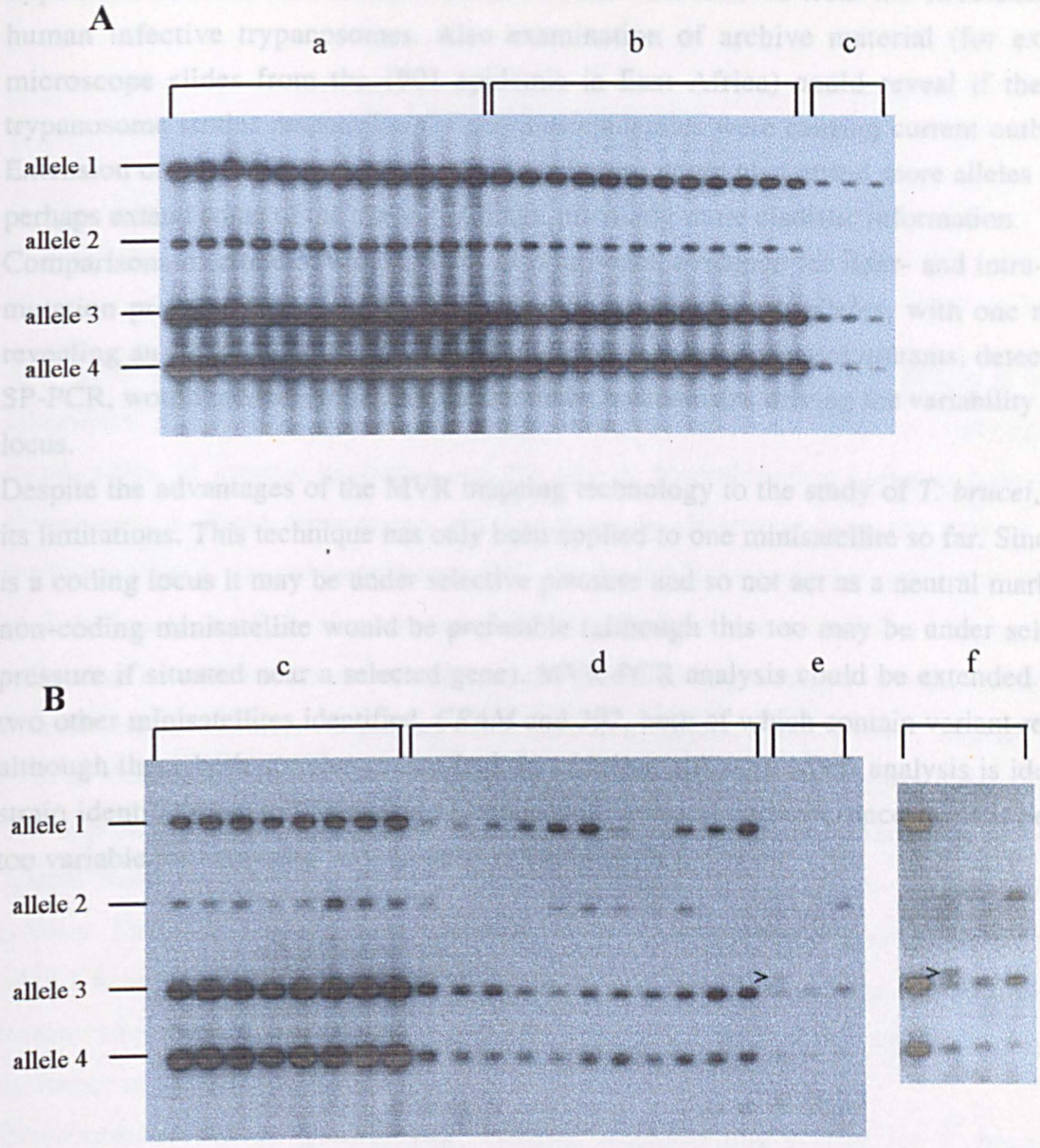


Figure 8.21. SP-PCR analysis of uncloned material derived from a STIB 386 x TREU 927/4 cross. DNA was extracted from an uncloned stabilated which was known to contain both parental and hybrid trypanosomes. The DNA was diluted, amplified by PCR to sub-visible levels, using primers *MS42*-W and *MS42*-F, fractionated on a 1% agarose gel, blotted and hybridized to the *MS42* repeat probe, as described in Materials and Methods. The material contained 4 alleles (numbered 1-4) present in the approximate ratio 3:1:12:27. Panel **A**, (a) 12 identical PCR reactions of 1/5 dilution of DNA, (b) 12 identical PCR reactions of 1/10 dilution of DNA, (c) 3 identical PCR reactions of 1/100 dilution of DNA. The autoradiograph was exposed for 24 hours. Panel **B**, (c) 8 identical PCR reactions of 1/100 dilution of DNA, (d) 11 identical PCR reactions of 1/1000 dilution of DNA, (e) 3 identical PCR reactions of 1/10000 dilution of DNA. The autoradiograph was exposed for 2 days. Novel sized mutant allele is marked by an arrowhead in (e) lane 1. (f) Longer exposure (4 days) of last 4 lanes to highlight the mutant allele. The number of amplifiable molecules in each track for each allele was calculated using the Poisson distribution decribed in Materials and Methods.

human infective trypanosomes from Zambia would allow the investigation of the hypothesis that Zambian human infective stocks were derived from the circulating non-human infective trypanosomes. Also examination of archive material (for example microscope slides from the 1901 epidemic in East Africa) could reveal if the same trypanosome strains responsible for previous epidemics were causing current outbreaks. Extension of the MVR analysis to other populations could also reveal more alleles and so perhaps extend some of the allele groupings, providing more cladistic information.

Comparisons of allele MVR maps has also provided evidence for inter- and intra-allelic mutation processes involved in the generation of new length alleles, with one mutant revealing an internal triplication and a truncation. Analysis of more mutants, detected by SP-PCR, would indicate which are the dominant mechanisms driving the variability at this locus.

Despite the advantages of the MVR mapping technology to the study of *T. brucei*, it has its limitations. This technique has only been applied to one minisatellite so far. Since this is a coding locus it may be under selective pressure and so not act as a neutral marker. A non-coding minisatellite would be preferable (although this too may be under selective pressure if situated near a selected gene). MVR-PCR analysis could be extended to the two other minisatellites identified, *CRAM* and *292*, both of which contain variant repeats, although these both are also coding loci. In addition, although MVR analysis is ideal for strain identification and identifying samples with a recent common ancestry it is perhaps too variable for analysing very distantly related samples.

Chapter 9

Discussion

The broad aim of this project was to study sexual recombination in *T. brucei* in both the laboratory and field. Laboratory crosses were examined in order to investigate the mechanisms of genetic exchange and the rate of recombination between homologous chromosomes. Field populations were analysed to determine both the role of sexual recombination in natural populations of *T. brucei* and investigate the population structure of *T. b. brucei* and *T. b. rhodesiense*. A further objective was to develop a typing system which can be used for disease tracking and aimed at distinguishing *T. b. rhodesiense* from *T. b. brucei* as well as investigating the evolution of human infectivity *i.e.* whether strains of *T. b. rhodesiense* from different foci have the same origin or have arisen independently. The approach employed to address these questions was to develop a sensitive genotyping system based on minisatellites and to use this system to study genetic crosses and field isolates. Discussion of the results is provided at the end of each results chapter and so what I have attempted here is to draw together the main conclusions and discuss their broader implications in relation to addressing some of the questions in the study of genetic exchange in trypanosomes.

Development of a genotyping system. A genotyping system for *T. brucei* was developed based on hypervariable mini- and microsatellite loci. Minisatellite markers were chosen as they are the most polymorphic loci identified to date in eukaryotic genomes. By using such a discriminating marker system, potentially, each individual stock could be identified; parental and progeny trypanosomes could be distinguished, and any differences between *T. b. rhodesiense* and *T. b. brucei* stocks should be revealed. The minisatellite loci used in this study were chosen because they were relatively small and so could be easily amplified by PCR, which provides several advantages over other typing systems. For example, as only small numbers of parasites are required for genotyping and the PCR primers are specific for trypanosomes, the parasites do not need to be removed from contaminating host DNA and can be genotyped directly without the need for growth

in laboratory rodents or in culture. Also, since the markers are single loci, the information obtained from them can be interpreted genetically, unlike classical DNA fingerprinting or RAPDs. Although each of the three minisatellites used as markers have heterozygosity levels of over 90%, they were all found in coding regions which may limit their variability. Non-coding minisatellites are less likely to be under selective pressure and so would provide neutral markers; however at the time of searching for minisatellite sequences in the database, the vast majority of sequences available were from coding regions. A library screen for minisatellite sequences, using Jeffreys' tandem repeat probes, could have been undertaken in order to isolate non-coding minisatellites, but this would have been a time-consuming exercise. It is also possible that *T. brucei* does not contain many non-coding single locus minisatellites, as intragenic sequences are usually short in this species. To circumvent the possible problem of selection, all three unlinked minisatellites were used in combination, as it is unlikely that all three markers are under the same selective pressure. Furthermore, this provides an increased level of discrimination between stocks.

Four microsatellites were also used throughout this study. One microsatellite, in the 3' untranslated region of the *TIM* gene, was identified from the published sequence, shown to be polymorphic and used to genotype parasites present in tsetse fly saliva. The three other microsatellite markers were, JS2, D2 and E5 which were all isolated and characterized by J. Sasse.

The identification of minisatellite loci in *T. brucei* is important not only because they provide a useful genotyping system but also because the high level of variation at these loci raises a series of questions about mutational mechanisms. Evidence from human studies indicates that the mutation events which generate new length alleles are most likely to occur during meiosis, with a 'hot spot' of recombination being found at one extensively studied minisatellite locus [Jeffreys *et al.*, 1998]. The high level of *T. brucei* minisatellite variability may also be a result of meiotic events. To study minisatellite instability in more detail and to develop an even more discriminating method for tracking strains, the MVR-PCR technique [Jeffreys *et al.*, 1991] was applied to the *MS42* minisatellite locus. This method digitally types *MS42* alleles using both allele size (without recourse to band size measurements) and the distribution patterns of variant repeats. This typing method has a major advantage over conventional genotyping methods in that individual alleles can be identified precisely. A total of 274 alleles were typed using this system providing an insight into the mutation processes involved in generating new length alleles. The conclusions from this analysis (Chapter 8) are:

1) *MS42* mutation processes do not appear to be very polar, in contrast to the circumsporozoite gene of *P. falciparum* and most of the human minisatellites analysed in this way. However, a less variable human minisatellite and several mouse minisatellites analysed to date do not show signs of polarity [Andreassen and Olaisen, 1998; Bois *et al.*, 1998].

2) There is evidence for possible crossover and gene conversion events, as well as internal duplications and deletions, suggesting that the mutation processes at this minisatellite probably involve inter- and intra-allelic events.

3) One cloned line has provided evidence for a reciprocal crossover event.

4) The examination of one F1 hybrid which contained a new length *MS42* allele indicated that the mutation was probably a multistep process involving an internal triplication and a truncation.

5) The maximum frequency of mitotic mutations which generate new length alleles was determined as being 0.00015 mutants/ progenitor molecules, while the meiotic mutation events occur at frequency of 0.0094 mutants/ meiosis (putative).

However both mutation frequency estimates are likely to be inaccurate as no mitotic mutants were detected and only one meiotic mutant allele was detected in 53 hybrid clones. If more hybrid clones become available for analysis, a more accurate mutation frequency estimate could be obtained. The possibility that meiotic events are more frequent than mitotic events could lead to the suggestion that meiotic-specific events may be involved in generating some mutations. Further circumstantial evidence of meiotic events being common comes from the similarity of the dispersion patterns of variant repeats within the *MS42* minisatellite to those described for human minisatellites, where meiosis-specific events are the major mechanisms for generating variation (variant repeats are randomly distributed throughout the allele). However one human minisatellite, MSY1, on the Y chromosome has a different dispersion pattern, whereby blocks of similar repeats are clustered [Jobling *et al.*, 1998; Bouzekri *et al.*, 1998]. The difference in dispersion patterns may be a reflection of the different mutation processes involved at this locus. The portion of the Y chromosome on which MSY1 is located does not undergo recombination and so can in effect be considered clonal with no inter-allelic recombination events occurring. The *T. brucei* minisatellite *MS42* does not have this dispersion pattern, suggesting meiotic processes may be involved in repeat turnover.

In order to strengthen this conclusion, it would be necessary to investigate the mutation events at this locus further. Three new length alleles have been detected from meiotic-enriched material by SP-PCR, although they were not analysed by MVR-PCR due to time

constraints. Analysis of these and other mutants may give an insight into complex mutation processes involved at this locus. The limiting factor in this investigation is the lack of pure meiotic material for SP-PCR. If parental lines differing in drug resistance were available (similar to those used by Gibson and Whittington, [1993], double drug resistant hybrid progeny could be selected for and so provide large numbers of pure hybrid material which could be analysed for mutations by SP-PCR and to isolate more meiotic mutants for comparison with mitotic mutants. By using allele-specific primers in the flanking DNA it should also be possible to detect mutation events which involved the exchange of flanking markers *i.e.* products of recombination. The analysis of crossovers on chromosome 1 has indicated a possible high rate of recombination near the *MS42* locus, albeit by analysis of a very small number of progeny and crossovers. It would be interesting to investigate this possible recombination 'hot spot', by analysing more progeny clones using further markers on chromosome 1. A search for additional polymorphic microsatellite markers is currently underway.

If SP-PCR analysis of pure hybrid material were available, the *T. brucei MS42* minisatellite would be a model for studying minisatellite biology, as it would be possible to study the effects of *cis* acting elements, currently believed to modulate mutation events in human mini- and microsatellites [Monckton *et al.*, 1994; Brock *et al.*, 1999], by deleting the regions of DNA flanking the repeats and thus provide a test not readily available in other organisms.

The function of the highly polymorphic minisatellites identified in *T. brucei* is still unknown. *CRAM* is believed to be a cell surface receptor localised to the flagellar pocket which is abundantly expressed in procyclics and at very low levels in bloodstream trypanosomes [Lee *et al.*, 1990]. The *292* gene is expressed at higher levels in bloodstream trypanosomes than in procyclics and is believed to encode a membrane associated protein with unknown function [Lee *et al.*, 1994]. Other proteins containing repeats have been found in membrane associated proteins of *T. brucei*, for example, the major surface protein, PARP [Mowatt and Clayton, 1988] and a microtubule-associated protein [Schneider *et al.*, 1988]. Membrane associated proteins of other parasitic protozoa have also been shown to contain tandem repeats, for example, the surface antigens of the *T. cruzi* [Macina *et al.*, 1989] and *P. falciparum* [Kemp *et al.*, 1987].

It is unclear, at present, if the *MS42* gene product is also located on the cell surface and what its function is. It would be interesting to characterize this protein, by knocking out the gene to determine if the gene product is essential and, by using specific antibodies, to determine its sub-cellular localisation.

The population studies presented in Chapter 8 have indicated that one *MS42* allele (allele 34) is associated with human serum resistance. It should be possible to transfect this unusual allele into a human serum sensitive strain to investigate if the allele is in some way involved in human serum resistance or if it is merely a marker for this phenotype.

Analysis of laboratory crosses. Having developed a highly informative genotyping system, the products of genetic crosses could then be analysed in order to dissect the process of mating in *T. brucei* (Chapter 5). The specific questions which were addressed in this chapter were:

1) At what life cycle stage does genetic exchange take place?

2) Does sexual recombination follow Mendel's laws of allele segregation at a locus and independent assortment of alleles at unlinked loci?

3) How common is triploidy?

4) Does pre-amplification in laboratory rodents select for/against hybrid trypanosomes?

5) Does the proportion of selfers, hybrids and parentals alter over time?

6) Does self-fertilisation occur only in the context of cross-fertilization?

7) What is the mechanism of genetic exchange?

8) What is the level of recombination between homologous chromosomes?

The approach taken to identify the life cycle stage at which mating takes place, was to genotype single trypanosomes from the relevant life cycle stages where mating could occur. To this end a genotyping system based on microsatellites was developed which could genotype single cells (Chapter 4). Using this technique it was possible to amplify a locus in 71% of single cells. To improve the utility of this method, it will be necessary to extent this analysis to several loci from a single trypanosome (multiplexing). Using this technique to analysis several loci from large numbers of cells (and controls) it should be possible to identify parentals, hybrids and tetraploid or haploid cells from the different life cycle stages within the tsetse fly. However, the main limitation was the fact that no tsetse flies with mixed salivary gland infections were available for dissection and analysis, and so this question remains unanswered.

To address the other questions, previously isolated progeny clones were analysed as well as new clones from previous crosses, increasing the total number of clones derived from crosses to 156, 81 of which were hybrid clones with 53 of these arising from unique events. Examination of these clones revealed some interesting features of *T. brucei* genetic exchange:

1) The minisatellite alleles, in the main, follow the laws of Mendelian inheritance in that they segregate in the progeny in broad agreement with Mendelian ratios and appear to be inherited independently.

2) The vast majority of hybrid clones are the equivalent of F1 hybrids, with only 2-6% being triploid, in marked contrast to the results of Gibson, who found that 66% of progeny clones were triploid [Gibson and Bailey, 1994]. It is possible that the extent of triploidy may be due to the different stocks involved in the different crosses. It would therefore be useful to study different crosses to investigate the 'normal' level of triploidy. The triploids identified in this analysis all involve extra copies of the *T. b. rhodesiense* stock STIB 386 chromosomes, with no triploids being identified from STIB 247 x TREU 927/4 crosses. The stock STIB 386 is quite different from the other two parental clones, in that it is a West African *T. b. gambiense* type 2 (i.e. a West African *T. b. rhodesiense*) whereas STIB 247 and TREU 927/4 are both from East Africa. The triploid clones identified by Gibson *et al.*, (1994) were also from a cross between an East African *T. b. rhodesiense* and a West African *T. b. brucei* stock, and all harboured extra chromosomes from the *T. b. rhodesiense* parental clone. Could this mean that *T. b. rhodesiense* is losing its ability to undergo meiosis? Another possibility is that a chemical signal which stimulates meiosis is not effectively received by the *T. b. rhodesiense* stock, perhaps because the parental stocks have diverged from each other. To test if the inability to undergo meiosis readily is an intrinsic property of *T. b. rhodesiense* it would be necessary to cross an East African *T. b. rhodesiense* stock with an East African *T. b. brucei* stock.

3) A new class of hybrids has been identified, trisomics. These clones appear to have inherited three copies of at least one chromosome but not other chromosomes. In this analysis, which was based on 4 markers on different chromosomes, 13% of F1 hybrid clones appeared to be trisomic for one of the markers. Assuming that each trisomy event is independent, it is possible to extrapolate that approximately 36% of hybrid clones are trisomic for a chromosome and 13% are trisomic for two chromosomes. The majority (6/7) of trisomies were for chromosome 1 and inherited the extra chromosome from STIB 386, suggesting that non-disjunction is common for STIB 386 for chromosome 1. Non-disjunction occurs frequently for chromosomes which have failed to undergo recombination and a failure of chromosomes to cross over is often due to a lack of homology [Koehler *et al.*, 1996]. It is interesting to note that the chromosome I homologues of STIB 386 are very different in size [Melville *et al.*, 1998] and MVR analysis of the *MS42* alleles in these homologues indicate that the alleles are different (one allele has variant repeats so distinct as to render MVR analysis with the current primers

impossible, data not shown). It is possible therefore that trisomy may be an aberrant product of meiosis due to sequence divergence of STIB 386 homologues. Moreover, analysis of crossing over on chromosome IV has revealed that STIB 386 has a lower frequency of recombination than TREU 927/4 [M. Hope, personal communication], although no such difference was detected for chromosome I.

One hybrid clone, which could be trisomic for chromosome 1, has a new length allele for the minisatellite *MS42*, indicating that either some form of DNA rearrangement has occurred and is associated with trisomy of that chromosome or a gene duplication event has resulted in a novel allele and so gives the clone three alleles and thus the appearance of trisomy.

4) A comparison of bloodstream and metacyclic clones derived from the same fly has revealed that the amplification of trypanosomes in mice favours the growth of hybrid genotypes (from one cross), indicating that selection can occur during growth in laboratory rodents and that the direct genotyping of trypanosomes from tsetse flies would be the best approach to assessing the parasite genotypes present in the salivary glands.

5) Examination of the frequency of crossing over on chromosome 1 using the minisatellite *MS42* and 2 microsatellites, gives genetic distance estimates ranging between 4.9kb/cM and 25 kb/cM, which is in the same order of magnitude as that from *P. falciparum* (15-30kb/cM) with an equivalent genome size [Walker-Jonah *et al.*, 1992]. This gives a crude estimate of the *T. brucei* recombination rate in the order of 6.7×10^{-4} Morgans/kb approximately 67 times higher than humans (1×10^{-5}) [Chakravarti *et al.*, 1984]. This high recombination rate may compensate for the non-obligatory nature of sexual recombination in *T. brucei.*

The three crosses which have been analysed in this study have been between stocks isolated from different geographical areas (Tanzania, Ivory Coast and Kenya) and, given the fact that sub-structuring in the *T. brucei* population due to geographical barriers has been demonstrated (Chapter 7), it would be unlikely that these strains would have come into contact with each other naturally. Perhaps a better strategy would have been to cross strains isolated from the same area at the same time, possibly generating fewer aberrant products. The three stocks in these crosses have different levels of human serum resistance (STIB 247 is sensitive, TREU 927/4 has intermediate resistance and STIB 386 is resistant) and analysis of the progeny from these crosses has been used in the study of the inheritance of human serum resistance [Lindergard, 1999]. However, if the hypothesis derived from the population studies suggesting that human infectivity arose independently in different geographical areas is correct, then the genetic component for human serum

resistance in the three stocks may be unrelated, and thus it may be unwise to pool data from all three crosses. The examination of a cross between a human infective and a non-human infective stock from the same area, for example Busoga, may be able to address the question of the inheritance of human infectivity. Progeny from a cross between a resistant stock carrying the unusual *MS42* allele (allele 34) which is associated with human infectivity, and a sensitive stock, could be analysed for human serum resistance and each minisatellite marker. In this way the association between human serum resistance and the *MS42* allele 34 could be examined.

Models of genetic exchange. A number of models of genetic exchange in *T. brucei* have been proposed since sexual reproduction was first demonstrated in the laboratory in 1986. The first model proposed by Paindavoine *et al.* in 1986 suggested that genetic exchange involved fusion of diploid cells followed by random elimination of DNA to return to a diploid state. This model attempts to explain the elevated DNA content found in some hybrids, but would predict non-Mendelian allele segregation, which has not been observed. Since this model does not fit the data presented in Chapter 5 nor the data from published analysis [Tait *et al.*, 1988; Sternberg *et al.*, 1989; Turner *et al.*, 1990; Gibson *et al.*, 1995], it can be disregarded. The observed segregation and independent assortment of alleles in *T. brucei* provides strong evidence that genetic exchange involves meiosis and syngamy. The majority of markers used in the analysis of crosses appear to be inherited in a Mendelian fashion and the observed ratios of hybrid genotypes is in general agreement with those predicted for a Mendelian system.

The inheritance of the kinetoplast is biparental for the minicircles [Gibson and Garside, 1990], but initial data [Sternberg *et al.*, 1989] presented for the maxicircles suggested that these were inherited uniparentally. However a more extensive analysis [Turner *et al.*, 1995] has indicated that maxicircles are also inherited biparentally but due to random segregation during post meiotic replication, fixation for one type of maxicircle rapidly occurs during the course of vegetative amplification of progeny clones.

Three basic models have been suggested based on meiosis and fusion and are discussed in turn. Sternberg and Tait, [1990] suggested two models: fusion after meiosis and fusion before meiosis, and Gibson *et al.*, (1995) proposed a second fusion before meiosis model. The first of these models (outlined in Figure 9.1) predicts that haploid gametes are produced (Figure 9.1A). Three types of haploid gametes could possibly be produced. If the divisions of nuclear DNA and kDNA are linked then haploid gametes with half the normal kDNA content would be produced. Alternatively the nuclear DNA could divide

without kDNA division so that half the gametes could have the normal kDNA content and half have no kDNA at all.

The fusion of haploid gametes which have half the normal kDNA content (1n 0.5k) will produce a diploid hybrid cell which has the normal kDNA content, (2n 1k) with the kDNA being biparentally inherited (Figure 9.1B). Alternatively, an akinetoplastic haploid cell and a kinetoplastic cell could fuse which would result in diploid cells with only one kinetoplast. However this model would predict that kDNA is uniparentally inherited. If only akinetoplastic and kinetoplastic cells could fuse then no mixed kDNA could arise. This does not fit the current data as hybrid trypanosomes with mixed kDNA have been observed [Turner *et al.*, 1995]. It is possible that some akinetoplastic and kinetoplastic cells could fuse, but also two kinetoplastic gametes could fuse, which would result in some hybrid cells containing mixed kDNA. Fusion of these haploid gametes is illustrated in Figure 9.1C, where the nuclei fuse as well as the kinetoplasts generating a 2n 2k cell. This cell may undergo division of kDNA without any nuclear division to produce a 2n 1k normal diploid cell, or the cell could undergo mitosis to produce two hybrid cells containing a 2n nuclear DNA content and double the normal kDNA content (2n 2k), which is then followed by gradual kDNA loss. Alternatively the fused cell undergoes mitosis to produce two diploid cells with normal kDNA content and two zoids which contain kDNA and no nuclei.

The prediction from the meiosis followed by fusion model (all variants) would be that the products of self-fertilisation would have kDNA derived from the same parent as the nuclear DNA, which could provide a test for this model if sufficient numbers of selfers can be generated.

Two variations of the second model of fusion before meiosis are presented in Figure 9.2 A and B. In the first, diploid cells fuse to produce a 4N 2k cell, which undergoes meiosis producing a cell with 8 nuclei and 4 kinetoplasts. The nuclei fuse in pairs generating hybrid nuclei and 'selfer' nuclei. Cell division results in hybrid and selfer cells some of which could have mixed kDNA (Figure 9.2A). Alternatively (Figure 9.2B) the 4N cell could undergo meiosis without DNA synthesis. Cell division would result in hybrid cells with mixed kDNA, and parental cells also with mixed kDNA (Figure 9.2B). However no selfers would be produced by this model. Selfers could only be generated by the fusion of two identical diploid cells and so would have kDNA only from that parental stock.

Another model, by Gibson *et al.* (1995) (outlined in Figure 9.2C), proposes the fusion of diploid parental cells, including the fusion of the kinetoplasts. The cell then undergoes meiosis to produce a multinucleate cell in which all but two of the haploid nuclei

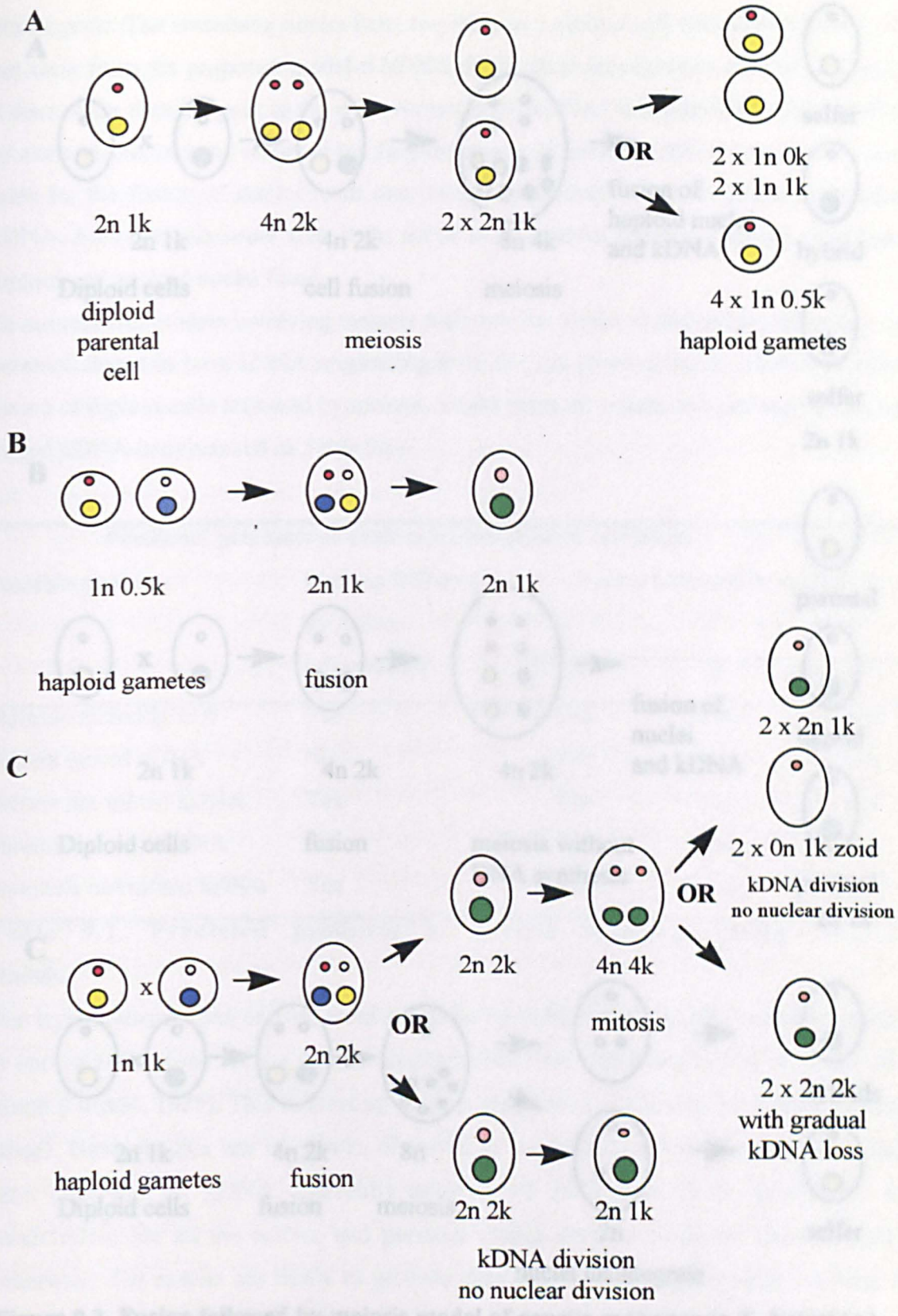


Figure 9.1. Meiosis followed by fusion model of genetic exchange in *T. brucei* (A) Meiosis generating haploid gametes. **(B and C)** Fusion of haploid gametes to produce hybrids.
Parent A nuclei in yellow, parent B nuclei in blue. Hybrid nuclei in green.
Parent A kDNA in red, parent B kDNA in white. Mixed kDNA in pink.

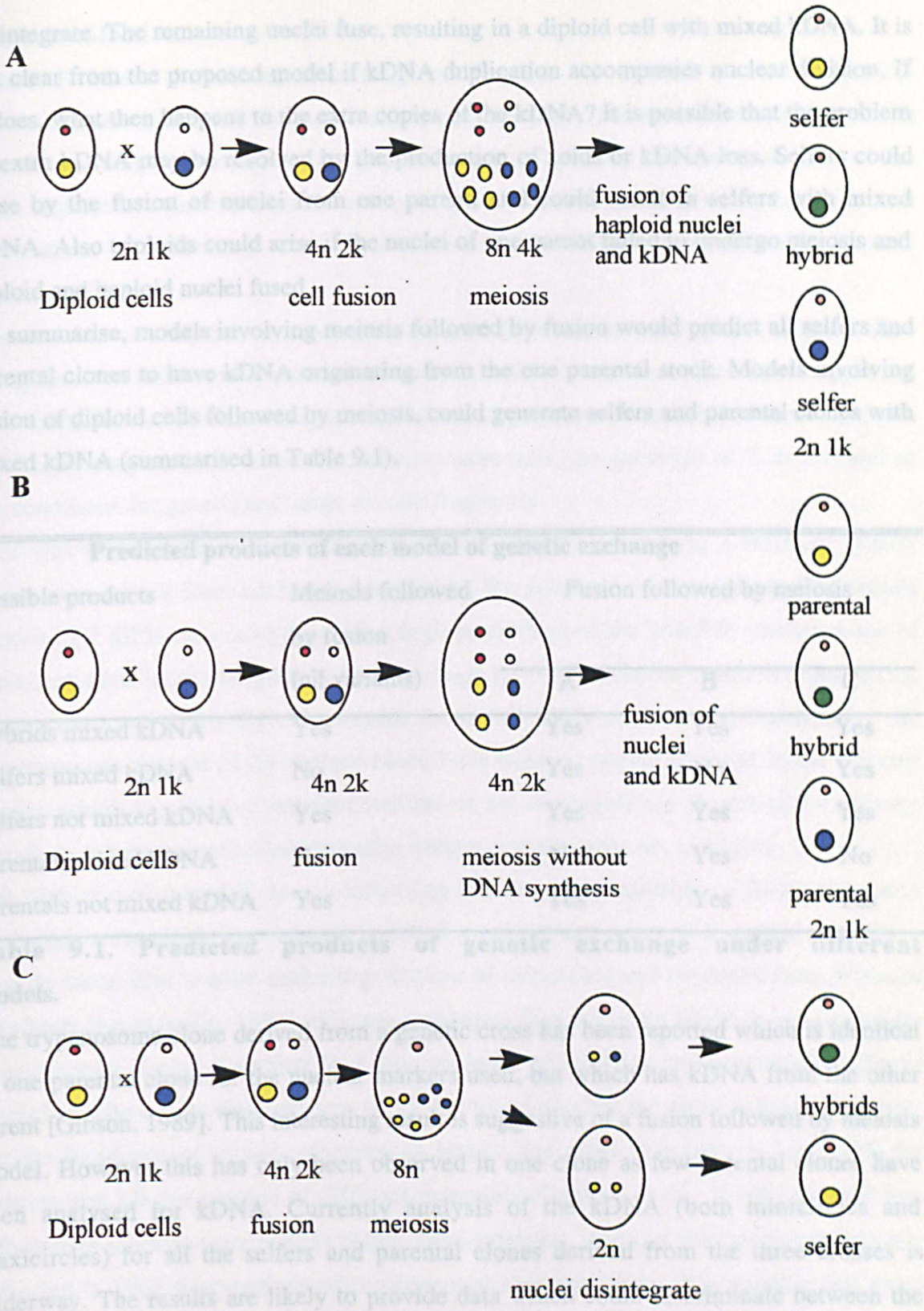


Figure 9.2. Fusion followed by meiosis model of genetic exchange in *T. brucei* (A) Diploid cells fuse, followed by meiosis to produce hybrids, parentals and selfers. **(B)** Diploid cells fuse, followed by meiosis without DNA synthesis to produce hybrids and parentals. **(C)** Model proposed by W.C. Gibson, taken from Gibson *et al.*, 1995.
Parent A nuclei in yellow, parent B nuclei in blue. Hybrid nuclei in green.
Parent A kinetoplast in red, parent B kinetoplast in white. Mixed kinetoplast in pink.

disintegrate. The remaining nuclei fuse, resulting in a diploid cell with mixed kDNA. It is not clear from the proposed model if kDNA duplication accompanies nuclear division. If it does, what then happens to the extra copies of the kDNA? It is possible that the problem of extra kDNA may be resolved by the production of zoids or kDNA loss. Selfers could arise by the fusion of nuclei from one parent, and could result in selfers with mixed kDNA. Also triploids could arise if the nuclei of one parent failed to undergo meiosis and diploid and haploid nuclei fused.

To summarise, models involving meiosis followed by fusion would predict all selfers and parental clones to have kDNA originating from the one parental stock. Models involving fusion of diploid cells followed by meiosis, could generate selfers and parental clones with mixed kDNA (summarised in Table 9.1).

Predicted products of each model of genetic exchange				
Possible products	Meiosis followed by fusion (all variants)	Fusion followed by meiosis		
		A	B	C
Hybrids mixed kDNA	Yes	Yes	Yes	Yes
Selfers mixed kDNA	No	Yes	No	Yes
Selfers not mixed kDNA	Yes	Yes	Yes	Yes
Parentals mixed kDNA	No	No	Yes	No
Parentals not mixed kDNA	Yes	Yes	Yes	Yes

Table 9.1. Predicted products of genetic exchange under different models.

One trypanosome clone derived from a genetic cross has been reported which is identical to one parental clone for the nuclear markers used, but which has kDNA from the other parent [Gibson, 1989]. This interesting result is suggestive of a fusion followed by meiosis model. However this has only been observed in one clone as few parental clones have been analysed for kDNA. Currently analysis of the kDNA (both minicircles and maxicircles) for all the selfers and parental clones derived from the three crosses is underway. The results are likely to provide data which could discriminate between the haploid gamete and cell fusion models.

Analysis of field isolates.

Mixed T. brucei infections in tsetse flies. The fact that sexual recombination can occur between different *T. brucei* stocks is beyond doubt. However, the extent to which genetic exchange occurs in natural populations is still highly controversial. One of the main arguments against a high frequency of genetic exchange has been the apparent lack of mixtures of different *T. brucei* infections in tsetse flies, a prerequisite for genetic exchange to occur.

Examination of uncloned trypanosomes isolated from the salivary glands of tsetse flies using the three minisatellite markers revealed that 36% of those samples from Lugala and 47% from Kiboko (both in Kenya) were of mixed genotype (Chapter 6), indicating that a significant proportion of tsetse flies harbour more than one genotype of *T. brucei* and so the conditions for genetic exchange are met frequently.

This was established unequivocally for two isolates by generating a series of cloned trypanosome lines from each and determining the genotype of each clone. One isolate contained 7 different genotypes with a high proportion of the possible combinations of alleles at each locus, suggesting that the high level of variation resulted from sexual recombination in this fly. If samples from this fly's midgut were available, the trypanosomes present in the midgut could have been compared to those in the salivary glands, which would have revealed whether or not the genotypes present in the salivary glands were recombinant. Unfortunately, midgut samples were not available.

The high rate of mixed *T. brucei* infections indicate the possibility of frequent genetic exchange in the field. One interesting question which has emerged from these results is how do tsetse flies acquire multiple genotypes of trypanosomes? Evidence from Welburn [Welburn and Maudlin, 1992] indicates that tsetse flies are more likely to be infected by trypanosomes during their first feed, the incidence of infection being markedly reduced thereafter. The results obtained here imply that the tsetse flies fed on mixed-infected hosts, but the incidence of mixed infections in human and other mammals is low (~3% for humans and 2% for a range of different mammals)[Godfrey *et al.*, 1990], although less informative markers were used to obtain these estimates. Because the tsetse samples were collected in 1969 from regions where there was abundant game, it is possible that game may have harboured mixed-infections, however no isolates from game were collected from this area at that time. Game have been suggested as a reservoir of a wide variety of different trypanosome genotypes [Mihok *et al.*, 1990], however, the depletion of game in East Africa has probably altered the feeding behaviour of tsetse flies with humans and cattle becoming frequent hosts. Since a more limited repertoire of trypanosome genotypes

is capable of infecting humans (and perhaps cattle), the diversity of trypanosomes and the number of mixed *T. brucei* infections in tsetse flies may be lower in areas where game is scarce. An alternative explanation is that in the field tsetse acquire infections from more than one host. To investigate this further, isolates from all possible hosts need to be collected and analysed directly using PCR amplification of hypervariable minisatellites. This approach would avoid the need to grow the samples in the laboratory and so avoid any selection.

The frequency of tsetse fly infections is approximately 0.1-1% [Welburn and Maudlin, 1997] and so if the ability of a tsetse fly to acquire an infection is independent of it acquiring a second infection, the predicted frequency of mixed infections is extremely low (0.01-0.0001%). Evidence from studies on *T. congolense*, which has a higher frequency of infection, ~3%, and so many more samples available for analysis, suggest that there is a decline in the rate of infection of tsetse flies with fly age and that there are resistant flies in the population [Woolhouse and Hargrove, 1998]. Although this work has not been extended to *T. brucei*, there is evidence from laboratory studies that the infection rates for non-teneral flies is reduced relative to teneral and that the presence of rickettsia-like organisms (RLOs) in the tsetse midgut alters the ability of flies to acquire infections [Welburn and Maudlin, 1997]. It is possible that the presence of RLOs in the midguts of wild tsetse flies, may cause a proportion of the tsetse fly population to become susceptible to *T. brucei* infections. It would be interesting to examine wild tsetse flies for the presence of RLOs to investigate if there is a correlation between mixed infections and the presence of RLOs, however this type of study would require very large numbers of tsetse samples.

T. brucei population structure. Although the conditions for genetic exchange are met, i.e. a significant number of tsetse infections are mixed, it is still unclear as to what extent genetic exchange occurs in the field. Three types of population structures have been proposed for *T. brucei* based on the extent of sexual recombination in the population: clonal (little genetic exchange), epidemic (some genetic exchange masked by clonal expansion of some strains) or panmictic (randomly mating).

Tibayrenc, who proposed the theory of clonality for a number of parasitic protozoa, including *T. brucei*, provided evidence from isoenzyme data for linkage disequilibrium [Tibayrenc *et al.*, 1990; Tibayrenc *et al.*, 1991]. However, there are several reasons, other than clonality, why a population may demonstrate linkage disequilibrium. A population with a mixture of different sub-populations (populations or species) each of which may be randomly mating or the existence of a population bottleneck could all result in linkage disequilibrium. Self-fertilization could also result in linkage disequilibrium, For example,

linkage disequilibrium for isoenzyme markers [Tibayrenc *et al.*, 1990] has been demonstrated for small numbers of samples of the malaria parasite *P. falciparum,* which has an obligatory sexual cycle in the mosquito vector [Walliker *et al.*, 1987]. This may be due to the high incidence of self-fertilization (resulting in parasites that are identical to the parental type), although a number of other studies have found no significant linkage disequilibrium [for example, Conway and McBride, 1991]. Analysis of oocysts isolated from mosquitoes in Papua New Guinea have shown that although self-fertilization may be common, revealed by an excess of homozygotes when compared to HW expectations, linkage disequilibrium is absent, suggesting that there is sufficient outbreeding to disrupt any linkage disequilibria [Paul *et al.*, 1995]. Since the mosquito vector acquires mixed infections from feeding on mixed-infected humans, the extent of self-fertilization is dependent on the number of mixed infected people and the number of different clones present in the human host, which varies according to the transmission intensity of the region. In regions of low transmission intensity self-fertilization is more common than in areas of high transmission intensity [Paul *et al.*, 1995].

Evidence for clonality in *T. cruzi* is more convincing with a few different genotypes being repeatedly sampled from a number of different countries [Tibayrenc *et al.*, 1986]. For *T. brucei,* some studies supporting the theory of clonality have been reported, for example Mihok analysed a large number of samples from Kenya and found deviation from Hardy Weinberg (HW) [Mihok *et al.*, 1990]. The clonal theory for the population structure of *T. brucei*, in which little sexual recombination is occurring, is diagrammatically represented in Figure 9.3A where a dendrogram of genotype similarity would resemble an evolutionary tree. Another population structure has been proposed in which sexual recombination is occurring but only between isolates from within the same lineage Figure 9.3B. This population structure should be revealed by analysing samples from a sympatric population and using polymorphic markers. Analysis with markers which are not very polymorphic could group samples into lineages, although this does not mean they are clonal lineages. Furthermore, if samples are combined from several different lineages, linkage disequilibrium would be observed.

The epidemic population structure proposed by Maynard-Smith is represented in Figure 9.3C, whereby one or two successful genotypes clonally expand, overshadowing the sexual nature of the underlying population.

In order to investigate the population structure of *T. brucei* and to investigate the relationship between *T. b. brucei* and *T. b. rhodesiense*, 116 field samples from 5 different

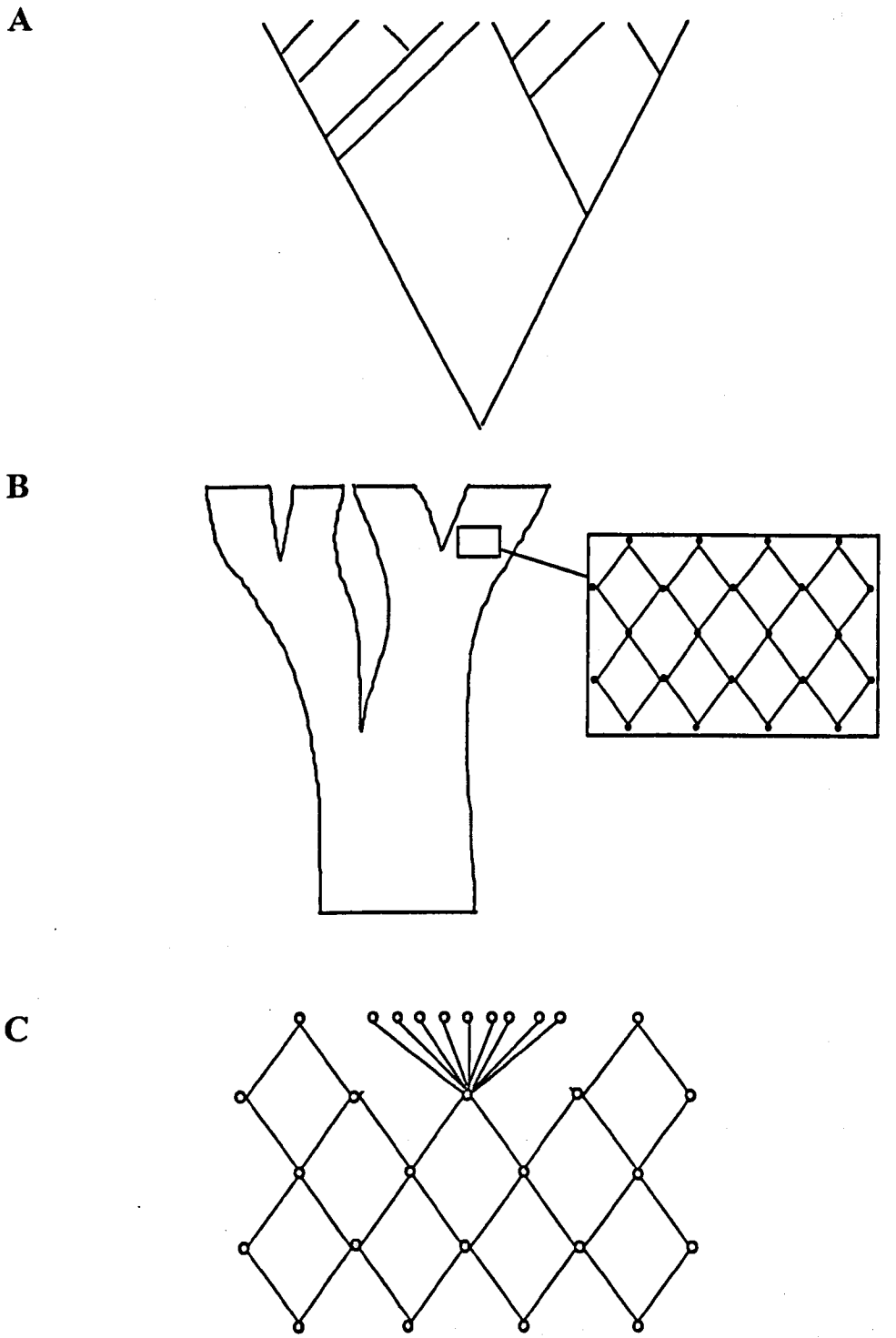


Figure 9.3. Representations of population structures from Maynard-Smith *et al.* (1993).
(A) A clonal population structure. No recombination occurs between isolates in the same or different lineages.
(B) Recombination occurs between isolates in the same lineage but not between isolates from different lineages. Therefore the structure within the branches is net-like, as represented by the expanded section of one of the branches.
(C). An epidemic population structure. There is frequent recombination between all members of the population, giving a net-like structure, but occasionally a 'successful' individual arises and increases rapidly to produce an epidemic clone.

geographical regions were analysed using hypervariable minisatellite markers (Chapter 7 and 8). The specific questions which were addressed were:

1) Are *T. b. brucei* populations panmictic, epidemic or clonal and do *T. b. rhodesiense* populations have the same structure?

2) Is there sub-structuring in *T. brucei* populations due to geographical barriers and host specificities?

3) Is it possible to differentiate between *T. b. rhodesiense* and *T. b. brucei* using a highly informative typing system? Are *T. b. rhodesiense* and *T. b. brucei* different sub-species?

4) Did human infectivity arise independently in different areas or do *T. b. rhodesiense* stocks constitute a monophyletic group?

6) Is it possible to track strains through populations?

The main conclusions from this analysis are presented and discussed below.

T. brucei has an epidemic population structure, whereby the underlying sexual nature of the population was obscured by the clonal expansion of one or two genotypes. The clonally expanded strains are associated with human infectivity, and appear to be highly related. This suggests that *T. b. rhodesiense* may have a clonal population structure and/or that *T. b. rhodesiense* stocks are restricted to sexual recombination within the *T. b. rhodesiense* lineage (resulting in homozygotes, which have been observed). The *T. b. brucei* stocks appear to be panmictic within the *T. b. brucei* lineage, with little sexual recombination between *T. b. rhodesiense* and *T. b. brucei* stocks in the Busoga focus (data is not available for other foci). Therefore it would appear that there is a different population structure for *T. b. rhodesiense* and *T. b. brucei* populations and by combining data from both populations and treating them as a single population an epidemic population structure emerges. These conclusions support the findings of Hide *et al.*, (1994). The results may explain the conclusions of clonal or epidemic population structures presented by other researchers as the majority of samples in most studies are from humans and are combined with some isolates from animals [Maynard-Smith *et al.*, 1993; Stevens and Wellburn, 1993; Truc and Tibayrenc, 1993], whereas the evidence for frequent mating was derived from samples isolated from tsetse flies, which are likely to harbour more *T. b. brucei* than *T. b. rhodesiense* stocks [Tait, 1980].

The limited genetic exchange between *T. b. rhodesiense* and *T. b. brucei* could arise during an epidemic as *T. b. rhodesiense* stocks may be co-transmitted frequently due to a predominant fly-human-fly cycle. In fact, the fly-host cycle has received limited attention in the debate about the population structure of *T. brucei* yet it would seem to be a key factor. An alternative explanation could be that *T. b. rhodesiense* and *T. b. brucei* are

becoming reproductively isolated due to a biological barrier to mating. Some evidence from laboratory experiments indicates that successful crosses between *T. b. rhodesiense* stocks and *T. b. brucei* stocks are the most difficult to obtain and there is a high incidence of trisomy and triploidy involved in such crosses [Chapter 5 and Gibson *et al.*, 1994], perhaps pointing to a partial genetic incompatibility between the two sub-species or a reduction in the ability of *T. b. rhodesiense* to undergo meiosis.

T. b. rhodesiense samples from Zambia and Uganda were quite distinct suggesting that *T. b. rhodesiense* is not a monophyletic group *i.e.* that *T. b. rhodesiense* stocks from Zambia and Uganda have not arisen from a single common ancestor which spread through East Africa, but that human infectivity arose independently in different lineages more than once. This poses a number of questions: how frequently does human infectivity arise, has human infectivity arisen independently in each focus and is the mechanism for human infectivity the same in each population? It is possible that the mechanisms for human infectivity in the Zambian population are different from those in Uganda as the nature of the disease is different in these countries (acute in Uganda and low virulence in Zambia [Hide, 1999]). These issues are of fundamental importance as human infectivity is the focus of much research. However, if trypanosomes in each focus have a different mechanism of human infectivity then samples from different foci must be examined. One mechanism of human infectivity was shown to involve the serum resistance-associated gene (*SRA*), which, when transfected into a *T. b. brucei* strain, can confer human serum resistance [Van Xong *et al.*, 1998]. However this is probably only one of several mechanisms of resistance as the *SRA* gene has been shown not to be expressed in at least two human infective strains of different geographical origins [Lindergard, 1999]. There remains another fundamental question to be addressed: once human infectivity has arisen in a lineage can this trait be spread to non-human infective samples by genetic exchange? From the data presented here, this is unlikely in the Busoga focus, but may not hold true in other foci.

Human infective genotypes have existed in the same focus for some time. Evidence from Busoga suggests that the same human infective genotype has been circulating in that area for many years, suggests that epidemics are caused by an upsurge in the numbers of pre-existing human infective stocks and are not caused by new genotypes. Genotypes which are associated with human infectivity have been observed in an area where no sleeping sickness has been reported (Kiboko), suggesting that human infective stocks are circulating in the area possibly posing a threat to humans in that area. It would be important to test each of these samples which have been identified as possible human

infective isolates, for human serum resistance using the blood incubation infectivity test (BIIT) to verify that minisatellite analysis can identify human infective trypanosomes in a population.

It is possible to distinguish *T. b. rhodesiense* from *T. b. brucei* for the Busoga population using the highly informative markers and in particular MVR-PCR of the minisatellite *MS42*. Since there are an average of 30 repeats in an allele and each one can be one of three types; a, g or n, approximately 3^{30} different alleles could be distinguished (if they existed) using this method. Therefore alleles that are identical are likely to be so because they are related. Using this system it was possible to identify a *T. b. rhodesiense* genotype for the Ugandan and Kenyan foci, however, a separate genotype is prevalent for the Zambian focus. Using this marker system it may be possible to identify the major human infective genotype for each focus and to use this to track the spread of individual foci. A similar MVR-PCR typing system has been developed for *P. falciparum,* although only a few samples have been analysed in this way [Arnot *et al.*, 1993]. The more popular typing system in *P. falciparum* is based on minisatellite sequences within the merozoite surface antigens *msp1* and *msp2* [Felger *et al.*, 1999].

From the results presented substantial population sub-structuring was observed, not only due to host specificities (humans in particular), but also due to geography, with few alleles being shared between populations from different geographical areas. For this reason, studies into the population structure of *T. brucei* in which samples are combined from different geographical areas [Tibayrenc *et al.*, 1990] are flawed.

The analysis of population samples presented here is based on a small number of isolates from five populations. Although some useful information can be obtained from such analyses, a more rigorous sampling strategy must be employed in order to obtain a full resolution to the questions posed. A large collection of samples needs to be obtained from a range of hosts and vectors from the same area at the same time and the same types of samples from different foci and endemic regions. In this way it should be possible to identify how many times human infectivity has arisen. It is clear that Ugandan and Zambian samples are distinct, is this also true of other areas such as Tanzania or West African *T. b. rhodesiense*? Evidence from a number of markers used to investigate a Tanzanian focus suggest that this focus is distinct from the Ugandan and Zambian foci and may have acquired human infectivity independently [Komba *et al.*, 1997].

In order to understand fully the factors involved in natural *T. brucei* populations information on the feeding preferences of tsetse flies for each focus would be valuable.

From such analysis it should be possible to determine which are the major reservoir hosts for human infective trypanosomes for each focus and whether strategies such as treating cattle for the disease to prevent the spread of human trypanosomiasis are viable. Since new strains are probably not the cause of new epidemics, other reasons for outbreaks need to be investigated, for example increases in fly numbers, or closer contact between tsetse and humans. In such analyses it should be possible to identify the most important factors in each focus.

In conclusion, the work presented here illustrates how mini- and microsatellites can be used to investigate questions regarding the basic biology of trypanosomes, which could be applied to other parasitic protozoa. Coupled with the sensitivity of PCR, minisatellite analysis can be used to genotype individual trypanosomes; distinguish hybrids and selfers from parental stocks; identify field isolates that are of mixed genotype; allow the identification and tracking of individual strains in populations and can provide insights into the evolution of these parasites. With adequate sampling, these markers can reveal the answers some of the many fundamental questions facing parasitologists today.

Appendix

Table A1.A STIB 247 x TREU 927/4

Clones	GUP no.	Lysate no.	Markers				Comments
			CRAM	*292*	*MS42*	JS2	
Reference 247			1-1	5-5	5-5	5-6	
Reference 927/4			3-4	3-4	3-4	3-4	
F124/28 bscl A1		68	1-3	3-5	3-5	3-5	Hybrid
F124/28 bscl B3		73	1-4	3-5	4-5	4-6	Hybrid
F124/28 bscl A6		72	1-3-4	3-4-5	3-4-5	nd	Mix or triploid
F124/28 bscl C5		71	1-4	3-5	4-5	4-6	Hybrid
F532/53 mcl 1	3114	107	1-3	3-5	3-4-6	4-5	Hybrid- mix or trisomic
F532/63 bscl 2		80	1-4	3-5	3-5	3-6	Hybrid
F532/63 bscl 3		78	1-4	4-5	4-5	4-5	Hybrid
F532/63 bscl 5		81	1-3	3-5	4-5	3-6	Hybrid
F532/63 bscl 7		79	1-4	4-5	4-5	3-6	Hybrid
F532/63 bscl 8		82	1-4	4-5	4-5	3-5	Hybrid
F532/72 mcl 1	3128, 4391	84	1-4	4-5	3-5	3-5	Hybrid
F532/72 mcl 2	3129	85	1-4	4-5	3-5	4-5	Hybrid
F532/72 mcl 3	3130	86	1-4	4-5	4-5	4-5	Hybrid
F532/72 mcl 4	3131	87	1-4	4-5	3-5	3-5	Hybrid
F532/72 mcl 5	3132, 4392	88	1-4	4-5	4-5	3-5	Hybrid
F532/72 mcl 6	3133, 4393	89	1-3	3-5	3-5	4-5	Hybrid
F532/72 mcl 7	3134	90	1-4	4-5	4-5	4-6	Hybrid
F532/72 mcl 9	3136, 4366	92	1-4	4-5	4-5	3-6	Hybrid
F532/72 mcl 10	3135, 4360	91	1-4	3-5	4-5	4-5	Hybrid
F974/70 mcl 4	3086	77	1-4	3-5	4-5	3-5	Hybrid
F124/28 bscl B1		69	3-4	3-4	3-4	nd	Parental
F124/28 bscl C3		75	3-4	3-4	3-4	nd	Parental
F124/28 bscl C2		70	3-4	3-4	3-4	nd	Parental
F974/70 mcl 1	3083	93	1-1	5-5	5-5	5-6	Parental
F974/70 mcl 2	3084	94	1-1	5-5	5-5	5-6	Parental
F974/70 mcl 3	3085	95	1-1	5-5	5-5	5-6	Parental
F974/70 mcl 6	3088	96	3-4	3-4	3-4	3-4	Parental
F974/70 mcl 7	3089	97	3-4	3-4	3-4	3-4	Parental
F974/70 mcl 8	3090	98	1-1	5-5	5-5	5-6	Parental
F974/78 mcl 1	3092	99	3-4	3-4	3-4	3-4	Parental
F974/78 mcl 2	3093	100	1-1	5-5	5-5	5-6	Parental
F974/78 mcl 3	3094	101	1-1	5-5	5-5	5-5	Selfer
F974/78 mcl 4	3095	102	3-4	3-4	3-4	3-4	Parental
F974/78 mcl 6	3096	105	1-1	5-5	5-5	5-6	Parental
F974/78 mcl 7	3097	106	1-1	5-5	5-5	5-6	Parental

Table A1.B STIB 386 x TREU 927 /4

Clones	GUP no.	Lysate no.	Markers				Comments
			CRAM	*292*	*MS42*	JS2	
Reference 386			1-2	1-2	1-2	1-2	
Reference 927/4			3-4	3-4	3-4	3-4	
F296/44 bscl 1	3199	2	2-3	2-3	1-3	1-3	Hybrid
F296/44 bscl 2	3200, 4261	3	1-2-3	1-2-4	1-2-3	1-2-3	Triploid
F296/44 bscl 3	3201	6	1-3	1-3	1-3	1-4	Hybrid
F296/44 bscl 4	3204	5	1-4	1-4	1-4	1-3	Hybrid
F296/44 bscl 7	3202	4	1-2-3	1-2-4	1-2-3	1-2-3	Triploid
F296/44 bscl 8	3205	7	1-4	1-4	1-4	1-3	Hybrid
F296/44 bscl 9	3206	8	1-4	1-4	1-4	1-3	Hybrid
F296/44 bscl 12	3211	9	1-3	2-3	1-2-4	2-3	Hybrid- mix or trisomic
F296/44 bscl 13	3209	10	1-4	1-4	1-4	1-3	Hybrid
F296/56 mcl 6	3215	1	2-3	2-3	1-3	1-3	Hybrid
F296/39 bscl 2			1-3	1-3	1-3	1-4	Hybrid
F296/39 bscl 5			1-3	1-3	1-3	1-4	Hybrid
F296/39 bscl 6			1-4	1-4	1-4	1-3	Hybrid
F296/39 bscl 7			1-4	2-3	1-3	2-3	Hybrid
F296/39 bscl 9			1-4	2-3	1-3	2-3	Hybrid
F296/39 bscl 12			1-3	1-3	1-3	1-4	Hybrid
F296/39 bscl 15/1			1-4	2-3	1-3	2-3	Hybrid
F296/39 bscl 17			1-3	2-3	1-2-4	2-3	Trisomic
F296/39 bscl 19			1-3	1-3	1-3	1-4	Hybrid
F296/39 bscl 27			1-4	1-4	1-4	1-3	Hybrid
F296/39 bscl 22/1			1-4	1-3	1-3	1-4	Hybrid
F296/39 bscl 24/1			1-4	2-3	1-3	2-3	Hybrid
F296/39 bscl 47			1-4	1-4	1-4	1-3	Hybrid
F296/42 bscl 26			1-3	1-3	1-3	1-4	Hybrid
F296/42 bscl 29			1-3	1-3	1-3	1-4	Hybrid
F296/42 bscl 41			1-3	1-3	1-3	1-4	Hybrid
F296/42 bscl 44			1-3	1-3	1-3	1-4	Hybrid
F296/42 bscl 47			1-3	1-3	1-3	1-4	Hybrid
F296/42 bscl 48			1-3	1-3	1-3	1-4	Hybrid
F296/42 bscl 49			1-3	1-3	1-3	1-4	Hybrid
F296/42 bscl 5	3203	22	3-4	3-4	3-4	3-4	Parental
F296/42 bscl 6	3210	23	3-4	3-4	3-4	3-4	Parental
F296/42 bscl 11	3208	24	3-4	3-4	3-4	3-4	Parental
F296/46 mcl 1	3196	12	3-4	4-4	3-4	3-4	Selfer
F296/46 mcl 2	3197	13	3-4	3-4	3-4	3-4	Parental
F296/56 mcl 1	3218	14	3-4	3-4	3-4	3-4	Parental
F296/56 mcl 2	3219	15	3-4	3-4	3-4	3-4	Parental
F296/56 mcl 3	3212	16	3-4	3-4	3-4	3-4	Parental
F296/56 mcl 4	3213	17	3-4	3-4	3-4	3-4	Parental
F296/56 mcl 5	3214	18	3-4	3-4	3-4	3-4	Parental
F296/56 mcl 7	3220	19	nd	3-4	3-4	3-4	Parental
F296/56 mcl 8	3216	20	3-4	3-4	3-4	3-4	Parental
F296/56 mcl 9	3217	21	3-4	3-4	3-4	nd	Parental
F296/39 bscl 1			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 4			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 8			nd	nd	3-4	3-4	Parental
F296/39 bscl 10			3-4	3-4	3-4	3-4	Parental

Table A1.B STIB 386 x TREU 927 /4 continued

Clones	GUP no.	Lysate no.	Markers				Comments
			CRAM	*292*	*MS42*	JS2	
F296/39 bscl 11			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 13			nd	3-4	3-4	3-4	Parental
F296/39 bscl 14			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 16			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 18			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 32			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 40			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 41			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 44			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 58			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 59			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 22/2			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 23/3			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 26/2			3-4	3-4	3-4	3-4	Parental
F296/39 bscl 57			nd	nd	3-4	3-4	Parental

Table A1.C STIB 247 x STIB 386

Clones	GUP no.	Lysate no.	Markers				Comments
			CRAM	*292*	*MS42*	JS2	
Reference 247			1-1	5-5	5-5	5-6	
Reference 386			1-2	1-2	1-2	1-2	
723VI-L		153	nd	nd	1-2-5	nd	Triploid
723CAB		154	nd	nd	1-2-5	nd	Triploid
F9/34 mcl 1		37	1-1	2-5	1-5	2-5	Hybrid
F9/45 mcl 2	3300	30	1-2	1-5	2-5	1-5	Hybrid
F9/45 mcl 7		32	1-2	1-5	1-5	1-5	Hybrid
F9/45 mcl 9	3290	33	1-1	2-5	1-5	1-6	Hybrid
F9/45 mcl 10	3291	34	1-1	2-5	1-5	1-6	Hybrid
F9/45 mcl 11	3287	35	1-2	1-5	1-5	1-5	Hybrid
F9/45 mcl 12	3288, 3296	36	1-1	2-5	1-2-5	1-5	Trisomic
F57/50 mcl 2		167	nd	nd	1-2-5	nd	Hybrid -mix or trisomic
F492/50 mcl 12	2843	39	1-2	1-5	1-5	2-6	Hybrid
F492/50 mcl 13	2856	40	1-1	2-5	1-5	1-6	Hybrid
F492/50 bscl 1			1-1	2-5	1-5	2-6	Hybrid
F492/50 bscl 2			1-1	1-5	1-5	2-6	Hybrid
F492/50 bscl 3			1-2	1-5	1-5	2-6	Hybrid
F492/50 bscl 4			1-1	2-5	1-5	2-6	Hybrid
F492/50 bscl 5/1			1-2	1-5	1-5	2-6	Hybrid
F492/50 bscl 6			1-2	1-5	1-2-5	1-5	Trisomic
F492/50 bscl 7			1-2	2-5	1-2-5	1-6	Hybrid -mix or trisomic
F492/50 bscl 8			1-1	2-5	1-5	1-6	Hybrid
F492/50 bscl 9			1-2	1-5	2-5	2-5	Hybrid
F492/50 bscl 11			1-2	1-5	1-5	2-5	Hybrid
F492/50 bscl 12			1-1	2-5	2-5	2-5	Hybrid
F492/50 bscl 13			1-1	2-5	1-5	1-6	Hybrid
F492/50 bscl 14			1-1	1-2-5	1-5	1-6	Trisomic
F492/50 bscl 15			1-1	2-5	1-5	1-6	Hybrid
F492/50 bscl 16			1-2	1-2-5	1-2-5	nd	Hybrid -mix or triploid

Table A1.C STIB 247 x STIB 386 continued

Clones	GUP no.	Lysate no.	Markers				Comments
			CRAM	*292*	*MS42*	JS2	
Reference 247			1-1	5-5	5-5	5-6	
Reference 386			1-2	1-2	1-2	1-2	
F492/50 bscl 17			1-2	1-5	1-5	2-5	Hybrid
F492/50 bscl 18			1-2	1-5	1-5	2-6	Hybrid
F492/50 bscl 19			1-1	2-5	1-5	2-6	Hybrid
F492/50 bscl 20			1-1	1-5	1-5	2-6	Hybrid
F492/50 bscl 21			1-1	1-5	1-5	2-6	Hybrid
F492/50 bscl 23			1-2	1-5	1-5	2-5	Hybrid
F492/50 bscl 25/1			1-1	2-5	1-5	2-6	Hybrid
F492/50 bscl 27			1-1	2-5	2-5	2-5	Hybrid
F9/28 mcl 3	2846	43	1-1	5-5	5-5	5-6	Parental
F9/34 mcl 2		159	1-1	5-5	5-5	5-6	Parental
F9/45 mcl 4	4295	31	1-1	5-5	5-5	5-6	Parental
F18/50 mcl 4		165	nd	nd	5-5	nd	Parental
F18/50 mcl 8		166	nd	nd	5-5	nd	Parental
F19/31 mcl 2		160	1-1	5-5	5-5	5-6	Parental
F19/31 mcl 3		161	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 1	2834	46	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 4	2838	49	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 5	2839	50	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 6	2840	51	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 7	2841	52	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 8	2847	56	1-1	5-5	5-5	5-5	Selfer
F492/50 mcl 9	2852	57	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 10	2853, 4312	58	1-1	5-5	5-5	5-5	Selfer
F492/50 mcl 11	2854, 4364	59	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 15	2857, 3365	61	1-1	5-5	5-5	5-5	Selfer
F492/50 mcl 16	2858, 3363	62	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 17	2859	63	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 18	2860, 3364	64	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 19		48	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 20	2861	65	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 21		55	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 22		53	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 23		54	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 24	2848	41	1-1	5-5	5-5	nd	Parental
F492/50 mcl 25	2855	60	1-1	5-5	5-5	5-6	Parental
F492/50 mcl 26	2862	66	1-1	5-5	5-5	6-6	Selfer
F492/50 bscl 22			1-1	5-5	5-5	5-6	Parental

Table A1.A-C. Minisatellite analysis of all trypanosome clones derived from genetic crosses. (**A**) Cross STIB 247 x TREU 927/4. (**B**)Cross STIB 386 x TREU 924/4. (**C**) Cross STIB 247 x STIB 386. Column 1 indicates the hybrid clone identification number used throughout this study. Columns 2 and 3 give additional identification numbers *i.e.* stabilate and lysate numbers. Columns 4-7 give the results of mini- and microsatellite analysis. Alleles are numbered 1-5 for the minisatellites, *CRAM*, *292* and *MS42*, and 1-6 for the microsatellite JS2. Column 8 gives the interpretation of results for each clone. Clones exhibiting three band patterns for one of the markers were considered either a mix or trisomic (lightly shaded rows). Clones which maintained their three band pattern after recloning were considered trisomic (shaded rows). Clones were considered triploid if a three band pattern was obtained for all markers analysed (shaded rows). GUP, Glasgow University Parasitology; nd, not determined.

Table A2.A

Minisatellite genotypes of *T. brucei* samples						
Zambia stocks	HS	Host	*MS42*	*CRAM*	*292*	Multilocus
z210	R	M	32/26b	68/45	55/32	1
z269	ND	DA	32/26b	68/45	55/32	1
z90 (Ch)	R	M	32/26b	59/59	32/32	2
z274	R	M	32/26b	68/45	55/32	1
z194 (Ch)	R	M	45/32	53/45	55/32	3
z218 (Ch)	R	M	47/42	53/45	55/51	4
z222	R	M	30f/26b	68/45	55/32	5
z208	R	M	32/26b	ND	ND	ND
z199	R	M	32/26b	68/45	55/32	1
z220	R	M	32/26b	68/45	55/32	1
z273	ND	DA	32/26b	ND	ND	ND
z203	R	M	32/26b	68/45	ND	ND
z221	R	M	32/26b	68/45	55/32	1
z231	R	M	32/26b	68/45	ND	ND
z244	R	M	32/26b	68/45	55/32	1
z267	ND	DA	32/26b	68/45	ND	ND
z212	R	M	32/26b	68/45	55/32	1
z270	ND	T	32/26b	68/45	55/32	1
z185 (Ch)	R	M	47/42	47/45	32/32	6

Table A2.B

Minisatellite genotypes of *T. brucei* samples

Nyanza stocks	HS	Host	*MS42*	*CRAM*	*292*	Multilocus
N 97	R	M	35a/34	61/55	76/44	7
N 96	R	M	35a/34	61/55	76/44	7
N 2340	R	M	35a/34 (38e,30e)	61/61	76/44	8
N 156	R	M	35a/34	61/55	76/44	7
N 95	R	M	35a/34	61/55	76/44	7
N 94	R	M	35a/34	61/55	76/44	7
N 106	R	M	35a/34	61/55	76/44	7
N 116	R	M	35a/34	61/55	76/44	7
N 7	ND	T	35c/28	50/36	51/44	9
N 110	R	M	35a/34	61/55	76/44	7
N 98	R	M	35a/34	61/61	76/44	8
N 111	R	M	34/33b	61/36	76/36	10
N 112	R	M	34/33b	61/36	36/36	11
N 115	R	M	34/33b	61/36	76/36	10
N 502	R	M	33b/33b	64/36	36/35	12
N 113	R	M	35a/34	61/55	76/44	7
N 120	R	M	35a/34	61/55	76/44	7
N 605	R	M	35a/34	61/55	76/44	7
N 609	R	M	35a/34	61/55	76/44	7
N 102	R	M	35a/33b	61/55	76/44	13
N 105	R	M	35a/34	61/55	76/44	7
N 149	R	M	33b/21	36/36	44/36	14
N 148	R	M	33b/33b	64/36	36/35	12
N 18	ND	T	33b/27d	45/36	76/44	15
N 118	R	M	ND	61/36	76/35	ND
N 103	R	M	35a/34	61/55	76/44	7

Table A2.C

Minisatellite genotypes of *T. brucei* samples

Busoga stocks	HS	Host	*MS42*	*CRAM*	*292*	Multilocus
B EA 174 un	R	M	35a/34	61/55	76/44	7
B EA 3 un	R	T	35a/34	61/55	76/44	7
B UTAR 3 un	R	M	35a/34	61/55	76/44	7
B UTAR 4 un	R	M	35a/34	61/55	76/44	7
B EA 2274 un	R	M	35d/34	61/61	76/44	16
B papol 60 un	S	C	38b/26a	115/55	76/76	17
B B76 un	S	C	33b/33b	55/50	44/44	18
B B135 un	S	C	33b/33b	ND	44/44	ND
B I155 un	S	C	36b/33b	36/36	44/44	19
B M12 un	S	C	33b/33b	55/50	44/44	18
B U89/8 un	R	M	34/34	61/55	76/76	20
B UgL un	R	M	35a/34	61/55	76/76	21
B UgA90 un	R	M	35a/34	61/55	76/76	21
B EO un	R	M	35a/34	61/55	76/76	21
B UgI un	R	M	35a/34	61/55	76/76	21
B UgK un	R	M	35a/34	61/55	76/76	21
BURI un	R	T	35a/34	61/55	76/76	21
B UgJ un	R	M	35a/34	61/55	76/76	21
B FLY 48 un	ND	T	33b/33b	61/55	76/76	22
B U89/2 un	R	M	35a/34	61/55	76/76	21
B UgM un	R	M	35a/34	61/55	76/76	21
B MAP un	R	M	35a/34	61/55	76/76	21
B 3194 un	R	M	35a/34	61/55	76/76	21
B 3196 un	R	M	35a/34	61/55	76/76	21
B 3200 un	R	M	35a/34	61/55	76/76	21
B 3202 un	R	M	35a/34	61/55	76/76	21
B 3203 un	R	M	35a/34	61/55	76/76	21
B M66 un	R	C	35a/34	61/55	76/76	21
B M80 un	S	C	36*/33b	55/36	44/29	27
B M85 un	ND	C	36*/33b	55/36	44/29	27
B papol 33 un	S	C	38d/26a	61/36	46/42	23
B papol 103un	R	C	35a/34	61/55	76/76	21
B S 14 un	S	C	33b/33b	ND	ND	ND
B 3205 un	R	M	35a/34	61/55	76/76	21
B 3206 un	R	M	35a/34	61/55	76/76	21

Table A2.C continued

Minisatellite genotypes of *T. brucei* samples						
Busoga stocks	HS	Host	*MS42*	*CRAM*	*292*	Multilocus
B B23 un	S	C	38c/26a	55/55	76/76	24
B B25 un	ND	C	38d/33b	36/36	44/44	25
B I147 un	S	C	35b/33b	36/36	44/44	26
B Mag 18 un	S	C	35b/33b	36/36	44/44	26
B Mag 40 un	S	C	35*/33b	36/36	44/44	possible 26
Busoga samples- mixed						
B M3 un	ND	C	38d/33b	55/43/36	44/26	NA
B S28 un	ND	C	38d/33b	55/43/36	44/26	NA
B S38 un	ND	C	38d/33b/26a	55/55	44/34	NA
B UgC90 un	ND	M	35a/34/33b	61/55/50	76/44	NA
B M31 un	ND	C	35*/34/26a	61/55	76/76	NA
B M32 un	ND	C	34*/30d/23	55/50	44/32	NA
B M42 un	ND	C	35*/34/26a	61/55	76/76	NA
B UgB90 un	ND	M	35a/34*	61/55/50	76/44	NA
B UgE90 un	ND	M	35a/34/33b	61/55/50	76/44	NA

Table A2.D

Minisatellite genotypes of *T. brucei* samples

Kiboko	HS	Host	*MS42*	*CRAM*	*292*	Multilocus
K 1337cl	ND	T	40/25a	83/83	48/44	28
K 1008cl	ND	T	35a/34	61/55	76/44	7
K 1027cl/un	ND	T	30a/17	61/52	60/50	29
K 927cl1B	ND	T	30c/20	68/61	41/50	30
K 927cl4B	ND	T	26c/25a	70/68	44/38	31
K 984cl	ND	T	30a/27c	73/55	41/50	32
K 854cl	ND	T	30a/27c	73/55	41/50	32
K 975cl	ND	T	35a/34	61/55	76/44	7
K 925cl	ND	T	29/27b	61/53	76/74	33
K 925 un	ND	T	38a/20	64/43	46/31	38
K 982cl	ND	T	35a/34	61/55	76/44	7
K 853cl	ND	T	35a/34	61/55	76/44	7
K 1009cl/un	ND	T	36a/25b	61/44	76/62	34
K 851cl	ND	T	35a/34	61/55	76/44	7
K 851 un	ND	T	30a/17	61/52	60/50	29
K 852cl	ND	T	35a/34	61/55	76/44	7
K 978cl	ND	T	30a/27c	73/55	51/39	35
K 978un	ND	T	26c/25b	63/55	53/40	39
K 997cl	ND	T	39a/25b	68/36	76/56	36
K 258cl	ND	T	30a/25c	64/43	56/56	40
K 936un	ND	T	35a/34	61/55	76/44	7
K 926un	ND	T	30a/27c	83/59	51/39	37
K 869 un	ND	T	35a/34	61/55	76/44	7
K 994cl/un	ND	T	35a/34	61/55	76/44	7
Kiboko uncloned samples- mixed						
K 981un	ND	T	39a/33a/30b/26c	73/59/52/43	76/60/53/44	NA
K 854un	ND	T	39b/26c/25a/20	73/70/64/61	51/44/41/38	NA
K 927un	ND	T	35a/34/30c/26c/25a/2C	70/68/61/59/52	76/50/44/41/38	NA
K 852un	ND	T	35a/34/30b/27b	64/52	76/74	NA
K 984 un	ND	T	39*/33*/30a/27c	73/59/52/43	76/60/53/41/35	NA
K 975 un	ND	T	35a/34	61/55	76/44/41/31	NA
K 997 un	ND	T	39a/30a/25a	73/61	76/56/55/53	NA
K 258 un	ND	T	30*/26*/25c	48/43	76/56/55/40	NA

Table A2.E

Minisatellite genotypes of *T. brucei* samples						
Lugala stocks	HS	Host	*MS42*	*CRAM*	*292*	Multilocus
L 834 un	ND	T	38d/27d	70/36	45/36	41
L 844 un	ND	T	35a/34	61/55	76/44	7
L 929 un	ND	T	33*/21	ND	45/44	ND
L 933 un	ND	T	33b/30d	50/36	76/45	42
L 934 un	ND	T	26a/22b	61/36	80/44	43
L 941 un	ND	T	38d/27d	70/36	45/36	41
L 832 un	ND	T	33b/26a	70/55	76/39	44
Lugala uncloned samples - mixed						
L 845 un	ND	T	31b/26a/21/16	98/73/70/50	76/50/44/11	NA
L 791 un	ND	T	33b/30d	70/57/53/50	76/44	NA
L 836 un	ND	T	33b/30d	70/57/53/50	76/44	NA
L 944 un	ND	T	31a/16	98/73	48/40/11	NA

Tables A2.A-E Minisatellite analysis of *T. brucei* samples. Column 1 indicates the isolate number of cloned (cl) and uncloned (un) samples. Samples were cloned unless otherwise stated. Column 2 indicates human serum sensitivity; R, resistant, S, sensitive and ND not determined. Column 3 gives details of the host from which each sample was isolated; M, man; C, cattle, DA, domestic animal and T, tsetse fly. Column 4 presents the genotype of each isolate for the marker *MS42* based on the number of repeats and on MVR internal maps of each allele (see chapter 8). * indicates that allele size (the number of repeats) only was used to genotype these isolates for *MS42*. These alleles were therefore not included in the MVR analysis in Chapter 8. Columns 5 and 6 indicate the *CRAM* and *292* minisatellite genotypes presented as an estimate of the number of repeats in each allele, respectively. Band size measurements used to estimate the number of repeat units within an allele were calculated using the fixed bin method, +/- 2 average standard deviations i.e. 2.3% of band size. Column 7 shows the multilocus genotype, *i.e.* the combined results from the three minisatellites. ND, not determined, NA, not applicable.

TABLE A3.A-E **RFLP analysis of *T. brucei* samples**
Table A3.A

	RFLP genotypes		
Zambian Stocks	***Bgl*11**	***Hin*c11**	***Mbo*II**
z210	+ -	+ -	+ -
z269	+ -	+ -	+ -
z90	+ -	+ -	+ -
z274	+ -	+ -	+ -
z194	+ +	- -	- -
z218	+ +	- -	- -
z222	+ -	+ -	ND
z208	ND	ND	ND
z199	+ -	+ -	ND
z220	+ -	+ -	- -
z273	ND	ND	ND
z203	ND	ND	ND
z221	+ -	+ -	ND
z231	ND	ND	ND
z244	+ -	+ -	ND
z267	ND	ND	ND
z212	+ -	+ -	ND
z270	+ -	+ -	ND
z185	+ +	- -	ND

Table A3.B

	RFLP genotypes		
Nyanza stocks	***Bgl*II**	***Hin*cII**	***Mbo*II**
N 97	+ -	+ -	+ -
N 96	+ -	+ -	+ -
N 2340	+ -	+ -	+ -
N 156	+ -	+ -	+ -
N 95	+ -	+ -	+ -
N 94	+ -	+ -	+ -
N 106	+ -	+ -	+ -
N 116	+ -	+ -	+ -
N 7	+ -	- -	+ -
N 110	+ -	+ -	- -
N 98	+ -	+ -	- -
N 111	- -	+ -	- -
N 112	- -	+ -	- -
N 115	- -	ND	- -
N 502	- -	- -	- -
N 113	+ -	+ -	+ -
N 120	+ -	+ -	- -
N 605	+ -	+ -	+ -
N 609	+ -	+ -	+ -
N 102	+ -	+ -	+ -
N 105	+ -	+ -	+ -
N 149	+ -	- -	- -
N 148	- -	- -	- -
N 18	+ -	- -	- -
N 118	ND	ND	ND
N 103	+ -	+ -	- -

Table A3.C

	RFLP genotypes		
Busoga stocks	***Bgl*11**	***Hin*cII**	***Mbo*II**
B EA 174	+ -	+ -	ND
B EA 3	+ -	+ -	ND
B UTAR 3	ND	+ -	ND
B UTAR 4	+ -	+ -	+ -
B EA 2274	ND	+ -	- -
B papol 60	+ -	+ -	+ -
B B76	- -	- -	ND
B B135	- -	- -	ND
B I155	- -	+ -	- -
B M12	- -	- -	- -
B UgC90	+ -	+ -	- -
B UgE90	+ -	+ -	+ -
B U89/8	+ -	+ -	+ -
B UgL	+ -	+ -	+ -
B UgA90	+ -	+ -	+ -
B EO	ND	ND	+ -
B UgI	+ -	+ -	+ -
B UgK	ND	ND	ND
BURI	ND	+ -	ND
B UgJ	+ -	+ -	+ -
B FLY 48	+ -	+ -	+ -
B U89/2	+ -	+ -	- -
B UgB90	+ -	+ -	- -
B UgM	+ -	+ -	+ -
B MAP	+ -	+ -	+ -
B 3194	+ -	+ -	+ -
B 3196	+ -	+ -	+ -
B 3200	+ -	+ -	+ -
B 3202	+ -	+ -	+ -
B 3203	+ -	+ -	+ -
B M31	+ -	+ -	+ -
B M32	- -	- -	+ -
B M42	+ -	+ -	+ -
B M66	+ -	+ -	+ -
B M80	- -	+ -	+ -
B M85	- -	+ -	+ -
B papol 33	+ -	+ -	+ -
B papol 103	+ -	+ -	+ -
B S 14	ND	ND	ND
B 3205	+ -	+ -	+ -
B 3206	+ -	+ -	+ -
B B23	+ -	ND	+ -
B B25	- -	+ -	+ -
B I147	- -	ND	+ -
B Mag 18	- -	ND	+ -
B Mag 40	- -	ND	- -

B M3	- -	ND	+ -
B S28	- -	ND	+ -
T168	- -	ND	+ -
B S38	+ -	ND	+ -

Table A3.D

	RFLP genotypes		
Kiboko stocks	***Bgl*II**	***Hin*cII**	***Mbo*II**
K 1337cl	+ +	- -	- -
K 1008cl	+ -	+ -	ND
K1027cl	+ +	- -	- -
K 927cl1B	+ +	- -	- -
K 927cl4B	+ +	- -	- -
K 984cl	+ +	- -	- -
K 854cl	+ +	- -	- -
K 975cl	+ -	+ -	+ -
K 925cl	+ +	- -	- -
K 925 un	+ +	- -	- -
K 982cl	+ -	+ -	+ -
K 853cl	+ -	+ -	+ -
K 1009cl	+ +	- -	- -
K 851cl	+ -	+ -	+ -
K 851 un	+ +	- -	- -
K 852cl	+ -	+ -	+ -
K 978cl	+ +	- -	- -
K 978un	+ +	- -	- -
K 997cl	+ +	- -	- -
K 258cl	+ -	- -	- -
K 936un	ND	ND	- -
K 926un	+ +	- -	- -
K 869 un	+ -	+ -	+ -
K 994cl	ND	ND	ND

Table A3.E

	RFLP genotypes		
Lugala stocks	***Bgl*II**	***Hinc*II**	***Mbo*II**
L 834	+ -	+ -	+ -
L 844	ND	+ -	+ -
L 929	ND	- -	ND
L 933	- -	- -	+ -
L 934	+ -	- -	- -
L 941	+ -	+ -	+ -
L944(mix)	- -	+ +	+ +
L 832	+ -	- -	- -

Tables A3.A-E RFLP genotypes of *T. brucei* samples. Column 1 indicates the isolate number of cloned and uncloned (unmixed) samples. Column 2, 3 and 4 presents the genotype of each isolate for the *Bgl*II, *Hinc*II and *Mbo*II RFLPs respectively.

Table A4

MS42 alleles defined by MVR mapping	Allele Frequencies for *MS42* for each of the *T. brucei* populations						
	populations						
	Kiboko (tsetse) n=65	Lugala (tsetse) n=23	Busoga n=97			Nyanza (human) n=52	Zambia (human) n=38
			HSR	HSS	ND		
16		0.087					
17	0.0308						
20	0.0462						
21		0.087				0.0192	
22a	0.0154						
22b		0.0435					
23					0.0103		
25a	0.0615						
25b	0.0462						
25c	0.0154						
26a		0.1304		0.0309	0.0309		
26b							0.42
26c	0.0615						
27a	0.0154						
27b	0.0308						
27c	0.0615						
27d		0.087				0.0192	
28						0.0192	
29	0.0154						
30a	0.1231						
30b	0.0308						
30c	0.0154						
30d		0.1304			0.0103		
30e						0.0192	
30f							0.0333
31a		0.0435					
31b		0.0435					
32							0.42
33a	0.0154						
33b		0.1739		0.1224	0.0918	0.1923	
34	0.1539	0.0435	0.268		0.0412	0.3654	
35a	0.1539	0.0435	0.2371		0.031	0.3269	
35b				0.0206			
35c						0.0192	
35d			0.0103				
36a	0.0154						
36b				0.0103			
38a	0.0154						
38b				0.0103			
38c				0.0103			
38d		0.087		0.0103	0.0408		
38e						0.0192	
39a	0.0303						
39b	0.0154						
40	0.0154						
42							0.0667
45							0.0333
47							0.0667

Table A4. The frequency of each *MS42* allele in each population. HSR, human serum resistant; HSS, human serum sensitive; ND, not determined. n= number of alleles sampled.

CRAM alleles defined by size	Allele Frequencies for *CRAM* for each of the *T. brucei* populations						
	populations						
	Kiboko (tsetse) n=67	Lugala (tsetse) n=26	Busoga n=99			Nyanza (mainly human) n=52	Zambia (mainly human) n=34
			HSR	HSS	ND		
36	0.0149	0.1538		0.1010	0.0505	0.1923	
43	0.0597				0.0202		
44	0.0149						
45						0.0192	0.4706
47							0.0294
48	0.0149						
50		0.1538		0.0202	0.0404	0.0192	
52	0.0896						
53	0.0149	0.0769					0.0588
55	0.1791	0.0769	0.2424	0.0606	0.1212	0.2885	
57		0.0769					
59	0.0597						0.0588
61	0.2238	0.0769	0.2626	0.0101	0.0606	0.4423	
63	0.0149						
64	0.0597					0.0385	
68	0.0299						0.3823
70	0.0299	0.2308					
73	0.0746	0.0769					
83	0.0448						
98		0.0769					
115				0.0101			

Table A5. The frequency of each *CRAM* allele in each population. HSR, human serum resistant; HSS, human serum sensitive; ND, not determined. n= number of alleles sampled.

292 alleles defined by size	Allele Frequencies for *292* for each of the *T. brucei* populations						
	populations						
	Kiboko (tsetse) n=71	Lugala (tsetse) n=25	Busoga n=96			Nyanza (mainly human) n=52	Zambia (mainly human) n=28
			HSR	HSS	ND		
11		0.08					
26					0.0208		
29				0.0104	0.0104		
31	0.0282						
32					0.0104		0.5357
34					0.0104		
35	0.0141					0.0577	
36		0.08				0.1346	
38	0.0282						
39	0.0282	0.04					
40	0.0282	0.04					
41	0.0563						
42				0.0104			
44	0.1831	0.24	0.0521	0.1562	0.1042	0.3846	
45		0.16					
46	0.0141			0.0104			
48	0.0141	0.04					
50	0.0704	0.04					
51	0.0422					0.0192	0.0357
53	0.0563						
55	0.0282						0.4286
56	0.0422						
60	0.0563						
62	0.0141						
74	0.0282						
76	0.2254	0.24	0.4687	0.0417	0.0937	0.4038	
80		0.04					

Table A6. The frequency of each *292* allele in each population. HSR, human serum resistant; HSS, human serum sensitive; ND, not determined. n= number of alleles sampled.

Figure A1

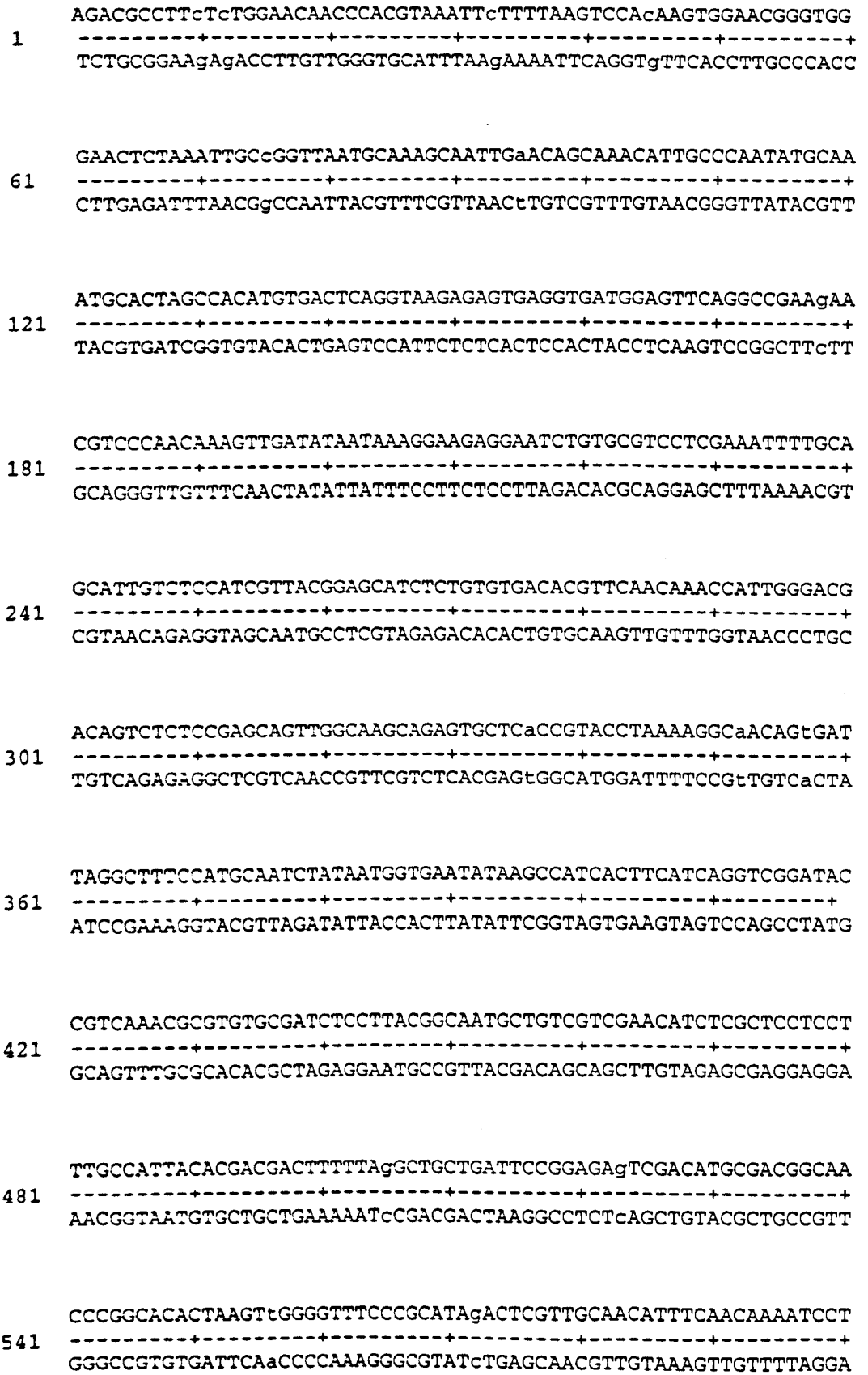

AGACGCCTTCTCTGGAACAACCCACGTAAATTCTTTTAAGTCCACAAGTGGAACGGGTGG
1
TCTGCGGAAGAGACCTTGTTGGGTGCATTTAAGAAAATTCAGGTGTTCACCTTGCCCACC
GAACTCTAAATTGCCGGTTAATGCAAAGCAATTGAACAGCAAACATTGCCCAATATGCAA
61
CTTGAGATTTAACGGCCAATTACGTTTCGTTAACTTGTCGTTTGTAACGGGTTATACGTT
ATGCACTAGCCACATGTGACTCAGGTAAGAGAGTGAGGTGATGGAGTTCAGGCCGAAGAA
121
TACGTGATCGGTGTACACTGAGTCCATTCTCTCACTCCACTACCTCAAGTCCGGCTTCTT
CGTCCCAACAAAGTTGATATAATAAAGGAAGAGGAATCTGTGCGTCCTCGAAATTTTGCA
181
GCAGGGTTGTTTCAACTATATTATTTCCTTCTCCTTAGACACGCAGGAGCTTTAAAACGT
GCATTGTCTCCATCGTTACGGAGCATCTCTGTGTGACACGTTCAACAAACCATTGGGACG
241
CGTAACAGAGGTAGCAATGCCTCGTAGAGACACACTGTGCAAGTTGTTTGGTAACCCTGC
ACAGTCTCTCCGAGCAGTTGGCAAGCAGAGTGCTCACCGTACCTAAAAGGCAACAGTGAT
301
TGTCAGAGAGGCTCGTCAACCGTTCGTCTCACGAGTGGCATGGATTTTCCGTTGTCACTA
TAGGCTTTCCATGCAATCTATAATGGTGAATATAAGCCATCACTTCATCAGGTCGGATAC
361
ATCCGAAAGGTACGTTAGATATTACCACTTATATTCGGTAGTGAAGTAGTCCAGCCTATG
CGTCAAACGCGTGTGCGATCTCCTTACGGCAATGCTGTCGTCGAACATCTCGCTCCTCCT
421
GCAGTTTGCGCACACGCTAGAGGAATGCCGTTACGACAGCAGCTTGTAGAGCGAGGAGGA
TTGCCATTACACGACGACTTTTTAGGCTGCTGATTCCGGAGAGTCGACATGCGACGGCAA
481
AACGGTAATGTGCTGCTGAAAAATCCGACGACTAAGGCCTCTCAGCTGTACGCTGCCGTT
CCCGGCACACTAAGTTGGGGTTTCCCGCATAGACTCGTTGCAACATTTCAACAAAATCCT
541
GGGCCGTGTGATTCAACCCCAAAGGGCGTATCTGAGCAACGTTGTAAAGTTGTTTTAGGA

Figure A1 continued

```
     TCTCAtGTACTACTGATGAAAAGGGCTGGTCgCGAAACAACAACCACCTAGCGTCAGATA
601  ---------+---------+---------+---------+---------+---------+
     AGAGTaCATGATGACTACTTTTCCCGACCAGcGCTTTGTTGTTGGTGGATCGCAGTCTAT

     ACCCGGCATAgGCGAGTTGCCgTGTGACAACATTGAGGGGAACACCACTTTCTGTCCTCT
661  ---------+---------+---------+---------+---------+---------+
     TGGGCCGTATcCGCTCAACGGcACACTGTTGTAACTCCCCTTGTGGTGAAAGACAGGAGA

     ctGGATACTCGATGGATCTAtAgAGGCGGCGAAAGAATCgCaTCAgCaCCtCaAAGGGtC
721  ---------+---------+---------+---------+---------+---------+
     gaCCTATGAGCTACCTAGATaTcTCCGCCGCTTTCTTAGcGtAGTcGtGGaGtTTCCCaG

     gAtGAgAcgACaTGAGCtCAGtcACtGCAAtGCAGCGAttCAAAAATGATAGCGCAAcTT
781  ---------+---------+---------+---------+---------+---------+
     cTaCTcTgcTGtACTCGaGTCagTGaCGTTaCGTCGCTaaGTTTTTACTATCGCGTTgAA

                                                    MS42-LR 5'GAA
     CTTCGCCACAAATCGACACTGGATCCAACACaGCTTTTACCACATTGTCGCTTCTCAGAA
841  ---------+---------+---------+---------+---------+---------+
     GAAGCGGTGTTTAGCTGTGACCTAGGTTGTGtCGAAAATGGTGTAACAGCGAAGAGTCTT
                                                          CTT

     CCTGTGAACCCTCCAAATTCG
     CCTGTGAACCCTCCAAATTCGAAATAACCGCGTCAACATATCGTTTTACATCACCCTCTa
901  ---------+---------+---------+---------+---------+---------+
     GGACACTTGGGAGGTTTAAGCTTTATTGGCGCAGTTGTATAGCAAAATGTAGTGGGAGAt
     GGACACTTGGGAGGTTTAAGC 5' MS42-L

     ACGGGGACAAGTCACAGCACCTTGTTAATATTGGTACCACATCCATTTCTGTAAAGTGAC
961  ---------+---------+---------+---------+---------+---------+
     TGCCCCTGTTCAGTGTCGTGGAACAATTATAACCATGGTGTAGGTAAAGACATTTCACTG

     CACCAAATACGTGCTCAACTTCAGCAGCAAATGCGGACGCGGCGTCACATAAAGTGCGAA
1021 ---------+---------+---------+---------+---------+---------+
     GTGGTTTATGCACGAGTTGAAGTCGTCGTTTACGCCTGCGCCGCAGTGTATTTCACGCTT

     CAAGGAAAGCTATCAATGATGTACGGCGCCTGTCCAGAATCGCGATACGCACAACACAGC
1081 ---------+---------+---------+---------+---------+---------+
     GTTCCTTTCGATAGTTACTACATGCCGCGGACAGGTCTTAGCGCTATGCGTGTTGTGTCG

     TGTTTTCGGCAACCCTCTTCAGTTCCTTGTATACATTGTCGGGCGTTAACTCCATGTCAT
1141 ---------+---------+---------+---------+---------+---------+
     ACAAAAGCCGTTGGGAGAAGTCAAGGAACATATGTAACAGCCCGCAATTGAGGTACAGTA
```

Figure A1 continued

```
     CTAATTCCCCCCTTTCCACAAGGCGAAAACTCAACTCACCCATATCTTTTTCCTTCAACC
1201 ---------+---------+---------+---------+---------+---------+
     GATTAAGGGGGGAAAGGTGTTCCGCTTTTGAGTTGAGTGGGTATAGAAAAAGGAAGTTGG

          MS42-RR 5'GGGCAGAGCGCATGTCCTCCAAT
     CACACTCTTCCGCAAGGGCAGAGCGCATGTCCTCCAAtGCACAGCCTGGACGAAGTACCA
1261 ---------+---------+---------+---------+---------+---------+
     GTGTGAGAAGGCGTTCCCGTCTCGCGTACAGGAGGTTaCGTGTCGGACCTGCTTCATGGT
                      CCCGTCTCGCGTACAGGAGGTTA 5' MS42-R

     ATGCCCGCATTACCTGTTGTGTCCCCcATGTGaAGAGACaTtgCAAGGGAAGAcCACTCG
1321 ---------+---------+---------+---------+---------+---------+
     TACGGGCGTAATGGACAACACAGGGGgTACACtTCTCTGtAacGTTCCCTTCTgGTGAGC

     TACAAGGGTGACCCCTTACCGCAGTAACGTTAATACAGcTGCGTTTcTTCCTTATTGCCT
1381 ---------+---------+---------+---------+---------+---------+
     ATGTTCCCACTGGGGAATGGCGTCATTGCAATTATGTCgACGCAAAgAAGGAATAACGGA

     CCCCCACCATTGCAGTAACGCCACTGGGGTTGCAtAACGGAACACCATTACTAtcaTAAA
1441 ---------+---------+---------+---------+---------+---------+
     GGGGGTGGTAACGTCATTGCGGTGACCCCAACGTaTTGCCTTGTGGTAATGATagtATTT

                         Ms42-QR 5'GAAGGGCGTTCAGGCATTCGTTC
     TAATTATGTTGTCCATATCAGTTAAATGAAGGGCGTTCAGGCATTCGTTCCGGAATGACT
1501 ---------+---------+---------+---------+---------+---------+
     ATTAATACAACAGGTATAGTCAATTTACTTCCCGCAAGTCCGTAAGCAAGGCCTTACTGA
                                    CTTCCCGCAAGTCCGTAAGCAAG 5'Ms42-Q

     CAAACACCTTtGGTGACGcTTTCTTCCACGTTAAAAAAAAGCGAACAGGAGCCTTCTTTT
1561 ---------+---------+---------+---------+---------+---------+
     GTTTGTGGAAaCCACTGCgAAAGAAGGTGCAATTTTTTTTCGCTTGTCCTCGGAAGAAAA

     CCAATGAACACGCGACCTCAACCACTACCTcGTTAGGAAGCTCCTCCTCAACATCGGAGC
1621 ---------+---------+---------+---------+---------+---------+
     GGTTACTTGTGCGCTGGAGTTGGTGATGGAgCAATCCTTCGAGGAGGAGTTGTAGCCTCG

                                                MS42-W 5'GGTGATTCATCGG
     AATTTTTCTTCCCGCTGATACATTCCGTTGCGTTTTTTGAAGTCAATGGTGATTCATCGG
1681 ---------+---------+---------+---------+---------+---------+
     TTAAAAAGAAGGGCGACTATGTAAGGCAACGCAAAAAACTTCAGTTACCACTAAGTAGCC
                                                CCACTAAGTAGCC

     CTCCCTTGCCA
     CTCCCTTGCCATTTTTTGCCGTCAcTGCATCCCCATCCCCGCCCTTAAACCCAgTAgGGA
1741 ---------+---------+---------+---------+---------+---------+
     GAGGGAACGGTAAAAAACGGCAGTgACGTAGGGGTAGGGGCGGGAATTTGGGTcATcCCT
     GAGGGAACGGA 5' MS42-WR
```

Figure A1 continued

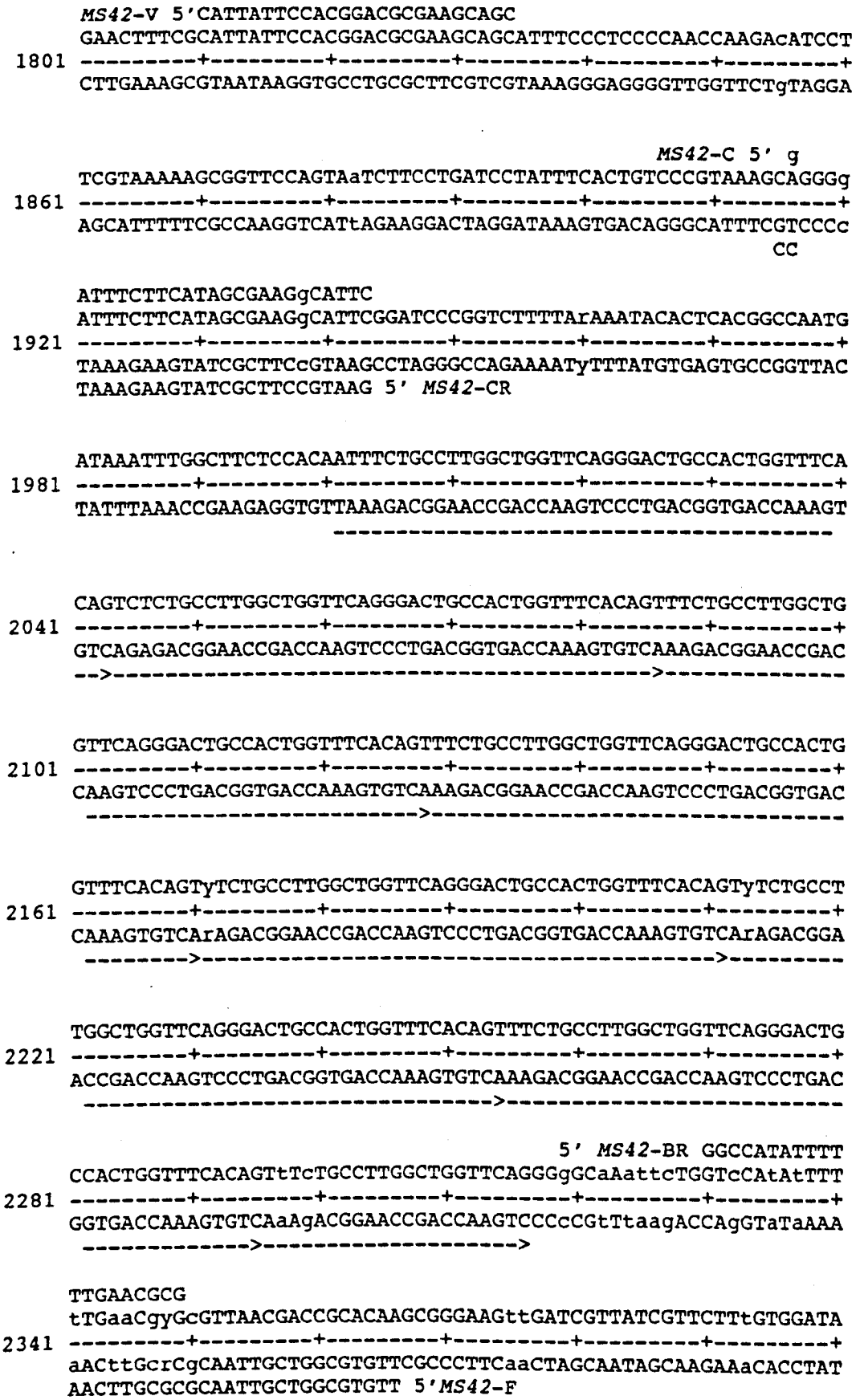


Figure A1 continued

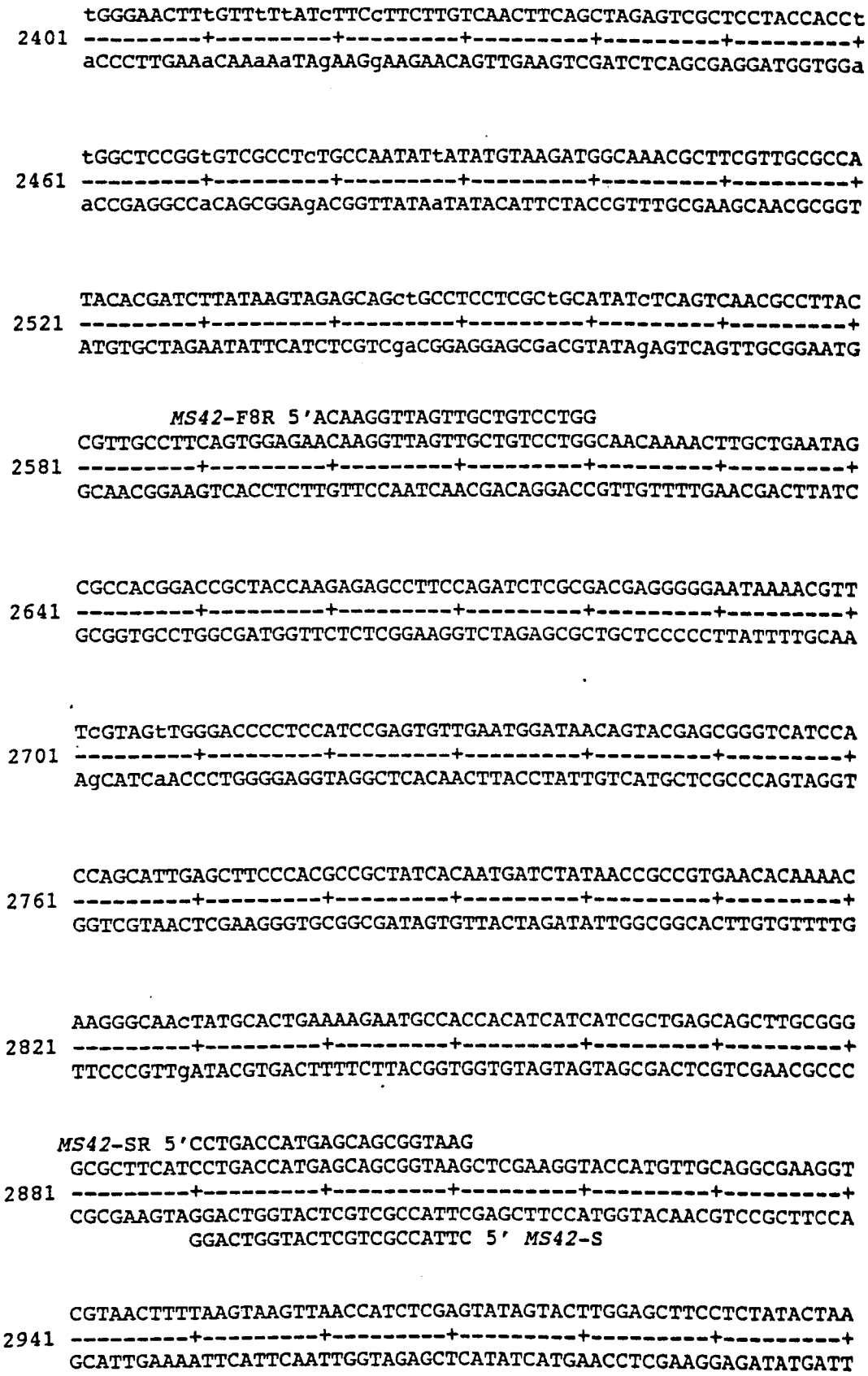


Figure A1 continued

```
     ATGAGCGGAGGGTACATTCATTAGGCCCCCTGTATACCTCTAACCtACACAAAAAATTGC
3001 ---------+---------+---------+---------+---------+---------+
     TACTCGCCTCCCATGTAAGTAATCCGGGGGACATATGGAGATTGGaTGTGTTTTTTAACG

     TGCGCATCGCCATGATGATAACTTTCCCCGAGCAGACAAATGAAAGGTAGAGTCACCCTG
3061 ---------+---------+---------+---------+---------+---------+
     ACGCGTAGCGGTACTACTATTGAAAGGGGCTCGTCTGTTTACTTTCCATCTCAGTGGGAC

     CGACTTCTGCAAAGAAGCTTGAAGGAAGAAGGGAGAAAATTTTCTGCAAGTGGTGAATAA
3121 ---------+---------+---------+---------+---------+---------+
     GCTGAAGACGTTTCTTCGAACTTCCTTCTTCCCTCTTTTAAAAGACGTTCACCACTTATT

     ACCGCTGCACGACTAAGCTTCACCCCTAGTTTCTCCACACAAACAAAAAGTATTCTGCAT
3181 ---------+---------+---------+---------+---------+---------+
     TGGCGACGTGCTGATTCGAAGTGGGGATCAAAGAGGTGTGTTTGTTTTTCATAAGACGTA

     TGTTAATAAAATAATCCACAGAATGAGATGGCACAACTTGAGGGAGGTGTGGGTTTCACA
3241 ---------+---------+---------+---------+---------+---------+
     ACAATTATTTTATTAGGTGTCTTACTCTACCGTGTTGAACTCCCTCCACACCCAAAGTGT

     GGCATATGTGAATGAACGCAAAAAAAAAAATAATTAAAAAACTGTATTTATGTTGTGCCCg
3301 ---------+---------+---------+---------+---------+---------+
     CCGTATACACTTACTTGCGTTTTTTTTTTATTAATTTTTTGACATAAATACAACACGGGc

     AAACGAGTGAAATCaTCAACAAAACCATCCCACAAGAAAGGAAGGGAGTaCAAGaGTATC
3361 ---------+---------+---------+---------+---------+---------+
     TTTGCTCACTTTAGtAGTTGTTTTGGTAGGGTGTTCTTTCCTTCCCTCAtGTTCtCATAG

     TTTCATTTTCATTACGCATGCATGTATCAaAgCCTGGGGAACCTGCAG
3421 ---------+---------+---------+---------+--------
     AAAGTAAAAGTAATGCGTACGTACATAGTtTcGGACCCCTTGGACGTC
```

Appendix Figure A1. Sequence of plasmid p42Sc3 containing the minisatellite *MS42*. Both strands of the nucleotide sequence of the insert of plasmid p42Sc3 are given. The sequence was determined using the ABI automated sequencing protocol. Seven full repeat units and part of one repeat unit were included and underlined. Small letters indicate that the consensus for a particular position was taken. ATG start site of the ORF is in bold. Some *MS42* primers are also given.

```
   1 MAMRSNFLCR LEVYRGPNEC TLRSFSIEEA PSTILEMVNL LKSYDLRLQH
  51 GTFELTAAHG QDEAPRKLLS DDDVVAFFSV HSCPCFVFTA VIDHCDSGVG
 101 SSMLVDDPLV LLSIQHSDGG VPTTKRFIPP RREIWKALLV AVRGAIQQVL
 151 LPGQQLTLFS TEGNGKALTE ICSEEAAALL IRSCMAQRSV CHLTYNIGRG
 201 DTGAKVVGAT LAEVDKKEDK NKVPISTKND NDQLPACAVV NARSKNMDQN
 251 LPPEPAKAET VKPVAVPEPA KAETVKPVAV PEPAKAETVK PVAVPEPAKA
       ------>--------------->--------------->-------------
 301 ETVKPVAVPE PAKAETVKPV AVPEPAKAET VKPVAVPEPA KAETVKPVAV
     >--------------->--------------->---------------->------
 351 PEPAKAEIVE KPNLSLAVSV FLKDRDPNAF AMKKSPALRD SEIGSGRLLE
     ------>
 401 PLFTKDVLVG EGNAASRPWN NAKVLPTGFK GGDGDAVTAK NGKGADESPL
 451 TSKNATECIS GKKNCSDVEE ELPNEVVVEV ACSLEKKAPV RFFLTWKKAS
 501 PKVFESFRNE CLNALHLTDM DNIIIYDSNG VPLCNPSGVT AMVGEAIRKK
 551 RSCINVTAVR GHPCTSGLPL QCLFTWGTQQ VMRALVLRPG CALEDMRSAL
 601 AEECGLKEKD MGELSFRLVE RGELDDMELT PDNVYKELKR VAENSCVVRI
 651 AILDRRRTSL IAFLVRTLCD AASAFAAEVE HVFGGHFTEM DVVPILTRCC
 701 DLSPLEGDVK RYVDAVISNL EGSQVLRSDN VVKAVLDPVS ICGEEVALSF
 751 LNRCIAVTEL MSSHRPFEVL MRFFRRLYRS IEYPERTESG VPLNVVTRQL
 801 AYAGLSDARW LLFRDQPFSS VVHEKDFVEM LQRVYAGNPN LVCRVAVACR
 851 LSGISSLKSR RVMAKEERDV RRQHCRKEIA HAFDGIRPDE VMAYIHHYRL
 901 HGKPNHCCLL GTVSTLLANC SERLSSQWFV ERVTQRCSVT METMLQNFED
 951 AQIPLPLLYQ LCWDVLRPEL HHLTLLPESH VASAFAYWAM FAVQLLCINR
1001 QFRVPTRSTC GLKRIYVGCS REGV
```

Appendix Figure A2. Predicted amino acid sequence of part of the *MS42* gene, beginning with the first methionine in the ORF. The repeat units are underlined.

References

Abaru, D.E. (1985). Sleeping sickness in Busoga, Uganda, 1876-1983. *Tropical Medicine and Parasitology* **36**:72-6.

Aksoy, S., Williams, S., Chang, S. and Richards, F.F. (1990). SLACS retrotransposon from *Trypanosoma brucei gambiense* is similar to mammalian LINEs. *Nucleic Acids Research* **18**:785-790.

Amarger, V., Gauguier, D., Yerle, M., Apiou, F., Pinton, P., Giraudeau, F., Monfouilloux, S., Lathrop, M., Dutrillaux, B., Buard, J. and Vergnaud, G. (1998). Analysis of distribution in the human, pig and rat genomes points towards a general subtelomeric origin of minisatellite structures. *Genomics* **52**:62-71.

Andreassen, R. and Olaisen, B. (1998). *De novo* mutations and allelic diversity at minisatellite locus *D7S22* investigated by allele-specific four-state MVR-PCR analysis. *Human Molecular Genetics* **7**:2113-2120.

Armour, J.A.L., Povey, S., Jeremiah, S. and Jeffreys, A.J. (1990). Systematic cloning of human minisatellites from ordered array charomid libraries. *Genomics* **8**:501-512.

Armour, J.A.L., Harris, P.C. and Jeffreys, A.J. (1993). Allelic diversity at minisatellite MS205 (D16S309): evidence for polarized variability. *Human Molecular Genetics* **2**:1137-1145.

Armour, J.A.L., Monckton, D.G., Neil, D.L., Tamaki, K., MacLeod, A., Allen, M., Crosier, M. and Jeffreys, A.J. (1993). Mechanisms of mutations at human minisatellite loci. *Genome Rearrangement and Stability* Eds. K. E. Davies and S. T. Warren. Cold Spring Harbour Laboratory Press. 43-57.

Armour, J.A.L., Anttinen, T., May, C.A., Vega, E.E., Sajantila, A., Kidd, J.R., Kidd, K.K., Bertranpetit, J., Paabo, S. and Jeffreys, A.J. (1996). Minisatellite diversity supports a recent African origin for modern humans. *Nature Genetics* **13**:154-160.

Arnot, D.E., Roper, C. and Bayoumi, R.A.L. (1993). Digital codes from hypervariable tandemly repeated DNA sequences in the *Plasmodium falciparum* circumsporozoite gene can genetically barcode isolates. *Molecular and Biochemical Parasitology* **61**:15-24.

Azen, E., Lyons, K.M., McGonigal, T., Barrett, N.L., Clements, L.S., Maeda, N., Vanin, E.F., Carlson, D.M. and Smithies, O. (1984). Clones from the human gene complex coding for salivary proline-rich proteins. *Proceedings of the National Academy of Science USA* **81**:5561-5565.

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

Barrett, M.P. and Le Page, R.W.F. (1993). A 6-phosphogluconate dehydrogenase gene from *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **57**:89-100.

Barrett, M., Bringaud, F., Doua, F., Melville, S. and Baltz, T. (1996). Hypervariability in gene copy number for the glucose transporter genes in trypanosomes. *Journal of Eukaryotic Microbiology* **43**:244-9.

Barrett, M.P., MacLeod, A., Tovar, J., Sweetman, J.P., Tait, A., Le Page, R.W.F. and Melville, S.E. (1997). A single locus minisatellite sequence which distinguishes between *Trypanosoma brucei* isolates. *Molecular and Biochemical Parasitology* **86**:95-99.

Barrett, M.P. (1999). The fall and rise of sleeping sickness. *The Lancet* **353**:1113-4.

Barry, J.D., Crowe, J.S. and Vickerman, K. (1983). Instability of the *Trypanosoma rhodesiense* metacyclic variable antigen repertoire. *Nature* **306**:699-701.

Barry, J.D.B. (1997). The biology of antigenic variation in African trypanosomes. *Trypanosomiasis and leishmaniasis: biology and control.* Eds. G. Hide, J. C. Mottram, G. H. Coombs and P. H. Holmes. Wallingford, UK, CAB International.

Bell, G.I., Selby, M.J. and Rutter, W.I. (1982). The highly polymorphic region near the human insulin gene is composed of simple tandemly repeated sequences. *Nature* **295**:31-35.

Benson, G. (1999). Tandem repeats finder: a program to analyse DNA sequences. *Nucleic Acid Research* **27**:573-580.

Bitter, W., Gerrits, R., Kieft, R. and Borst, P. (1998). The role of transferrin-receptor variation in the host range of *Trypanosoma brucei*. *Nature* **391**:499-502.

Bois, P., Stead, J.D.H., Bakshi, S., Williamson, J., Neumann, R., Moghadaszadeh, B. and Jeffreys, A.J. (1998). Isolation and characterization of mouse minisatellites. *Genomics* **50**:317-330.

Borst, P., Fase-Flowler, F. and Gibson, W.C. (1981). Quantitation of genetic difference between *Trypanosoma brucei gambiense, rhodesiense* and *brucei* by restriction enzyme analysis of kinetoplast DNA. *Molecular and Biochemical Parasitology* **3**:117-131.

Borst, P., Van der Ploeg, M., Van Hoek, J. F. M., Tas, J. and James, J. (1982). On the DNA content of trypanosomes. *Mol. Biochem. Parasitol.* **6**:13-23.

Borst, P., Bitter, W., Blundell, P., Cross, M., McCulloch, R., Rudenko, G., Taylor, M.C. and Van Leeuwen, F. (1997). The expression sites for variant surface glycoproteins of *Trypanosoma brucei. Trypanosomiasis and leishmaniasis: biology and control.* Eds. G. Hide, J. C. Mottram, G. H. Coombs and P. H. Holmes. Wallingford, UK, CAB International.

Bouzekri, N. and Taylor, P.G., Hammer, M.F. and Jobling, M.A. (1998). Novle mutation processes in the evolution of a haploid minisatellite, MSY1: array homogenization without homogenization. *Human Molecular Genetics* **7**:655-659.

Brock, G.J.R., Anderson, N.H. and Monckton, D.G. (1999). *Cis*-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. *Human Molecular Genetics* **8**:1061-7.

Brun, R. and Jenni, L. (1987). Human serum resistance of metacyclic forms of *Trypanosoma brucei brucei, T. brucei rhodesiense* and *T. brucei gambiense*. *Parasitology Research* **73**:218-23.

Buard, J. and Vergnaud, G. (1994). Complex recombination events at hypervariable minisatellite CEB1 (D2S90). *The EMBO Journal* **13**:3203-10.

Buard, J. and Jeffreys, A.J. (1997). Big, bad minisatellites. *Nature Genetics* **15**:327-328.

Capon, D.J., Chen, E.Y., Levinson, A.D., Seeburg, P.H. and Goeddel, D.V. (1983). Complete nucleotide sequence of the T24 human bladder carcinoma oncogene and its normal homologue. *Nature* **302**:33-37.

Chakravarti, A., Buetow, K.H., Antonarakis, S.E., Waber, P.G., Boehm, C.D. and Kazazian, H.H. (1984). *American Journal of Human Genetics* **36**:1239-58.

Church, G.M. and Gilbert, W. (1984). Genomic sequencing. *Proceedings of the National Academy of Science USA* **81** :1991-95.

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

Cibulskis, R.E. (1992). Genetic variation in *Trypanosoma brucei* and the epidemiology of sleeping sickness in the Lambwe Valley, Kenya. *Parasitology* **104**:99-109.

Cleveland, L.R. (1956). Brief accounts of the sexual cycles of the flagellates of Cryptocerus. *Journal of Protozoology* **3**:161-180.

Conway, D.J. and McBride, J.S. (1991). Population genetics of *Plasmodium falciparum* within a malaria hyperendemic area. *Parasitology* **103**:7-16.

Cox, F.E.G. (1993) Parasitic protozoa. Modern Parasitology. Ed F.E.G. Cox. Oxford, UK. Blackwell Scientific Publications.

Cui, X., Li, H., Goradia, T.M., Lange, K., Kazazian, H.H., Galas, D.J. and Arnheim, N. (1989). Single sperm typing: Determination of genetic distance between the G-gamma globulin and parathyroid hormone loci. *Proceedings of the National Academy of Science* **86** :9389-9393.

Deane, M.P., Sousa, M.A., Pereira, N.M., Goncalves, A.M., Momen, H. and Morel, C.M. (1984). *Trypanosoma cruzi* innoculation schedules and re-isolation methods select individual strains from doubly infected mice as demonstrated by schizodeme analysis. *Journal of Protozoology* **31** :276-280.

Degan, R., Pospichal, H., Enyaru, J. and Jenni, L. (1995). Sexual compatibility among *Trypanosoma brucei* isolates from an epidemic area in south-eastern Uganda. *Parasitology Research* **81** :253-257.

De Greef, C., Imberechts, H., Matthyssens, G., Van Meirvenne, N. and Hamers, R. (1989). A gene expressed only in serum-resistant variants of *Trypanosoma brucei rhodesiense*. *Molecular and Biochemical Parasitology* **36** :169-176.

De Greef, C. and Hamers, R. (1994). The serum-resistance associated (SRA) gene of *Trypanosoma brucei rhodesiense* encodes a VSG-like protein. *Molecular and Biochemical Parasitology* **68** :277-284.

Devlin, B. and Risch, N. (1992). A note on Hardy-Weinberg equilibrium of VNTR data by using the Federal Bureau of investigation's fixed-bin method. *American Journal of Human Genetics* **51** :549-553.

Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. and Weissenbach, J. (1996). A comprehensive genetic map of the human based on 5,264 microsatellites. *Nature* **380**:152-154.

Dutton, J.E. (1902). Perliminary note on the trypanosomes occurring in the blood of man. *British Medical Journal* **2**:881.

El-Sayed, N.M.A., Alarcon, C.M., Beck, J.C., Sheffield, V.C. and Donelson, J.E. (1995). cDNA expressed sequence tags of *Trypanosoma brucei rhodesiense* provide new insights into the biology of the parasite. *Molecular and Biochemical Parasitology* **73** :75-90.

Enyaru, J.C.K. (1993). Isoenzyme characterization of *Trypanozoon* isolates from two sleeping sickness areas in south eastern Uganda. *Acta Tropica* **55** :97-115.

Enyaru, J.C.K., Rwendeire, A.J.J. and Stevens, J.R. (1997). Genetic diversity in *Trypanosoma (Trypanozoon) brucei* isolates from mainland and Lake Victoria island populations in south-eastern Uganda: epidemiological and control implications. *Annals of Tropical Medicine and Parasitology* **91** :107-113.

Felger, I., Irion, A., Steiger, S. and Beck, H.P. (1999). Genotypes of merozoite surface protein 2 of *P. falciparum* in Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **93** :S1/3-S1/9.

Gendler, S.J., Lancaster, C.A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E.-N. and Wilson, D. (1990). Molecular cloning and expression of human tumour-associated polymorphism epithelial mucin. *Journal of Biochemical Chemistry* **265**:15286-15293.

Gibson, W.C., Mehlitz, D., Lanham, S.M. and Godfrey, D.G. (1978). The identification of *Trypanosoma brucei gambiense* in Liberian pigs and dogs by isoenzymes and by resistance to human plasma. *Tropenmedizin und Parasitologie* **29** :335-345.

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

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

Gibson, W.C., Osinga, K.A., Michels, P.A. and Borst, P. (1985). Trypanosomes of subgenus Trypanozoon are diploid for housekeeping genes. *Molecular and Biochemical Parasitology* **16** :231-242.

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

Gibson, W.C. (1986). Will the real *Trypanosoma b. gambiense* please stand up. *Parasitology Today* **2**:255-257.

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

Gibson, W. and Garside, L. (1990). Kinetoplast DNA minicircles are inherited from both parents in genetic hybrids of *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **42** :45-54.

Gibson, W.C. and Garside, L.H. (1991). Genetic exchange in *Trypanosoma brucei brucei*: variable chromosomal location of housekeeping genes in different trypanosome stocks. *Molecular and Biochemical Parasitology* **42** :45-54.

Gibson, W.C., Garside, L. and Bailey, M. (1992). Trisomy and chromosome size changes in hybrid trypanosomes from a genetic cross between *Trypanosoma brucei rhodesiense* and *T. b. brucei*. *Molecular and Biochemical Parasitology* **51** :189-199.

Gibson, W.C. and Whittington, H. (1993). Genetic exchange in *Trypanosoma brucei*: Selection of hybrid trypanosomes by introduction of genes conferring drug resistance. *Molecular and Biochemical Parasitology* **60** :19-26.

Gibson, W.C. and Bailey, M. (1994). Genetic exchange in *Trypanosoma brucei*: evidence for meiosis from analysis of a cross between drug-resistant transformants. *Molecular and Biochemical Parasitology* **64** :241-252.

Gibson, W. (1995). The significance of genetic exchange in trypanosomes. *Parasitology Today* **11** :465-468.

Gibson, W.C., Kanmogne, G. and Bailey, M. (1995). A successful backcross in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **69** :101-110.

Gibson, W., Winters, K., Mizen, G., Kearns, J. and Bailey, M. (1997). Intraclonal mating in *Trypanosoma brucei* is associated with outcrossing. *Microbiology* **143**:909-920.

Gilbert, D.A., Lehman, N., O'Brien, S.J. and Wayne, R.K. (1990). *Nature* **344**:764.

Gill, P., Jeffreys, A.J. and Werrett, D.J. (1985). Forensic application of DNA 'fingerprints'. *Nature* **318**:577.

Godfrey, D.G. and Kilgour, V. (1976). Enzyme electrophoresis in characterising the causative organism of Gambian trypanosomiasis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **70**:219-224.

Godfrey, D.G., Scott, C.M., Gibson, W.C., Mehlitz, D. and Zillmann, U. (1987). Enzyme polymorphism and the identity of *Trypanosoma brucei gambiense*. *Parasitology* **94**:337-347.

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

Goebloed, E., Ligthart, G.S., Minter, D.M., Wilson, A.J., Dar, F.K. and Paris, J. (1973). Comparisons of antigenic types of *Trypanosoma brucei* subgroup organisms isolated from wild tsetse flies. *Annals of Tropical Medicine and Parasitology* **67**:31-43.

Gottesdiener, K., Garcia-Anoveros, J., Lee, M.G. and Van der Pleog, L.H.T. (1990). Chromosomal organisation of the protozoan *T. brucei*. *Molecular and Cellular Biology* **10**:6079-6083.

Hagelberg, E., Sykes, B. and Hedges, R. (1990). Ancient bone DNA amplified. *Nature* **342**:485.

Hager, K.M. and Hajduk, S.L. (1997). Mechanism of resistance in African trypanosomes to cytotoxic human HDL. *Nature* **385**:823-6.

Hartl, D.L. and Clark, A.G. (1997). *Principles of population genetics*. Sunderland, MA, Sinauer Associates.

Heisch, R.B., McMahon, J.P. and Manson-Bahr, P.E.C. (1958). The isolation of *Trypanosoma rhodesiense* from a bushbuck. *British Medical Journal* **ii**: 1203-4.

Hide, G., Cattand, P., Le Ray, D., Barry, J.D. and Tait, A. (1990). The identification of *Trypanosoma brucei* subspecies using repetitive DNA sequences. *Molecular and Biochemical Parasitology* **39**:213-226.

Hide, G., Welburn, S.C., Tait, A. and 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**:95-111.

Hide, G., Tait, A., Maudlin, I. and Welburn, S.C. (1996). The origins, dynamics and generation of *Trypanosoma brucei rhodesiense* epidemics in East Africa. *Parasitology Today* **12**:50-55.

Hide, G. (1999). History of sleeping sickness in East Africa. *Clinical Microbiology Reviews* **12**:112-125.

Hill, A.V.S. and Jeffreys, A.J. (1985). Use of minisatellite DNA probes for determination of twin zygosity at birth. *Lancet* **i**:1394-1395.

Hoare, C.A. (1972). *The trypanosomes of mammals*. Oxford: Blackwell Scientific Publications.

Hope, M., MacLeod, A., Melville, S., Tait, A. and Turner, C.M.R. Maintenance of diploidy (in megabase chromosomes) in *Trypanosoma brucei* after genetic exchange. *in press*.

Jeffreys, A.J., Brookfield, J.F.Y. and Semeonoff, R. (1985). Positive identification of an immigration test-case using human DNA fingerprints. *Nature* **317**:818-819.

Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985). Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**:67-73.

Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985). Individual-specific "fingerprints" of human DNA. *Nature* **316**:76-79.

Jeffreys, A.J., Wilson, V., Kelly, R., Taylor, B.A. and Bulfield, G. (1987). Mouse DNA 'fingerprints': analysis of chromosome localisation and germline stability of hypervariable loci in recombinant inbred strains. *Nucleic Acid Research* **15**:2823-2836.

Jeffreys, A.J., Wilson, V., Neumann, R. and Keyte, J. (1988). Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. *Nucleic Acids Research* **16**:10953-10971.

Jeffreys, A.J., MacLeod, A., Neumann, R., Povey, S. and Royle, N. (1990). 'Major minisatellite loci' detected by minisatellite clones 33.6 and 33.15 correspond to the cognate loci *D1S111* and *D7S437*. *Genomics* **7**:449-452.

Jeffreys, A.J., Neumann, R. and Wilson, V. (1990). Repeat unit sequence variation in minisatellites: a novel source of polymorphism for studying allelic variation and mutation by single molecule analysis. *Cell* **60**:473-485.

Jeffreys, A.J., MacLeod, A., Tamaki, K., Neil, D.L. and Monckton, D.G. (1991). Minisatellite repeat coding as a digital approach to DNA typing. *Nature* **354**:204-209.

Jeffreys, A.J., Turner, M. and Debenham, P. (1991). The efficiency of multilocus DNA fingerprint probes for individualization and the establishment of family relationships, determined from extensive casework. *American Journal of Human Genetics* **48**:824-840.

Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L. and Armour, J.A.L. (1994). Complex gene conversion events in germline mutation at human minisatellites. *Nature Genetics* **6**:136-145.

Jeffreys, A.J., Murray, J. and Neumann, R. (1998). High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Molecular Cell* **2**:267-273.

Jenni, L., Marti, S., Schweizer, J., Betschart, B., Le Page, R.W.F., Wells, J.M., Tait, A., Paindavoine, P., Pays, E. and Steinert, M. (1986). Hybrid formation between African trypanosomes during cyclical transmission. *Nature* **322**:173-175.

Jobling, M.A., Bouzekri, N. and Taylor, P.G. (1998). Hypervariable digital DNA code for human paternal lineages: MVR-PCR at the Y-specific minisatellite MSY1 (DYF155S1). *Human Molecular Genetics* **7**:643-653.

Kable, M.L., Seiwert, S.D., Heidmann, S. and Stuart, K. (1996). RNA editing: a mechanism for gRNA-specified uridylate insertion into precursor mRNA. *Science* **273**:1189-95.

Kemp, D.J., Coppel, R.L. and Anders, R.F. (1987). Repetitive proteins and genes of malaria. *Annual Review of Microbiology* **41**:181-208.

Kigotho, A.W. (1997). Southern Sudan hit by epidemic of sleeping sickness. *Lancet* **350**:502.

Kimmel, B.E., Ole-Moiyoi, O.K. and Young, J.R. (1987). *Ingi*, a 5.2kb dispersed sequence element from *Trypanosoma brucei* that carries half of a small mobile element at either end and has homology with mammalian LINEs. *Molecular and Cellular Biology* **7**:1465-175.

Koehler, K.E., Hawley, R.S., Sherman, S. and Hassold, T. (1996). Recombination and nondisjunction in humans and flies. *Human Molecular Genetics* **5**:1495-1504.

Koerner, T., de Raadt, P. and Maudlin, I. (1995). The 1901 Uganda sleeping sickness epidemic revisited: a case of mistaken identity? *Parasitology Today* **11**:303-306.

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

Kooy, R.F., Hirumi, H., Moloo, S.K., Nantulya, V.M., Dukes, P., Van der Linden, P.M., Duijudan, W.A.L., Janse, C.J. and Overdulve, J.P. (1989). Evidence for diploidy in metacyclic forms of African trypanosomes. *Proceeding of The National Academy of Science USA* **86**:5469-5472.

Le Page, R.W.F., Wells, J.M., Prospero, T.D. and Sternberg, J. (1988). Genetic analysis of hybrid *T. brucei*. *Current communications in molecular biology: molecular genetics of parasitic protozoa.* Eds. M. J. Turner and D. Arnott. Cold Spring Harbor, Cold Spring Harbor Laboratory Press. 65-71.

Lee, M.G.-S., Bihain, B.E., Russell, D.G., Deckelbaum, R.J. and Van der Ploeg, L.H.T. (1990). Characterization of a cDNA encoding a cysteine-rich cell surface protein located in the flagellar pocket of the protozoan *Trypanosoma brucei*. *Molecular and Cellular Biology* **10**:4506-4517.

Lee, M.G.-S., Russell, D.G., D'Alesandro, P.A. and Van der Ploeg, L.H.T. (1994). Identification of membrane-associated proteins in *Trypanosoma brucei* encoding an internal, EARLRAEE amino acid repeat. *Journal of Biological Chemistry* **269**:8408-8415.

Letch, C.A. (1984). A mixed population of Trypanozoon in *Glossina palpalis palpalis* from the Ivory Coast. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **78**:627-630.

Lewis, P.O. and Zaykin, D. (1999). Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d12). Free program distributed by the authors over the internet from the GDA Home Page at http://chee.unm.edu/gda/.

Li, H., Gyllensten, U.B., Cui, X., Saiki, R-K., Erlich, H.A. and Arnheim, N. (1988). Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* **355**:414-417.

Lindergard, G. (1999). Ph.D. Thesis. University of Glasgow.

Macedo, A., Martins, M.S., Chiari, E. and Pena, S.D.J. (1992). DNA fingerprinting of *Trypanosoma cruzi*: a new tool for characterization of strains and clones. *Molecular and Biochemical Parasitology* **55**:147-154.

Macedo, A.M., Melo, M.N., Gomes, R.F. and Pena, S.D.J. (1992). DNA fingerprints: a tool for identification and determination of the relationships between species and strains of Leishmania. *Molecular and Biochemical Parasitology* **53**:63-70.

Macina, R.A., Affranchino, J.L., Pollevick, G.D., Jazin, E.E. and Frasch, A.C.C. (1989). Variable number of repeat units in genes encoding *Trypansoma cruzi* antigens. *FEB Letters* **257**:3655-368.

MacLeod, A., Turner, C.M.R. and Tait, A. (1997). Detection of single copy gene sequences from single trypanosomes. *Molecular and Biochemical Parasitology* **84**:267-270.

MacLeod, A., Turner, C.M.R. and Tait, A. A high level of mixed *Trypanosoma brucei* infections in tsetse flies detected by three hypervariable minisatellites. *Molecular and Biochemical Parasitology, in press.*

McNamara, J.J., Mohammed, G. and Gibson, W.C. (1994). *Trypanosoma* (Nannomona) *godfreyi* sp. nov, from tsetse in The Gambia: biological and biochemical characterization. Parasitology, **109**, 496-509.

Masiga, D.K., Smyth, A.J., Hayes, P., Bromidge, T.J. and Gibson, W.C. (1992). Sensitive detection of trypanosomes in tsetse flies by DNA amplification. *Int. J. Parasitol.* **22**:909-918.

Mathieu-Daude, F. and Tibayrenc, M. (1994). Isoenzyme variability of *Trypanosoma brucei s.l.*: genetic, taxonomic, and epidemiological significance. *Experimental Parasitology* **78**:1-19.

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

May, C.A., Jeffreys, A.J. and Armour, J.A.L. (1996). Mutation rate heterogeneity and the generation of allele diversity at the human minisatellite MS205 (D16S309). *Human Molecular Genetics* **5**:1823-1833.

Maynard-Smith, J., Smith, N.H., O'Rourke, M. and Spratt, B.G. (1993). How clonal are bacteria? *Proceedings of the National Academy of Science* **90**:4384-4388.

Mehlitz, D., Zilmann, U., Scott, C.M. and 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. *Tropenmedizin und Parasitologie* **33**:113-118.

Melville, S.E., Leech, V., Gerrard, C.S., Tait, A. and Blackwell, J.M. (1998). The molecular karyotype of megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers. *Molecular and Biochemical Parasitology* **94**:155-173.

Melville, S.E., Gerrard, C.S. and Blackwell, J.M. Multiple causes of size variation in the diploid megabase chromosomes of African trypanosomes. *in press.*

Michels, P.A.M., Hannaert, V. and Bakker, B.M. (1997). Glycolysis in Kinetoplastida. *Trypanosomiasis and leishmaniasis: biology and control.* Eds. G. Hide, J. C. Mottram, G. H. Coombs and P. H. Holmes. Wallingford, UK, CAB International.

Mihok, S., Otieno, L.H. and Darji, N. (1990). Population genetics of *Trypanosoma brucei* and the epidemiology of human sleeping sickness in the Lambwe Valley, Kenya. *Parasitology* **100**:219-33.

Molyneux, D.H. (1997). Current public health status of the trypanosomiases and leishmaniases. *Trypanosomiasis and leishmaniasis: biology and control.* Eds. G. Hide, J. C. Mottram, G. H. Coombs and P. H. Holmes. Wallingford, UK, CAB International.

Monckton, D., Tamaki, K., MacLeod, A., Neil, D.L. and Jeffreys, A.J. (1993). Allele-specific MVR-PCR analysis at minisatellite *D1S8*. *Human Molecular Genetics* **2**:513-519.

Monckton, D.G., Neumann, R., Guram, T., Fretwell, N., Tamaki, K., MacLeod, A. and Jeffreys, A.J. (1994). Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. *Nature Genetics* **8**:162-170.

Moser, D.R., Cook, G.A., Ochis, D.E., Bailey, C.P., McKane, M.R. and Donelson, J.E. (1989). Detection of *Trypanosoma congolense* and *Trypanosoma brucei* subspecies by DNA amplification using the polymerase chain reaction. *Parasitology* **99**:57-66.

Mowatt, M.R. and Clayton, C.E. (1988). Polymorphism in the procyclic acidic repetitive protein gene family of *Trypanosoma brucei*. *Molecular and Cellular Biology* **8**:4055-4062.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**:1616-1622.

Nantulya, V.M. (1991). Molecular diagnosis of parasites. *Experientia* **47**:142-145.

Neil, D.L. and Jeffreys, A.J. (1993). Digital DNA typing at a second hypervariable locus by minisatellite variant repeat mapping. *Human Molecular Genetics* **2**:1129-1135.

Oliveira, R.P., Broude, N.E., Macedo, A.M., Cantor, C.R., Smith, C.L. and Pena, S.D.J. (1998). Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. *Proceedings of the National Academy of Science USA* **95**:3776-80.

Paindavoine, P., Zampetti-Bosseler, F., Pays, E., Schweizer, J., Guyaux, M., Jenni, L. and Steinert, M. (1986). Trypanosome hybrids generated in tsetse flies by nuclear fission. *EMBO Journal* **5**:3631-3636.

Paindavoine, P., Zampetti-Bossler, F., Coquelet, H., Pays, E. and Steinert, M. (1989). Different allele frequencies in *Trypanosoma brucei brucei* and *Trypanosoma brucei gambiense* populations. *Molecular and Biochemical Parasitology* **32**:61-72.

Paul, R.E.L., Packer, M.J., Walmsley, M., Lagog, M., Ranford-Cartwright, L.C., Paru, R. and Day, K.P. (1995). Mating patterns in malaria parasite populations of Papua New Guinea. *Science* **269**:1709-11.

Pepin, J. and Milford, F. (1994). The treatment of African trypanosomiasis. *Advances in Parasitology* **33**:1-47.

Piper, M.B., Bankier, A.T. and Dear, P.H. (1998). A HAPPY map of *Crytosporidium parvum*. *Genome Research* **8**:1299-1307.

Ranford-Cartwright, L.C., Balfe, P., Carter, R. and Walliker, D. (1991). Genetic hybrids of *Plasmodium falciparum* identified by amplification of genomic DNA from single oocysts. *Mol. Bioch. Parasitol.* **49**:239-244.

Raper, J., Nussenzweig, V. and Tomlinson, S. (1996). The main lytic factor of *Trypanosoma brucei brucei* in normal human serum is not high density lipoprotein. *Journal of Experimental Medicine* **183**:1023-9.

Rickman, L.R. and Robson, J. (1970). The blood incubation infectivity tests: a simple test which may serve to distinguish *Trypanosoma brucei* from *T. rhodesiense*. *Bulletin of the World Health Organisation* **42**:650-1.

Rifkin, M.R. (1978). Identification of the trypanocidal factor in normal human serum: high density lipoprotein. *Proceedings of the National Academy of Science* **75**:3450-3454.

Rifkin, M.R., De Greef, C., Jiwa, A., Landsberger, F.R. and Shapiro, S.Z. (1994). Human serum-sensitive *Trypanosoma brucei rhodesiense*: a comparison with serologically identical human serum-resistant clones. *Molecular and Biochemical Parasitology* **66**:211-220.

Robson, J., Rickman, L.R., Allsopp, R. and Scott, D. (1972). The composition of the *Trypanosoma brucei* sub-group in non-human reservoirs in the Lambwe Valley, Kenya, with particular reference to the distribution of *T. rhodesiense*. *Bulletin of the World Health Organisation* **46**:

Roditi, I., Carrington, M. and Turner, M. (1987). Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei*. *Nature* **325**:272-274.

Royle, N.J., Clarkson, R.E., Wong, Z. and Jeffreys, A.J. (1987). Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics* **3**:352-360.

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

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular cloning: a laboratory manual.* Cold Spring Harbor, Cold Spring Harbor Laboratory Press.

Sasse, J. (1998). Ph.D Thesis. University of Cambridge. The development of genetic markers for the *Trypanosoma brucei* genome.

Schneider, A., Hemphill, A., Wyler, T. and Seebeck, T. (1988). Large microtubule-associated protein of *T. brucei* has tandemly repeated, near identical sequences. *Science* **241**:459-462.

Schweizer, J., Tait, A. and Jenni, L. (1988). The timing and frequency of hybrid formation in African trypanosomes during cyclical transmission. *Parasitology Research* **75**:98-101.

Schweizer, J. and Jenni, L. (1991). Hybrid formation in the life cycle of *Trypanosoma brucei*: detection of hybrid trypanosomes in a mid-gut derived isolate. *Acta Tropica* **48**:319-321.

Schweizer, J., Pospichal, H., Hide, G., Buchannan, N., Tait, A. and Jenni, L. (1994). Analysis of a new genetic cross between two East African *Trypanosoma brucei* clones. *Parasitology* **109**:83-93.

Scott, A.G., Tait, A. and Turner, C.M.R. (1996). Characterisation of cloned lines of *Trypanosoma brucei* expressing stable resistance to MelCy and Suramin. *Acta Tropica* **60**:251-262.

Scott, C., Frezil, J.-L., Toudic, A. and Godfrey, D.G. (1983). The sheep as a potential reservoir of human trypanosomiasis in the Republic of the Congo. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **77**:397-401.

Scribner, K.T., Arntzen, J.W. and Burke, T. (1994). Comparative analysis of intra- and interpopulation genetic diversity in *Bufo bufo* , using allozyme, single-locus microsatellite, minisatellite, and multilocus minisatellite data. *Molecular and Biological Evolution* **11**:737-748.

Seebeck, T., Whittaker, P.A., Imboden, M.A., Hardman, N. and Braun, R. (1983). Tubulin genes of *Trypanosoma brucei*: a tightly clustered family of alternating genes. *Proceedings of the National Academy of Science USA* **80**:4634-38.

Shapiro, S.Z., Naessens, J., Liesegang, B., Moloo, S.K. and Magondu, J. (1984). Analysis by flow cytometry of DNA synthesis during the life cycle of African Trypanosomes. *Acta Tropica* **41**:313-323.

Signer, E.N. and Jeffreys, A.J. (1993). Application of human minisatellite probes to the development of informative DNA fingerprints and isolation of locus-specific markers in animals. *DNA fingerprinting: state of the science.* Eds. S. D. J. Pena, R. Chakraborty, J. T. Epplen and A. J. Jeffreys. Basel, Switzerland, Birkhauser Verlag. 421-428.

Simon, M., Phillips, M. and Green, H. (1991). Polymorphism due to variable number of repeats in human involucrin gene. *Genomics* **9**:576.

Simpson, L. (1987). The mitochondrial genome of kinetoplastid protozoa: genome organization, transcription, replication and evolution. *Annual Review of Microbiology* **41**:363-82.

Stephens, J.W.W. and Fantham, H.B. (1910). On the peculiar morphology of a trypanosome from a case of sleeping sickness and the possibility of its being a new species (*T. rhodesiense*). *Annual Tropical Medical Parasitology* **4**:343-350.

Sternberg, J., Tait, A., Haley, S., Schweizer, J. and Jenni, L. (1987). The inheritance of chromosomal and kinetoplast variants in *T. brucei* spp. *Molecular strategies of parasitic invasion, UCLA symposia on molecular and cellular biology.* Eds. N. Agabian, H. Goodman and N. Nogeira. New York, Liss, A.R. 253-258.

Sternberg, J., Tait, A., Haley, S., Wells, J.M., Le Page, R.F.W., Schweizer, J. and Jenni, L. (1988). Gene exchange in African trypanosomes: characterisation of a new hybrid genotype. *Molecular and Biochemical Parasitology* **27**:191-200.

Sternberg, J., Turner, C.M.R., Wells, J.M., Ranford-Cartwright, L.C., Le Page, R.F.W. and Tait, A. (1989). Gene exchange in African trypanosomes: frequency and allelic segregation. *Molecular and Biochemical Parasitology* **34**:269-280.

Sternberg, J. and Tait, A. (1990). Genetic exchange in *Trypanosoma brucei*. *Trends in Genetics* **6**:317-322.

Stevens, J.R. and Welburn, S.C. (1993). Genetic process within an epidemic of sleeping sickness in Uganda. *Parasitology Research* **79**:421-427.

Stevens, J.R., Mathieu-Dande, F., McNamara, J.J., Mizen, V.H. and Nzila, A. (1994). Mixed populations of *Trypanosoma brucei* in wild *Glossina palpalis palpalis*. *Tropical Medicine and Parasitology* **45**:313-318.

Stevens, J.R. and Tibayrenc, M. (1995). Detection of linkage disequilibrium in *Trypanosoma brucei* isolated from tsetse flies and characterized by RAPD analysis and isoenzymes. *Parasitology* **110**:181-186.

Su, X.-Z. and Wellems, T.E. (1996). Towards a high-resolution *Plasmodium falciparum* linkage map: polymorphic markers from hundreds of simple sequence repeats. *Genomics* **33**:430-444.

Swinkels, B.W., Gibson, W.C., Osmga, K.A., Kramer, R., Veneman, G.H., van Boom, J.H. and Borst, P. (1986). Characterisation of the gene for the microbody (glycosomal) triosephosphate isomerase of *Trypanosoma brucei*. *EMBO J* **5**:1291-1298.

Swindle, J. and Tait, A. (1996). Trypanosomatid genetics. *Molecular Biology of Parasitic Protozoa* Eds. D. F. Smith and M. Parsons. New York, Oxford University Press Inc. 6-34.

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

Tait, A. (1983). Sexual processes in the Kinetoplastida. *Parasitology* **86**:29-57.

Tait, A., Babiker, E.A. and LeRay, D. (1984). Enzyme variation in *T. brucei spp.* I. Evidence for the sub-speciation of *T. b. gambiense*. *Parasitology* **89** :311-326.

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

Tait, A., Sternberg, J. and Turner, C.M.R. (1988). Genetic Exchange in *Trypanosoma brucei*: allelic segregation and reassortment. *Molecular Genetics of Parasitic Protozoa.* Eds. M. J. Turner and D. Arnot. Cold Spring Harbor Laboratory.

Tait, A., Turner, C.M.R., Le Page, R.W.F. and Wells, J.M. (1989). Genetic evidence that metacyclic forms of *Trypanosoma brucei* are diploid. *Molecular and Biochemical Parasitology* **37** :247-255.

Tait, A. (1990). Genetic exchange and evolutionary relationships in protozoan and helminth parasites. *Parasitology* **100**:75-87.

Tait, A., Buchanan, N., Hide, G. and Turner, M. (1993). Genetic recombination and karyotype inheritance in *T. brucei* species. *Genome analysis of protozoan parasites* Ed. S. P. Mozaria. Nairobi, Kenya, ILRAD publication. 93-107.

Tait, A., Buchanan, N., Hide, G. and Turner, C.M.R. (1996). Self-fertilisation in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **76** :31-42.

Tamaki, K., Monckton, D.G., MacLeod, A., Neil, D.L., Allen, M. and Jeffreys, A.J. (1992). Minisatellite variant repeat (MVR) mapping: analysis of 'null' repeat units at D1S8. *Human Molecular Genetics* **1**:401-406.

Thacker, J., Webb, M.B.T. and Debenham, P.G. (1988). Fingerprinting cell lines: uses of human hypervariable DNA probes to characterise mammalian cell cultures. *Somatic Cell Molecular Genetics* **14** :519-525.

Thein, S.L., Jeffreys, A.J. and Blacklock, H.A. (1986). Identification of a post-transplant cell population by DNA fingerprint analysis. *Lancet* **ii**:37.

Thein, S.L., Jeffreys, A.J., Gooi, H.C., Cotter, F., Flint, J., O'Connor, N. and Wainscoat, J.S. (1987). Detection of somatic changes in human cancer DNA by DNA fingerprint analysis. *British Journal of Cancer* **55** :353-356.

Tibayrenc, M., Kjellberg, F. and 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 taxonomic consequences. *Proceedings of the National Academy of Science, USA* **87** :2414-2418.

Tibayrenc, M., Ward, P., Moya, A. and Ayala, F.J. (1986). Natural populations of *Trypanosoma cruzi*, the agent of Chagas' disease, have a complex multiclonal structure. *Proceedings of the National Academy of Science USA* **83** :115-119.

Tibayrenc, M. and Ayala, F.J. (1991). Towards a population genetics of micro-organisms: the clonal theory of parasitic protozoa. *Parasitology Today* 7:228-232.

Tibayrenc, M., Kjellberg, F., Arnaud, J., Oury, B., Breniere, F.S., Darde, M.L. and Ayala, F.J. (1991). Are eukaryotic micro-organisms clonal or sexual? A population genetics vantage. *Proceedings of the National Academy of Science USA* **88** :5129-33.

Tibayrenc, M., Neubauer, K., Barnabe, C., Guerrini, F., Skarecky, D. and Ayala, F.J. (1993). Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. *Proceedings of the National Academy of Science USA* **90** :1335-39.

Tomlinson, S. and Raper, J. (1996). The lysis of *Trypanosoma brucei brucei* by human serum. *Nature Biotechnology* **14** :717-21.

Truc, P. and Tibayrenc, M. (1993). Population genetics of *Trypanosoma brucei* in Central Africa: taxonomic and epidemiological significance. *Parasitology* **106**:137-149.

Truc, P., Bailey, J.W., Duoa, F., Laveissiere, C. and Godfrey, D.G. (1994). A comparison of parasitological methods for the diagnosis of gambian trypanosomiasis in an area of low endemicity in Cote d'Ivoire. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **88**:419-421.

Turner, C.M.R. and Barry, J.D. (1989). High frequency of antigenic variation in *Trypanosoma brucei* rhodesiense infections. *Parasitology* **99**:67-75.

Turner, C.M.R., Sternberg, J., Buchanan, N., Smith, E., Hide, G. and Tait, A. (1990). Evidence that the mechanism of gene exchange in *Trypanosoma brucei* involves meiosis and syngamy. *Parasitology* **101**:377-386.

Turner, C.M.R., Hide, G., Buchanan, N. and Tait, A. (1995). *Trypanosoma brucei*: Inheritance of kinetoplast DNA maxicircles in a genetic cross and their segregation during vegetative growth. *Experimental Parasitology* **80**:234-241.

Urquhart, A. and Gill, P. (1993). Tandem-repeat internal mapping (TRIM) of the involucrin gene: repeat number and repeat-pattern polymorphism within a coding region in human populations. *American Journal of Human Genetics* **53**:279.

Van der Ploeg, L.H.T., Cornelissen, A.W.C.A., Barry, J.D. and Borst, P. (1984). Chromosomes of the kinetoplastida. *EMBO Journal* **3**:3109-3115.

Van Xong, H., Vanhammme, L., Chamekh, M., Chimfwembe, C.E., Van Den Abbeele, J., Pays, A., Van Meirvenne, N., Hamers, R., De Baetselier, P. and Pays, E. (1998). A *VSG* expression site-associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. *Cell* **95**:839-846.

Vickerman, K. and Preston, T.M. (1970). Spindle microtubules in the dividing nuclei of trypanosomes. *Journal of Cell Science* **6**:365-384.

Vickerman, K. (1985). Developmental cycles and biology of pathogenic trypanosomes. *British Medical Bulletin* **41**:105-14.

Vickerman, K. (1994). The evolutionary expansion of the Trypanosomatid flagellates. *International Journal for Parasitology* **24**:1317-31.

Vickerman, K. (1997). Landmarks in trypanosome research. *Trypanosomiasis and leishmaniasis: biology and control.* Eds. G. Hide, J. C. Mottram, G. H. Coombs and P. H. Holmes. Wallingford, UK, CAB International.

Vogelstein, B., Fearon, E.R., Kern, S.E., Hamilton, S.R., Preisinger, A.C., Nakamura, Y. and White, R. (1989). Allelotype of colorectal carcinomas. *Science* **244**:207-211.

Walker-Jonah, A., Dolan, S.A., Gwadz, R.W., Panton, L.J. and Wellems, T.E. (1992). *Molecular and Biochemical Parasitology* **51**:313-20.

Walliker, D., Quakyi, I.A., Wellems, T.E., McCutchan, T.F., Szarfman, A., London, W.T., Corcoran, L.M., Burkot, T.R. and Carter, R. (1987). Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236**:1661-66.

Wang, Z. and Webber, J.L. (1992). Continuous linkage map of human chromosome 14 short tandem repeat polymorphisms. *Genomics* **13**:532-536.

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

Webster, P. and Russell, D.G. (1993). The flagellar pocket of trypanosomatids. *Parasitology Today* **9**:201-206.

Welburn, S.C. and Maudlin, I. (1992). The nature of the teneral state in *Glossina* and its role in the acquisition of trypanosome infections in tsetse. *Annals of Tropical Medicine and Parasitology* **86** :529-36.

Welburn, S. and Maudlin, I. (1997). Current trends in parasite vector interactions. *Trypanosomiasis and leishmaniasis: biology and control.* Eds. G. Hide, J. C. Mottram, G. H. Coombs and P. H. Holmes. Wallingford, UK, CAB International.

Wells, J.M., Prospero, T.D., Jenni, L. and Le Page, R.W.F. (1987). DNA contents and molecular karyotypes of hybrid *Trypanosoma brucei. Molecular and Biochemical Parasitology* **24** :103-116.

WHO (1995). Tropical Disease Research, progress 1975-94, highlights 1993-94. UNDP/World Bank/WHO.

WHO African trypanosomiasis control. D. o. C. o. T. Diseases. http://www.who.int/ctd/html/trypanoepidat.html

Willett, K.C. and Fairbairn, H. (1955). The Tinde experiment: a study of *Trypanosoma rhodesiense* during eighteen years of cyclical transmission. *Annals of Tropical Medicine and Parasitology* **49** :278.

Wong, Z., Wilson, V., Jeffreys, A.J. and Thein, S.L. (1986). Cloning a selected fragment from a human DNA 'fingerprint' : isolation of an extremely polymorphic minisatellite. *Nucleic Acids Research* **14** :4605-4616.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A.J. (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA. *Annual Human Genetics* **51** :269-288.

Woolhouse, M.E.J. and Hargrove, J.W. (1998). On the interpretation of age-prevalence curves for trypanosome infections of testes flies. *Parasitology* **116**:149-156.

Woodward, R., Carden, M.J. and Gull, K. (1994). Molecular characterisation of a novel, repetitive protein of the paraflagellar rod in *Trypanosoma brucei. Molecular and Biochemical Parasitology* **67** :31-39.

Wyman, A. and White, R. (1980). A highly polymorphic locus in human DNA. *Proceedings of the National Academy of Science USA* **77** :6754.

Zampetti-Bosseler, F., Schweizer, J., Pays, E., Jenni, L. and Steinert, M. (1986). Evidence for haploidy in metacyclic forms of *Trypanosoma brucei. Journal of Cell Science* **95** :57-59.

Zhang, L., Cui, X., Schmitt, K., Hubert, R. and Navidi, W. (1992). Whole genome amplification from a single cell: Implications for genetic analysis. *Proceedings of the National Academy of Science* **89** :5847-5851.

ELSEVIER

Molecular and Biochemical Parasitology 86 (1997) 95–99

MOLECULAR AND BIOCHEMICAL PARASITOLOGY

Short communication

A single locus minisatellite sequence which distinguishes between *Trypanosoma brucei* isolates[1]

Michael P. Barrett [a,*], Annette MacLeod [b], Jorge Tovar [d], Justin P. Sweetman [c], Andrew Tait [b], Richard W.F. Le Page [c], Sara E. Melville [c]

[a] *Laboratory of Biochemical Parasitology, Division of Infection and Immunity, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK*

[b] *Wellcome Unit of Molecular Parasitology, The Anderson College, 56 Dumbarton Road, Glasgow G11 6NU, UK*

[c] *Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK*

[d] *Department of Medical Parasitology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK*

Received 6 December 1996; received in revised form 2 January 1997; accepted 4 January 1997

Keywords: *Trypanosoma brucei*; Variable number of tandem repeat sequence ; Minisatellite; Genetic exchange

Restriction fragment length polymorphism (RFLP) analysis has become the method of choice in the analysis of genetic crosses in *Trypanosoma brucei* [1–3], although alternative markers which may be identified by polymerase chain reaction (PCR), particularly the so called variable number of tandem repeat (VNTR) or minisatellite and microsatellite sequences, have superseded RFLP analysis in mammalian genetic analysis [4]. We report the identification of a single locus, hypervariable, tandemly repeated sequence in the genome of *T. brucei* which can be used as a genetic marker to follow inheritance of chromosomes in genetic crosses and to identify DNA polymorphisms in the field. Variability of tandem repeat copy number between alleles has been detected and can be identified by PCR analysis.

The repetitive DNA sequence, tbms42, was cloned fortuitously in experiments designed to isolate the 6-phosphogluconate dehydrogenase gene, *gnd*, of *T. brucei* [5]. The plasmid pTGR3 was found to contain a 1.6 Kb insert, comprised almost entirely of the tbms42 repeat (Accession

Abbreviations: PFGE, Pulsed field gel electrophoresis; VNTR, Variable number of tandem repeat; PGK, Phosphoglycerate kinase; PCR, polymerase chain reaction; RFLP, restriction fragment length polymorphism; EATRO, East African Trypanosomiasis Research Organisation; STIB, Swiss Tropical Institute Basel; TREU, Trypanosomiasis Research Organisation, Edinburgh University; Mb, Megabase pairs.

* Corresponding author.

[1] *Note*: Nucleotide sequence data reported in this paper have been submitted to the EMBL data base with the accession number X70187.

PII S0166-6851(97)02832-6

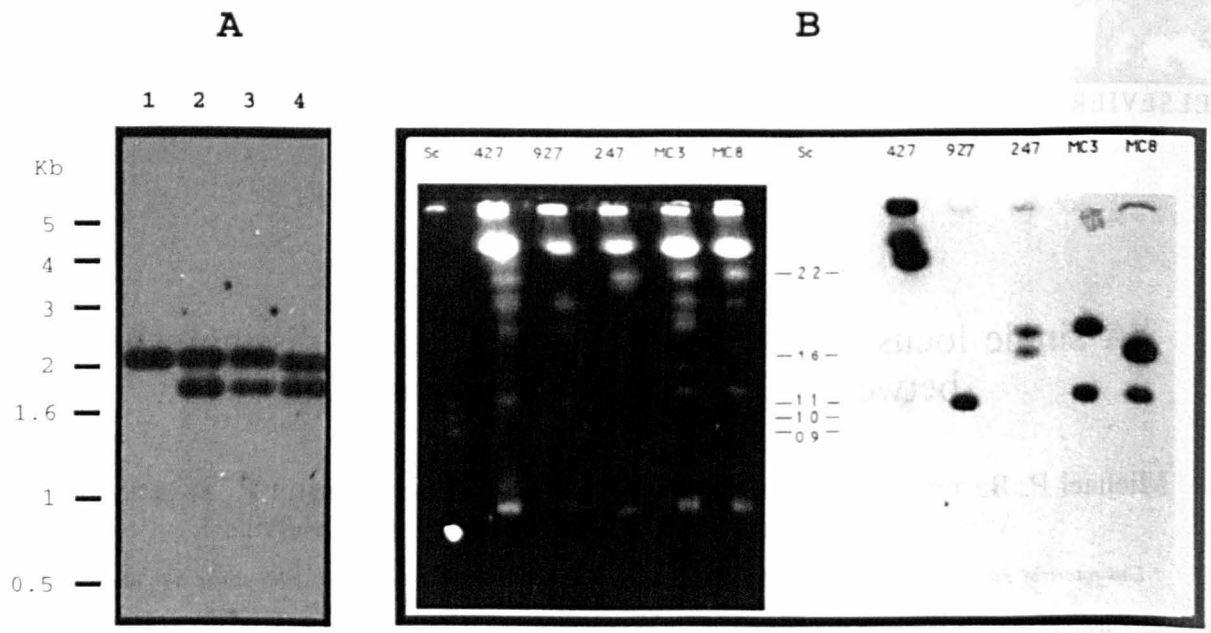


Fig. 1. Southern blot analysis of *T. brucei* genomic DNA probed with tbms42. (A) *Rsa*I digested DNA probed with tbms42. DNA from *T. brucei* stocks 247 (1), MC3 (2), MC8 (3) and 927 (4), was digested to completion with restriction endonuclease *Rsa*I. Fragments separated by electrophoresis and blotted onto a membrane were probed at high stringency with tbms42. DNA molecular mass size markers were the BRL 1 Kb ladder. Southern blot analysis and hybridization were carried out according to standard protocols [20]. (B) Chromosomal localisation of tbms42. Chromosomes from *T. brucei* strains 427, 247, 927, MC3 and MC8 were separated by PFGE. Chromosomes of *Saccharomyces cerevisiae* (Sc) were run as size markers (given in megabase pairs). Blocks containing a total of 5×10^7 procyclic organisms were prepared as described [21]. Chromosomes were separated in 1.2% NA agarose on a Biorad CHEF DRII box, using pulse times ramped down from 1200–700 s, for 168 h at 82 V (2.5 V cm^{-1}) and 15°C in 1 × TBE (0.09 M tris–borate, 0.2 mM EDTA pH 8.0). The gel was stained with ethidium bromide (left), blotted and probed with the insert from tbms42 prior to autoradiography (right).

No. X70187). The structure of this repeated sequence has many features in common with minisatellite or VNTR sequences of higher eukaryotes. The repeat is 60% GC-rich, unlike non-coding DNA in *T. brucei* which contains on average 40% G and C residues, or coding DNA which contains 50% G and C residues [6]. The repeat unit is not homologous to any sequences in the nucleic acid data bases and, like most other eukaryotic minisatellite sequences [7], it is not precisely reiterated over its entire length. When used to probe genomic DNA of *T. brucei* isolate East African Trypanosmiasis Research Organisation (EATRO) 427 digested with various restriction endonucleases, tbms42 hybridized to two bands of different sizes (data not shown) and also to two homologous chromosomes in blots of pulsed field gel electrophoresis (PFGE) separations of chromosomes from isolate EATRO 427; These data suggest that tbms42 is a diploid single copy minisatellite sequence in which different alleles differ in the number of copies of repeats.

This polymorphic marker has been used to analyse the segregation of alleles in genetic crosses by probing blots of restriction enzyme digested *T. brucei* DNA blots of parental isolates (Swiss Tropical Institute Basal (STIB) 247 and Trypanosomiasis Research Organisation (TREU) 927) and two of their progeny clones (MC3 and MC8; described in [8]). The results showed that the number of repeats differed between the parental stocks, based on the differences in size of restriction fragments detected in each of the parental clones (Fig. 1A). The repeat copy number was estimated by densitometric scanning of autoradiographs of DNA blots to determine the

relative numbers of repeat in each allele and by counting repeats on sequencing gel autoradiographs. Isolate STIB 247 appeared to be homozygous for tbms42 (estimated 26 copies of the repeat in each allele) while isolate TREU 927 was heterozygous (each allele containing 25 and eight copies respectively). Southern blots of restriction digests of DNA from progeny clone MC3 revealed one allele with 26 and one with 18 copies while MC8 had one allele with 18 copies and at least one with 26 copies (Fig. 1A). Size-fractionated chromosomes from the same strains were also probed with tbms42 (Fig. 1B). In parental isolate STIB 247 the probe hybridised to two bands with estimated sizes of 1.6 and 1.4 Mb while in the parental isolate TREU 927, it hybridised to a single band estimated as 1.2 Mb in size. Analysis of the progeny clones MC3 and MC8 showed that tbms42 hybridised to bands of 1.4 Mb and 1.2 Mb indicating that the progeny inherited one homologue from each parent. The intensity of hybridization to the larger band (1.4 Mb) in clone MC8 is higher than the equivalent band in clone MC3 and is consistent with the presence of two copies this allele (see also lane 3, Fig. 1A). Phosphoglucose isomerase (PGI), α-tubulin and phosphoglycerate kinase (PGK) gene probes all hybridised to the same chromosomes (data not shown), showing that the sequence described here is linked to these genes in the *T. brucei* genome. In addition, both PGK and tbms42 probes hybridised to identical clones from a phage P1 library of genomic DNA from *T. brucei* isolate 927, suggesting tight linkage between these loci (data not shown). A polymorphic region downstream of the PGK cluster, has previously been described [9] and may correspond to tbms42.

As tbms42 has the structure of a minisatellite sequence, it could be highly polymorphic and therefore a useful marker for the analysis of field populations. Primers were designed to the sequences flanking tbms42 and used to amplify the repeat from DNA extracted from a collection of cloned isolates from Zambia, Kenya and Uganda (described in [10,11]). A total of thirteen distinct alleles were identified on the basis of differences in size in the PCR amplified products as illustrated in Fig. 2. Each isolate is heterozygous for repeat length in tbms42 and the data show that this locus is highly polymorphic and can therefore be used to distinguish between field isolates of *T. brucei*. While it is difficult to draw firm conclusions about epidemiological relationships based on the use of a single marker, the results obtained have been compared with the more extensive multiple locus analysis undertaken on these isolates previously [8,12]. It had been shown that isolates of *T. b. rhodesiense* from Zambia formed a closely related

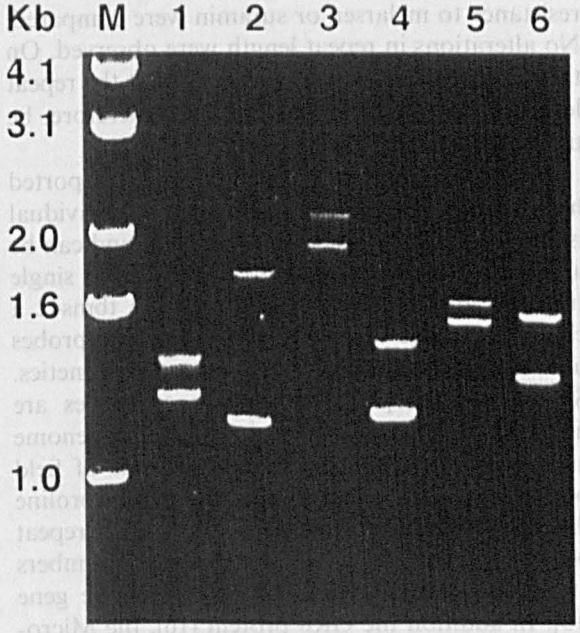


Fig. 2. PCR amplification of a range of tbms42 alleles from different *T. brucei* isolates. PCR products were separated on a 1% Seakem agarose gel and visualised by ultra violet transillumination. 1, TREU 984; 2, TREU 997; 3, Z218; 4, Z210; 5, EATRO 1051; 6, EATRO 7. Lane M contains a 1 Kb marker ladder (BRL). PCR reactions were performed in 10 μl reaction volume in 45 mM Tris–HCl pH 8.8, 11 mM $(NH_4)_2SO_4$, 4.5 mM $MgCl_2$, 6.7 mM 2-mercaptoethanol, 4.4 μM EDTA pH 8.0, 113 $\mu g\ ml^{-1}$ BSA, 1 mM each of the four deoxyribonucleotide triphosphates, 1 μM of each oligonucleotide primer, 1 unit of Amplitaq Polymerase (Perkin Elmer, Cetus USA) and 1 unit of Taq start antibody (Clontech). The reactions were carried out in a Robocycler gradient 96 (Stratagene) using primers designed from sequences flanking tbms42 (Oligo Ms-5′Cggatttcttcatagcgaaggcattc3′ and oligo Ms-D 5′aacttcccacttgtgcggtcgtta3′) under the following conditions: 96°C for 50 s, 66°C for 50 s and 70°C for 90 s, for a total of 32 cycles.

group which was distinct from the populations of *T. b. rhodesiense* isolated in Kenya and Uganda. PCR analysis of tbms42 presented in this study supported these conclusions in that the Zambian stocks contained four alleles and three genotypes (data not shown).

Reports from other organisms [7] have shown that minisatellites can alter the number of repeats by gene conversion. In order to determine whether this occurred at an appreciable frequency with tbms42 in *T. brucei*, 30 cloned lines derived from single cloned stocks which had been multiply passaged (bloodstream or procyclic stages), transmitted through tsetse flies or selected for resistance to melarsen or suramin were compared. No alterations in repeat length were observed. On the basis of these data we conclude that the repeat length is mitotically stable and can, therefore, be used as a genetic marker.

PCR amplification across the repeat reported here provides a means of distinguishing individual isolates without Southern blot analysis and can be applied to the amplification of DNA from single trypanosomes [13]. These features give tbms42 a clear advantage over other single-copy probes used in RFLP analysis in trypanosome genetics. Similar polymorphic minisatellite sequences are likely to occur at other loci within the genome and could be developed for the analysis of field populations. For example the glutamate–proline dipeptide repeat of the procyclic acidic repeat protein (PARP) [14] shows differences in numbers of repeat present in different alleles of the gene [15]. In addition the GR6 protein [16], the Microtubule Associated Repeat Protein (MARP 1) [17], the Gb4 protein [18] and a novel protein found in the paraflagellar rod [19], all contain tandemly repeated sequences. If variation in repeat length occurs between strains, such loci could also be used as genetic markers by the PCR amplification approach described in this study.

Acknowledgements

This work was funded by the Wellcome Trust (MPB, AT, AM, RWFLP) and the Medical Research Council (SEM).

References

[1] Jenni, L., Marti, S., Schweizer, J., Betschart, B., Le Page, R.W.F., Wells, J.M., Tait, A., Paindavoine, P., Pays, E. and Steinert, M. (1986) Hybrid formation between African trypanosomes during cyclical transmission. Nature 322, 173–175.

[2] Sternberg, J. and Tait, A. (1990) Genetic exchange in African trypanosomes. Trends Genet. 6, 317–322.

[3] Tait, A. and Turner, C.M.R. (1990) Genetic exchange in *Trypanosoma brucei*. Parasitol. Today 6, 70–75.

[4] Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) Hypervariable minisatellite regions in human DNA. Nature 314, 67–73.

[5] Barrett, M.P. and Le Page, R.W.F. (1993) A 6-Phosphogluconate dehydrogenase gene from *Trypanosoma brucei*. Mol. Biochem. Parasitol. 57, 89–100.

[6] Clayton, C.E. (1988) The molecular biology of the kinetoplastidae. Genet. Eng. 7, 2–25.

[7] Jeffreys A.J., Tamaki, K., Macleod, A., Monckton, D.G., Neil, D.L. and Armour, J.A.L. (1994). Complex gene conversion events in germline mutation at human minisatellites. Nature Genet. 6, 136–145.

[8] Turner, C.M.R, Sternberg, J., Buchanan, N., Smith, E., Hide, G. and Tait, A. (1990) Evidence that the mechanism of gene exchange in *Trypanosoma brucei* involves meiosis and syngamy. Parasitology 101, 377–386.

[9] Gibson, W. C., Swinkels, B.W. and Borst, P. (1988) Post-transcriptional control of the differential expression of phosphoglycerate kinase genes in *Trypanosoma brucei*. J. Mol. Biol. 201, 439–447.

[10] Goedbloed, E., Ligthart, G.S., Minter, D.M., Wilson, A.J.Dar, F. K. and Paris, J. (1973) Serological studies of trypanosomiasis in East Africa. II: Comparison of antigenic types of *Trypanosoma brucei* sub-group organisms isolated from wild tsetse flies. Ann. Trop. Med. Parasitol. 67, 31–43.

[11] Hide, G., Buchanan, N., Welburn, S.C., Maudlin, I., Barry, J.D. and Tait, A. (1991) *Trypanosoma brucei rhodesiense*: Characterisation of stocks from Zambia, Kenya and Uganda using repetitive DNA probes. Exp. Parasitol. 72, 430–439.

[12] Hide, G., Welburn, S.C., Tait, A. and 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, 95–111.

[13] Macleod, A., Turner, C.M.R. and Tait, A. Detection of single copy gene sequences from single trypanosomes. Mol. Biochem. Parasitol. (in press).

[14] Mowatt, M.R. and Clayton, C.E. (1987) Developmental regulation of a novel repetitive protein of *Trypanosoma brucei*. Mol. Cell Biol. 7, 2838–2844.

[15] Mowatt, M.R. and Clayton, C.E. (1988) Polymorphism in the procyclic acidic repetitive protein family of *Trypanosoma brucei*. Mol. Cell. Biol. 8, 4055–4062.

[16] Muller, N., Hemphill, A., Imboden., M., Duvallet, G., Dwinger, R.H. and Seebeck, T. (1992) Identification and characterization of two repetitive non-variable antigens from African trypanosomes which are recognized early during infection. Parasitology 104, 111–120.

[17] Schneider, A., Hemphill, A., Wyler, T. and Seebeck, T. (1988) Large microtubule-associated protein of *T. brucei* has tandemly repeated, near identical sequences. Science 241, 459–462.

[18] Rindisbacher, L., Hemphill, A. and Seebeck, T. (1993) A repetitive protein for *Trypanosoma brucei* which caps the microtubules at the posterior end of the cytoskeleton. Mol. Biochem. Parasitol. 58, 83–96.

[19] Woodward, R., Carden, M.J. and Gull, K. (1994) Molecular characterisation of a novel, repetitive protein of the paraflagellar rod in *Trypanosoma brucei*. Mol. Biochem. Parasitol. 67, 31–39.

[20] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: A laboratory manual. Cold Spring Harbour Laboratory Press, Cold Spring Harbour.

[21] Van der Ploeg, L.H.T., Cornellisen, A.W.C.P., Barry, J.D. and Borst, P. (1984) Chromosomes of Kinetoplastida. EMBO J. 3, 3109–3115.

ELSEVIER

Molecular and Biochemical Parasitology 84 (1997) 267–270

MOLECULAR AND BIOCHEMICAL PARASITOLOGY

Short Communication

Detection of single copy gene sequences from single trypanosomes

Annette MacLeod[a,*], C. Michael R. Turner[b], Andy Tait[a]

[a] *Wellcome Unit of Molecular Parasitology, Anderson College, University of Glasgow, 56 Dumbarton Road, Glasgow G12 6NU, UK*
[b] *Division of Infection and Immunity, I.B.L.S., Joseph Black Building, Glasgow University, Glasgow G12 8QQ, UK*

Received 22 August 1996; revised 28 November 1996; accepted 29 November 1996

Keywords: Dinucleotide repeats; PCR amplification; *Trypanosoma brucei*

The ability to determine the genotype of single cells by polymerase chain reaction (PCR) amplification of specific genes has been used extensively in human genetics [1,2], and is a powerful technique for the analysis of a range of genetic and biological questions. With African trypanosomes, PCR amplification of the multicopy satellite sequence from different species of trypanosomes has been used to detect trypanosomes both from infected blood [3] and tsetse flies [4]. It has been shown that a PCR signal can be obtained from DNA or trypanosome extracts diluted to less than a single genome equivalent [3,4]. While these reports show that amplification of multiple copy sequences can be used to detect trypanosomes to a high degree of sensitivity, to date PCR based techniques using single copy sequences have not been used to detect or genotype single trypanosomes. Studies with *Plasmodium falciparum* have shown that single copy gene sequences can be amplified from single cells [5], although the life cycle stage used (oocyst) contains many nuclei. In this paper we report the development of a method for determining the genotype of single trypanosomes with a high degree of efficiency using a target single copy gene sequence. This technique should have applications for the study of the laboratory genetics, population genetics and genomic rearrangements in trypanosomes.

The single copy triosephosphate isomerase gene (TIM) sequence of *Trypanosoma brucei* contains, within the 3′ untranslated region, 14 repeats of the dinucleotide TA in the sequence obtained from stock EATRO 427 [6]. Using the published se-

Abbreviations: EATRO, East African Trypanosomiasis Research Organisation; ORF, open reading frame; PCR, polymerase chain reaction; STIB, Swiss Tropical Institute Basel; TIM, triosephosphate isomerase gene; TREU, Trypanosome Research Edinburgh University.

* Corresponding author. Tel.: +44 141 3078072; fax: +44 141 3305422; e-mail: gvwa08@udcf.gla.ac.uk

PII S0166-6851(96)02810-1

quence, several pairs of primers were designed. Initially, two of these, E and F (Fig. 1a), were used to amplify (by PCR) samples of DNA from stocks STIB 386 and TREU 927 [7] of *T. brucei*. The products were run out on a 4% Nusieve agarose gel and visualised with ethidium bromide (Fig. 1b). The sizes of the fragments differ in each of the stocks (STIB 386–100bp; TREU 927–80bp) indicating a polymorphism in the lengths of each repeat array which has been confirmed by sequence analysis (data not shown). The lower band is due to primer dimers and present in the absence of any DNA template (track 3, Fig. 1b).

Initial experiments using dilutions of extracted DNA equivalent to a single genome (0.12 pg) per reaction showed that, using nested primers, it was technically feasible to amplify a single copy sequence to visible levels on an ethidium bromide stained gel (data not shown). To determine whether it was possible to amplify the TIM locus from a single trypanosome, bloodstream trypanosomes were isolated from mice which had been infected with either a mixture of stocks STIB 386 and TREU 927 or solely with TREU 927. The trypanosome mixture was diluted in guinea pig serum and single parasites were isolated optically using an inverted microscope. Each cell was then transferred to a thin walled microtube by the addition of 10 μl of PCR buffer, which has been described elsewhere [8], and subjected to PCR amplification.

Using the bloodstream trypanosomes of a single stock (TREU 927), 16 single trypanosomes were subjected to PCR amplification (using primer pairs C/D followed by A/B-Fig. 1) and five gave an ethidium stained band of the predicted size (Fig. 2a; tracks 4–19). The inability to amplify a fragment from all samples may be attributed to either a failure to transfer single cells to the PCR tubes or a failure of the first pair of primers to successfully amplify one of the two allelic copies. In order to provide a rigorous negative biological control, ten (individual) drops from the infected blood which contained no trypanosomes by microscopic inspection, were also subjected to PCR amplification, as before, using primers C/D followed by A/B (data not shown). One of these drops produced an amplified fragment which is presumed to be due to overlooking a trypanosome in this drop. However, given only 1/10 of the drops lacking trypanosomes gave a positive signal compared to 7/18 drops containing trypanosomes, (Fig. 2a: tracks 1, 2 and 4–19) we conclude that the amplified product is dependent on the presence of trypanosomes rather than extraneous contaminating DNA. The positive controls used in this experiment were drops

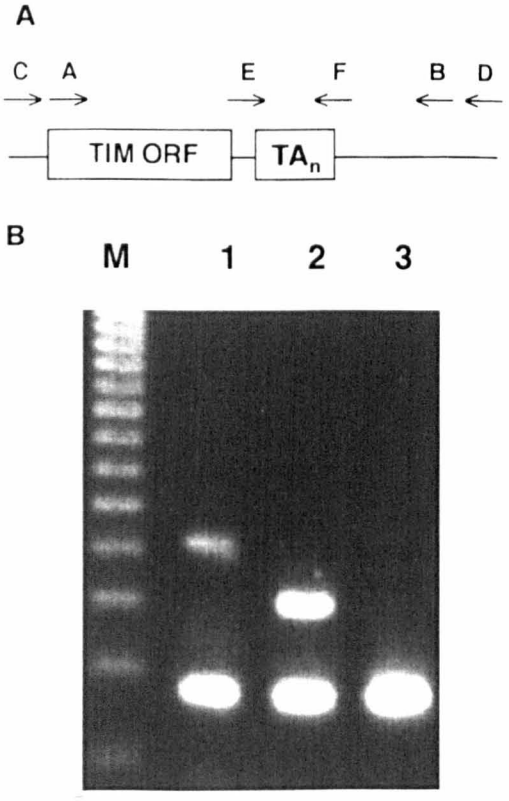


Fig. 1. PCR amplification of the TIM locus from DNA isolated from stocks STIB 386 and TREU 927. (A) The schematic diagram indicates the TIM locus with the approximate positions of the TA dinucleotide repeat and the primers (A–F) used for PCR amplification. (B) Ethidium bromide stained agarose gel separation of the PCR amplified products obtained from genomic DNA stocks STIB 386 and TREU 927 using the primers E and F. Track 1, STIB 386; track 2, TREU 927; track 3, zero DNA control; M, 20bp ladder (Advanced Biotechnologies). Primer sequences: A, gcgtagtggcctccacctttgttc; B, aacacccccctattgttccctctcc; C, caacttactggggacgctgctatc; D ctacactctcttttcctctcccag; E, tgccgttgagtgggtgaagatagc; F, ctcctgctacctgtctttacatc.

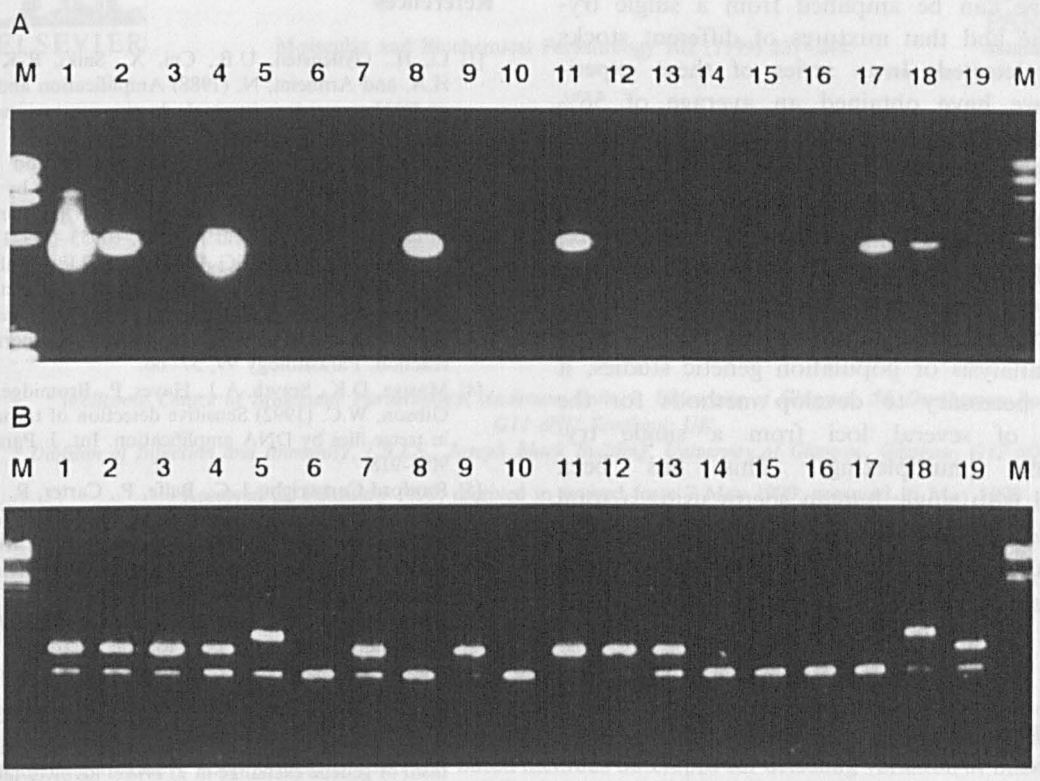


Fig. 2. PCR amplification of the TIM locus from single trypanosomes using a series of nested primers. (A) 16 single bloodstream stage trypanosomes were isolated optically, transferred to a PCR microtube by the addition of 10 μl of PCR buffer, described elsewhere [8], overlaid with mineral oil and heated to 96°C for 5 min (to disrupt the cell). The samples were then subjected to PCR amplification, by the addition of primers C and D to a final concentration of 0.2 μM, under the following cycling conditions: 96°C for 1 min, 64°C for 1 min and 70°C for 90 s for a total of 26 cycles. A 1 μl sample of the product was transferred to a fresh PCR microtube and subjected to a second PCR amplification using primer pair A and B under the same conditions for a further 26 cycles. The final products were separated on 1% Seakem agarose, stained with ethidium bromide and visualised by ultra violet illumination. Tracks 1 and 2, each contained more than one trypanosome; track 3, zero trypanosome control; tracks 4–19, drops each containing a single trypanosome; M, $\phi \times 174$ *Hae*III markers. (B) Ethidium bromide stained 4% Nusieve agarose separation of the PCR products of single drops containing a single trypanosome isolated from a mixture of stocks TREU 927 and STIB 386. The conditions for PCR amplification were identical to those described in (A) except that primer pair A/B was used for the initial amplification and primer pair E/F for the second. Tracks 1–19, single trypanosomes in a single drop; M, $\phi \times 174$ *Hae*III markers.

containing more than one trypanosome and all drops produced an amplified fragment (Fig. 2a, tracks 1 and 2). In order to test whether mixtures of trypanosome stocks could be detected, single trypanosomes were isolated (as described above) from an infection containing two stocks of the parasite (TREU 927 and STIB 386) which had different lengths of the dinucleotide repeat and so could be distinguished after PCR amplification. A total of 19 drops containing single trypanosomes, isolated from a mixed infection, were PCR amplified and the results obtained are shown in Fig. 2b. A total of 12 out of 19 drops containing single trypanosomes gave an ethidium bromide stained fragment of either 80 bp (TREU 927) or 100 bp (STIB 386) but none gave two fragments indicating that none of the drops contain two trypanosomes (the lower size band on these gels is due to primer dimers). These data support the conclusion that a single

copy gene can be amplified from a single trypanosome and that mixtures of different stocks can be detected. In a series of these experiments, we have obtained an average of 56% amplified signals from drops containing single trypanosomes, which compares well with the success rate for developing clones in mice injected with single trypanosomes. The results presented provide a reliable technique for defining the genotype of single trypanosomes. To improve the utility of this technique in laboratory genetic analysis or population genetic studies, it will be necessary to develop methods for the analysis of several loci from a single trypanosome (multiplexing). This has been achieved with single human sperm using a total of five loci [2] and experiments are currently being undertaken to extend this approach to single trypanosomes using a range of polymorphic loci.

Acknowledgements

We would like to thank Michelle Hope for her help and contribution to this work and the Wellcome Trust for financial support. CMRT is a Royal Society University Research Fellow.

References

[1] Li, H., Gyllensten, U.B., Cui, X., Saiki, R-K., Erlich, H.A. and Arnheim, N. (1988) Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature, 355, 414–417.

[2] Jeffreys, A.J., Wilson, V., Neumann, R. and Keyte, J. (1988) Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. Nucleic Acid Res. 16, 10953–10971.

[3] Moser, D.R., Cook, G.A., Ochis, D.E., Bailey, C.P., McKane, M.R. and Donelson, J.E. (1989) Detection of *Trypanosoma congolense* and *Trypanosoma brucei* subspecies by DNA amplification using the polymerase chain reaction. Parasitology 99, 57–66.

[4] Masiga, D.K., Smyth, A.J., Hayes, P., Bromidge, T.J. and Gibson, W.C. (1992) Sensitive detection of trypanosomes in tsetse flies by DNA amplification. Int. J. Parasitol. 22, 909–918.

[5] Ranford-Cartwright, L.C., Balfe, P., Carter, R. and Walliker, D. (1991) Genetic hybrids of *Plasmodium falciparum* identified by amplification of gemonic DNA from single oocysts. Mol. Biochem. Parasitol. 49, 239–244.

[6] Swinkels, B.W., Gibson, W.C., Osmga, K.A., Kramer, R., Veneman, G.H., van Boom, J.H. and Borst, P. (1986) Characterisation of the gene for the microbody (glycosomal) triosephosphate isomerase of *Trypanosoma brucei*. EMBO J. 5, 1291–1298.

[7] Turner, C.M.R., Sternberg, J., Buchanan, N., Smith, E., Hide, G. and Tait, A. (1990) Evidence that the mechanism of genetic exchange in *T. brucei* involves meiosis and syngamy. Parasitology 101, 377–386.

[8] Jeffreys, A.J., Neumann, R. and Wilson, V. (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying allelic variation and mutation by single molecule analysis. Cell 60, 473–485.

ELSEVIER

Molecular and Biochemical Parasitology 102 (1999) 237–248

MOLECULAR AND BIOCHEMICAL PARASITOLOGY

www.elsevier.com/locate/parasitology

A high level of mixed *Trypanosoma brucei* infections in tsetse flies detected by three hypervariable minisatellites

Annette MacLeod [a,*], C. Michael R. Turner [b], Andy Tait [a]

[a] *Wellcome Centre of Molecular Parasitology, Anderson College, University of Glasgow, 56, Dumbarton Road, Glasgow, G11 6NU Scotland, UK*

[b] *Division of Infection and Immunity, I.B.LS., Joseph Black Building, University of Glasgow, Glasgow, G12 8QQ Scotland, UK*

Received 22 February 1999; received in revised form 7 May 1999; accepted 11 May 1999

Abstract

The issue of whether genetic exchange occurs at a significant frequency in natural populations of *Trypanosoma brucei* is controversial and one of the arguments against a high frequency has been the apparent lack of host infections with mixtures of trypanosome genotypes. Three minisatellite markers (*MS42*, *CRAM*, *292*) within the coding regions of three genes have been identified and PCR based methods developed for detecting variation at these loci using crude lysates of infected blood as templates. Initial PCR analysis, using primers flanking the repeats, of DNA from two cloned stocks of the parasite has shown that two DNA fragments of different size were amplified from each stock. Analysis of the inheritance of these fragments into the F1 progeny of crosses demonstrated that the different size fragments were alleles that segregated in a Mendelian manner. The alleles at each of the three loci segregated independently consistent with their localisation on three different chromosomes. Analysis of a series of cloned isolates from tsetse flies showed that these loci were highly variable giving heterozygosities of 94% and the identification of 12 distinct alleles in a sample of 17 cloned isolates. In order to determine whether isolates are heterogeneous in terms of trypanosome genotype, the allelic variation at these three loci was examined in uncloned samples from tsetse flies isolated in Kiboko, Kenya and Lugala, Uganda. A significant proportion of the isolates (36% in Lugala and 47% in Kiboko) contained more than two alleles at one or more of the loci thus demonstrating that a high proportion of tsetse flies were infected with more than one genotype of trypanosomes. This was established, unequivocally, for two isolates by generating a series of cloned trypanosome lines from each and determining the genotype of each clone; one isolate (927) contained seven different genotypes with a high proportion of the possible combinations of alleles at each locus. These results indicate the possibility of frequent genetic exchange in the field, they imply that a significant proportion of mammalian hosts must contain mixtures of different trypanosome genotypes and they demonstrate the advantages of using minisatellite markers for the analysis of the population structure of *T. brucei*.

Keywords: *Trypanosoma brucei*; Minisatellites; *CRAM*; *292*; *MS42*; Mixed infections

Abbreviations: *CRAM*, cysteine-rich acidic integral membrane protein; EATRO, East African Trypanosomiasis Research Organisation; ORF, open reading frame; PCR, polymerase chain reaction; STIB, Swiss Tropical Institute Basel; TREU, Trypanosome Research Edinburgh University; VNTR, variable number of tandem repeat.

* Corresponding author. Tel.: +44-141-3303579; fax: +44-141-3305422.

E-mail address: gvwa08@udcf.gla.ac.uk (A. MacLeod)

0166-6851/99/$ - see front matter

PII: S0166-6851(99)00101-2

1. Introduction

Although genetic exchange in *Trypanosoma brucei* has been shown to occur in the laboratory [1], the importance of this sexual process in natural populations of trypanosomes remains controversial. Three models of the population structure of trypanosomes have been proposed, firstly that trypanosomes undergo frequent, if not random, mating [2–4], secondly that trypanosomes have a clonal population structure [5–7] and thirdly that mating occurs at high frequency but is masked by the expansion of a few genotypes in a short time frame—an epidemic population structure [8,9]. An understanding of the extent and importance of genetic exchange in *T. brucei* populations has practical implications, as sexual reproduction would facilitate the spread and inheritance of traits of medical and economic importance as well as generating a population with a high degree of genetic diversity that would make the definition of common strains causing particular disease patterns difficult.

Much of the information obtained to date has been generated from relatively small sample sizes which are often diverse in time, host species sampled and geographical location and therefore assume that little or no population sub-structuring occurs. Furthermore, limited attention has been paid to intra-isolate heterogeneity (mixed genotype samples), which is a critical issue if genetic exchange is occurring at any significant level. However, some studies have reported isolates containing more than one genotype of parasite [10–12] from both tsetse flies and mammalian hosts. The restraint on examining larger population samples has been the need to expand isolates in laboratory rodents to prepare pure parasites in sufficient quantity for analysis by isoenzyme and randomly amplified polymorphic DNA (RAPD) markers, while the detection of intra-isolate heterogeneity usually requires parasite cloning due to the difficulties in applying genetic interpretations to some of the markers used. On this basis there is a need to develop new markers that allow parasites to be genotyped without recourse to growth in the laboratory thus facilitating the analysis of large sample sizes and avoiding any potential selection during vegetative growth. Additionally, genetically interpretable markers with high levels of polymorphism would be particularly advantageous in examining whether population sub-structuring occurs, whether predominant genotypes are stable in time and place and whether individual isolates are heterogeneous in terms of parasite genotype.

Hypervariable minisatellites, or variable number tandem repeat (VNTR) loci, are very useful genetic markers as they often have a high degree of heterozygosity and many different allelic states based on variation in the number of repeat units in the tandem array. Such minisatellites have been used extensively in human genetics for individual identification, paternity testing [13] and linkage mapping [14] but have only been used to a limited extent in the analysis of parasite genomes. The trypanosome minisatellite locus, *MS42* [15], varies in a strain specific manner allowing genotype and allele frequencies to be determined, as well as providing a means of identifying and tracking individual strains. The use of locus-specific primers to PCR amplify minisatellite markers should enable the genotyping of trypanosomes even when contaminated with large quantities of host DNA, in addition to allowing the analysis of small quantities of DNA as demonstrated by the detection and genotyping of single trypanosomes by PCR amplification of single copy genes [16]. Because of their high level of polymorphism, minisatellite markers are particularly useful in determining variation between populations due to geographical barriers, defining mating systems and detecting heterogeneity within a sample.

In this paper we describe the identification of two further hypervariable minisatellite sequences that can be PCR amplified directly from *T. brucei*-infected blood lysates and the use of these for the analysis of genotype heterogeneity in parasite isolates. The degree of genetic diversity in two collections of trypanosomes isolated from wild tsetse flies has been determined in order to examine the extent to which these flies harboured mixed *T. brucei* infections within their salivary glands. *T. brucei* undergoes genetic exchange in tsetse salivary glands [17] and requires the presence of at least two genotypes to initiate mating

so the detection of mixed infections would indicate the extent to which this critical pre-condition for sexual recombination in *T. brucei* occurs in the field.

2. Materials and methods

2.1. Trypanosome stocks

The 28 stocks used in this study were isolated by Goebloed et al. [18] from infected salivary glands of tsetse flies and passaged three to ten times in mice before transfer to our laboratory (Table 1). The isolate numbers are the stabilate designations given by the Centre for Tropical Veterinary Medicine, University of Edinburgh (TREU) except for isolate 258 from the London School of Hygiene & Tropical Medicine (LUMP). The uncloned stabilates were used to infect irradiated (600 rads) MF1 mice and infected blood harvested at peak parasitaemia for subsequent PCR analysis. Stabilates of these first peak parasitaemias were used to infect mice for the genera-

Table 1
Numbers of distinct minisatellite alleles detected in uncloned *T. brucei* isolates[a]

Isolate	Year of isolation	No. of different sizes PCR products detected for each minisatellite marker			Minimum no. of genotypes present
		MS42	*CRAM*	*292*	
K 936	1969	2	2	2	1
K 926	1969	2	2	2	1
K 869	1969	2	2	2	1
K 981	1969	4	4	3	2
K 994	1970	2	2	2	1
K 1027	1969	2	2	2	1
K 984	1969	4	4	5	3
K 854	1969	4	4	4	2
K 975	1969	2	2	4	2
K 925	1970	2	2	2	1
K 1009	1970	2	2	2	1
K 851	1970	2	2	2	1
K 852	1969	4	2	2	2
K 978	1969	2	2	2	1
K 997	1970	3	2	4	2
K 258	1970	3	2	4	2
K 927	1970	6	6[b]	5	3
L 836	1969	2	4	2	2
L 791	1969	2	4	2	2
L 944	1969	2	2	3	2
L 834	1969	2	2	2	1
L 929	1969	2	2	2	1
L 933	1969	2	2	2	1
L 941	1970	2	2	2	1
L 832	1970	2	2	2	1
L 934	1969	2	2	2	1
L 844	1970	2	2	2	1
L 845	1969	4	4	4	2

[a] The table lists the isolates used in the genotype analysis with all isolates obtained from *Glossina pallidipes* salivary glands, except for L941 which was obtained from *Glossina fuscipes*. The prefix K indicates isolates from tsetse flies captured in Kiboko, Kenya while the prefix L indicates those from Lugala, Uganda. Isolates with greater than two PCR bands were considered to harbour mixed trypanosome genotypes.

[b] Fifth and sixth bands visible after high resolution electrophoresis.

tion of parasite clones. The derivation of stocks STIB 386 and TREU 927 and the F1 progeny clones from a genetic cross of these two stocks has been described elsewhere [19].

2.2. Optical cloning

Clones were made by direct observation of single parasites derived from infected mice from a drop of blood diluted in guinea pig serum in a well of a humidified Terasaki plate. Immediately after the presence of a single cell was detected, the trypanosome was removed in 20 μl of 50% PBSG (phosphate buffered saline [20]/1% glucose) /50% guinea pig serum and injected into an irradiated MF1 mouse.

2.3. Crude lysates and DNA preparation

Crude lysates from infected mouse blood were prepared as follows: 500 μl of phosphate buffered saline [20] was added to 500 μl of infected mouse blood and centrifuged at 2500 × *g* for 5 min, after which 750 μl of the supernatant was discarded. This washing process was repeated three times. The final parasite/blood pellet was then resuspended in 50 μl lysis buffer (50 mM Tris–HCl pH 8, 100 mM EDTA 0.5% SDS, 0.64 mg ml^{-1} proteinase K) and incubated overnight at 56°C. The lysates were then diluted 1/100 in deionised water and the proteinase K was heat inactivated at 95°C for 5 min. One microlitre of lysate was then used as a template for each of the subsequent PCR reactions. Purified parasites were prepared from infected blood and DNA extracted from them as described by Turner et al. [19].

2.4. PCR analysis

Using the published sequences for *CRAM* and *292* [21,22], specific primer pairs flanking the repeated minisatellite within the coding region were designed. Sequence of the DNA flanking the *MS42* repeated region was determined from a plasmid clone of genomic DNA from a STIB 247 × TREU 927/4 F1 hybrid; F532/72mcl7 (data not shown). From this sequence, two specific *MS42* primers were designed. The primer sequences were:
CRAM-G, 5′ CTGCTGATGCCGTACATGATGATTTC; *CRAM*-H, 5′ AACTCCCTCCCGATCGATCACAAC; *292*-G, 5′ ACACCCCCTCTCCACTTCAGATAC; *292*-H, 5′ GCTGAACCTGTGGGCCCCTCAATTG; *MS42*-F, 5′TTGTGCGGTCGTTAACGCGCGTTCAA, *MS42*-W, 5′GGTGATTCATCGGCTCCCTTACCA. All PCR reactions were performed in 10 μl reaction volumes in 45 mM Tris–HCl pH 8.8, 11 mM $(NH_4)_2SO_4$, 4.5 mM $MgCl_2$, 6.7 mM 2-mercaptoethanol, 4.4 μM EDTA, 113 μg ml^{-1} BSA, 1 mM each of the four deoxyribonucleotide triphosphates, 1 μM of each oligonucleotide primer and one unit of Amplitaq Polymerase (Perkin Elmer, Cetus USA) using a template of either 5 ng of genomic DNA or 1 μl of crude lysate. Zero trypanosome DNA controls and mouse DNA controls were included in every set of PCR reactions. The reactions were carried out in a Robocycler gradient 96 (Stratagene). The cycling conditions, which were identical for the amplification of all three genes, were as follows: 96°C for 50 s, 64°C for 50 s and 70°C for 180 s, for a total of 28 cycles. All PCR products were separated by electrophoresis on a 1% Seakem agarose gel and visualised by ethidium bromide staining and UV illumination.

3. Results

3.1. Allelic variation in CRAM and 292

The sequence of the single copy gene from *T. brucei*, encoding the cysteine-rich acidic integral membrane protein, *CRAM*, has been described previously from EATRO 427 and shown to contain 66 copies of a 36-bp motif [21]. Using the published sequence, a pair of primers flanking the repeated region were designed. These primers were used initially to amplify (by PCR) samples of DNA from stocks STIB 386 and TREU 927 (Fig. 1, lanes 2 and 3). Two amplified fragments of different sizes were detected in each stock and presumed to represent allelic variation in the number of repeat units, with

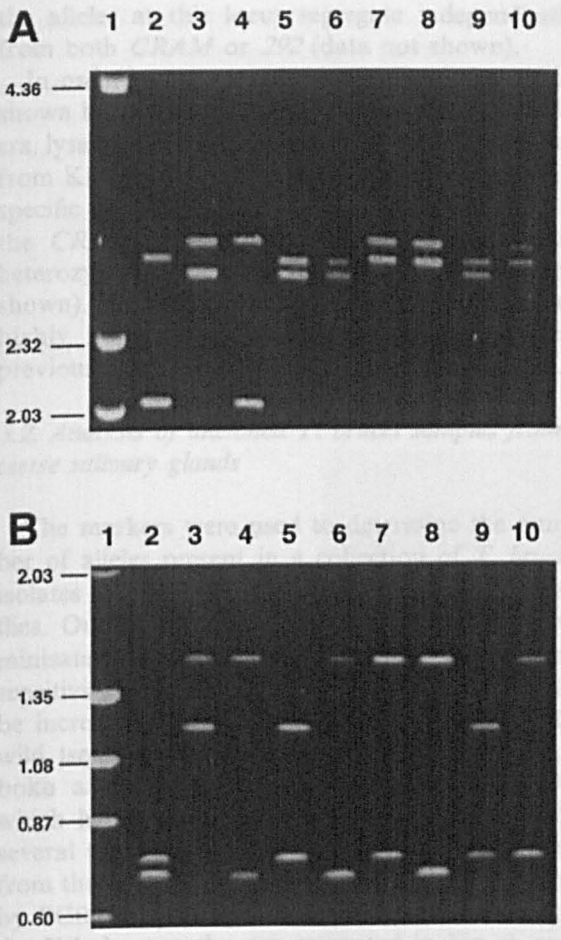


Fig. 1. PCR amplification of the variable regions of the *CRAM* (A) and *292* (B) genes from DNA isolated from stocks STIB 386, TREU 927/4 and hybrid clones derived from a STIB 386 TREU 927 cross. PCR conditions were as described in Section 2. PCR products were separated on a 1% Seakem agarose gel and visualised by ethidium bromide staining. STIB 386 and the cloned line TREU 927/4 are the parental stocks while the stocks designated F296 are the F1 progeny. Lane 1: Lambda *Hin*d III and ϕx *Hae* III markers (Advanced Biotechnologies). Lanes 2–10, PCR amplification of the loci from: STIB 386; TREU 927; F296/44bcl1; F296/44bcl4; F296/39bcl7; F296/44bcl3; F296/44bcl12; F296/44bcl8; F296/39bcl2. Zero trypanosome DNA controls and mouse DNA controls were negative (not shown).

each stock being heterozygous for different sized alleles. In order to demonstrate that these bands were different alleles, F1 progeny clones from a cross between STIB 386 and TREU 927 were analysed for the *CRAM* marker and the results are shown in Fig. 1A; lanes 4–10. It is clear that each F1 progeny clone has inherited one band from each parent. These results confirm the two bands as alleles that segregate in the F1 progeny and are therefore allelic size variants. A similar approach was taken in the analysis of the gene encoding the membrane-associated protein *292* [22]. The cloned gene from *T. brucei* (EATRO 427) contained a repeated region consisting of 47 copies of a 24-bp repeat unit. By designing primers flanking the repeats, the repeated region of this locus was amplified by PCR from stocks STIB 386 and TREU 927 (Fig. 1B; lanes 2 and 3). As with the *CRAM* locus, each stock showed two bands differing in size consistent with being allelic length variants at this locus. PCR amplification of this variable region from the same F1 progeny of a cross between stocks STIB 386 and TREU 927 also demonstrates that these variable size PCR products are different alleles which are inherited in a classical Mendelian manner (Fig. 1B; lanes 4–10). It is also clear from the patterns of allele inheritance (Fig. 1A, B) that the alleles for the two minisatellite loci are inherited independently from each other. For example, hybrids F296/44bcl4 and F296/39bcl7 have inherited the same alleles for the *CRAM* locus, the upper allele of STIB 386 and the lower allele of TREU 927/4 (Fig. 1A; lanes 5 and 6), but for the *292* minisatellite, these two hybrids have inherited different parental alleles (Fig. 1B; lanes 5 and 6), indicating that the two loci are inherited independently. This independent assortment of alleles is expected for these loci, as they are located on different chromosomes [21,22]. A similar analysis, for both loci, was undertaken with the progeny from crosses between stocks STIB 247 and TREU 927 and STIB 247 and STIB 386. The data (not shown) are again consistent with Mendelian segregation, further supporting the conclusion that the size differences in the repeat regions of these genes represent allelic variation at each of the two loci. Similar allelic polymorphisms due to variation in the number of repeats within a tandem array have been described for the *MS42* minisatellite locus [15] and

the alleles at this locus segregate independently from both *CRAM* or *292* (data not shown).

In order to estimate the level of polymorphism shown by the *CRAM* and *292* minisatellite markers, lysates of cloned lines derived from 17 isolates from Kiboko were PCR amplified using the locus specific primers. Twelve distinct alleles for each of the *CRAM* and *292* loci were detected, giving heterozygosities of 94% for each locus (data not shown). These results showed that these loci are highly polymorphic as had been demonstrated previously for the *MS42* minisatellite locus [15].

3.2. Analysis of uncloned T. brucei samples from tsetse salivary glands

The markers were used to determine the number of alleles present in a collection of *T. brucei* isolates derived from the salivary glands of tsetse flies. Our prediction was that by using all three minisatellite loci (*MS42*, *292* and *CRAM*) the sensitivity of detecting different genotypes would be increased. The samples were derived from 28 wild tsetse flies of the genus *Glossina*, from Kiboko and Lugala (isolated in 1969 and 1970), which had been amplified in mice and passaged several times [18]. Parasite DNA or crude lysates from the infected mouse blood were then analysed by PCR using all three markers. The results for the Kiboko samples are presented in Fig. 2, and for all samples in Table 1. From the genetic analysis and the genotypes exhibited by cloned trypanosome stocks, one genotype will either contain two different sized alleles (heterozygous) or a single allele (homozygous). A number of the isolates show two alleles at each of the three minisatellite loci, for example tracks 1, 2 and 3 in Fig. 2A–C, and could represent a single genotype heterozygous at all three loci. Formally, such genotypes could represent a mixture of two homozygotes but the high levels of heterozygosity for these loci argue against this. A second group of isolates show more than two alleles (e.g. lanes 4, 7, 8, Fig. 2A–C) up to a maximum of six alleles (lane 17, Fig. 2C) indicating that they contain a mixture of different trypanosome genotypes. Thus in both sets of isolates a significant proportion of the tsetse flies harboured more than one genotype of *T. brucei*, generating more than two PCR products (summarised in Table 1). Assuming that all the trypanosomes are heterozygous for alleles at each locus and that multiple alleles at more than one locus do *not* assort independently, the minimum number of genotypes present in each isolate can be determined (Table 1). In total, eight out of 17 (47%) of the Kiboko samples contained more than one genotype, and four out of 11 (36%) of the Lugala samples. Most mixed samples probably contained a minimum of two different genotypes, as four alleles at one or more loci were detected. However, two samples must have contained at least three different genotypes, as isolate TREU 984 contained five different *292* alleles and isolate TREU 927 contained six different *MS42* alleles. It is clear that by using a combination of different highly variable minisatellites rather than relying on one marker, the sensitivity of the analysis was greatly increased. Despite this, the number of mixed infections is probably an underestimate as some trypanosome genotypes could be lost during the amplification in mice and repeated passaging. Analysis of the alleles at the three loci in the Kiboko population of isolates shows that there are at least 13 distinct alleles of *CRAM*, 17 distinct alleles of *292*, and 12 alleles of *MS42*, i.e. a very high level of allelic variation. Analysis of all the Kiboko isolates using the three loci shows that out of 17 isolates, 16 are genotypically distinct (tracks 6 and 12 are identical, Fig. 2) demonstrating a high level of polymorphism in the trypanosomes within this field population.

3.3. Cloned parasites from mixed infections

The estimates of the number of genotypes in the isolates showing more than two alleles at two or more loci is a minimum. As the loci are unlinked, the alleles at each locus would be expected to assort independently and so many more distinct genotypes could be present in such isolates. As little is known (in trypanosomes) as to the mechanism by which the repeat length variants are generated, it is possible that they have arisen by mutation during mitosis. To address both these issues, a series of cloned trypanosomes lines were

established from two different samples, 927 and 845, both of which appeared to harbour multiple trypanosome genotypes. Each cloned line was then analysed by PCR for all three loci. The results from the thirteen 927 clones for one marker, *292*, are shown in Fig. 3. Four distinct banding patterns, each containing two bands, clearly demonstrate that more than one genotype of *T. brucei* was present in the original tsetse salivary glands. These genotypes account for four of the alleles observed in the uncloned isolate (lane 2, Fig. 3) but the fifth and sixth alleles are not found in any of the cloned lines, implying that further genotypes are present but have not been cloned. The fact that all the cloned lines show only two alleles yet have undergone multiple rounds of mitosis during clonal growth (from 1 to 10^8 trypanosomes) suggests that length variants at this locus are not generated at an appreciable frequency during mitosis. The clones obtained were also characterised with respect to the two other minisatellite markers (*CRAM* and *MS42*)

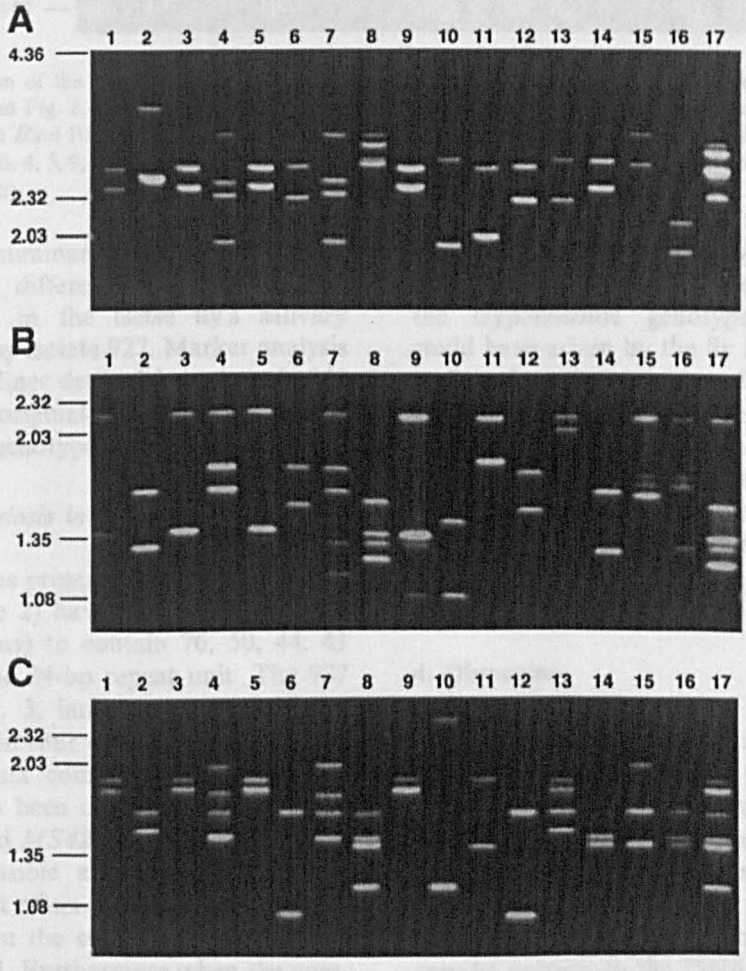


Fig. 2. PCR amplification of three minisatellite markers from uncloned tsetse derived trypanosomes. One microlitre of 1/100 dilution of a crude lysate made from infected blood was used in each PCR reaction. All reactions were performed under the same PCR conditions as described in Section 2. PCR products were separated on a 1% Seakem agarose gel and visualised by ethidium bromide staining. Lanes 1–17: stocks 936; 926; 869; 981; 994; 1027; 984; 854; 975; 925; 1009; 851; 852; 978; 997; 258; 927, using primers specific for the: (A) *CRAM* locus; (B) *292* locus; (C) *MS42* locus. No PCR products were generated in the control samples, i.e. zero trypanosome DNA controls and mouse DNA controls (not shown).

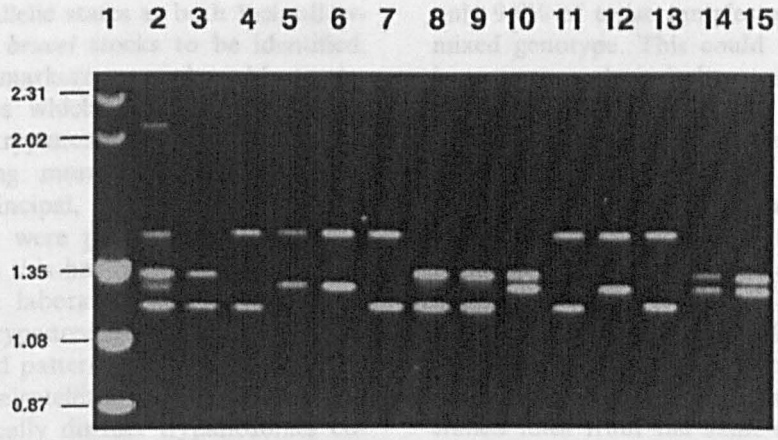


Fig. 3. PCR amplification of the *292* locus from TREU 927 derived clones. PCR reactions were performed on crude lysates as described in Section 2 and Fig. 2. PCR products were separated on a 1% Seakem agarose gel and visualised by ethidium bromide staining. Lane 1: Lambda *Hin*d III and ϕx *Hae* III markers (Advanced Biotechnologies); lane 2: uncloned tsetse derived isolate 927; lanes 3–15: 927 clones, 4B, 4, 5, 9, 10, 12, 13, 14, 18B, 20, 22, 23 and 24. Zero trypanosome DNA controls and mouse DNA controls were negative (not shown).

and the results are summarised in Table 2. There are at least seven different genotypes of trypanosomes present in the tsetse fly's salivary glands represented by isolate 927. Marker analysis of the eight cloned lines derived from sample 845 also indicate that original sample contained at least two different genotypes of trypanosome.

3.4. Evidence for meiosis in field sample 927

The five *292* alleles present in the uncloned 927 sample (Fig. 3, lane 2) have been calculated (by band size estimations) to contain 76, 50, 44, 41 and 38 copies of the 24-bp repeat unit. The 927 derived clones (Fig. 3, lanes 3–14) contain the four smaller alleles in four different combinations out of a possible six combinations. This allele assortment has also been observed for the other markers, *CRAM* and *MS42* (Table 2), where three out of the six possible allele combinations of *CRAM* and four out of six possible *MS42* combinations are found in the clones, excluding those alleles not recovered. Furthermore when the complete multilocus genotypes of each clone are considered seven of the 13 clones are distinct showing different combinations of alleles at the three loci. It seems likely that the analysis of further clones would identify further allelic combinations and a high level of multiple genotypes within one tsetse fly. These data raise the question of the origins of the trypanosome genotypes identified. These could have arisen by the fly feeding on one mammalian host containing all the different trypanosome genotypes or these genotypes could represent the products of mating as a result of the tsetse feeding on a mammal containing two or more genotypes. It would be difficult to formally distinguish between these possibilities without further analysis.

4. Discussion

In order to detect the full range of trypanosome genotypes in field isolates, three highly variable minisatellite loci have been analysed; one of these, *MS42*, has already been described [15], and here we describe two further minisatellite loci, within the coding regions of two previously described structural genes, *CRAM* and *292*. By designing specific primers to the DNA flanking the tandem repeats of these genes, it was possible to amplify, by PCR, the hypervariable regions from a range of *T. brucei* samples including those contaminated with host DNA. These single locus markers proved to be extremely informative due to the

large number of allelic states at both loci, allowing individual *T. brucei* stocks to be identified. The minisatellite markers were also able to detect those samples which contained more than one genotype of trypanosome by virtue of these samples containing more than two alleles at each locus. In principal, this could suggest that the trypanosomes were polyploid, however this is very unlikely as this has only been reported in the progeny from laboratory crosses and genotyping of single trypanosomes from field isolates only shows diploid patterns.

Using these minisatellites we have found evidence that genetically distinct trypanosomes coexist in salivary glands of a significant proportion of tsetse flies; 47% of Kiboko and 36% of Lugala samples. This is in contrast to the isoenzyme data of Godfrey et al. [11] where only 9.6% of tsetse samples were shown to be of mixed genotype. This could reflect the fact that isoenzyme analysis is less sensitive for detecting mixtures due to a lower level of variation. Isoenzyme analysis of a large collection of samples from mammalian hosts from across Africa also detected infections of more than one genotype, in both human (3%) and a range other mammals (2%) [11]. Stevens et al. [10] examined the isoenzyme patterns of trypanosomes from the midguts of three tsetse flies and found two flies containing five and nine different *T. brucei* phenotypes respectively. Further analysis of cloned lines from the same samples, by RAPD analysis, revealed that even more genotypes were present and that all three tsetse flies harboured mixed infections [12], thus suggesting that previous studies based on isoenzyme analysis alone

Table 2
The minisatellite genotypes of cloned lines derived from the uncloned isolates K927 and L845[a]

Isolate number	Genotype given as number of repeats per allele			Multilocus genotype
	MS42	*CRAM*	*292*	
Uncloned K927	35/34/30/26/25/20	70/68/62/61/52	76/50/44/41/38	Mixed
Clone K927/4B	26/25	70/68	44/38	A
Clone K927/4	25/20	68/61	50/38	B
Clone K927/5	30/20	68/61	50/41	C
Clone K927/9	30/20	68/61	50/41	C
Clone K927/10	25/20	68/61	50/38	B
Clone K927/12	26/25	68/61	44/38	D
Clone K927/13	26/25	68/61	44/38	D
Clone K927/14	26/25	70/68	44/41	E
Clone K927/18B	30/26	62/61	50/38	F
Clone K927/20	30/20	68/61	50/41	C
Clone K927/22	25/20	68/61	50/38	B
Clone K927/23	25/20	68/61	44/41	G
Clone K927/24	25/20	68/61	44/41	G
Uncloned L845	31/26/21/16	98/73/70/50	75/49/43/12	Mixed
Clone L845/1	31/16	98/73	49/12	H
Clone L845/2	31/16	98/73	49/12	H
Clone L845/3	31/16	98/73	49/12	H
Clone L845/4	31/16	98/73	49/12	H
Clone L845/5	31/16	98/73	49/12	H
Clone L845/6	31/16	98/73	49/12	H
Clone L845/7	31/16	98/73	49/12	H
Clone L845/8	26/21	70/50	75/43	J

[a] Each multilocus genotype is determined from the combination of alleles at the three minisatellite loci and is assigned a different letter.

may under-represent the degree of trypanosome genetic diversity and the number of mixed tsetse infections present in natural populations.

It is possible that all estimates of mixed infections recorded to date, may not reflect the true levels of genetic diversity and the frequency of mixed infections, due to the influence of sample bias. All samples analysed, except those from the midgut of tsetse [12], have been grown in rodents and passaged to greater or lesser extent before analysis. It is probable that the repeated passaging of trypanosome populations in laboratory rodents may serve to filter out less virulent trypanosome types, especially as different genotypes have different rates of growth in rodents [23]. Another factor which may influence the reliability of sampling is selection due to host infectivity. Host selection has been reported; for example approximately 63% of trypanosomes isolated from the salivary glands of tsetse flies in Kiboko were unable to infect rodents [18], and 35% of samples from the Ivory Coast were lost as they failed to grow in mice [24]. While the problems associated with selection operating on any sampling procedure involving growth or culture is well recognised, it is assumed that such selection does not bias the analysis of markers which have no obvious phenotype on which selection would operate. Given the nature of the markers used to date, it has been impossible to test this assumption, however the markers described in this paper could readily be used to address this question.

Tsetse salivary glands are the probable site at which genetic exchange takes place between *T. brucei* stocks [17] and a prerequisite for genetic exchange to occur is that there are at least two different strains of *T. brucei* present in the salivary glands at the same time [1,19]. We have demonstrated that a significant proportion of tsetse flies harbour mixed *T. brucei* infections in their salivary glands, suggesting that genetic exchange could be occurring in the field. The detection of seven distinct yet highly related genotypes in one fly isolate (927) and the allele assortment which was demonstrated for the three minisatellite loci, cannot easily be explained without genetic exchange being involved. The most direct way of testing this would have been to analyse the genotypes of trypanosomes present in this fly's midgut to determine what genotypes had infected the fly and, from this analysis, deduce whether the genotypes in the salivary gland were recombinant. Unfortunately midgut samples were not collected. An alternative approach, would be to analyse the frequencies and nature of the different genotypes in the population of isolates to estimate the probability that the fly had ingested the seven genotypes detected. This would require the generation of multiple clones from each isolate and has not been undertaken. Similar numbers of genotypes have been isolated from fly midguts [10] presumably reflecting genotypes ingested when the flies feed on infected mammals and so the observation of multiple genotypes in the salivary glands does not per se indicate that these are generated by genetic exchange. However, given that the available evidence shows that salivary gland infections primarily arise as a result of the first teneral blood meal [25], the results presented show that a high proportion of mammalian hosts must be infected with more than one genotype of trypanosomes. The data obtained by analysing the trypanosome genotypes in the midgut of the tsetse could arise as a result of post-teneral blood meals and therefore not result in maturation to the salivary gland. Thus, the results presented clearly show that the main prerequisite for mating, namely the ingestion and maturation of more than one genotype of trypanosome, is satisfied in a high proportion of tsetse flies.

The high levels of mixed infections detected in this study and the recombinant genotypes detected in trypanosomes from one fly, together provide evidence that genetic exchange does occur. However the data presented in this paper do not address how frequently genetic exchange occurs and to what extent it is involved in generating diversity. It is likely that population diversity has been underestimated due to sampling selection and the use of markers with low heterozygosity. This can now be assessed by

analysing trypanosomes, directly from their source, without the additional step of growth in laboratory rodents, using the highly polymorphic and informative minisatellite markers described here.

Acknowledgements

This work was supported by a grant from the Wellcome Trust to A. Tait.

References

[1] Jenni L, Marti S, Schweizer J, et al. Hybrid formation between African trypanosomes during cyclical transmission. Nature 1986;322:173–5.

[2] Tait A. Evidence for diploidy and mating in trypanosomes. Nature 1980;287:536–8.

[3] Tait A. Genetic exchange and evolutionary relationships in protozoan and helminth parasites. Parasitology 1990;100:75–87.

[4] Tait A, Buchanan N, Hide G, Turner M. Genetic recombination and karyotype inheritance in *T. brucei* spp. In: Mozaria SP, editor. Genome Analysis of Protozoan Parasites. Nairobi: ILRAD, 1993:93–107.

[5] Tibayrenc M, Kjellberg F, Ayala FJ. A clonal theory of parasitic protozoa: the population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas* and *Trypanosoma* and their medical and taxonomic consequences. Proc Natl Acad Sci USA 1990;87:2414–8.

[6] Tibayrenc M, Ayala FJ. Towards a population genetics of micro-organisms: the clonal theory of parasitic protozoa. Parasitol Today 1991;7:228–32.

[7] Mathieu-Daude F, Stevens J, Welsh J, Tibayrenc M, McClelland M. Genetic diversity and population structure of *Trypanosoma brucei*: clonality versus sexuality. Mol Biochem Parasitol 1995;72:89–101.

[8] Hide G, Welburn SC, Tait A, Maudlin I. Epidemiological relationships of *Trypanosoma brucei* stocks from South East Uganda: evidence for different population structures in human infective and non-human infective isolates. Parasitology 1994;109:95–111.

[9] Maynard Smith J, Smith NH, O'Rouke M, Spratt BG. How clonal are bacteria? Proc Natl Acad Sci USA 1993;90:4384–8.

[10] Stevens JR, Mathieu-Daude F, McNamara JJ, Mizen VH, Nzila A. Mixed populations of *Trypanosoma brucei* in wild *Glossina palpalis palpalis*. Trop Med Parasitol 1994;45:313–8.

[11] Godfrey DG, Baker RD, Rickman LR, Mehlitz D. The distribution, relationships and identification of enzymatic variants within the subgenus *Trypanozoon*. Adv Parasitol 1990;29:1–74.

[12] Stevens JR, Tibayrenc M. Detection of linkage disequilibrium in *Trypanosoma brucei* isolated from tsetse flies and characterized by RAPD analysis and isoenzymes. Parasitology 1995;110:181–6.

[13] Wong Z, Wilson V, Patel I, Povey S, Jeffreys AJ. Characterization of a panel of highly variable minisatellites cloned from human DNA. Ann Hum Genet 1987;51:269–88.

[14] Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 1987;235:1616–22.

[15] Barrett MP, MacLeod A, Tovar J, et al. A single locus minisatellite sequence which distinguishes between *Trypanosoma brucei* isolates. Mol Biochem Parasitol 1997;86:95–100.

[16] MacLeod A, Turner CMR, Tait A. Detection of single copy gene sequences from single trypanosomes. Mol Biochem Parasitol 1997;84:267–70.

[17] Gibson WG, Bailey M. Genetic exchange in *Trypanosoma brucei*-evidence for meiosis from analysis of a cross between drug-resistant transformants. Mol Biochem Parasitol 1994;64:241–52.

[18] Goebloed E, Ligthart GS, Minter DM, Wilson AJ, Dar FK, Paris J. Comparisons of antigenic types of *Trypanosoma brucei* subgroup organisms isolated from wild tsetse flies. Ann Trop Med Parasitol 1973;67:31–43.

[19] Turner CMR, Sternberg J, Buchanan N, Smith E, Hide G, Tait A. Evidence that the mechanism of gene exchange in *Trypanosoma brucei* involves meiosis and syngamy. Parasitology 1990;101:377–86.

[20] Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1989.

[21] Lee MG-S, Bihain BE, Russell DG, Deckelbaum RJ, Van der Ploeg LHT. Characterization of a c-DNA encoding a cysteine-rich cell surface protein located in the flagellar pocket of the protozoan *Trypanosoma brucei*. Mol Cell Biol 1990;10:4506–17.

[22] Lee MG-S, Russell DG, D'Alesandro PA, Van der Ploeg LHT. Identification of membrane-associated proteins in *Trypanosoma brucei* encoding an internal, EARLRAEE amino acid repeat. J Biol Chem 1994;269:8408–15.

[23] Turner CMR, Aslam N, Dye C. Replication, differentiation, growth and the virulence of *Trypanosoma brucei* infections. Parasitology 1995;111:289–300.

[24] Mehlitz D, Zilmann U, Scott CM, Godfrey DG. Epidemiological studies on the animal reservoir of *Gambi-*

ense sleeping sickness. Part III. Characterization of *Trypanozoon* stocks by iso-enzymes and sensitivity to human serum. Tropen Parasitol 1982;33:113–8.

[25] Maudlin I. Transmission of African trypanosomes: interactions among tsetse immune system, symbionts and parasites. Adv Dis Vector Res 1991;7:117–47.