Linkage mapping and genetic analysis of *Trypanosoma brucei*

Anneli Clare Cooper

Submitted in fulfillment of the requirements for
the Degree of Doctor of Philosophy

Division of Infection & Immunity
Faculty of Biomedical & Life Sciences
University of Glasgow

December 2009

Abstract

Trypanosoma brucei is a protozoan parasite of major public health and economic importance in sub-Saharan Africa, where it is the causative agent of sleeping sickness in man and Nagana in cattle. The complete genome sequence of *T.brucei* is now available and the diploid genetic system has recently been demonstrated to be Mendelian. This opens up the possibility of using a classical genetic approach to identify genetic loci that determine important phenotypic traits in this parasite, such as host specificity, drug resistance, and pathogenicity. A genetic map of the non human-infective subspecies, *T.b.brucei*, has already been assembled and successfully used in quantitative trait analysis of a number of traits specific to this pathogen.

This thesis describes the construction of a separate genetic map for the sub-species responsible for > 90% of human African trypanosomiasis infections, *T.b.gambiense,* which differs significantly from *T.b.brucei* in many key phenotypes. The genetic linkage map was constructed from the analysis of 119 polymorphic microsatellite markers in a population of 38 F1 progeny, obtained from the genetic cross of a *T.b.gambiense* group 2 strain, STIB 386, with a *T.b.brucei* strain, STIB 247. Eleven major linkage groups were resolved, one for each of the megabase chromosomes, resulting in a total genetic map length of 733 cM, and an average map unit size of 24 Kb/cM. The map provides a 90% probability of a marker being within 268 Kb of any genetic locus. A comparative analysis of the *T.b.gambiense* and *T.b.brucei* genetic maps revealed synteny and marker order to be conserved between the two sub-species. However, variation was observed in the location of regions of high and low recombination frequency (hot and cold spots) in the two maps. The genetic linkage map presented here is the first available for *T.b.gambiense* and can now be utilised to find the location within the genome of genes responsible for phenotypic traits in this clinically important sub-species. These traits include human infectivity, tsetse transmissibility and virulence, in addition to sensitivity to the trypanocidal drug, pentamidine, for which phenotypic variation between the parents was characterised both *in vitro* and *in vivo* in this thesis.

The ability of the *T.brucei* genetic maps to pinpoint loci underlying phenotypic variation is limited by the number of recombination events, and therefore progeny, available for analysis. To increase the utility of this approach for future studies, an improved method for progeny isolation from uncloned genetic cross populations was also developed. This *in vitro* bloodstream cloning procedure is

scalable and efficient, and replaces a time consuming and technically demanding *in vivo* method. Twelve new progeny clones were isolated by this approach during the trial and incorporated into the analysis, representing a step toward a higher resolution second-generation genetic map.

Finally, whilst undertaking genotyping analysis with microsatellite markers the development of spontaneous chromosome 10 abnormalities was observed. A detailed investigation identified seven laboratory-adapted *T.brucei* lines in which loss of heterozygosity appeared to have occurred. These alterations to the karyotype significantly exceeded the well-characterised genomic rearrangements of subtelomeric regions that are frequently associated with antigenic variation in African trypanosomes. Microsatellite analysis, pulsed field gel electrophoresis and Illumina next generation sequencing demonstrated these changes to be the product of mitotic recombination events in the chromosome core, resulting in an extensive loss of heterozygosity of up to 75% of the chromosome and correlated with an improved growth phenotype. Further work is now required to determine the extent and frequency with which these abnormalities might occur, however these findings do highlight the potential instability of the molecular karyotype of *T.brucei* in prolonged *in vitro* culture.

Table of contents

Abstract ii
Table of contents iv
List of tables viii
List of figures x
Declaration xii
Publications xiii
Acknowledgements xiv
Abbreviations xv

1 Introduction 1
1.1 General introduction to *T.brucei* 2
1.2 Life cycle 6
1.3 Antigenic variation 8
1.4 Mechanisms of VSG switching 9
1.5 Pathology of Human African Trypanosomiasis 11
1.6 *T.brucei* in the laboratory 13
1.7 Surveillance and control of *T.brucei* 14
1.8 Chemotherapy 15
1.8.1 Human African trypanosomiasis drugs 15
1.8.1.1 Suramin 15
1.8.1.2 Pentamidine 16
1.8.1.3 Melarsoprol 16
1.8.1.4 Eflornithine (DFMO) 17
1.8.2 Veterinary drugs 18
1.8.2.1 Homidium 18
1.8.2.2 Diminazene aceturate (Berenil) 18
1.8.2.3 Isometamidium chloride (Samorin) 19
1.9 Drug resistance 19
1.9.1 Drug resistance in African trypanosomes 19
1.9.2 Mechanisms of drug resistance in *T.brucei* 22
1.9.2.1 P2/*TbAT1* 23
1.9.2.2 Other resistance mechanisms 24
1.10 Trypanosome genetics 25
1.10.1 The structure of the *T.brucei* genome 25
1.10.2 The genome sequence 26
1.11 Genetic mapping 29
1.12 Genetic exchange in *T.brucei* 31
1.13 Identifying molecular markers in *T.brucei* 35
1.13.1 Isoenzymes 36
1.13.2 Restriction fragment length polymorphism (RFLP) 36
1.13.3 Random Amplified Polymorphic DNA (RAPD) 38
1.13.4 Amplified Fragment Length Polymorphism (AFLP) 39
1.13.5 Microsatellites 40
1.13.6 Single nucleotide polymorphisms (SNPs) 41
1.14 Parasite Genetic maps 43
1.14.1 *T.brucei* genetic map 45
1.15 Project Objectives 49

2 Materials and Methods 50

2.1 Chemicals, enzymes and reagents 51
2.2 Buffers and Solutions 52
2.2.1 Molecular Biology 52
2.2.2 Trypanosome *in vitro* and *in vivo* growth 53
2.3 Trypanocidal Drugs 54
2.4 Mice stocks 54
2.5 *Trypanosoma brucei* strains 55
2.5.1 Parental strains 55
2.5.2 F1 progeny clones 55
2.5.3 Uncloned STIB 386 x STIB 247 populations 56
2.6 Molecular Biology Techniques 59
2.6.1 DNA extraction from blood 59
2.6.2 DNA extraction from bloodstream form culture and procyclic form culture cells 59
2.6.2.1 High quality DNA 59
2.6.2.2 Crude lysates 60
2.6.2.3 DNA isolation from 96 well cultures 60
2.6.2.4 Preparation of DNA samples for PCR amplification 60
2.6.3 General PCR procedure 61
2.6.4 Modifications to the general PCR protocol for cloning and DNA sequencing 62
2.6.4.1 Proof-reading Taq polymerase 62
2.6.4.2 Post-PCR addition of 3' adenine residues by Taq polymerase 62
2.6.5 Gel electrophoresis 62
2.6.6 Oligonucleotides 63
2.6.7 Microsatellite primer design 63
2.6.8 Genotyping *T.brucei* strains 63
2.6.9 Construction of the *T.b.gambiense* STIB 386 genetic map ... 64
2.6.9.1 Microsatellite screening for inheritance of STIB 386 alleles in F1 progeny clones 64
2.6.9.2 Linkage analysis 64
2.6.10 Cloning and sequencing 65
2.6.10.1 DNA ligation 65
2.6.10.2 Transformation of *E.coli* competent cells 65
2.6.10.3 Analysis of transformants 65
2.6.10.4 Digestion of DNA by restriction enzyme 66
2.6.10.5 DNA quantification 66
2.6.11 High throughput whole genome sequencing of *T.brucei* strains 66
2.6.11.1 *T.brucei* strain DNA 66
2.6.11.2 Illumina sequencing 66
2.6.11.3 Data analysis 67
2.6.11.4 SNP-RFLPs 67
2.6.11.5 DNA Sequencing 68
2.6.11.6 Sequence analysis 68
2.6.12 Pulsed Field Gel Electrophoresis (PFGE) 68
2.6.12.1 Preparation of chromosome blocks for PFGE 68
2.6.12.2 PFGE 69
2.6.13 Southern blotting 69
2.6.13.1 Preparation of DNA probes 70
2.6.13.2 Hybridization 70
2.6.13.3 Stripping and reprobing blots 71

2.7 *In vitro T.brucei* culture techniques 71
2.7.1 *T.brucei in vitro* cell culture 71
2.7.2 Adaptation of trypanosomes from blood to procyclic form culture 71
2.7.3 Adaptation of trypanosomes from blood to bloodstream form culture 72
2.7.4 Adaptation of trypanosomes from bloodstream form culture to procyclic form culture 72
2.7.5 Preparation of bloodstream culture stabilates 72
2.7.6 Preparation of procyclic culture stabilates 73
2.7.7 *In vitro* cloning of new STIB 386 x STIB 247 hybrid clones 73
2.7.8 Procyclic form culture growth assays 74
2.7.9 *In vitro* bloodstream form culture drug sensitivity assays 74
2.8 *T.brucei* infections in mice 75
2.8.1 Routine growth of *T.brucei* in mice 75
2.8.2 Preparation of blood stabilates 75
2.8.3 *In vivo* Minimum Curative Dose (MCD) assays 76

3 The construction of a genetic map for *Trypanosoma brucei gambiense* and evaluation of drug sensitivity phenotypes for linkage analysis 77
3.1 Introduction 78
3.2 Results 82
3.2.1 Identification of heterozygous markers and the genotyping of F1 progeny 82
3.2.2 Construction of the STIB 386 genetic linkage map 84
3.2.3 Marker segregation proportions 88
3.2.4 Variation in recombination between chromosomes 88
3.2.5 Comparison of the genetic maps of *T.b.gambiense* and *T.b.brucei* and the physical map of *T.b.brucei* 105
3.2.6 Mutation frequency 106
3.2.7 In vivo drug sensitivity assay 108
3.2.8 In vitro drug sensitivity assay 112
3.3 Discussion 114

4 The isolation of new progeny clones by *in vitro* bloodstream cloning 122
4.1 Introduction 123
4.1.1 Introduction to *T.brucei* crosses and methods of hybrid isolation 123
4.1.1.1 The isolation of hybrids by *in vivo* cloning 123
4.1.2 The isolation of clones by *in vitro* procyclic cloning 128
4.1.3 The isolation of clones by *in vitro* bloodstream cloning 128
4.2 Results 130
4.2.1 The isolation of clones by *in vitro* bloodstream cloning 130
4.2.2 Addition of the new progeny clones to the STIB 386, *T.b.gambiense* genetic map 139
4.3 Discussion 145

5 Loss of Heterozygosity in laboratory-adapted *T.brucei* strains ... 148
5.1 Introduction 149
5.2 Results 155

5.2.1 Microsatellite genotyping of LOH lines 155
5.2.2 Karyotype analysis of LOH isolates by PFGE 166
5.2.3 Analysis of the LOH boundaries by Illumina sequencing and SNP-RFLP..... 173
5.2.3.1 Illumina sequence results..... 174
5.2.4 RFLP analysis..... 178
5.2.5 Sequencing the LOH boundary of TREU 927 LOH line 1 185
5.2.6 Growth assays..... 192
5.3 Discussion 194

6 Conclusions and future research directions 204

7 Appendices 215

List of tables

Chapter 1

1.1 A comparison of the main characteristics of some commonly used molecular markers..... 37

1.2 Information on the origins of the three parental *T.brucei* stocks available for investigation in the laboratory..... 46

Chapter 2

2.1 Unique progeny genotyping data..... 57

Chapter 3

3.1 Differences in phenotype between the two parental stocks used to generate the STIB 386 genetic map..... 80

3.2 Characteristics of the genetic linkage maps of *T.b.gambiense*..... 87

3.3 Susceptibility of STIB 247 and STIB 386 to trypanocidal drugs..... 109

Chapter 4

4.1 A summary of reported STIB 247 x STIB 386 crosses..... 126

4.2 Summary of the composition of *in vitro* derived clones for the four different fly populations..... 137

4.3 Genotype of the 12 new unique hybrids isolated from the STIB 247 x STIB 386 cross by *in vitro* cloning..... 138

4.4 Characteristics of the genetic linkage maps of *T.b.gambiense* with the new in vivo culture derived progeny clones added..... 143

Chapter 5

5.1 Genotype of the standard parent lines and lines exhibiting allele drop-out by PCR..... 157

5.2 Details of the seven LOH lines identified by routine microsatellite genotyping in the laboratory..... 165

5.3 Predicted sizes of normal and LOH chromosome 10 homologues by PFGE..... 168

5.4 PCR primers for probes used in Southern blot analysis of LOH..... 169

5.5 SNP-RFLPs used to refine the LOH boundary in TREU 927 LOH line 1..... 182

5.6 SNP-RFLPs used to refine the LOH boundary in TREU 927 LOH line 2..... 183

5.7 SNP-RFLPs used to refine the LOH boundary in progeny clone F29/46 bcl 1..... 184

5.8 SNP Sequence results for both homologues at the predicted

chromosome LOH boundary in TREU 927 LOH line 1 compared to standard TREU 927..... 187
5.9 Clones with SNP haplotypes that deviated from the majority pattern seen in Table 5.8..... 188

Appendices

1 Primer information and amplification conditions for PCR reactions in Chapter 5..... 216
2 Primer information for all the microsatellite markers used in the generation of the STIB 386 genetic map..... 218
3 Primers used to sequence the LOH boundary of TREU 927 LOH line 1 (Section 5.2.5)..... 223
4a-k The segregation data for all of the microsatellite markers on chromosome 1 in the STIB 386 genetic map..... 224
5 Recombination events in every linkage group of every individual for progeny clones 20-38..... 236
6 Genotype of all clones isolated by *in vitro* limiting dilution cloning of bloodstream culture in Chapter 4..... 238
7 Segregation data for the new progeny clones for all microsatellite markers on each of the 11 megabase chromosomes in the STIB 386 genetic map..... 245

List of figures

Chapter 1

1.1 Distribution of human African trypanosomiasis in sub-Saharan Africa..... 5

1.2 The life cycle of *T.brucei* in the tsetse fly vector and mammalian host..... 7

1.3 Genetic linkage map of *T.b.brucei*..... 48

Chapter 3

3.1 A representative microsatellite marker used for mapping analysis..... 83

3.2 Genetic linkage maps corresponding to the 11 megabase chromosomes of *T.b.gambiense*..... 85

3.3 Genotype segregation proportions..... 89

3.4 The genetic size of each linkage group (cM) relative to its physical size (Mb)..... 91

3.5 Comparison of the *T.b.gambiense* genetic maps with the physical and genetic maps of *T.b.brucei*..... 93

3.6 Evidence for a mutation in the STIB 386 allele of marker TB6/15 in progeny clone F28/46 bscl 4..... 107

3.7 48 hour growth inhibition assay with pentamidine..... 113

Chapter 4

4.1 Methods used for isolating clonal products from a *T.brucei* genetic cross..... 127

4.2 The protocol for generating STIB 386 x STIB 247 hybrids by *in vitro* cloning..... 131

4.3 The predicted Poisson distribution of *T.brucei* cloning plates prepared at different dilutions in 96 well plates..... 132

4.4 PCR microsatellite screening of *in vitro* bloodstream culture clones derived from STIB 247 x STIB 386 genetic crosses..... 133

4.5 PCR amplification of the seven molecular markers for STIB 386 and STIB 247 parental strains used for genotyping *in vitro* clones..... 134

4.6 The genetic linkage map of *Trypanosoma brucei gambiense* reconfigured for the addition of 11 new progeny clones..... 140

4.7 Segregation distortion on chromosome 1..... 144

Chapter 5

5.1 Proposed mechanisms by which a LOH genotype may be generated following a DSB..... 153

5.2 The approximate position of seven commonly used genotyping markers on the 11 megabase chromosomes of *T.brucei* TREU 927..... 156

5.3 LOH in a TREU 927 procyclic form culture line at one

	microsatellite locus.....	159
5.4a	Microsatellite evidence for LOH in TREU 927 procyclic culture lines.....	161
5.4b	Microsatellite evidence for LOH in a STIB 247 x STIB 386 progeny clone F29/46 bcl 1 procyclic culture line.....	163
5.5a	The 1-6 Mb molecular karyotype of standard TREU 927 and LOH lines.....	171
5.5b	The 1-6 Mb molecular karyotype of STIB 247 x STIB 386 progeny clone F29/46 bcl 1 standard and LOH lines.....	172
5.6	Illumina SNP distribution in TREU 927 LOH Line 1.....	176
5.7	TREU 927 LOH line 1 chromosome 10 SNP plot.....	177
5.8	Example of five SNP-RFLPs in TREU 927 wild type and TREU 927 LOH line 2.....	180
5.9	Mapping the LOH boundary in TREU 927 LOH line 1.....	189
5.10	Summary of the LOH boundaries on chromosome 10 in four LOH lines.....	190
5.11	Growth comparison of standard TREU 927 and 927 LOH procyclic form culture lines *in vitro*.....	193

Declaration

I declare that this thesis and the results presented within it are entirely my own work, with the following exceptions. The microsatellite marker screening in Chapter 3 was performed with the assistance of Lindsay Sweeney. The *in vivo* Minimum Curative Dose assays in Chapter 3 were performed with the assistance of Dr Caroline Clucas. Jacqueline McQuillan of the Pathogen Sequencing Unit, Wellcome Trust Sanger Institute performed the Illumina sequence analysis in Chapter 5.

No part of this thesis has been previously submitted for a degree at any other institution.

Anneli Clare Cooper

Publications

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

Acknowledgements

First and foremost I would like to thank my supervisors, Dr Annette MacLeod and Prof. Mike Turner for giving me the opportunity to undertake my PhD with them and for their invaluable advice, guidance and support over the past four years. I am also indebted to Prof. Andy Tait for kindly sharing his considerable wisdom and useful ideas during the course of this project. Thanks go to all members of the Turner Tait MacLeod group, past and present, for their help, support and friendship. I would especially like to acknowledge Lindsay Sweeney for sharing her tissue culture expertise and helping with marker screening, and Dr Caroline Clucas for her assistance with the drug sensitivity phenotyping. Special thanks also go to Dr Liam Morrison for the helpful discussions, perspective and banter.

Beyond Glasgow University I would like to acknowledge the Pathogen Sequencing Unit at the Wellcome Trust Sanger Institute, particularly Christiane Hertz-Fowler and Jacqueline McQuillan, for providing the Illumina sequence and data analysis, and also for their friendly hospitality during my visit to Hinxton.

Finally I would like to thank my family for their love and support and Richard for making the whole process possible.

Abbreviations

ABC	ATP-Binding Cassette
AFLP	amplified fragment length polymorphisms
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
BES	bloodstream expression site
BIR	break-induced replication
bp	base pair
BSA	bovine serum albumin
CBSS	Carter's Balanced Salt Solution
cM	centiMorgan
CNS	central nervous system
CQR	chloroquine resistance
CSF	cerebrospinal fluid
DALY	disability adjusted life years
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
DFMO	D,L-a-difluoromethyl ornithine
DGC	directional gene clusters
dGTP	deoxyguanosine triphosphate
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
D-PBS	Dulbecco's phosphate buffered saline
DSB	double strand break
dTTP	deoxythymidine triphosphate
EDTA	ethylenediaminetetraacetic acid
ES	expression site
ESAG	expression site associated gene
FCS	foetal calf serum
g	gram
GFP	green fluorescent protein
GUTat	Glasgow university trypanozoon antigen type
HAPT1	high affinity pentamidine transporter 1
HAT	human African trypanosomiasis
hr	hour
i.p	intraperitoneal
IC_{50}	50% Inhibitory Concentration
IMDM	Iscove's modified Dulbecco's medium
Kb	kilobase
kDNA	kinetoplast DNA
kg	kilogram
km	kilometre
L	litre
LAPT1	low affinity pentamidine transporter 1
LB	Luria-Bertani medium
LOD	log of the odds
LOH	loss of heterozygosity
M	molar
Mb	megabase
MCD	minimum curative dose
mg	milligram

ml	millilitre
mM	millimolar
mRNA	messenger RNA
ng	nanogram
nm	nanometre
nM	nanomolar
ODC	ornithine decarboxylase
ORF	open reading frame
PCR	polymerase chain reaction
PDT	population doubling time
PfCRT	*P.falciparum* chloroquine resistance transporter
PFGE	pulsed field gel electrophoresis
QN	quinine
QTL	quantitative trait loci
RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
RFP	red fluorescent protein
RNA	ribonucleic acid
rpm	revolutions per minute
SDM	semi-defined medium
SDS	sodium dodecyl sulphate
SnoRNA	small nucleolar RNA
SNP	single nucleotide polymorphism
SSC	saline sodium citrate
STIB	Swiss Tropical Institute, Basal
TbAT1	*T.brucei* adenosine transporter 1
TBE	tris-borate EDTA
TE	Tris-EDTA
TIGR	The Institute for Genome Research
TREU	Trypanosome Research Edinburgh University
tRNA	transfer RNA
UV	ultraviolet
v/v	volume per volume
VAT	variable antigen type
VSG	variant surface glycoprotein
w/v	weight per volume
WHO	World Health Organisation
WTSI	Wellcome Trust Sanger Institute
yr	year
µg	microgram
µl	microliter
µM	micromolar

Chapter 1

Introduction

1.1 General introduction to *T.brucei*

The family Trypanosomatidae, of the order Kinetoplastidae are unicellular flagellated protists, which are principally parasitic. Within the nine genera of this family there are two, *Trypanosoma* and *Leishmania*, which contain major medically and economically important pathogens of humans and domestic animals. These include the vector-borne parasites responsible for Chagas disease in the Americas, leishmaniasis in the tropics and subtropics, as well as African trypanosomiasis in sub-Saharan Africa. The causative agents of African trypanosomiasis are members of the genus *Trypanosoma*, which are found in a wide range of vertebrates and except for a few examples (*T.evansi* and *T.equiperdum*) are transmitted from one vertebrate to another via an invertebrate vector. The mode of transmission used by the parasite further defines the genus into two broad categories: the Stercoraria, where faecal contamination transmits the parasites and the Salivaria, in which trypanosomes are injected into the vertebrate host during a blood meal by a biting insect. The first transmission mode is exemplified by the reduviid-transmitted *Trypanosoma cruzi*, which causes Chagas disease and the second by the tsetse-transmitted Africa trypanosomes.

Human and animal African trypanosomiasis (also known as sleeping sickness and nagana respectively) are endemic to sub-Sahara Africa, the distribution of which is restricted by the range of the tsetse fly (*Glossina*) vector. Current estimates put this distribution at 8.5 million km^2, covering 40% of the total land mass of 36 countries, where trypanosomes infect a range of mammalian hosts (reviewed in Allsopp, 2001). *Trypanosoma brucei* is the causative agent of human African trypanosomiasis (HAT) in man and is one of the species responsible for animal trypanosomiasis in domestic livestock. *T.brucei* is comprised of three subspecies that are morphologically identical, but differ in terms of host range, pathogenicity and geographical distribution. The *Trypanosoma brucei brucei* subspecies is non-pathogenic to man, but is able to infect a wide range of wild and domestic animal hosts throughout sub-Saharan Africa where along with the other *Trypanosoma* species, *T.congolense* and *T.vivax*, it is responsible for animal trypanosomiasis.

Trypanosoma brucei gambiense and *Trypanosoma brucei rhodesiense* are the other two subspecies of *T.brucei,* which are both human-infective and are the causative agents of HAT. *T.b.rhodesiense* is responsible for an acute form of HAT, found in eastern and southern African which totals < 10% of the reported cases of

sleeping sickness (WHO, 2006a). This parasite is zoonotic and a number of wild and domestic animal species are known to act as reservoirs for transmission (Heisch *et al.*, 1958; Onyango *et al.*, 1966; Geigy *et al.*, 1971). *T.b.gambiense* is the sub-species found in central and western Africa, where it is responsible for > 90% of reported cases of HAT (WHO, 2006a) and causes a more chronic form of the disease. Although this parasite has not traditionally been considered zoonotic, with humans acting as the main reservoir for disease transmission, *T.b.gambiense* has been detected in a number of animal species (pig, primate, rodent) in West Africa suggesting a possible role for an animal reservoir in its epidemiology (Mehlitz *et al.*, 1982; Njiokou *et al.*, 2006; Simo *et al.*, 2006).

The epidemiology of HAT is related to complex interactions between humans, parasites, the tsetse fly vector and any reservoir hosts. Consequently, the behaviour and distribution of the vector plays an important part in sustaining endemicity and in the re-emergence of epidemics. Despite the widespread abundance of the vector, disease is not present throughout the entire range of the tsetse fly but distributed in discrete foci. The low infection rates of tsetse with *T.brucei*, usually < 1% (Maudlin *et al.*, 1998), limits the epidemic potential of African trypanosomes and as a result, the frequency and intensity of human contact with tsetse are considered major determinants for risk of infection (Méda and Pepin, 2001).

In West and Central Africa, the *palpalis* group of tsetse fly, which feeds frequently on man (Njiokou *et al.*, 2004; Simo *et al.*, 2008b) are the main vector of *T.b.gambiense* (Krafsur, 2009) and vector distribution is closely linked to human occupation patterns. Humans are believed to be the main reservoir for infection so control relies predominantly on the chemotherapeutic treatment of infected persons. Conversely, *T.b.rhodesiense* is known to be zoonotic and mainly transmitted by the *morsitans* group of tsetse fly which live in open woodland and on the outskirts of savannah, primarily away from villages (Krafsur, 2009). They feed preferentially on game animals and domestic stock (Clausen *et al.*, 1998) so endemic outbreaks of human infections tends to be sporadic and it is usually an occupational disease of young active males entering tsetse infested areas. The major risk of human infection occurs if the disease becomes epidemic, more domestic animal - man and man - man transmission occurs and infection becomes peridomestic (Smith *et al.*, 1998). This situation creates the potential for epidemic outbreaks so chemoprophylaxis of the domestic animal reservoir is also important in control.

There are an estimated 60 million people currently living at risk of HAT, discontinuously distributed in approximately 250 endemic foci (WHO, 1998). The majority of these foci of disease are caused by the *T.b.gambiense* form of the parasite, although *T.b.rhodesiense* remains active in certain endemic countries (WHO, 2006a). The most recent World Health Organisation (WHO) update of the disease situation, following a renewed period of active surveillance and control indicates an overall decline in the reported number of people infected from estimates of around 300,000 a decade ago (WHO, 1998) to approximately 70,000 today, with around 17,000 new cases reported annually (WHO, 2006a).

The countries identified with a high prevalence of infection (> 100 cases/yr) are Tanzania and Uganda for *T.b.rhodesiense* and Angola, Democratic Republic of Congo, Sudan, Uganda, Central African Republic, Chad, and Congo for *T.b.gambiense* (Figure 1.1). The geographical distribution of the two forms of the disease does not presently overlap, with only a single country, Uganda, in which foci of both parasites are found. Currently these foci remain discrete (Picozzi *et al.*, 2005), separated by the Great Rift Valley. However there is concern that the movement of infected livestock may present a risk for convergence, which would complicate the diagnosis and treatment of disease, which relies on subspecies specific chemotherapy.

Tsetse-transmitted trypanosomes represent one of the most debilitating limitations on development and agricultural production in rural sub-Sahara Africa. As a major cause of human depopulation, decimation of cattle, and land abandonment the annual economic losses in livestock and crop production alone are estimated to be US$1-5 billion (ILRAD, 1993; ILRI, 1999). In public health terms, the impact of the disease can be measured in Disability Adjusted Life Years (DALYs). The WHO have estimated the burden of HAT to be 1.53 million DALYs which compares to 0.67 million for Chagas disease, which infects 16-18 million people and 2.09 million for leishmaniasis, which infect 12 million people, indicating the substantial burden this disease places on a community (WHO, 2004b).

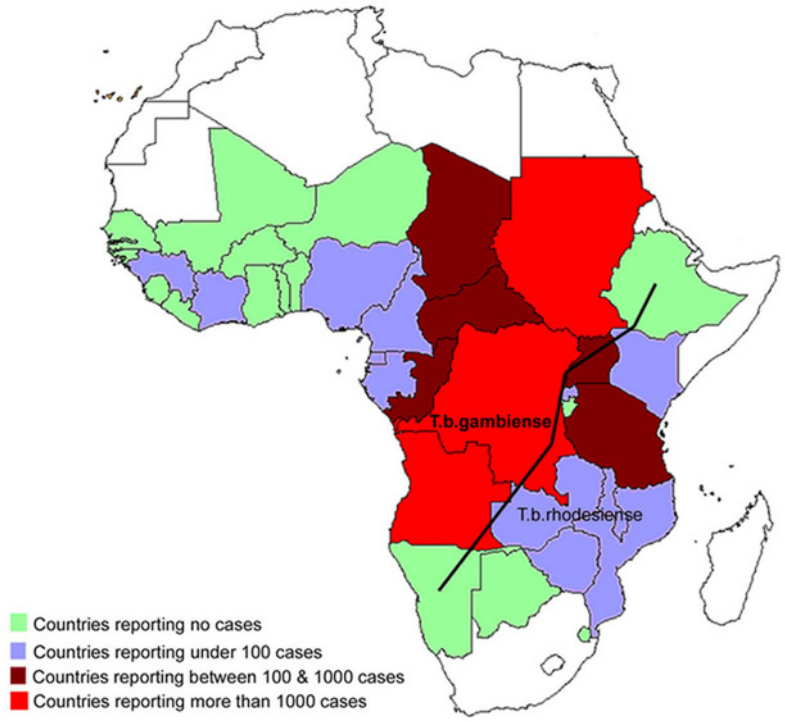


Figure 1.1 Distribution of human African trypanosomiasis in sub-Saharan Africa.

Thirty-six countries in sub-Saharan Africa are considered endemic for HAT caused by either the *T.b.gambiense* or *T.b.rhodesiense* parasite. *T.b.gambiense* is responsible for 97% of infections (WHO, 2006a) and foci are distributed in West and Central Africa (to the left of the dividing line). *T.b.rhodesiense* foci are responsible for the remaining 3% of infections and foci are distributed in East and South Africa (to the right of the line). The number of new cases of both forms of the disease in each endemic country, reported to WHO each year is indicated by one of four colours (Figure reproduced from Simarro *et al.*, 2008 doi:10.1371/journal.pmed.0050055.g003).

1.2 Life cycle

*Trypanosoma brucei s*pecies have a complex life cycle involving both a mammalian host and a tsetse fly vector. The life cycle is completely extracellular and develops through six morphologically distinct stages as the parasites transition between the differing environments of the vector midgut and salivary glands and mammalian host bloodstream (reviewed in Vickerman, 1985; Vickerman *et al.*, 1988; Matthews, 2005). As depicted in Figure 1.2, the parasite alternates between proliferative stages, dividing by binary fission, and non-proliferative stages associated with the necessary adaptive changes as the parasite transitions between host environments.

Infection in the mammalian host begins when metacyclic trypomastigotes are intradermally injected with saliva by the tsetse fly during a blood meal, where they rapidly transform into long slender trypomastigotes. These cells express bloodstream stage-specific variant surface glycoprotein (VSG) coat proteins (Vickerman, 1969). The kinetoplast (an organelle unique to the Kinetoplastidae) is located in a posterior position and the mitochondrial activity is relatively repressed (Vickerman, 1965). Trypomastigotes divide by binary fission at the bite site and then enter the bloodstream via the lymphatics where they proliferate further. As parasite numbers increase in the bloodstream, a significant number of them differentiate to a non-dividing morphologically short stumpy form which has a more developed mitochondrion and is preadapted for survival in the tsetse fly (Vickerman, 1965).

Stumpy form parasites are ingested by a tsetse fly during blood feeding from an infected host. Upon uptake, procyclic forms are generated which multiply in the fly midgut and replace the VSG coat with a procyclic-specific surface coat composed of EP and GPEET procyclins (Acosta-Serrano *et al.*, 1999). The kinetoplast is repositioned and the mitochondrion is activated (Matthews *et al.*, 1995). After multiplying for ~10 days the division arrests in a percentage of cells and they migrate towards the anterior portion of the gut, and then to the salivary glands. Here they transform into epimastigotes, attach to salivary gland tissue by their flagella, and undergo further rounds of proliferation. Finally the epimastigotes differentiate into non-dividing, infective metacyclic trypomastigotes approximately 20 days into infection. This stage is preadapted for survival in the mammalian host having acquired a metacyclic specific VSG coat (Turner *et al.*, 1988). The metacyclics are positioned freely in the tsetse saliva

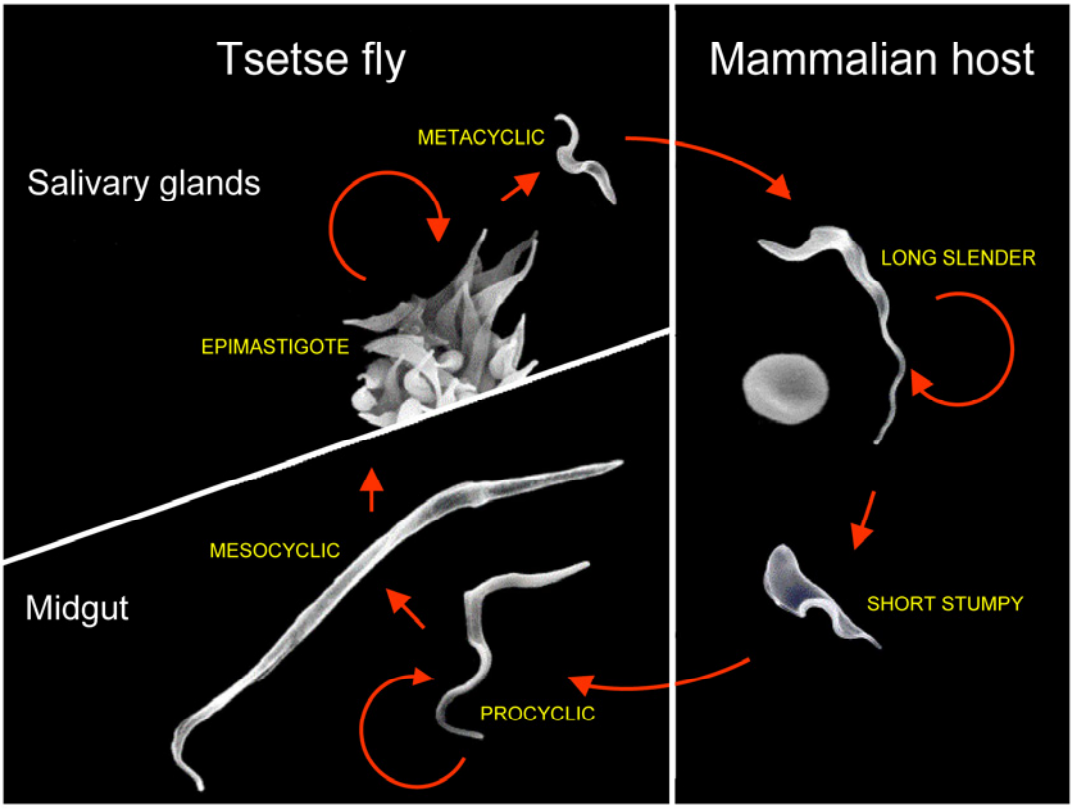


Figure 1.2 The life cycle of *T.brucei* in the tsetse fly vector and mammalian host.

Scanning electron micrographs of the six different *T.brucei* lifecycle stages are shown to scale. There are three proliferative stages (indicated by circular arrows); the long slender stage in the bloodstream of the mammalian host, the procyclic stage in the tsetse midgut and the epimastigote stage in the tsetse salivary gland. The short stumpy stage in the host and the mesocyclic and metacyclic stages in the tsetse fly are non-proliferative. Electron micrographs were produced by Dr Laurence Tetley and the figure adapted from Barry and McCulloch (2001). Reproduced with the permission of D.Barry and L.Tetley.

where they can be transmitted to a new mammalian host when the fly takes its next blood meal.

T.brucei has developed unique mechanisms to facilitate its survival in these very different environments, in particular the changes that occur to the mitochondrion and the parasites cell surface, as well as more obvious morphological changes. Alterations in the mitochondrion occur to allow successful transfer of the metabolism from the glucose and nutrient-rich mammalian host bloodstream where the trypomastigote mitochondrion is repressed and the parasite rely on simple aerobic glycolysis, to the invertebrate vector. Within the tsetse, nutrients are more limited and the procyclic form mitochondrion becomes activated to allow the parasites to metabolise the amino acid proline from which the parasite must derive its energy (Matthews, 2005).

1.3 Antigenic variation

Another important adaptation of the African trypanosomes, as an extracellular parasite living freely in the bloodstream of the mammalian host, is the highly evolved and elaborate strategy for immune evasion, termed antigenic variation (reviewed in Barry and McCulloch, 2001; McCulloch, 2004; Morrison *et al.*, 2009a). Characteristic fluctuations in parasitemia are brought about by an interaction of the host's immune system with bloodstream form trypomastigotes that are able to periodically change their antigenic surface coat. This surface coat is composed of approximately 5 x 10^6 copies (Auffret and Turner, 1981) of identical VSG proteins (Cross, 1975) which form a densely packed monolayer 12-15 nm thick (Vickerman, 1969). The VSG coat is only present in the infective metacyclic and mammalian stages of the life cycle and functions to protect the underlying invariant antigens from the immune system.

As parasites proliferate in the bloodstream, the host's immune system recognises and responds to the dominant VSG types in the population with antibodies that facilitate the neutralization and killing of trypanosomes expressing these Variable Antigenic Types (VATs). Trypanosomes that have switched their VAT so that they express a different, antigenically distinct VSG coat however will be able to evade detection. These antigenically distinct trypanosomes give rise to a new population and cause a recrudescence in the parasitemia. The immune system will once again respond to this proliferated population by producing a new set of antibodies that mount an effective assault on that new dominant VSG in the population. Once

again, VSG switching among a small portion of the trypanosomes will render them undetectable, and they successfully evade the host immune response, prolonging an infection that can last for many years. Only one VSG is expressed at any one time by the parasite and switching of this expression to a new VSG is a spontaneous event estimated to occur at rates of up to 10^{-3} switches/cell/division (Turner and Barry, 1989).

A silent repertoire in excess of 1600 *VSG* genes has been identified in the trypanosome genome (Marcello and Barry, 2007). The VSGs are positioned largely in subtelomeric arrays on the megabase chromosomes (Berriman *et al.*, 2005) but also at the ends of the smaller multiple copy minichromosomes (Williams *et al.*, 1982). Only a small proportion of the *VSG* genes present in silent arrays are actually predicted to be fully functional VSGs (~5%) with the majority being pseudogenes which provide an archive for recombination reactions to make fully functional mosaic VSGs (Marcello and Barry, 2007).

In order for a *VSG* gene to be expressed in the mammalian host, it must be located in a specialised subtelomeric transcription unit, termed the bloodstream expression sites (BES). These BESs are positioned proximal to the telomeres on one or both ends of the megabase chromosomes and also on a subset of smaller nuclear chromosomes termed the intermediate chromosomes (Hertz-Fowler *et al.*, 2008). Bloodstream expression sites occur as polycistronic transcription units which contain expression site associated genes (ESAGs) in addition to the active VSG, and only one BES is under active transcription at any one time (Johnson *et al.*, 1987; Kooter *et al.*, 1987; Berriman *et al.*, 2002). The actual number of these sites varies between strains, but is around 20 in the most studied trypanosome strain, *T.b.brucei* Lister 427 (Becker *et al.*, 2004).

1.4 Mechanisms of VSG switching

African trypanosomes have evolved a number of different mechanisms by which the switching of active VSG expression can occur (reviewed in Barry and McCulloch, 2001; Morrison *et al.*, 2009a). These mechanisms can be broadly divided into two basic categories; *in situ* transcriptional switching, in which transcription from the currently active expression site is silenced and an alternative expression site is concomitantly activated, and the more frequent recombinational switching, in which the VSG in the active BES is replaced with a silent VSG from the archive, but the same BES remains active.

The transcriptional *in situ* mechanism involves a synchronized switch in which the active BES becomes transcriptionally silenced and transcription at a different one of the 20 or so BES in the genome becomes activated. In contrast to the other mechanisms this does not require the use of a recombination reaction and is thought to occur considerably less often than the recombinational switching mechanisms (< 10^{-5} events/cell/generation) (Ulbert *et al.*, 2002). The molecular mechanism responsible for *in situ* switching however has not yet been determined (Barry and McCulloch, 2001; Borst and Ulbert, 2001).

Recombination-based mechanisms are the most frequently reported strategy used by *T.brucei* for VSG switching. Three different pathways have been described; duplicative transposition, reciprocal telomere exchange and segmented gene conversion, which differ in the VSG-containing source DNA and whether the DNA is duplicated or exchanged into the BES. All three pathways however are believed to exploit common homologous recombination machinery with the necessary homology provided by repetitive flanking regions, conserved sequence up or downstream of the VSG or the VSGs themselves.

The most frequently observed mechanism experimentally (Robinson *et al.*, 1999) is duplicative transposition which involves the duplication of genes from the silent VSG repertoire into the active BES, for example by gene conversion or break-induced replication (BIR). This results in the deletion of the currently expressed *VSG* gene from the BES but leaves the template DNA intact. A variety of different sites can act as a template for this type of recombination including the VSGs from subtelomeric arrays or mini chromosomes or those in expression sites on megabase and intermediate chromosomes. The amount of DNA copied can also vary. Duplicative transposition normally extends beyond the VSG sequence to region of homology up and downstream of the *VSG* gene. The upstream boundary is usually a set of imperfect 70 bp repeats (Liu *et al.*, 1983) that have been observed to flank > 90% of VSGs genes (Marcello and Barry, 2007), but can extend beyond that to other genes in the BES (Kooter *et al.*, 1988). The downstream limits of gene conversion can also vary, from conserved sequences at the 3' end of the VSG (Borst and Cross, 1982) all the way to the end of the chromosome (De Lange *et al.*, 1983).

The second recombination pathway, reciprocal telomere exchange, involved a straightforward chromosomal crossover between homologous sequence in the active BES and a telomeric VSG sequence on another chromosome. A variety of

different homologous sequences can be targeted upstream of the VSG by this approach (Pays *et al.*, 1983b; Pays *et al.*, 1985), and thus the size of the exchanged DNA fragments can vary, but the reciprocal exchange involves the entire region downstream to the telomere. This type of recombination mechanism is limited to telomerically located *VSG* genes and is thus believed to be a less commonly used mechanism than duplicative transposition, as the large VSG arrays are not suitable templates for reciprocal exchange.

As the infection progresses, a third recombination pathway is used to generate novel *VSG* genes through a segmental gene conversion reaction combining multiple VSG pseudogenes and gene fragments (Pays *et al.*, 1983c; Roth *et al.*, 1989). The VSG open reading frame (ORF) sequence provides the homology or microhomology for the gene conversion reaction rather than the flanking sequences that predominate in the other two mechanisms. VSG pseudogenes have been observed to make up the majority of the VSG archive (Marcello and Barry, 2007), and thus segmented gene conversions are postulated to play a major role in *T.brucei* antigenic variation, particularly later in infection when the mammalian host is likely to have generated an immune response to many of the conventional VSGs (Thon *et al.*, 1990; Barbet and Kamper, 1993).

1.5 Pathology of Human African Trypanosomiasis

Human African trypanosomiasis caused by either *T.b.rhodesiense* or *T.b.gambiense,* is traditionally divided into two stages: early (haemolymphatic), in which the parasite proliferates in the blood and lymph and late (encephalitic), when the parasite has crossed the blood brain barrier and invaded the central nervous system (CNS). Although there is no clear clinical distinction between the two stages, a differentiation is made principally for ease of treatment, as the recommended chemotherapy is dependant on whether the infection has reached the CNS stage.

Starting at the bite site, parasites multiply in the tissue resulting in localised inflammation that culminates in the development of a lesion known as a chancre at the bite site 4 - 14 days following inoculation. From this site parasites spread to the draining lymphatic and blood systems where they divide by binary fission. Increasing parasitemia causes symptoms such as malaise, fever, headache, anaemia and generalised oedema and itching. Some early swelling of cervical lymph glands, known as Winterbottom's sign, may be evident in *T.b.gambiense*

but generally symptoms are fairly non-specific and may be easily confused with other diseases such as malaria (Stich *et al.*, 2002).

Central nervous system involvement may start weeks to years after the initial infection as parasites cross the blood brain barrier and invade brain and cerebrospinal fluid (CSF). As parasites multiply in the CNS a range of neurological signs and symptoms gradually appear. The late stage is characterised by progressive neurological disturbance to cognitive, motor, sensory and circadian rhythms. Symptoms include irritability, loss of concentration, changes in personality, lassitude and anxiety. Later this progresses to violent mood swings, delirium, hallucinations, suicidal tendency and manic episodes and cerebral ataxia leads to difficulty with walking and speech. Sleep disturbance is the central feature of gambiense, and to a lesser extent, rhodesiense sleeping sickness, characterised by a reduction in alertness, daytime somnolence and night time restlessness which in later stages progresses to coma and eventually death (Stich *et al.*, 2002; Kennedy, 2004).

Differences in HAT caused by the two human-infective subspecies occur largely in the speed at which the disease progresses. *T.b.gambiense* causes a chronic infection in which a person may be infected for months or even years before symptoms emerge. However once clinical disease is evident the infection tends to be quite advanced. *T.b.rhodesiense* is more virulent and disease develops rapidly with symptoms emerging in weeks (WHO, 2006b). This difference has important implications for control of the disease. Diagnosis of infection usually involves detection of parasites in lymph or blood followed by lumbar puncture to determine if there is CNS involvement, as different therapeutic interventions are needed depending on the disease progression (Kennedy, 2004).

Traditionally, infection with either of the causative agents of HAT is considered to involve a natural and inevitable progression from early stage to late stage infection, finally resulting in a fatal outcome if chemotherapeutic intervention is not received. However, the possibility of self-resolving or asymptomatic chronic carriage, for *T.b.gambiense* at least, has been postulated as a rare outcome of infection (Checchi *et al.*, 2008).

1.6 *T.brucei* in the laboratory

The study of *T.brucei* is valuable for understanding both the disease and biology of the parasite, but also has much wider applications as a lower eukaryotic model organism for gaining insights into molecular biology, biochemical and cell biology processes. Much of this understanding has come about from the analysis of *T.brucei* in the laboratory, and consequently both *in vivo* and *in vitro* systems have been development for studying the parasite and its interaction with the mammalian host and tsetse fly vector.

Animal models of African trypanosomiasis, most usually rodent, have been developed to investigate the pathology of *T.brucei* in the mammalian host (Poltera *et al.*, 1980; Keita *et al.*, 1997; Kennedy, 2006), investigate biological processes such as antigenic variation (Turner, 1997; Morrison *et al.*, 2005) and to test novel drug compounds *in vivo* (Kodama *et al.*, 2008; Thuita *et al.*, 2008; Barker *et al.*, 2009). In combination with laboratory colonies of tsetse flies, the full lifecycle can also be completed in the laboratory allowing all stages to be observed and providing the resources for genetic crosses to be performed (Jenni *et al.*, 1986). In addition, an axenic culture system has also been developed for two of the dividing lifecycle stage of *T.brucei*; the mammalian bloodstream stage (Hirumi *et al.*, 1977; Hirumi and Hirumi, 1989) and the tsetse midgut stage procyclic stage (Brun and Schonenberger, 1979) allowing a range of *in vitro* studies to be conducted.

T.brucei strains, which are adapted to growth in the laboratory and retain the ability to differentiate from replicating long slender forms into non-dividing short stumpy forms in the mammalian host, are known as pleomorphic lines (Vickerman, 1965). These lines are not ideal for all laboratory applications as they tend to grow to only a low density in axenic culture and rodents (Turner *et al.*, 1995), characterised by an oscillating parasitemia in the latter and undergo frequent VSG switching events (Turner and Barry, 1989). Although this makes them less amenable for certain genetic and biological studies, for investigating differentiation and development in the life cycle and performing genetic crosses in the tsetse fly, pleomorphic lines are essential. For this reason a pleomorphic strain of *T.b.brucei*, Trypanosome Research Edinburgh University (TREU) 927, was selected as the reference strain for the *T.brucei* genome sequence (van Deursen *et al.*, 2001; Berriman *et al.*, 2005).

The majority of research into the genetic and biochemical processes of *T.brucei* however, has been conducted in a monomorphic *T.brucei* line, Lister 427. Such monomorphic lines have been adapted to laboratory manipulation through serial syringe passage in rodents and have lost the ability to differentiate into stumpy form parasites. The frequent passage of monomorphic lines such as Lister 427 has also resulted in the loss of their ability to complete the life cycle in the tsetse fly vector and thus they are no longer tsetse transmissible (Peacock *et al.*, 2008). However, Lister 427 is antigenically stable (Barry, 1997), amenable to genetic manipulation (McCulloch *et al.*, 2004), and highly adapted to growth in laboratory systems, growing to a density of 5×10^6 cells/ml in culture and 1×10^9 cells/ml in rodents (Hesse *et al.*, 1995), making them ideal for experimental use.

Lister 427, was the first *T.brucei* strain to be adapted to *in vitro* culture and has been widely disseminated to laboratories around the world (Cross and Manning, 1973). Much of our current understanding into the process of antigenic variation, DNA repair and recombination in *T.brucei* has been determined using this strain. Recent evidence however, has suggested that Lister 427 has undergone substantial chromosomal alterations during its passage history indicating that trypanosome lines can undergo both phenotypic and genomic changes when adapted to laboratory culture (Peacock *et al.*, 2008).

1.7 Surveillance and control of *T.brucei*

The long duration of asymptomatic infection is considered a key determinant behind the endemic and epidemic potential of HAT. Sleeping sickness is thus one of few infectious diseases where proactive systematic population screening is essential for control, especially in *T.b.gambiense* which has a large human reservoir and longer incubation period (WHO, 2004a). Early detection of cases can thus have a dramatic effect on transmission rates and a combination of active case detection and treatment is the cornerstone of the WHO strategy of prevention and disease control (WHO, 2004a). For *T.b.rhodesiense*, because reservoir hosts also plays an important part in sustaining endemicity and of re-emergence of epidemics the chemotherapy of domestic animals is also important for control (Fevre *et al.*, 2005).

1.8 Chemotherapy

Because of the parasite's ability to undergo antigenic variation, there is little prospect for a *T.brucei* vaccine. Consequently, chemotherapy remains the cornerstone of disease control for the foreseeable future. However this approach relies on a very limited number of drugs with the disadvantages of high toxicity, limited efficacy, undesirable routes of administration and emerging treatment failure/drug resistance. There are currently only four drugs available for treating HAT, all but one of which were developed more than 50 years ago. These are pentamidine and suramin for early stage and eflornithine and melarsoprol for late stage disease treatment. In addition to considering the clinical stage of the disease, chemotherapy choice is also dependent on which subspecies, *T.b.gambiense* or *T.b.rhodesiense*, is thought be causing the infection. Three drugs are available for the treatment of animal trypanosomiasis, isometamidium chloride (Samorin), diminazene aceturate (Berenil) and homidium. In contrast to the drugs used to treat HAT, one of these drugs, isometamidium chloride, can be used prophylactically in cattle to prevent infection while the other two are used therapeutically on a case-by-case basis.

1.8.1 Human African trypanosomiasis drugs

1.8.1.1 Suramin

Suramin is a polysulphonated symmetrical naphthalene derivative which was introduced for use against early stage trypanosomiasis in 1922 (Wang, 1995). Although it may be used for the treatment of the haemolymphatic stage of either form of disease, due to reported treatment failures in the 1950s for the *T.b.gambiense* subspecies, it is the drug of choice for early *T.b.rhodesiense* infections only (Pepin and Milord, 1994). The drug is highly negatively charged and thus readily soluble in water, but cannot pass through biomembranes and is consequently unable to cross the blood brain barrier (Fairlamb, 2003). This limits the application of suramin to the treatment of early stage disease only. However its negative charge does allow it to bind to many serum proteins with high avidity such as low density lipoprotein through which it may enter trypanosomes by receptor mediated endocytosis (Vansterkenburg *et al.*, 1993). The mode of action of suramin is not clear but its high negative charge may allow it to bind to and inhibit many metabolic enzymes (Docampo and Moreno, 2003) including transport

of glycolytic enzymes into glycosomes (Wang, 1995) and many different dehydrogenases and kinases including dihydrofolate reductase and thymidine kinase (Chello and Jaffe, 1972). Resistance to suramin in the field is not currently considered a significant problem, however suramin resistant lines can be readily generated in the laboratory by subcurative drug treatment (Scott *et al.*, 1996).

1.8.1.2 Pentamidine

Pentamidine is one of the diamidine class of drugs, which have a long history of involvement in the treatment of protozoan infections. It was first introduced for the treatment of HAT in 1937 and remains the first line treatment of choice for early stage *T.b.gambiense* infections. The drug is highly protonated at physiological pH and because it is slowly excreted, daily doses can accumulate in cells to a intracellular concentration as high as 1 mM (Damper and Patton, 1976). Pentamidine diffuses only slowly across biomembranes, requiring transporter-mediated uptake so is ineffective in treating late stage infection. Pentamidine uptake has been extensively characterised and demonstrated to be taken up by various transporters. These include the aminopurine P2 transporter (Carter *et al.*, 1995), which also mediates uptake of Berenil (de Koning *et al.*, 2004) and the melaminophenyl arsenicals (Carter and Fairlamb, 1993), and the activity of a High Affinity Pentamidine Transporter (HAPT1) and Low Affinity Pentamidine Transporter (LAPT1) have also been detected (De Koning, 2001b).

The precise mechanism of action of pentamidine is unknown but as a polycation it has the ability to interact electrostatically with cellular anions such as nucleic acids with possible effects on trans-splicing, RNA editing, kinetoplast DNA (kDNA) replication and polyamine synthesis (Fairlamb, 2003). Examples of pentamidine resistance in the field have been reported (Dukes, 1984) and both *in vivo* and *in vitro* pentamidine resistance lines can be readily generated in the laboratory (Berger *et al.*, 1995; Bernhard *et al.*, 2007; Bridges *et al.*, 2007). In addition, some examples of these have cross-resistance both to other diamidines and the melaminophenyl class of drugs (Bernhard *et al.*, 2007; Bridges *et al.*, 2007).

1.8.1.3 Melarsoprol

Melarsoprol is a melaminophenyl based organic arsenical that was introduced as a trypanocidal agent in 1949 (Fairlamb, 2003). It can be used to treat late stage disease caused by either subspecies of parasite and it is the only effective

treatment of late stage *T.b.rhodesiense* infection. It has the highest toxicity of all the trypanocides, causing a serious reactive encephalopathy in 5-10% patients, half of whom will die as a result of the treatment (Pepin and Milord, 1994). The prodrug is amphipathic and so has the capability to diffuse across biological membranes. However evidence suggests that the prodrug is rapidly converted to the highly hydrophilic active metabolite Melarsen oxide (Melox), within 15 minutes of administration, thus requiring alterative mechanisms for uptake (Nok, 2003). One method of transporter-mediated uptake has already been identified, via the same P2 aminopurine transporter used by the diamidine class of drugs (Carter and Fairlamb, 1993). Research also suggests that other transporters exist, one route of which is likely to be the HAPT1 transporter (Matovu *et al.*, 2003).

The mode of action of melarsoprol is unknown, but cells are lysed quickly *in vitro* and various different possibilities have been postulated. Arsenicals are known to form a stable interaction with thiols so it may react with trypanothione, a low molecular weight thiol unique to trypanosomes, to form MelT a competitive inhibitor of trypanothione reductase which is essential in maintaining the intracellular thiol-redox balance (Fairlamb *et al.*, 1989). It may also block glycolysis in parasites through inhibition of pyruvate kinase, as trypanosomes rely solely on glycolysis for energy provision, leading to cell lysis (Wang, 1995).

The average treatment failure rate for melarsoprol is historically less than 10% (Barrett, 2001), which is thought to be due to a number of pharmokinetic and administration factors. Over the last decade however, treatment failure has increasingly being reported from epidemic foci at much higher levels including 30% in North West Uganda (Legros *et al.*, 1999), 21% in Sudan (Brun *et al.*, 2001) and 25% in Angola (Stanghellini and Josenando, 2001) indicating a potential drug resistance problem.

1.8.1.4 Eflornithine (DFMO)

Eflornithine or D,L-α-difluoromethyl ornithine (DFMO) is an analogue of ornithine which acts as a specific suicide inhibitor of ornithine decarboxylase (ODC). It was first registered for use against trypanosomes in the 1990s, following initial development as an anticancer drug (Burri and Brun, 2003) and is the only new drug that has been registered for the treatment of sleeping sickness in the last 50 years. Eflornithine is effective in treating both early and late stages of *T.b.gambiense* infection and has a better safety profile than melarsoprol

(Chappuis *et al.*, 2005) but is not effective in treating *T.b.rhodesiense* and is presently very costly.

This is the only drug whose mode of action has been determined (reviewed in Wang, 1995; Fairlamb, 2003). It is a selective irreversible inhibitor of ODC, which is a key enzyme in polyamine synthesis. *T.brucei* has limited capacity to scavenge the trace amounts of polyamines found in the serum of the mammalian host thus inhibition of its polyamine synthesis cycle rapidly depletes polyamine levels in the trypanosomes. This results in the disturbance of S-adensyl methionine metabolism, reduction in DNA, RNA and protein synthesis and brings them to a trypanostatic status, which requires an intact humeral immune response to kill. Eflornithine has a similar affinity for both mammalian and trypanosomal ODC but specificity arises due to differences in enzyme turnover rates. Mammalian ODC is degraded rapidly and replaced while *T.b.gambiense* ODC is stable and once irreversibly inhibited cannot be replaced. This is also responsible for the natural insensitivity of *T.b.rhodesiense* to eflornithine that, like mammalian ODC, has a more rapid turnover. The high number of reported melarsoprol-refractory cases in some foci and the improved safety record of eflornithine has led some health centres to switch to eflornithine for first line treatment of *T.b.gambiense*, but currently the cost is prohibitive for most rural centres (Burri and Brun, 2003).

1.8.2 Veterinary drugs

1.8.2.1 Homidium

Homidium is a phenanthridium compound first introduced 50 years ago and used extensively in the 1960s and 1970s. Its mode of action is not fully understood but may work by interfering with glycosomal function, trypanothione metabolism or kDNA minicircle replication (Wang, 1995). Homidium chloride, homidium bromide and ethidium are still employed but their utility has considerably reduced due to widespread resistance (Codjia *et al.*, 1993) and the drug has the disadvantage of being a potent mutagen which may increase the occurrence of resistance (Wang, 1995).

1.8.2.2 Diminazene aceturate (Berenil)

Berenil is an aromatic diamidine related to pentamidine and with similar biochemical properties. It was first introduced in 1955 as a trypanocide and

babesiacide agent and is the most commonly used veterinary therapeutic agent for animal African trypanosomiasis (Anene *et al.*, 2001). Like pentamidine, the drug is a cation that diffuses only slowly across biological membranes so utilises the P2 aminopurine transporter to enter cells (de Koning *et al.*, 2004). Evidence suggests Berenil does not accumulate via the HAPT1 and LAPT1 transporters used by pentamidine but that additional uptake may also occur via a P2-independent method that has yet to be quantified (de Koning *et al.*, 2004). The mode of action is thought to work in a similar method to pentamidine by binding to the minor groove in DNA where it may inhibit kinetoplast topoisomerase II, RNA editing and RNA trans-splicing (Wang, 1995). Resistance to Berenil is an increasing problem in the control of trypanosomiasis in domestic animals in Africa (Afewerk *et al.*, 2000) and both drug resistant and cross-resistant lines have been observed in the field and generated in the laboratory (Scott, 1995; McDermott *et al.*, 2003).

1.8.2.3 Isometamidium chloride (Samorin)

This drug is a conjugate of homidium and part of the Berenil molecule, and is used exclusively as a veterinary trypanocide both prophylactically and therapeutically. Isometamidium chloride can be used prophylactically due to its ability to stay in circulation for up to three months post treatment (Matovu *et al.*, 2001a). Resistance is a problem in many parts of Africa with cross-resistance to homidium (Afewerk *et al.*, 2000) and it is now being recommended that the drug be administered as a sanative pair with periodic Berenil treatment (de Koning, 2001a). The mode of action is unknown but the drug has been shown to accumulate in the kinetoplast and it has been suggested that the main mode of action may involve cleavage of kDNA topoisomerase complexes (Shapiro and Englund, 1990).

1.9 Drug resistance

1.9.1 Drug resistance in African trypanosomes

Resistance to the melaminophenyl arsenicals (Pospichal *et al.*, 1994; Scott *et al.*, 1996; Brun *et al.*, 2001) and diamidines (Berger *et al.*, 1995; McDermott *et al.*, 2003; Bridges *et al.*, 2007) as well as suramin (Payne *et al.*, 1994; Scott *et al.*, 1996), Eflornithine (Phillips and Wang, 1987), isometamidium chloride (Afewerk *et al.*, 2000; McDermott *et al.*, 2003) and homidium (Na'Isa, 1967; Codjia *et al.*, 1993; Mulugeta *et al.*, 1997) has been reported either in the field or in laboratory

generated *in vivo* and *in vitro* strains. More worryingly, cross-resistance between the melaminophenyl arsenicals (e.g. melarsoprol), and the diamidines, (e.g. the HAT drug pentamidine and the veterinary drug Berenil) (Williamson and Rollo, 1959; Pospichal *et al.*, 1994; Bernhard *et al.*, 2007; Bridges *et al.*, 2007) has been reported despite very different proposed mechanisms of action (reviewed in Wang, 1995).

There are very few drugs available for the effective treatment of human and animal trypanosomiasis in sub-Saharan Africa, and only limited prospective new drugs on the horizon. An understanding of the mechanisms of acquired and potential resistance to these drugs is therefore crucial to circumvent existing treatment failure problems and reduce the selection of resistance to the next generation of drugs. It can also potentially lead to the identification of novel drug targets, provide insight into the mode of action of drugs and provide a rationale for new chemotherapeutic strategies to overcome or reverse resistance such as combination chemotherapy.

The repeated use of chemotherapy inevitably leads to the development of resistance in the target organism. The speed at which drug resistance develops however is influenced by the contribution of selection pressure factors such as the extent of drug use, under-dosing, self-administration and improper prophylactic use (Geerts *et al.*, 2001). Exposure of parasites to mass prophylactic treatment and sub-therapeutic drug concentrations are considered important factors for development of resistance and exert a strong selective pressure for the emergence of resistant clones pre-existing in the population (Boyt, 1986; Geerts, 1998).

The drugs used to treat HAT have several advantages that have slowed down the development of resistance. These include the limited distribution and number of cases; historically ~300,000 cases/yr compared to 300 million for malaria, the fact that HAT drugs are not given prophylactically, and that the administration of drugs is parenteral and occurs in a clinical setting thus reducing the potential for misuse (Barrett, 2001).

However despite these factors, resistant lines can be readily selected in the laboratory and treatment failures have been reported, particularly to melarsoprol, the principal drug used against late stage HAT infection and only drug available for treating late stage *T.b.rhodesiense* infection. The concern for HAT is that even

in instances where parasites in humans are not exposed to classical risk factors for developing drug resistance, a contributing factor that may be increasing the speed of resistance is the observed cross-resistance with veterinary drugs (Barrett, 2001).

Isometamidium chloride, diminazene and homidium have been in use for more than 35 years and an estimated 35 million doses/yr are currently used in sub-Saharan Africa (Geerts, 1998). Trypanocidal drugs remain the principle method of control for animal African trypanosomiasis in most affected countries for both a prophylaxis and treatment as they are easily available, affordable and effective. Under dosing of trypanosomes is thought to occur frequently though and their free availability increases the risks of misuse (Geerts, 1998), resulting in a growing concern that their future effectiveness may be severely affected by widespread drug resistance. Resistance to one or more of the three main veterinary drugs has been reported in more than 13 countries in sub-Saharan Africa (Peregrine, 1994; Afewerk *et al.*, 2000) and even multi-drug resistance in several foci (Codjia *et al.*, 1993; Mulugeta *et al.*, 1997; Afewerk *et al.*, 2000) and this is likely to be an under estimation because of a lack of reliable data at a regional and national level (Geerts, 1998).

Although different compounds are used in the treatment of human and animal African trypanosomiasis, some of the drugs are chemically related, for example the HAT drug pentamidine and the veterinary drug Berenil are both diamidines, and there is evidence for the potential of cross-resistance between the arsenical and diamidines classes of drugs. In particular pentamidine/Berenil/melarsoprol cross-resistance may have profound implications on the development of drug resistance in human sleeping sickness in the field (Barrett, 2001).

The epidemiology and effective chemotherapy of HAT is complicated by the zoonotic nature of the parasites and the possibility of genetic exchange between trypanosomes. Although *T.brucei* has been demonstrated to be capable of genetic exchange in the laboratory (Jenni *et al.*, 1986), it is not yet clear how often this process might occur in natural populations. In general the prevalence of *T.brucei* infection of tsetse flies is thought to be relatively low, which might preclude such interactions, however examples of mixed infections has been reported in at least two tsetse fly populations, suggesting the opportunity for genetic exchange to occur does exist (MacLeod *et al.*, 1999).

Both *T.b.rhodesiense* and *T.b.gambiense* are also, to a differing extent, zoonotic and it is unclear how the interactions of human-infective and non human-infective strains might affect the spread of drug resistance in the field. In *T.b.gambiense*, experimental infections of a variety of animals is possible and in particular the role of the domestic pig as a possible reservoir host has been implicated (Nkinin *et al.*, 2002) but little is still known regarding any role they may play in the epidemiology of infection. Conversely *T.b.rhodesiense* is zoonotic and known to have a wide range of domestic and wild animal reservoir hosts, which have a greater impact on the disease epidemiology.

Research conducted in Uganda and Kenya found that up to 20-44% of trypanosomes isolated from cattle were human-infective (Matovu *et al.*, 1997; Hide, 1999), suggesting that human-infective strains of *T.brucei* have the potential to be present in domestic animals and thus may be exposed to treatment with veterinary drugs. This may have important implications offering two possible routes of increasing drug resistance to human infective parasites; 1) resistance to veterinary drugs develops in human-infective strains present in animals during treatment, which provides cross-resistance to diamidine and arsenical drugs if transmitted to humans, 2) human-infective strains may acquire resistance from a non-human-infective strain as a result of genetic exchange during co-infection of a tsetse fly (Barrett, 2001). There is currently no evidence to suggest whether either of these processes occurs in natural populations but the theoretical possibility does exist. The relationship between drug, host and parasite and cross-resistance between veterinary and human trypanocides thus needs to be taken into account when considering drug resistance and appropriate chemotherapy in areas where human and animal trypanosomiasis co-exists.

1.9.2 Mechanisms of drug resistance in *T.brucei*

Drug resistance can arise due to a variety of mechanisms that include; a reduction in drug accumulation through increased drug efflux or reduced drug uptake, drug inactivation via sequestration or metabolism, alteration of the drug target or use of an alternative metabolic pathway. Some progress has been made in understanding mechanisms of drug resistance in African trypanosomes and to date these have been largely associated with changes in drug uptake.

1.9.2.1 P2/*TbAT1*

Transport studies have observed that *T.brucei* contains two high affinity transport systems for adenosine (Carter and Fairlamb, 1993). These purine transporters are important to the cells as they cannot synthesize their own purines *de novo* and must scavenge nucleic acids directly from the host. The adenosine transporters, P1 and P2 have been identified for this purpose, distinguished by their inhibition with different competitive substrates. P1 is a broad specificity purine transporter that transports inosine and adenosine whereas the P2 transporter transports adenosine and adenine. In addition to the uptake of aminopurines, the P2 transporter is also able to transport a range of other substrates that include isometamidium chloride (de Koning, 2001a), the melaminophenyl arsenicals (Carter and Fairlamb, 1993) and the diamidine classes of trypanocidal drugs (Carter *et al.*, 1995). This ability to transport apparently dissimilar compounds such as adenosine, the diamidines and melaminophenyl arsenicals was resolved with the identification of a common recognition motif, which is shared by all the substrates (Barrett and Fairlamb, 1999; de Koning and Jarvis, 1999).

A reduction in drug uptake through the disruption of the P2 transporter activity or the coding gene, *TbAT1* (Maser *et al.*, 1999), has been implicated in several laboratory-generated drug resistant lines to pentamidine, Berenil and melaminoarsenicals (Maser *et al.*, 1999; Bridges *et al.*, 2007). In addition, a number of point mutations in the *TbAT1* gene identified in a laboratory-generated melarsoprol resistance line (Maser *et al.*, 1999) have also been found in field isolates and melarsoprol refractory cases in Uganda (Matovu *et al.*, 2001b). The association however is not clear-cut as *TbAT1* mutations have not been found in all melarsoprol-refractory cases and not all strains exhibiting *TbAT1* mutations are melarsoprol-refractory.

In the laboratory, genetic disruption of *TbAT1*, creating a null mutant in which all P2 activity is lost has been investigated (Matovu *et al.*, 2003). Although loss of P2 appears sufficient for relatively high level resistance to Berenil (19-fold), it results in only low levels of resistance to pentamidine and the melaminophenyl arsenicals of approximately 2-3 fold (Matovu *et al.*, 2003). The resistance attributed to loss of P2 is considerably lower than that observed in many of the drug resistant lines (Carter and Fairlamb, 1993; Scott *et al.*, 1996; Bridges *et al.*, 2007) indicating the probable accumulation of multiple mutations during the course of selection and the likelihood that other resistance mechanisms exist.

Studies performed by de Koning and others have suggest that P2 is not the only transporter of these drugs and that additional transporters exist for pentamidine, the melaminoarsenicals and to a lesser extent, Berenil (Carter *et al.*, 1995; Maser *et al.*, 1999; De Koning, 2001b). Two additional pentamidine transporters have been biochemically characterised in *T.brucei*, which are estimated to be responsible for 30-50% pentamidine uptake, although the genes and natural substrates have not yet been identified. These include a high affinity pentamidine transporter (HAPT1) (De Koning, 2001b; de Koning and Jarvis, 2001), which has also been shown capable of transporting melaminophenyl arsenicals (Matovu *et al.*, 2003) and a low affinity transporter (LAPT1) (De Koning, 2001b; de Koning and Jarvis, 2001) without apparent cross-reactivity to other trypanocides. Loss of either or both of these transporters has been proposed as possible mechanisms of resistance in *T.brucei*. The combined loss of both P2 and HAPT1 activity has been observed in both a Cymelarsan-selected STIB 386 resistance line and a pentamidine-selected *TbAT1 -/-* resistant line (Bridges *et al.*, 2007). The exact contribution made by each of the transporters to the drug resistance profile, both in isolation, and in combination remains to be established however. Loss of LAPT1 function, which does not appear to transport the arsenicals, has also been postulated as a possible mechanism of pentamidine resistance but has not been experimentally demonstrated (Bridges *et al.*, 2007).

1.9.2.2 Other resistance mechanisms

Although most of the characterised laboratory-derived resistance has been associated with alterations in transporter related uptake, there is at least one report of separate pentamidine and melarsoprol resistance not due to a reduction in drug uptake. Berger *et al* (1995) selected a pentamidine resistance mutant with a fully functional P2 but with a five-fold increase in *in vivo* drug resistance. It did induce two-fold increased resistance to melarsoprol but was not found to be cross-resistant to other diamidines and the mechanism of resistance has not yet been determined.

Drug export is also a proposed method of drug resistance with the ABC transporter superfamily as prime candidates. These large membrane proteins have ATP binding cassettes that allow active drug efflux against a concentration gradient. Further to the role of thiol groups as a potential target for melarsoprol, over expression of a metal thiol conjugate transporter *TbMRPA*, a member of the multidrug resistance associated family, has been linked with a non-transport

mediated melarsoprol resistant line in *T.b.brucei* (Shahi *et al.*, 2002). Over expression of *TbMRPA* was found to lead to a 100-fold increase in mRNA levels resulting in a 10-fold increase in the minimum inhibitory concentration of melarsoprol with no cross-resistance to the diamidines. At present it is unclear how relevant this mechanism for resistance may be to selected drug resistance in the laboratory or treatment failure in the field.

Although recent advances have been made in our understanding of the role played by the P2 and HAPT1 transporters in drug resistance to the melaminophenyl arsenicals and diamidine classes of drug in *T.brucei*, our knowledge is far from complete and we still lack a full understanding of the mode of action of these drugs and the key determinants of resistance. The recent publication of the genome sequence for *T.brucei* (Berriman *et al.*, 2005) and on-going sequencing of *T.b.gambiense* raises hopes for a new era of discovery both in understanding the mechanism of action and drug resistance mechanisms of current drugs as well as revealing novel drug targets for the development of new chemotherapeutic approaches for targeting the parasites.

1.10 Trypanosome genetics

1.10.1 The structure of the *T.brucei* genome

As a member of the order Kinetoplastida, *T.brucei* possesses a two-unit genome consisting of a nuclear genome plus the kinetoplast, a mitochondrial genome contained within the single mitochondrion. The kDNA is composed of a network of topologically interlinked mini and maxi circles (reviewed in Shlomai, 2004; Liu *et al.*, 2005) and can account for as much as 20% of the total cellular DNA content of the trypanosome (Simpson, 1972). The nuclear genome of *T.brucei* on the other hand is linear and has undergone intensive scrutiny over the last 25 years in order to determine chromosomal organisation, karyotype and composition. The chromosomes of *T.brucei* do not condense during mitosis, but the nuclear karyotype has been observed by separating chromosomes using pulsed field gel electrophoresis (PFGE) (Van der Ploeg *et al.*, 1984; Gottesdiener *et al.*, 1990; Melville *et al.*, 1998). Unusually the nuclear genome consists of three classes of chromosomes; the mini, intermediate and megabase chromosomes, which are categorized by size based on their migration in an electric field (Van der Ploeg, 1984).

The mini and intermediate chromosomes are of uncertain ploidy and both their size and number varies between strains. The mini chromosomes (50-150 Kb), number around one hundred and are transcriptionally silent (Van der Ploeg, 1984). The chromosomes are organised around a palindromic core of 177 bp repeats (Wickstead *et al.*, 2004) and serve to expand the repertoire of VSG repeats involved in antigenic variation (Williams *et al.*, 1982). The intermediate sized chromosomes are larger (200-900 Kb) but less numerous (1-7 copies) than the mini chromosomes (Van der Ploeg, 1984) and share similar repetitive elements at their core (Wickstead *et al.*, 2004). The intermediate chromosomes also contribute to immune evasion by providing bloodstream expression sites for the *VSG* genes which are involved in antigenic variation (Van der Ploeg, 1984).

The 11 megabase chromosomes sized 1-6 Mb (Van der Ploeg, 1984; Gottesdiener *et al.*, 1990; Melville *et al.*, 1998) are diploid and encode all the housekeeping genes in addition to VSGs, VSG expression sites and other related proteins associated with the evasion of the immune response (Johnson and Borst, 1986; Melville *et al.*, 1998; Berriman *et al.*, 2005). Analysis by PFGE has demonstrated this chromosome class to be extremely size polymorphic both between and within strains, with homologous chromosomes varying in size by as much as two-fold within strains (Gottesdiener *et al.*, 1990) and four-fold between strains (Melville *et al.*, 1998; Melville *et al.*, 2000). This size plasticity exceeds that reported in many other size variable protozoan genomes (Blaineau *et al.*, 1991; Janse, 1993; Lanzer *et al.*, 1993; Lanzer *et al.*, 1994; Henriksson *et al.*, 1995; Ravel *et al.*, 1996; Wincker *et al.*, 1996) and can result in regions of hemizygosity between pairs of differentially sized paired homologues. The cause of this variation has been attributed to copy number polymorphisms affecting all regions of the chromosomes, but particularly those associated with antigenic variation (Melville *et al.*, 1999; Callejas *et al.*, 2006).

1.10.2 The genome sequence

The publication of the *T.brucei* genome sequence in 2005 (Berriman *et al.*, 2005), alongside a comparative analysis of the two related kinetoplastids, *T.cruzi* and *L.major* ('the tritryps'), marked a significant advance in our understanding of the Trypanosomatid parasites (El-Sayed *et al.*, 2005). The *T.brucei* sequencing project was the culmination of a decade long, multi-centre collaboration to sequence the 11 chromosomes of the megabase nuclear genome. A *T.b.brucei* isolate was the preferred subspecies, for laboratory handling purposes, and the strain selected for

sequencing was TREU 927/4 antigen type (GUTat) 10.1. This strain expresses a single VAT derivative while remaining pleomorphic in mammalian hosts, readily fly transmissible and amenable to genetic manipulation (van Deursen *et al.*, 2001).

Chromosomes 1, 9, 10 and 11 were sequenced by the Wellcome Trust Sanger Institute (WTSI) using a whole chromosome shotgun strategy. In this approach, small insert libraries were prepared and sequenced for each chromosome from a mixture of both PFGE-separated homologues. Sequence data was therefore derived from both homologues at all positions and the resulting assembly is a haploid consensus of the two. An alternative "bacterial artificial chromosome (BAC) walking" strategy was used by The Institute for Genome Research (TIGR) for sequencing chromosomes 2-8. In this approach a whole genome library of end-sequenced BAC clones was used to map a seed clone to each chromosome. From this seed clone a minimal tile path of overlapping BACs was "walked" to either end of the chromosome. Each large BAC was then fully sequenced by creating individual small insert libraries in a similar strategy to the whole chromosome shotgun approach. The resulting haploid sequence is composed of BAC-sized tiles, each derived from one or other of the homologues. This difference in sequencing strategy used for the two groups of chromosomes has implications for detecting sequence polymorphisms in the megabase genome that are discussed in further detail in section 1.13.6. The smaller chromosome classes of the nuclear genome were not included in this analysis. However separate mini and intermediate chromosome sequence projects are ongoing using a whole chromosome shotgun approach.

The resulting sequence data reveals a total haploid genome size for the megabase chromosomes of 26 Mb. This is composed of approximately 900 pseudogenes and 9068 protein coding genes of which 1700 are specific to *T.brucei* (Berriman *et al.*, 2005). The *T.brucei* genome shows a conservation of synteny and genome organisation with the other tritryps, with the same 6200 core genes present in all three pathogens, despite key differences in lifecycle, disease and pathology (Ghedin *et al.*, 2004; El-Sayed *et al.*, 2005).

These genes exist in large syntenic clusters, termed directional gene clusters (DGC), which are non-overlapping and directionally transcribed (El-Sayed *et al.*, 2003; Hall *et al.*, 2003). The arrangement of the DGCs is such that those containing conserved housekeeping genes are present towards the chromosome core and those containing species-specific genes are present in the subtelomeric

regions or at the break points between syntenic blocks (El-Sayed *et al.*, 2005). In contrast to higher eukaryotes, transcription in trypanosomes occurs polycistronically with post-transcriptional modification in the form of trans-splicing and polyadenylation to generate mRNAs for translation (reviewed in Ivens *et al.*, 2005). This organisation of genes into polycistronic units and reliance on post-transcription modification may be responsible for shaping the evolution of the tritryp genomes through gene family expansion, divergence and loss, particularly in the subtelomeric regions associated with immune evasion (Ghedin *et al.*, 2004; El-Sayed *et al.*, 2005).

Approximately 20% of the genome is given over to coding subtelomeric genes, in the region between the terminal array of telomeric repeats and first housekeeping gene (Berriman *et al.*, 2005). The majority of this region is species specific and relates to antigenic variation either in the VSG associated proteins present towards the chromosome core, arrays of *VSG* genes and pseudogenes or VSG expression sites positioned towards the telomeres (Barry *et al.*, 2005; Horn and Barry, 2005; Hertz-Fowler *et al.*, 2008). Because of the repetitive nature of sequence in this region, the subtelomeric regions have been particularly difficult to assemble and at present the genome sequence is incomplete. The current GeneDB (release four) (GeneDB, 2005) covers the "core" genes and extend into the subtelomere for each chromosome, but only two of the megabase chromosomes have been fully assembled from telomere to telomere (chromosome 1 and chromosome 10) (Hall *et al.*, 2003; Hertz-Fowler *et al.*, 2007). Projects to clone and sequence the subtelomeric regions of the remaining chromosomes are on-going so that a complete picture of the genome including the genes and expression sites associated with antigenic variation can be made accessible.

Also on-going are separate genome projects to sequence another *T.brucei* subspecies, *T.b.gambiense*, as well as the animal trypanosome strains *T.vivax* and *T.congolense*. The *T.brucei* genome strain TREU 927/4 strain is a *T.b.brucei*, isolated from an area in Kenya in which human African trypanosomiasis has not been reported. Due to its reduced pathogenicity, this is the preferred subspecies for laboratory handling and as *T.b.rhodesiense* is closely related to *T.b.brucei* the genome sequence of *T.b.brucei* will provide information on both genomes. However it is the *T.b.gambiense* subspecies which is responsible for the majority of the current HAT infections in sub-Saharan Africa (WHO, 2006a). Although *T.b.gambiense* is related to *T.b.brucei*, it is genetically distinct and differs profoundly in many phenotypic and genotypic characteristics such as human

infectivity. In order to study the mechanisms of disease in pathogenic human infective strains a separate *T.b.gambiense* sequence is desirable. The sequencing of a West African *T.b.gambiense* strain, DAL972/1, which was originally isolated from a human patient in Cote d'Ivoire in 1986 is currently ongoing at WTSI. A 5-fold partial shotgun sequencing approach has been used, for which preliminary sequence reads and partially assembled contigs are now available from the GeneDB resource (GeneDB, 2005).

Strains of the animal trypanosomes *T.vivax* and *T.congolense* are also currently being sequenced at WTSI using a partial shotgun sequencing strategy to 5-fold and 1-fold coverage respectively. Using the completed *T.b.brucei* sequence as a framework, it is hoped that once completed, comparison of the four trypanosome genomes may also help to uncover the genetic basis of the differences between them and provide information on where genetic material has been lost or retained.

1.11 Genetic mapping

The recent published draft of the *T.brucei* genome sequence identified 9068 predicted protein-coding genes of which approximately 50% have not yet been characterised or do not show sufficient homology to any known genes to have been assigned a putative function. One of the ongoing challenges for the research community is therefore how to utilise this vast amount of new sequence data to elucidate gene function and identify those genes involved in key biological processes.

There is an array of both traditional and new emerging molecular tools that can be utilised in the analysis of *T.brucei* genetics. These include microarray and transcriptomics techniques to dissect gene expression, reverse genetics tools such as RNAi and transfection based gene knock-in and knock-out approaches to examine the effect of altered gene expression on phenotype and traditional genetic mapping strategies to identify loci associated with a particular phenotype.

This last approach comprises generating a large number of progeny from a genetic cross of two different parental strains. Polymorphic molecular markers that can be scored at numerous loci are used to establish a unique pattern of heritable alleles for each individual progeny clone. This pattern of inheritance can then be used to establish the position and relationship of marker alleles to each other to construct

a genetic linkage map. The use of a genetic map can help identify genetic loci responsible for a trait of interest through a simple analysis of the co-segregation of specific marker alleles with a defined phenotype in each progeny clone. In contrast to biochemical approaches, genetic mapping thus requires no prior knowledge of the biochemical processes that gave rise to the phenotype.

The principles on which genetic linkage mapping are based derive from the segregation of parental alleles during meiosis via chromosome recombination and independent assortment and their subsequent distribution in the progeny. The nuclear genome of a eukaryotic organism is composed of linear chromosomes. During meiosis, these chromosomes assort independently in the progeny such that the segregation of genes on one chromosome will be independent of that of any other chromosome. Similarly genes that are present on the same chromosome are physically linked and thus will remain together and not segregate independently during meiosis. However this linkage between genes on the same chromosome can be disrupted by recombination or "crossing over" of homologous chromosomes which occurs during the early stages of meiosis. During this process the homologous chromosome synapse and reciprocal strand exchange occur at specialist sites along the length of the chromosomes. The resolution of these crossovers results in new recombinant chromosomes segregating in the progeny, which differ from either of the original parental chromosomes.

The likelihood of a crossover occurring between any two loci on a chromosome is related to their location. Genes that are physically closer together on a chromosome are less likely to be disrupted by recombination and thus will be inherited together more frequently than genes located far apart. The relationship between genes can therefore be established by comparing how often they are inherited together in the progeny. In the context of generating a genetic map this is achieved through the use of polymorphic molecular markers, evenly distributed through the genome. Markers are first identified in the parental strains of a cross followed by genotype analysis of the mapping population to determine the inheritance of the parental alleles in the progeny.

A genetic linkage map establishing the relationship and genetic distance between molecular markers can then be determined, even if the genomic locations of the markers are unknown. The level of recombination that has occurred between markers during crossing over and meiosis is determined by analyzing the allelic combination of all progeny and computing the frequencies of every marker allele

with every other marker allele in all combinations. Alleles of markers that are on the same chromosome and close together will be inherited together more often. If enough of these markers are available, detailed linkage analysis can be performed to identify which markers are together in linkage groups or chromosomes and precisely how close together these markers are. This information allows for the construction of detailed linkage maps of the genome.

The genetic distance between markers is a reflection of the recombination frequency and is measured in centiMorgans (cM). One centiMorgan is the approximate equivalent of a 1% recombination frequency. For example if between two markers, 5 out of 100 progeny have undergone a recombination event (5% or a recombination frequency of 0.05), these markers would be approximately 5 cM apart. When the recombination frequency between markers is small < 0.1 or 10%, the conversion to cM is constant. However for larger recombination frequencies it is not so and a mapping function is required for an accurate translation. The two main mapping functions are Haldane (Haldane, 1919) and Kosambi (Kosambi, 1944), which differ in respect to their consideration of interference. Interference is the effect a crossover in one region of a chromosome has in reducing or physically interfering with the probability of another crossover occurring nearby. The Haldane mapping function assumes an absence of interference between crossovers while the Kosambi mapping function adjusts for its possibility. Kosambi map distances are therefore usually slightly smaller than Haldane mapping distances.

Once generated such maps can then be utilized in comparative strain analysis, to gain insight into the genetics of sexual recombination and as an important first step to localizing genes and quantitative trait loci (QTLs) associated with phenotypes of interest. In order to use the classical genetic approaches such as linkage analysis however, the demonstration of an ability to undergo normal genetic exchange within the confines of standard Mendelian genetics is a crucial requirement.

1.12 Genetic exchange in *T.brucei*

For many years the trypanosomes were believed to be completely asexual, replicating only by binary fission throughout their lifecycle. However a capacity for genetic exchange has now been demonstrated in all three of the “tritryps” from laboratory genetic crosses in the tsetse fly (Jenni *et al.*, 1986), sand fly

(Akopyants *et al.*, 2009) and mammalian cell culture (Gaunt *et al.*, 2003) for *T.brucei*, *L.major* and *T.cruzi* respectively. The frequency with which sexual recombination might occur in natural populations of all three pathogens however remains controversial.

Evidence that sexual exchange may occur in *T.brucei* first originated from the observation of isoenzyme variation and hybrid-like banding patterns in different trypanosome strains and populations in the field (Gibson *et al.*, 1978; Gibson *et al.*, 1980; Tait, 1980). Then, in 1986, Jenni *et al* demonstrated the first laboratory genetic cross of *T.brucei* by the co-transmission of two genetically different *T.brucei* strains through tsetse flies (Jenni *et al.*, 1986). The products of the cross were cloned in mice and characterised by isoenzyme and restriction fragment length polymorphism (RFLP) analysis, which identified them as hybrids of the parental lines.

Since then, additional genetic crosses have been conducted both within and between strains belonging to all three *T.brucei* subspecies. With regards to *T.b.gambiense*, crosses have so far only been possible with one of the two clades of *T.b.gambiense* parasites. Group two parasites, characterised by virulence in rodents and a variable expression of human serum resistance, are more similar to *T.b.rhodesiense* and *T.b.brucei*, and genetic crosses have been demonstrated between isolates of this group and the other subspecies (Sternberg *et al.*, 1988; Sternberg *et al.*, 1989; Turner *et al.*, 1990; Gibson *et al.*, 1997). *T.b.gambiense* group 1 strains form the larger group and are characterised by slow growth in rodents and stable expression of human serum resistance. So far, crosses using strains from this group have not been reported because they do not transmit easily through the usual laboratory tsetse fly vector, *Glossina morsitans*.

Generating *T.brucei* crosses can be difficult and technically demanding. This is because genetic exchange in the tsetse fly is non–obligatory and therefore not all co-infected flies will result in the production of progeny clones. This is compounded by the fact that the products of a cross have traditionally needed to be isolated and cloned in mice before analysis. This is a time consuming and labour intensive process that requires large numbers of animals for clonal expansion. As a result, until recently only small numbers of progeny clones have been isolated from any cross, with the consequence of severely limiting any ability to perform statistically significant genetic analysis. In spite of these difficulties, to date ~11 genetic crosses have now been reported (reviewed in Gibson and

Stevens, 1999; MacLeod, 2007) and the products analysed, which has allowed considerable progress to be made in understanding genetic exchange in *T.brucei*.

Because no chromosomal condensation has been observed in any life cycle stage, analysis of the genetic exchange process has focused on two molecular techniques; *in situ* analysis of crosses between fluorescent trypanosomes to determine where genetic exchange occurs in the tsetse fly, and analysis of the products of crosses with molecular markers to determine the inheritance of parental alleles in the progeny.

When a tsetse fly is co-infected with two different strains of *T.brucei*, it has been observed that hybrid forms could be readily isolated from the salivary glands but not the midgut suggesting that the salivary glands were the likely site of genetic exchange (Gibson and Whittington, 1993). This was recently experimentally validated in the laboratory of Wendy Gibson through the development of parental lines transfected with either red fluorescent protein (RFP) or green fluorescent protein (GFP) (Bingle *et al.*, 2001). When crosses were performed between these fluorescent lines, the segregation of parental alleles in the hybrids resulted in abundant yellow progeny clones (the combined product of RFP and GFP genes) in addition to red and green trypanosomes, that could be visualised by fluorescence microscopy of dissected salivary glands (Gibson *et al.*, 2006). No yellow hybrid progeny were visualised in the midgut or migratory stages between the proventriculus and salivary gland confirming the salivary glands as the only location in the tsetse fly in which genetic exchange occurs.

The evidence from fluorescent crosses is supported by molecular marker analysis of clones isolated from the midgut and salivary glands in which genotypes consistent with hybrid clones have only been identified in the salivary gland stage of the life cycle (Gibson and Whittington, 1993; Gibson and Bailey, 1994; Gibson *et al.*, 1997). Detailed examinations of progeny clones isolated from the original cross performed by Jenni *et al* (1986) and other early crosses by PFGE, RFLP and isoenzyme analysis, has demonstrated that during a cross three types of progeny are produced; those expressing a parental genotypes consistent with transmission without mating, hybrid clone genotypes of the two parents lines and occasionally the products of a self-fertilisation event (Gibson, 1989; Sternberg *et al.*, 1989; Turner *et al.*, 1990; Tait *et al.*, 1996). This confirmed that genetic exchange in tsetse flies, in addition to being a non-obligatory process, can involve self as well as cross-fertilisation.

Analysis of the DNA content of hybrids from different crosses has shown that some of the progeny clones had elevated DNA levels compared to the parents consistent with the progeny being triploid or trisomic (Gibson *et al.*, 1992; Gibson and Bailey, 1994; Gibson and Stevens, 1999). This suggested that in some cases non-disjunction events or the fusion of haploid and diploid gametes is occurring, and led to a variety of models being proposed to explain genetic exchange. These theories range from conventional Mendelian to a random fusion and loss hypothesis in which the fusion of two diploid genomes is followed by random chromosome loss to restore diploidy (Paindavoine *et al.*, 1986a). The numbers of triploid hybrids varies between the crosses, but those which produce a significant proportion of triploid progeny tend to originate in crosses between subspecies. In crosses of the same subspecies, the majority of progeny isolated from crosses has so far been diploid (Turner *et al.*, 1990; Hope *et al.*, 1999; MacLeod *et al.*, 2005a). In addition triploidy has not been reported in the field (MacLeod *et al.*, 2000) suggesting that this process might not occur naturally or might be more frequent in laboratory crosses between strains that are not fully genetically compatible.

The recent publication of the *T.brucei* genome opened up the opportunity for a much more detailed analysis of the products of genetic exchange. Microsatellite markers for each chromosome were identified from the genome sequence and screened against the parental lines to identify those which were heterozygous for two crosses in the laboratory carried out between *T.b.brucei* and *T.b.gambiense* strains (MacLeod *et al.*, 2005a). A large number of progeny (79 clones) have been isolated from these crosses (Sternberg *et al.*, 1989; Turner *et al.*, 1990; MacLeod *et al.*, 2005a) which allowed genetic analysis of ratios of marker inheritance to be performed. One informative (heterozygous) marker was identified on every megabase chromosomes and the inheritance of parental alleles examined in the 79 progeny clones. The result of this analysis formally demonstrated that independent assortment and segregation consistent with Mendelian inheritance and meiosis occurs in *T.brucei* (MacLeod *et al.*, 2005a).

Genetic exchange between *T.brucei* strains has now been unequivocally demonstrated in the laboratory but the exact mechanism by which it takes place has yet to be formally demonstrated. Several models to explain the genetic exchange mechanism in *T.brucei* have been proposed over the years, including a fusion/loss model, a traditional haploid gamete system and a fusion with meiosis hypothesis.

The fusion-loss model proposed by Paindavoine *et al* (1986a) suggests that the two diploid nuclei fuse to form a tetraploid cell which then progressively loses DNA to regenerate a diploid cell with non-Mendelian segregation of parental alleles. A recent experimental laboratory cross of *T.cruzi* has demonstrated the ability of this parasite to undergo genetic exchange with evidence for a similar chromosome fusion and loss mechanism (Gaunt *et al.*, 2003). In *T.brucei* however the proven Mendelian inheritance of markers does not support this model, leaving two other possibilities that are supported by the available evidence.

In the fusion model (Gibson *et al.*, 1995) the diploid parental cells fuse to form an intermediate polyploid cell, with intact separate nuclei which then undergo meiosis to yield eight haploid nuclei. Fusion of two of these with disintegration of the remaining nuclei results in diploid progeny. As the fused nuclei undergo meiosis separately in this model, Mendelian inheritance would be retained and the resulting progeny could be the result of either self or cross-fertilisation.

The alternative standard model proposed by Tait and Sternberg (1990) involved a traditional mechanism in which meiosis takes place to generate haploid gametes followed by fusion of these haploid gametes from two different strains to create a new diploid progeny. At present no data can formally distinguish between the two models however as neither the predicted polyploid or haploid intermediate/gamete forms have been observed.

Genetic exchange in a third trypanosomatid, *Leishmania major*, appears more likely to resemble the mechanism of *T.brucei* than *T.cruzi*. Recently experiments demonstrated that genetic exchange can occur with a laboratory cross of two *L.major* strains in the sand fly vector (Akopyants *et al.*, 2009). As with the other Trypanosomatids no gamete forms have been identified but the inheritance patterns of the nuclear genome were compatible with a Mendelian system.

The formal demonstration of genetic exchange in *T.brucei* in combination with an abundance of data emerging from the genome sequence project opens up opportunities to use genetic linkage analysis in these parasites.

1.13 Identifying molecular markers in *T.brucei*

A huge variety of molecular markers have become available for the study of eukaryotic genetics, and more particularly, genetic linkage analysis, since the first

linkage map, of *Drosophila melanogaster*, was published nearly a century ago. These include isoenzymes, RFLP, Amplified Fragment Length Polymorphisms (AFLP), Random Amplified Polymorphic DNA (RAPD), microsatellites and single nucleotide polymorphisms (SNPs). These markers differ from each other in many respects (Table 1.1), including the type of polymorphism and their distribution in the genome, dominance as well as the ease, cost and utility of the genotyping technologies used in their detection. Over the last 30 years, many of these marker types have been utilised for different genetic analyses of *T.brucei.*

1.13.1 Isoenzymes

Isoenzymes were among the earliest markers to be applied to the study of trypanosome genetics. They were first used in the mid 1970s in epidemiology surveys and for the taxonomic characterisation of trypanosome stocks (Gibson *et al.*, 1978; Kilgour, 1980; Godfrey *et al.*, 1987), but have also proved useful in the analysis of the products of some of the early genetic crosses (Jenni *et al.*, 1986; Sternberg *et al.*, 1988; Sternberg *et al.*, 1989). This approach involves the electrophoresis separation of enzyme containing cell extracts in starch or polyacrylamide gels. The relative motility of each isoenzyme through the gel is dependent on the size, shape and charge of the protein and therefore in turn its amino acid composition. Polymorphisms can be observed between allozymes if sequence variation has resulted in any amino acid changes that affect these variables and ultimately result in a detectable alteration in relative mobility. Isoenzymes however are limited in number, life cycle variable and can only be detected when sequence variation results in a significant alteration to the protein mobility. As a result isoenzymes have been progressively replaced by a series of increasingly powerful DNA based molecular techniques that are capable of detecting finer levels of genetic variation with smaller amount of sample material.

1.13.2 Restriction fragment length polymorphism (RFLP)

RFLP markers were introduced in the mid-1970s and detect DNA polymorphisms between and within isolates by the differential digestion of double stranded DNA with restriction endonucleases followed by Southern blot analysis. Polymorphism in the RFLP assay originates from differences in genomic DNA sequence causing the loss, gain or relocation of restriction enzyme recognition sites. The RFLP technique is reproducible, easy to score, locus specific and does not require any knowledge of the target organism sequence, providing a probe library is available.

	Isoenzymes	RFLPs	AFLPs	RAPDs	Microsatellites	SNPs
Method of analysis	Protein gel electrophoresis	Restriction digest, Southern blotting	Restriction digest, Ligation of adapters Adapter specific PCR	PCR with arbitrary primers	PCR with locus specific primers	Direct sequencing, variety of methods based on allele distinction including hybridization, restriction digest, oligonucleotide ligation
Type of polymorphism	Non-synonymous substitutions, insertions, deletions	Single base substitutions, insertions, deletions	Single base substitutions, insertions, deletions	Single base substitutions, insertions, deletions	Tandem repeat number variations	Single base substitutions
Dominance	Co-dominant	Co-dominant	Dominant	Dominant	Co-dominant	Co-dominant
Abundance	Low	High	High	High	Medium	V.high
Number of loci analysed	Single	Single	Multiple	Multiple	Single	Single
DNA sequence information required	No	No	No	No	Yes	Yes
Amount of DNA required	n/a	High	Medium	Low	Low	Low
Reproducibility	High	High	High	Intermediate/low	High	High
Major applications	Linkage mapping Population studies	Linkage mapping	Linkage mapping Fingerprinting for population studies	Fingerprinting for population studies Linkage mapping	Linkage mapping Population studies	Linkage mapping Association studies

Table 1.1 A comparison of the main characteristics of some commonly used molecular markers.

As a result, RFLP assay, have been widely and extensively used in measuring genetic diversity and constructing genetic maps including the first human genome map in the 1980s (Donis-Keller *et al.*, 1987) as well as the early genetic maps for *Toxoplasma gondii* (Sibley *et al.*, 1992), *Plasmodium falciparum* (Walker-Jonah *et al.*, 1992) and *Plasmodium chabaudi chabaudi* (Grech *et al.*, 2002). However the Southern blotting required by this type of assay does have its drawbacks. It requires a large amount of high quality DNA and can be time consuming, expensive and laborious. For these reasons the use of RFLPs in Trypanosomes have been mainly restricted to genotyping the products of genetic crosses (Jenni *et al.*, 1986; Sternberg *et al.*, 1989; Turner *et al.*, 1990), and where sufficient genetic material allows, population genetics (Paindavoine *et al.*, 1986b; Paindavoine *et al.*, 1989). In consequence there has been increasing interest in PCR based approaches, which can amplify any sequence of interest from nanogram amounts of DNA, often without the need for DNA extraction, as well as allowing the direct visualization of the amplified PCR product.

1.13.3 Random Amplified Polymorphic DNA (RAPD)

The RAPD marker system (Williams *et al.*, 1990) utilised a derivative of PCR, termed arbitrarily primed PCR, which unlike traditional PCR analysis does not require any specific knowledge of the DNA sequence of the target organism. This is achieved through the use of a single, short arbitrary oligonucleotide primer. Successful DNA amplification with this primer is dependant on the presence of primer binding sites on opposite strands of the template, within an amplifiable distance of each other. Polymorphisms result from variations in the DNA sequence, that, under stringent PCR conditions, allow the primer to bind to the template DNA at one locus but inhibit primer binding at another. This selective amplification can then be visualised as the presence or absence of a band of a particular molecular weight in the fragment fingerprint following gel electrophoresis.

RAPD analysis has been applied to analysing parasite populations of *Trypanosoma* (Stevens and Tibayrenc, 1995; Lun *et al.*, 2004; Li *et al.*, 2005), *Leishmania* (Martinez *et al.*, 2003), and *T.gondii* (Guo and Johnson, 1995) in the absence of genomic sequence data and contributed genetic markers for construction of the linkage map of *Eimeria tenella* (Shirley and Harvey, 2000). RAPD markers are however considered to have greater problems of reproducibility than other systems. A further limitation in that nearly all RAPD markers are dominant and

thus in diploid organism it is not possible to distinguish between amplification from a dominant homozygous locus or heterozygous locus individuals, which limits their suitability for some genetic mapping approaches.

1.13.4 Amplified Fragment Length Polymorphism (AFLP)

The amplified fragment length polymorphism (AFLP) method is a PCR based fingerprinting technique which combines the principles of restriction enzyme site polymorphisms utilized in the RFLP approach with a more efficient PCR based protocol (Vos *et al.*, 1995). Restriction fragments are first generated by digestion of genomic DNA with two different restriction enzymes, a rare six base cutter and a more frequent four base cutter. The frequent cutter ensures the generation of fragments of an appropriate size for electrophoresis while the rare cutter provides specificity for detection. Adapters are then ligated to either end of the digested fragments to supply the primer binding sites for PCR amplification. Selectivity is provided by labelling only the primer which recognizes the rare base cutter adapter site for downstream detection and by varying the bases of a 3' nucleotide extension of 1-3 bp on each primer. Polymorphisms in the DNA sequence which alter the restriction sites, primer binding sites or the size of digest fragment can be detected as differences in the AFLP fingerprint profile, visible after electrophoresis and autoradiography or automated detection.

This approach, as with the application of RAPD and RFLP markers does not require any prior knowledge of the genome sequence. It is also highly reproducible and able to amplify polymorphic markers from multiple independent restriction sites within the genome simultaneously. As a result AFLP markers have been widely employed for DNA fingerprinting and genetic mapping of a range of bacterial, fungal and protozoan genomes (Janssen *et al.*, 1996; Masiga and Turner, 2004; Meudt and Clarke, 2007), including the genetic maps of the apicomplexa parasites *E.tenella* (Shirley and Harvey, 2000) and *P.c.chabaudi* (Martinelli *et al.*, 2005a).

However AFLPs are not without their problems when used as molecular markers. Firstly the AFLP technique requires high quality purified DNA samples. Secondly with this fingerprint technique it can be difficult to distinguish between mixed and clonal populations. Lastly, and more particularly related to its use in genetic mapping analysis of diploid organisms, is the difficulty in distinguishing dominant homozygous from heterozygous loci. This a problem common to both the AFLP and RAPD marker techniques caused by the marker dominance. Polymorphisms are

detected between isolates as the presence or absence of a band of a particularly size in the fingerprint profile. This marker dominance means that any isolate that is heterozygous at a particular locus would produce an identical AFLP profile to an isolate that is homozygous for the presence of the band.

Because of this, heterozygosity in the parental stocks of a genetic cross can only be inferred by also screening a sufficient number of the resulting progeny clones. If a particular band is present in one parent and absent in another, all the progeny will inherit the "absence" allele from one parent and will then either all inherited the "presence" allele if the other parent in homozygous or if the parent in heterozygous, absence/presence will segregate approximately equally in the progeny. It is only these segregating markers that can be utilized for linkage analysis. Despite these drawbacks, the relative abundant of polymorphic loci in trypanosomes has established AFLP fingerprinting as a useful marker system over recent years for assessing the genetic diversity of trypanosome isolates and identifying markers for subspecies characterization (Agbo *et al.*, 2002; Masiga *et al.*, 2006; Simo *et al.*, 2008a).

1.13.5 Microsatellites

Microsatellites and the larger minisatellites are tandemly repetitive sequences composed of 2-6 bp (microsatellites) or 10-100 bp (minisatellites) DNA repeats. Microsatellites are the more abundant of the two and are widely dispersed, in both coding and non-coding sequences in most eukaryotic genomes. Microsatellites have a relatively high degree of polymorphism both within and between genomes caused by replication slippage and repair that can alter the number of microsatellite repeats at each locus. These polymorphisms can be detected by a simple PCR assay using primers that flank the tandem repeat region, and visualized as a difference in PCR product size by gel electrophoresis or automated screening with fluorescent primers.

Unlike the markers so far mentioned however, knowledge of the genome sequence is required for primer design. If the genome sequence data from the target organisms has been determined it is a theoretically straightforward process to identify microsatellite repeat regions in the DNA sequence and design primers to the conserved flanking regions. If sequence data is not available however a more laborious process is required in which a genomic library must first be created,

screened with simple repeat probes to identify potential microsatellites and then sequenced to allow primer design.

Microsatellites have several advantages that have made them a standard in current linkage map construction. In addition to being highly abundant, easy to score and reproducible, the PCR assay requires only a small amount of DNA, an important feature for studies in which only a limited amount of sample is available. Microsatellites are also locus specific so can be interpreted genetically and importantly, unlike AFLP and RAPD markers, they are co-dominant. This has facilitated their use in genetic linkage mapping projects where they have been key to constructing genetic maps for the protozoan parasite *P.falciparum* (Su *et al.*, 1999) and the human blood fluke, *Schistosoma mansoni* (Criscione *et al.*, 2009).

A high degree of polymorphism combined with co-dominance allows microsatellites to detect heterozygosity as well as mixed infections, making them ideally suited for high resolution molecular fingerprinting and population genetics studies, in addition to genetic mapping. As a result these markers have been used extensively in genetic analyses in many organisms including *T.brucei* and other trypanosome species for population genetic studies (Biteau *et al.*, 2000; MacLeod *et al.*, 2000; MacLeod *et al.*, 2001b; Koffi *et al.*, 2007; Koffi *et al.*, 2009), detecting mixed infections (MacLeod *et al.*, 1999; Balmer and Caccone, 2008), and genotyping the products of genetic crosses (Hope *et al.*, 1999; MacLeod *et al.*, 2005a; Gibson *et al.*, 2006).

1.13.6 Single nucleotide polymorphisms (SNPs)

Single nucleotide polymorphisms (SNPs) are variations in the base present at a single DNA sequence position. These sequence differences, caused by point mutations, are present in both the coding and non-coding regions of DNA and are the most abundant form of genetic variation in the genome. Although in principle, four possible nucleotide variants can be present at each SNP locus (A, C, T or G), in practice mutation bias means that SNPs are usually bi-allelelic (either the two pyrimidines C/T or two purines G/A). The value of SNPs as molecular markers based on their polymorphism is therefore lower compared to the highly informative multi-allelic microsatellites markers. However their abundance has made them a desirable molecular marker for applications in which a very high density of molecular markers are required. In addition to their high abundance,

SNPs are also co-dominant, genetically stable, selectively neutral (in the non-coding regions), and amenable to automated detection.

SNP polymorphisms are the form of variation underlying many of the molecular markers previously described in this chapter included the RFLPs, AFLPs and RAPDs methods. However these technologies were developed in the absence of sequence information and are only able to detect these polymorphisms through indirect means and in relatively small numbers. In recent years the increasing availability of genome sequence data has made the direct analysis of genetic variation at the DNA sequence level possible. In addition the development of automated and high-throughput genotyping technologies is also making rapid large-scale analyses of markers resulting from SNP discovery a reality.

Of all the molecular markers, the technology for genotyping samples for SNPs is the most diverse (reviewed in Gut, 2001; Vignal *et al.*, 2002), both in allele distinction methodology and detection. Methods available for distinguishing SNPs include allele-specific hybridization, oligonucleotide ligation, invasive cleavage, primer extension, direct sequencing, and restriction site cleavage. Most of these genotyping methods can also be analyzed in a variety of formats, the choice of which will depend on a number of factors including scale, cost, and the number of samples versus the number of markers to be genotyped. For high throughput sample analysis a number of new technologies are available including next generation sequencing and high density microarray. Many of these large scale high-throughput SNP detection and genotyping platforms are dependent on expensive equipment however, which can make SNP discovery and genotyping a costly process.

SNPs markers are useful tools for the high resolution mapping of complex traits by QTL analysis, association studies and the study of recombination. Their potential has already been demonstrated in the recent publication of high density genetic maps for several model organisms including *D.melanogaster* (Hoskins *et al.*, 2001), *Caenorhabditis elegans* (Davis and Hammarlund, 2006) and *Candida albicans* (Forche *et al.*, 2004) as well as the human genome where several million SNPs markers have been identified at a frequency of one per Kb (Frazer *et al.*, 2007).

In *T.brucei*, the megabase genome of strain TREU 927 was jointly sequenced at TIGR and WTSI using different strategies, as discussed previously, and the first draft made available in 2005 (Berriman *et al.*, 2005). Sequencing of chromosomes 2-8 was performed at TIGR using a BAC clone by clone walking approach in which

one BAC, derived from one of the two homologues was sequenced at any one locus. The consequence of this approach is a haploid mosaic sequence for which sequence data at each nucleotide location is derived from only one homologue and thus polymorphisms cannot be detected. For the sequencing of chromosome 1, 9-11 performed by Sanger a different whole chromosome shotgun approach was used. Small-insert libraries were prepared and sequenced for each chromosome from a mixture of both homologues with the resulting assembly, a haploid consensus of the two. Information on polymorphisms in these chromosomes therefore has been retained. For the smallest chromosome, chromosome 1, for which the homologues could be partially separated by PFGE, the chromosomes were sequenced disproportionately allowing the specific assembly of one of the homologues. Polymorphisms detected in reads derived from the other homologue were then annotated in genome sequence release. In theory polymorphisms can also be identified between the homologue of the much larger chromosomes 9-11 from analysis of the sequence reads (although only chromosome 9 was partially separated by PFGE) and as a result of ongoing analysis SNP data for these chromosomes may be made available in future data releases.

1.14 Parasite Genetic maps

In the last decade, genetic maps have been constructed for a number of medically important pathogens including the human helminth *S.mansoni* (Criscione *et al.*, 2009) and the protozoa parasites *E.tenella* (Shirley and Harvey, 2000), *T.gondii* (Sibley *et al.*, 1992), *P.c.chabaudi* (Martinelli *et al.*, 2005a) and *P.falciparum* (Su *et al.*, 1999). This has provided a framework for genome projects and enabled localisation of genes and QTL loci underlying important phenotypic traits such as drug resistance (Su *et al.*, 1997; Ferdig *et al.*, 2004), virulence (Su *et al.*, 2002; Saeij *et al.*, 2006; Taylor *et al.*, 2006; Saeij *et al.*, 2007) and strain-specific immunity (Martinelli *et al.*, 2005b).

T.gondii is a widespread apicomplexan parasite of mammals and cause of opportunistic disease in man. Genetic analysis has revealed that the majority of strains fall into one of three clonal lineages. These clonal types exhibit strong phenotypic differences of interest, particularly in terms of their virulence in mice. A *T.gondii* linkage map was generated from several different genetic crosses performed between the different lineages through co-transmission in cat hosts. Initially this was constructed from the inheritance of 64 RFLP markers in 19 progeny by Southern blot and then improved following the release of expressed

sequence tag data and the isolation of further progeny clones (Sibley *et al.*, 1992). Data from the inheritance of 250 PCR-RFLP markers in 71 progeny from the three crosses were combined into a single analysis to define 14 linkage groups with an average recombination unit of 104 Kb/cM (Khan *et al.*, 2005). This relatively large map unit size equates to a low recombination frequency in *T.gondii* and thus identifies the number of progeny rather than the number of molecular markers as the limiting factor for linkage analysis. The identification of further progeny and a convergence with the independent screening of mappable traits by other collaborators led to the identification of several genes in the rhoptry family of secretory proteins associated with virulence (Su *et al.*, 2002; Taylor *et al.*, 2006) (Saeij *et al.*, 2006; Saeij *et al.*, 2007; Boyle *et al.*, 2008) and strain specific differences in host cell transcription between the lineages (Saeij *et al*., 2007; Boyle *et al*., 2008.).

Linkage analysis has also been performed with genetic maps of both the rodent malaria parasite *P.c.chabaudi* and the human malaria parasite *P.falciparum* to map loci contributing to traits such as drug resistance (Carlton *et al.*, 1998; Fidock *et al.*, 2000; Ferdig *et al.*, 2004), erythrocyte invasion (Hayton *et al.*, 2008), strain specific immunity (Martinelli *et al.*, 2005b), folate metabolism (Wang *et al.*, 1997) and sexual development (Vaidya *et al.*, 1995).

P.falciparum has a complex life cycle involving obligatory sexual stages in the midgut of female anopheles mosquitoes where genetic recombination occurs. A genetic cross can be achieved in the laboratory, however this requires the use of a chimpanzee host and can be laborious and expensive. As a result, only three *P.falciparum* laboratory crosses have as yet been performed (Walliker *et al.*, 1987; Wellems *et al.*, 1990; Hayton *et al.*, 2008). One of these crosses was between a chloroquine resistant clone (Dd2) and a chloroquine sensitive clone (HB3) (Wellems *et al.*, 1990; Ferdig and Su, 2000; Fidock *et al.*, 2000) for which 16 independent recombination progeny were generated. A preliminary genetic map generated from the inheritance data of 85 RFLP markers in these progeny localised the chloroquine resistance (CQR) phenotype to a single locus within a 500 Kb segment of chromosome 7 (Wellems *et al.*, 1991). This was further refined by constructing a higher resolution genetic map using a panel of 900 microsatellite markers derived from available genome sequences and an improved number of progeny clones. Fourteen linkage groups were inferred from this analysis corresponding to the 14 nuclear chromosomes of the *P.falciparum* genome (Su *et al.*, 1999).

Comparison of the physical and genetic maps indicated that the recombination frequency in *P.falciparum* was high (17 Kb/cM) and comparatively evenly distributed throughout the parasite genome. A relatively small number of progeny (35 independent clones) were thus able to provide adequate resolution to localised the chloroquine resistance determinant to a 36 Kb segment of *P.falciparum* chromosome 7 (Su *et al.*, 1997). Analysis of this 36 Kb region identified a number of candidate genes of which a previously unknown transporter, PfCRT (*P.falciparum* CQ resistance transporter), was eventually proven to play a key role in CQR in both the Dd2 CQR line and the field (Fidock *et al.*, 2000). Furthermore progeny from this same cross were also used to map a defect in male gamete development (Vaidya *et al.*, 1995), folate metabolism (Wang *et al.*, 1997) and employed to map a QTL contributing to quinine susceptibility (Ferdig *et al.*, 2004).

1.14.1 *T.brucei* genetic map

The development of forward genetics as a functional system in *T.brucei* is an important tool to gain understanding of the genetic system of *T.brucei* and opens up the possibility of identifying genes responsible for phenotypes of interest in these strains. Three *T.brucei* lines are available in the laboratory (Table 1.2) that have phenotypes for which this technique is ideally suited. Swiss Tropical Institute, Basal (STIB) 247 and TREU 927 are strains of the *T.b.brucei* subspecies while STIB 386 has been characterised as a group two *T.b.gambiense*. They were each isolated from different regions of Africa and different host species and differ in a range of important phenotypes such as human infectivity, host range and pathogenicity (Tait *et al.*, 2002). Genetic crosses have already been undertaken in all pair wise combinations between the three stocks of *T.brucei*, STIB 247, TREU 927 and STIB 386 (Turner *et al.*, 1990) allowing the genetic basis of these differences to be analysed.

Strains were co-transmitted through tsetse flies and the resulting metacyclic and bloodstream populations were then analysed for novel hybrid genotypes using a series of genetic markers that differ between the two parental stocks. The initial crosses and isolation of progeny were performed on a relatively small scale in the 1980s and early 1990s utilising PFGE, isoenzymes, and RFLP markers to identify hybrids and investigate the process of genetic exchange (Jenni *et al.*, 1986; Sternberg *et al.*, 1988; Sternberg *et al.*, 1989; Turner *et al.*, 1990). The identification of several polymorphic mini and microsatellites in the late 1990s

and two concerted cloning efforts in the late 1990s and early 2000s resulted in a total of ~ 40 independent progeny clones for each of the first two crosses (STIB 247 x TREU 927 and STIB 247 x STIB 386) (MacLeod *et al.*, 2005a). This provided enough progeny for a statistically significant chi-squared distribution analysis to be undertaken, which formally proved a conventional Mendelian inheritance (MacLeod *et al.*, 2005a), thus confirming that genetic mapping is a suitable approach for *T.brucei*.

Parental stock	Place of isolation	Host	Year of Isolation	Species
STIB 247	Tanzania	Hartebeest	1971	*T.b.brucei*
STIB 386	Ivory Coast	Man	1978	*T.b.gambiense,* group 2
TREU 927	Kenya	Tsetse fly	1969	*T.b.brucei*

Table 1.2 Information on the origins of the three parental *T.brucei* stocks available for investigation in the laboratory.

Following the availability of genome sequence contigs of *T.brucei* in the early 2000s, 810 microsatellite sequences were identified which consisted of more than 12 repeat units and were roughly equally distributed along each of the 11 chromosomes. Because genetic mapping is based on the segregation of alleles, it requires at least one of the parental genotypes to be heterozygous for each molecular marker used, so all three parent lines were screened for heterozygosity for a number of markers by PCR amplification. TREU 927 was found to be the most heterozygous with ~20% (182) markers, STIB 247 the least with only 1% and STIB 386 somewhere between the two. The implication of this difference in heterozygosity and thus the number of markers available for analysis, is that it will be significantly easier to create a genetic map for the *T.b.brucei* sequence strain (TREU 927) and *T.b.gambiense* isolate (STIB 386) than for the more homozygous STIB 247 strain.

A genetic map was generated for the *T.b.brucei* strain TREU 927 (MacLeod *et al.*, 2005b) by genotyping all 39 progeny from the TREU 927 x STIB 247 cross with 182 microsatellite markers and analysing allele inheritance in Map Manager (Manly *et al.*, 2001), a genetic linkage map program (Figure 1.3). Sixteen linkage groups were identified with the 11 main groups corresponding to the 11 megabase

chromosomes but with chromosomes 6, 9, 10 and 11 having an additional one or two separate unlinked smaller linkage groups. The genetic map spans 1157.5 cM in total, with 122 map intervals and an average physical distance between markers of ~150 Kb. This marker coverage means there is a 90% probability that at least one marker is within 11 cM of any given gene. The recombination frequency was found to be high, with the average physical size of each recombination unit of 15.6 Kb/cM. Thus *T.b.brucei* and *P.falciparum* have recombination frequencies that are more similar than *T.gondii*, which helps to mitigate the problems associated with small sample sizes, and provide adequate resolution for genetic mapping. In contrast to *P.falciparum* however there is considerable variation in size of the recombination unit (cM) within and between chromosomes with both hot and cold spots of recombination present.

This completed genetic map allows both single marker and QTL analysis of any segregating phenotype in the TREU 927 cross by the linking of hybrid phenotype and genotype. Based on haplotype analysis, it is possible to identify the genomic location of candidate open reading frames. If the region between the flanking markers is relatively large (> 100 Kb) and crossovers occur between the flanking markers, then fine mapping within this defined region can be undertaken in order to narrow the region further and so reduce the number of candidate genes involved in determining the trait of interest. At this point a reverse genetics approach can potentially be undertaken, involving allelic disruption of each candidate gene in turn or transfection of candidate genes into sensitive strains followed by determining any changes in phenotype compared to the parental stock and identifying genes of interest in the sequence strain *T.b.brucei* TREU 927.

As previously mentioned, it is important to note that *T.b.gambiense* is responsible for the majority of the current HAT infections in sub-Saharan Africa and although *T.b.gambiense* is related to *T.b.brucei*, it is genetically distinct and differs profoundly in many phenotypic and genotypic characteristics. In order to study the mechanisms of disease in pathogenic human-infective strains a separate *T.b.gambiense* genetic map is desirable to map specific phenotypes found in this subspecies, such as their ability to resist lysis in human serum and differences in pathogenicity. As *T.b.gambiense* is currently being sequenced, a genetic map for this subspecies will also provide an additional resource for analysis and can be used for linkage mapping segregating phenotypes in the STIB 386 x STIB 247 cross we have available in the laboratory.

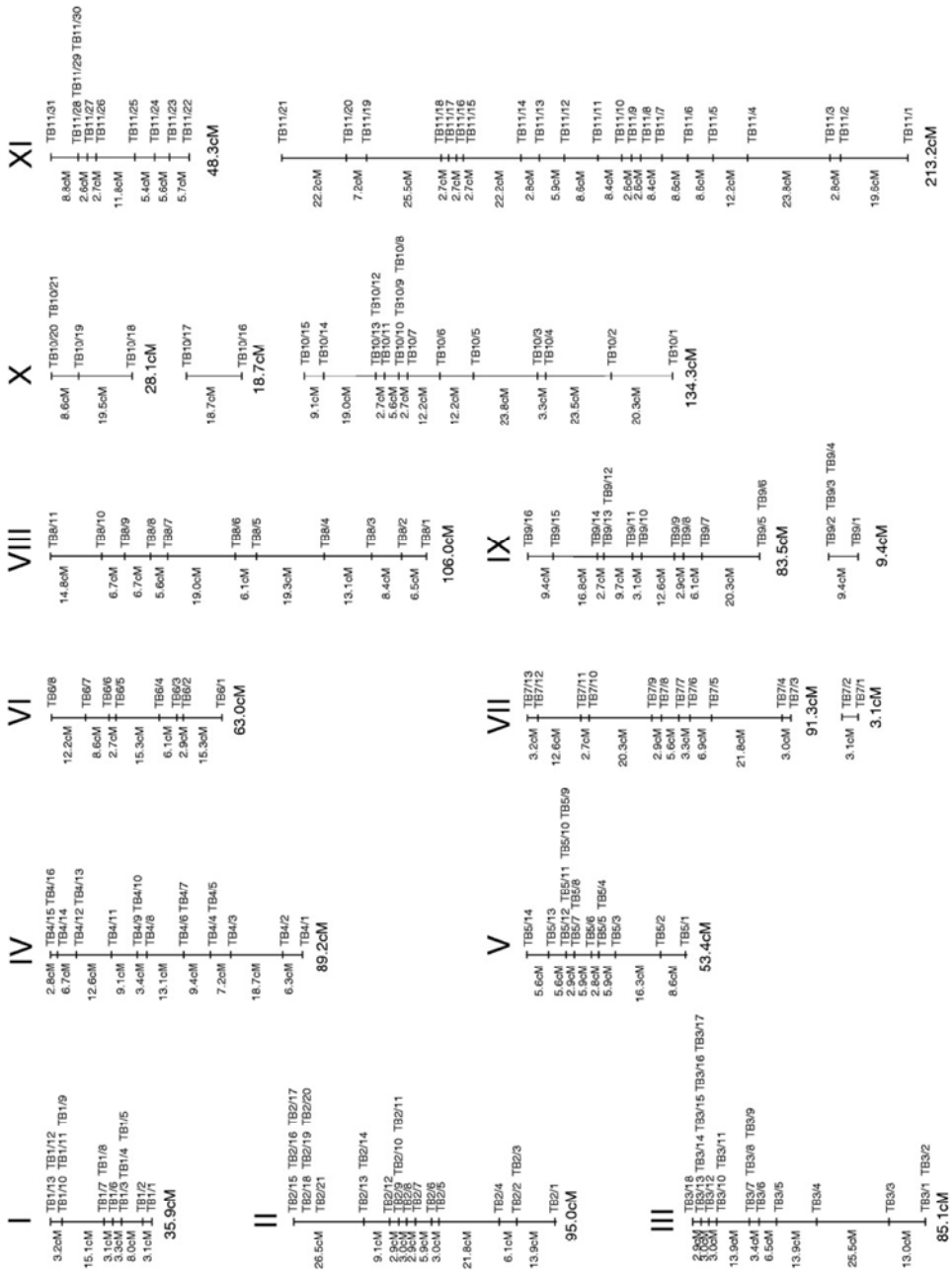


Figure 1.3 Genetic linkage map of *T.b.brucei*.

Genetic linkage map of *T.b.brucei*, genome strain TREU 927, corresponding to the 11 megabase chromosomes of *T.brucei.* Microsatellite markers are indicated to the right in each map. The genetic distance between each marker is given in centiMorgans, Haldane corrected and the genetic size of the linkage groups is given below each linkage group. Figure reproduced from MacLeod *et al* (2005b).

1.15 Project Objectives

The overall aim of this project was to develop genetic mapping tools for *T.b.gambiense* with which to perform genetic analysis, linkage mapping, and comparative strain analysis. The research presented in this thesis can be described in four specific objectives.

1. Construction of a *T.b.gambiense* genetic linkage map using a mapping population derived from a STIB 386 *T.b.gambiense* x STIB 247 *T.b.brucei* genetic cross, and comparative analysis with the *T.b.brucei* TREU 927 genetic map (Chapter 3).

2. Determination of the sensitivity of the parental lines to a number of trypanocidal drugs, and evaluation of the differences in drug sensitivity as potential phenotypes for dissection by a linkage mapping approach (Chapter 3).

3. Development of a novel *in vitro* cloning technique for the isolation of new and unique progeny clones from *T.brucei* genetic crosses, as a substitute for *in vivo* amplification of parasites (Chapter 4).

4. Investigation of the genetic basis of spontaneous loss of heterozygosity in several *T.brucei* culture strains, from an initial observation by molecular marker analysis, and assessment of an associated growth phenotype (Chapter 5).

Chapter 2

Materials and Methods

2.1 Chemicals, enzymes and reagents

BioRad	CHEF DNA markers of *Hansenula wingei, Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*
Cambrex Bioscience	SeaKem LE agarose, NuSieve GTG agarose
CP Pharmaceuticals	Heparin
GE Healthcare	AlkPhos Direct Labelling and Detection System with CDP-*Star,* Amersham Hybond N+ nylon membrane, Amersham Hyperfilm ECL
GIBCO	Dulbecco's phosphate buffered saline (D-PBS), foetal calf serum (FCS), Iscove's modified Dulbecco's medium (IMDM) with Glutamax, penicillin-streptomycin solution
Invitrogen	SDM-79 powdered media
Minitüb	Triladyl reagent
MWG-Biotech	Custom oligonucleotides
New England Biolabs (NEB)	100 bp DNA Ladder, 1 Kb DNA Ladder, all restriction endonucleases
Qiagen	Genomic-tip 100/G, Genomic DNA Buffer Set, QIAamp DNA Blood Mini Kit, QIAprep Spin Miniprep Kit, Qiaquick Gel Extraction Kit.
SAFC Biosciences	Serum Plus medium supplement
Smith Medical International Ltd	Portex tubing (0.75 mm x 30 M)
StrataClone	*Pfu* DNA polymerase, StrataClone PCR Cloning Kit

Thermo Scientific Custom PCR MasterMix, *Taq* DNA polymerase

Tyco healthcare 23G aluminium hub blunt needles

All other chemicals and regents were purchased from Fisher Scientific UK Ltd, Invitrogen Ltd, Sigma-Aldrich Ltd or BDH Ltd.

2.2 Buffers and Solutions

2.2.1 Molecular Biology

Crude lysis buffer – 50 mM Tris-HCl pH 8, 100 mM EDTA pH 8, 0.5% (w/v) SDS. Add 0.5 mg/ml proteinase K immediately before use.

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

Hybridisation buffer - 0.5 M NaCl, 4% (w/v) blocking regent, in AlkPhos direct labelling hybridisation buffer solution. Store in 25 ml aliquots at -20°C.

LB agar - 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar (Autoclaved).

LB medium - 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl (Autoclaved).

NDS buffer - 0.5 M EDTA, 10 mM Tris, 1% (w/v) lauroyl sarcosine, pH 8.0 and pH 9.0. Filter-sterilised and stored at 4°C.

Primary wash buffer – 2 M Urea, 0.1% (w/v) SDS, 50 mM sodium phosphate pH 7, 150 mM NaCl, 1 mM $MgCl_2$, 0.2% (w/v) blocking reagent. Stored at 4°C for up to 1 week.

Proteinase K stock solution – 20 mg/ml in 50 mM Tris-HCl (pH 8.0), 5 mM calcium acetate. Stored at 4°C.

20x Saline sodium citrate (SSC) – 3 M NaCl, 0.3 M tri-sodium citrate pH 7.0.

20x Secondary wash buffer – 1 M Tris, 2 M NaCl, pH 10.0. Stored at 4°C. Dilute $^1/_{20}$ in dH_20 and add 2 mM $MgCl_2$ immediately before use.

Sodium Dodecyl Sulphate (SDS) - 10% (w/v) stock solution (Autoclaved).

Southern blot denaturing solution - 1.5 M NaCl, 0.5 M NaOH.

Southern blot neutralising solution – 3 M NaCl, 0.5 M Tris-HCl pH 7.5.

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

T(0.1)E buffer - 10 mM Tris–HCl pH 8, 100 µM EDTA pH 8 (Autoclaved).

5x Tris-Borate-EDTA Buffer (TBE) - 450 mM Tris, 450 mM Boric acid, 10 mM EDTA, pH 8. Dilute 1/10 in dH_20 before use.

10x Tris-Borate-(0.1)EDTA Buffer (TB(0.1)E) – 890 mM Tris, 890 mM Boric acid, 2 mM EDTA, pH 8.0 (Autoclaved). Dilute 1/10 in dH_20 before use.

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

2.2.2 Trypanosome *in vitro* and *in vivo* growth

Bloodstraw stabilate solution - 25% (v/v) Triladyl in sterile ddH_20, Stored at 4°C.

Bloodstream culture stabilate solution - 10% (v/v) DMSO in Serum Plus. Filter-sterilised. Stored at 4°C.

CBSS (Fairlamb *et al.*, 1992)/heparin – 25 mM HEPES, 120 mM NaCl, 5.4 mM KCl, 0.55 mM $CaCl_2$, 0.4 mM $MgSO_4$, 5.6 mM Na_2HPO_4, 11.1 mM D-Glucose, 100 U/ml heparin. Filter-sterilised and stored at 4°C.

Cis-aconitate solution – 30 mM cis-aconitate in ddH_20, pH 7.5. Filter-sterilised. Stored in 1 ml aliquots at -20°C.

Cyclophosphamide solution - 31.25 mg/ml in ddH_20. Stored in 5 ml aliquots at -20°C.

HMI-9 media (Hirumi and Hirumi, 1989) modified (Paul Voorheis, personal communication) – IMDM supplemented with; 1 mM hypoxanthine, 50 µM

bathocuproinedisulfonic acid, 1.5 mM cysteine, 2 mM sodium pyruvate, 160 µM thymidine, 200 µM 2-mercaptoethanol, 1.4 mM glucose, 125 µM adenosine, 125 µM guanosine, 30 µg/ml kanamycin, 1 mg/ml methyl cellulose, 20% (v/v) Serum Plus. Filter sterilised. Stored in 500 ml aliquots at 4°C.

Procyclic culture stabilate solution - 20% (v/v) glycerol in FCS. Filter-sterilised. Stored at 4°C.

SDM-79 media (Brun and Schonenberger, 1979) – One custom made media powder pack/5 L ddH_20/10 g $NaHCO_3$, pH 7.3. Filter-sterilised. Stored in 450 ml aliquots at 4°C.
Each aliquot of media was supplemented before use with 15% (v/v) FCS, 100 units/ml penicillin and 100 µg/ml streptomycin.

2.3 Trypanocidal Drugs

Cymelarsan (Rhone Merieux)
Diluted as appropriate in sterile H_20 immediately prior to use.

Suramin sodium salt (Sigma S2671)
Diluted as appropriate in sterile H_20 immediately prior to use.

Diminazene Aceturate (Berenil) (Sigma D7770)
Diluted as appropriate in sterile H_20 immediately prior to use.

Pentamidine isoethionate salt (Aldrich 439843)
100 mg/ml in sterile H_20 stock for *in vivo* assays stored in 20 µl aliquots at -20°C.
1 mM in sterile H_20 stock for *in vitro* assays stored in 20 µl aliquots at -20°C.

Drugs were diluted to the appropriate concentration in sterile H_20 for *in vivo* experiments and the culture stage appropriate media for *in vitro* experiments.

2.4 Mice stocks

Adult female ICR mice were supplied by Harlan Laboratories.

2.5 *Trypanosoma brucei* strains

2.5.1 Parental strains

Three cloned stocks of *T.brucei* were used as parental lines in this investigation. These were the *T.b.brucei* strains STIB 247 and TREU 927 and the *T.b.gambiense* group 2 strain STIB 386. The origin of these lines is described in Chapter 1 (Table 1.2). Further information concerning the pedigree of these lines has also been collated by the George Cross laboratory and is available as an electronic resource (Cross, 2009).

2.5.2 F1 progeny clones

Progeny clones derived from crosses between STIB 386 and STIB 247, previously performed in the laboratory, were utilised for this analysis. The progeny and their derivation has been described previously (Turner *et al.*, 1990; Tait *et al.*, 2002; MacLeod *et al.*, 2005a). Briefly, the trypanosome stocks were amplified in ICR mice and then tsetse flies were co-infected with a mixture of the two bloodstream stage parental trypanosomes. After developing within the flies to the metacyclic stage, the populations of trypanosomes from each fly were monitored for the presence of the products of mating, grown up in mice, and then cryopreserved in liquid nitrogen from the first peak of parasitaemia. The stabilates were designated by the fly number and the day (post-fly infection) on which the trypanosomes were sampled (F9/45, etc.). In several cases, the same fly was sampled more than once (F9/45, F9/56, etc.). Cloned lines were then established either by directly cloning metacyclic stage trypanosomes in individual immuno-suppressed mice or by optically cloning from bloodstream stage infections derived directly from feeding infected tsetse on a mouse. The resulting metacyclic and/or bloodstream, cloned populations from six mixed infected flies (F 8, 19, 28, 29, 80 and 492) were then genotyped with two microsatellite markers Ch5/JS2 (TB5/4) (Sasse, 1998) and Ch2/PLC (TB2/20) (MacLeod *et al.*, 2005b) and three minisatellites markers, Ch1/MS42 (TB1/1) (Barrett *et al.*, 1997), Ch10/CRAM (TB10/14), and Ch3/292 (TB3/13) (MacLeod *et al.*, 1999) that were heterozygous in one or both of the two parental stocks. This resulted in the identification of 38 independent F1 progeny clones from the cross, each of a different and unique genotype. A list of all 38 hybrids and their genotypes is provided in Table 2.1.

An additional set of 12 new progeny clones were generated by *in vitro* cloning during the course of this study. The generation of these clones and their inclusion into the STIB 386 genetic map is detailed in Chapter 4 and the protocol for this technique is detailed in section 2.7.7.

2.5.3 Uncloned STIB 386 x STIB 247 populations

Uncloned populations derived from several of the STIB 386 x STIB 247 crosses described above, from which F1 progeny clones had been previously identified, were used as sources of material for the *in vitro* cloning experiment detailed in Chapter 4. Stabilates were designated by tsetse fly number and the day (post fly infection) in which trypanosomes were sampled. A number of these uncloned bloodstream populations, derived from different genetic crosses in different flies, are available as stabilates. Four of these were used in the *in vitro* cloning experiment (F28/46, F29/42, F9/31, F492/50), all of which have had several progeny clones isolated previously.

Hybrids and parents	Genotyping marker scores				
	TB5/4	TB2/20	TB1/1	TB3/13	TB10/14
STIB 247 parental stock	5-6	5-5	5-5	5-5	1-1
STIB 386 parental stock	1-2	1-2	1-2	1-2	1-2
F9/45 mcl 2	1-5	1-5	2-5	1-5	1-2
F9/45 mcl 10	1-6	2-5	1-5	2-5	1-1
F9/45 mcl 11	1-5	2-5	1-5	1-5	1-2
F9/45 mcl 12	1-5	2-5	-	2-5	1-1
F9/34 mcl 1	2-5	1-5	1-5	2-5	1-1
B80 cl 2	1-5	2-5	2-5	2-5	1-2
F492/50 bscl 1	2-6	2-5	1-5	2-5	1-1
F492/50 bscl 6	1-5	1-5	-	1-5	1-2
F492/50 bscl 8	1-6	2-5	1-5	2-5	1-1
F492/50 bscl 9	2-5	2-5	2-5	1-5	1-2
F492/50 bscl 12	2-5	1-5	2-5	2-5	1-1
F492/50 bscl 14	1-6	1-5	1-5	1-5	1-1
F492/50 bscl 21	2-6	2-5	1-5	1-5	1-1
F492/50 bscl 23	2-5	2-5	1-5	1-5	1-2
F492/50 bscl 5/1b	2-6	2-5	1-5	1-5	1-2
F9/41 bscl 5	1-5	2-5	2-5	2-5	1-2
F9/41 bscl 7	1-5	1-5	2-5	2-5	1-2
F9/41 bscl 9	2-6	2-5	1-5	1-5	1-1
F29/46 bscl 3	2-6	2-5	1-5	1-5	1-1
F29/46 bscl 4	2-6	1-5	1-5	1-5	1-1
F19/31 clone 1	1-6	1-5	1-5	2-5	1-2
F 19/31 bscl 11	1-5	2-5	1-5	1-5	1-1
F28/46 bscl 6	2-5	2-5	1-5	1-5	1-2
F28/46 bscl 11	2-5	2-5	2-5	2-5	1-2
F29/46 bscl 2	2-6	2-5	1-5	1-5	1-1
F28/46 bscl 1	2-5	1-5	1-5	1-5	1-1
F28/46 bscl 4	2-6	2-5	1-5	1-5	1-2
F28/46 bscl 7	2-5	1-5	1-5	2-5	1-1
F28/46 bscl 8	2-6	2-5	2-5	1-5	1-2
F29/46 bscl 1	1-5	1-5	1-5	2-5	1-1
F9/41 bscl 1	1-6	1-5	2-5	2-5	1-1
F9/41 bscl 2	1-6	1-5	1-5	1-5	1-1
F9/41 bscl 8	1-5	1-5	1-5	2-5	1-1
F9/41 bscl 11	1-5	2-5	1-5	1-5	1-2
F19/31 bscl 5	1-6	2-5	1-5	1-5	1-1
F19/31 bscl 10	2-6	2-5	1-5	1-5	1-2
F19/31 bscl 8	1-6	2-5	1-5	1-5	1-2
F492/50 bscl 7	2-5	2-5	1-5	1-5	1-2

Table 2.1 Unique progeny genotyping data.

The name and relevant genotypes of the parental strains and 38 unique F1 progeny derived from the STIB 386 x STIB 247 crosses that were analysed for the construction of the *T.b.gambiense* linkage map. Inheritance of marker alleles from both parents for two microsatellites (TB5/4 and TB2/20) and three minisatellites (TB10/14, TB3/13 and TB1/1) that were used as genotyping markers.

2.6 Molecular Biology Techniques

The methods used during the course of this work followed standard molecular biological procedures, which have largely been described elsewhere (Maniatis *et al.*, 1989). 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, where relevant, within each results chapter.

2.6.1 DNA extraction from blood

Trypanosomes were isolated from 200 μl of mouse blood at a parasitemia of ~10^7-10^8 trypanosomes/ml by centrifuging at 250 *g* for seven minutes. The trypanosome-containing supernatant and buffy coat were carefully transferred to a fresh tube, centrifuged at 900 *g* to pellet parasites and the supernatant removed. The cell pellet was resuspended in 200 μl D-PBS and DNA was extracted from the cells using the Qiaamp DNA blood minikit according to the manufacturers instructions.

2.6.2 DNA extraction from bloodstream form culture and procyclic form culture cells

2.6.2.1 High quality DNA

High quality DNA was prepared from cultured *T.brucei* cells for Illumina sequencing and genetic mapping studies.

Large volumes of high molecular weight DNA for use as starting material for Illumina sequencing were typically prepared from 200 ml of procyclic culture in log phase growth at a density of > 1 x10^7 trypanosomes/ml. The cell culture was centrifuged at 2000 *g* for seven minutes, the supernatant discarded and replaced with D-PBS and the centrifuge step repeated. Up to 100 μg of genomic DNA was extracted from the cell pellet using a Genomic tip 100/G column with the Genomic DNA Buffer Set in accordance with the manufacturer's instructions.

For genetic mapping and genotyping protocols, smaller volumes of DNA were prepared from 200 μl procyclic form culture cells in log phase of growth at a

density of ~10^7 cells/ml, or 1 ml of bloodstream form culture cells at a density of ~10^6 cells/ml. Cell culture samples were spun at ~2000 *g* for seven minutes, washed with one volume of D-PBS, the centrifuge step repeated and then the cell pellets resuspended in 200 μl D-PBS. DNA was extracted from the pelleted cells using the Qiaamp DNA blood minikit in accordance with the manufacturer's instructions.

2.6.2.2 Crude lysates

For rapid genotyping of trypanosome cells from culture and *in vitro* assays, crude lysates were prepared from both procyclic and bloodstream form culture cells. An aliquot of 0.2-1 ml of bloodstream or procyclic culture at a log phase of growth was centrifuged at 2000 *g* for five minutes at room temperature, the supernatant discarded and the pellet resuspended in 200 μl D-PBS. The cells were centrifuged again, supernatant discarded and the D-PBS wash step repeated. The pellet was resuspended in 50 μl crude lysis buffer supplemented with 0.5 mg/ml proteinase K and incubated at 56°C overnight. Lysates were stored at –20°C and before use diluted $^1/_{50}$ in sterile water and incubated at 95°C for five minutes to inactivate the proteinase K.

2.6.2.3 DNA isolation from 96 well cultures

A quick boil lysis technique was adopted for DNA extraction from small volume of cultures, for example, the wells of a 96-well microtitre plate. One hundred microlitres of bloodstream culture was isolated from each well and centrifuged at 2000 *g* for five minutes. The supernatant was then discarded, the cell pellet resuspended in 100 μl D-PBS and the centrifuge step repeated. The supernatant was discarded again, the pellet resuspended in 10 μl sterile H_20 and heated at 95°C for five minutes. This lysate was then stored at –20°C.

2.6.2.4 Preparation of DNA samples for PCR amplification

For all DNA samples types, the DNA concentration was diluted for PCR amplification by $^1/_5$ – $^1/_{5000}$ in molecular biology quality water. One microlitre of this diluted DNA was used as a template for subsequent PCR reactions

2.6.3 General PCR procedure

Because of the sensitivity of PCR, precautions were taken to ensure that reagents and equipment used for PCR were kept free of contaminating DNA. Aliquots of PCR dedicated regents and tips were used and kept separate from post-PCR products.

All PCR reactions, unless otherwise stated were prepared in 10 μl reaction volumes containing: 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, 1mM dATP, 1mM dGTP, 1mM dTTP and 1mM dCTP (Custom PCR MasterMix), 1 μM each oligonucleotide primer, 0.5 units *Taq* DNA polymerase and 1 μl DNA template. Reactions were overlaid with mineral oil to prevent evaporation and amplification carried out in a Robocycler gradient 96 (Stratagene).

The cycling conditions for all microsatellite markers were as follows;

30 cycles of:

Denature	95°C for 50 seconds
Anneal	50°C for 50 seconds
Extension	65°C for 50 seconds

The cycling conditions for the three minisatellites (TB1/1, TB3/13 and TB10/14) were as follows;

30 cycles of:

Denature	95°C for 50 seconds
Anneal	60 °C for 50 seconds
Extend	65°C for 3 minutes

The cycling conditions for any additional PCR reactions are listed with the relevant oligonucleotides in Appendix 1.

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

2.6.4.1 Proof-reading Taq polymerase

PCR amplification of DNA for downstream cloning and sequencing reactions were modified by the addition of the proofreading polymerase, *pfu* DNA polymerase, to the PCR reaction in a 1:10 ratio with *Taq* polymerase.

2.6.4.2 Post-PCR addition of 3' adenine residues by Taq polymerase

Taq polymerase catalyses the addition of an adenine residue to the 3' ends of both strands of the PCR product during amplification, which may be paired with a modified uridine overhang on the ends of the Strataclone PCR cloning vector pSC-A-amp/kan for ligation. PCR products synthesised with proofreading DNA polymerases such as *pfu* do not contain this 3' adenine overhang and thus an additional step is required to correct for this before proceeding to the ligation reaction.

PCR reactions were kept on ice and 0.1 μl *Taq* polymerase added. This was transferred to a robocycler preheated to 72°C and incubated for 10 minutes. Following this step, the PCR reactions were keep on ice before proceeding to the ligation reaction.

2.6.5 Gel electrophoresis

Standard PCR products, restriction digest reactions and minisatellites PCR products were separated by gel electrophoresis on a 1% (w/v) Seakem LE agarose gel.

Microsatellite PCR products were separated by gel electrophoresis on a 3% (w/v) Nusieve GTG agarose gel.

Where appropriate, a 100 bp or 1 Kb DNA ladder was loaded on the gel alongside the PCR and restriction digest products to allow the estimation of DNA fragment sizes.

All gel electrophoresis was performed in 0.5x TBE buffer containing 50 ng/ml ethidium bromide and visualised by UV illumination using an UVIpro platinum 1.1 Gel documentation system.

2.6.6 Oligonucleotides

Oligonucleotide primers were designed using the PRIDE primer design service (PRIDE, 1998) and where appropriate the *T.brucei* reference sequence accessed though the *T.brucei* GeneDB resource (GeneDB, 2005). All oligonucleotides were synthesised by MWG-Biotech. The sequence and genomic location of all primers in this thesis is given in Appendices 1-3.

2.6.7 Microsatellite primer design

Microsatellite markers were identified from the *T.brucei* genome sequence (Berriman *et al.*, 2005) accessed though the *T.brucei* GeneDB resource (GeneDB, 2005) with the Tandem Repeat Finder program (Benson, 1999). Candidate markers were identified as sequences containing > 10 copies of a repeat motif of 2-6 nucleotides with > 70% sequence identity. Primer pairs were then designed for each microsatellite marker in the unique sequence flanking each repeat region using the PRIDE primer design program (PRIDE, 1998).

Primers for 810 markers, evenly distributed throughout the 11 chromosomes of the *T.brucei* genome, which had been previously designed for screening the TREU 927 x STIB 247 cross during construction of the TREU 927 *T.b.brucei* map, were available (MacLeod *et al.*, 2005b). Primers for an additional 215 new markers were designed specifically for the construction of the STIB 386 map.

2.6.8 Genotyping *T.brucei* strains

A core panel of five microsatellite markers (TB5/4, TB2/20, TB1/1, TB10/14 and TB3/13 had been previously identified for which a unique genotype has been assigned for all parents and hybrids (Table 2.1). This panel of markers was thus used routinely to identify and confirm the genotype matched that of the original clone for all trypanosome laboratory stocks and DNA samples in use in this project.

A set of new progeny bloodstream clones were generated during the course of this project by limited dilution *in vitro* cloning of uncloned bloodstream populations derived from mice infected with metacyclic stage trypanosomes from mixed infected tsetse flies. Comparisons were made between the genotypes of any potential new progeny for these five markers (as well as two additional heterozygous markers identified during the generation of the STIB 386 genetic map, TB10/28 and TB3/1 (Chapter 3), to the existing hybrid genotypes derived from the same uncloned fly population. New clones with an identical genotype to existing clones were considered duplicates and discarded. Clones with a different genotype to all progeny previously isolated from the same fly were considered as new and unique STIB 386 x STIB 247 progeny clones.

2.6.9 Construction of the *T.b.gambiense* STIB 386 genetic map

2.6.9.1 Microsatellite screening for inheritance of STIB 386 alleles in F1 progeny clones

Microsatellites were used to screen the parental STIB 386 and STIB 247 genomic DNA by PCR to identify loci that were heterozygous for allele size in STIB 386 and so would segregate in the progeny. Markers were identified as heterozygous for STIB 386 based on a visible product size difference by gel electrophoresis. Amplification of visibly heterozygous markers was repeated for all 38 progeny from the STIB 386 x STIB 247 cross. Following gel electrophoresis, progeny were assigned a binary score based on the inherited STIB 386 allele at each microsatellite locus as described in detail in Chapter 3. All gels were independently scored by a second individual to ensure progeny genotypes were correctly assigned.

Marker screening was performed with the assistance of Mrs Lindsay Sweeney.

2.6.9.2 Linkage analysis

A genetic map of STIB 386 was generated, based on the segregation of marker alleles in the F1 progeny, for loci heterozygous in the STIB 386 parent. The allele segregation data was analysed using the Map Manager QTX software (Manly *et al.*, 2001) with a Haldane map function and the highest level of significance for linkage criteria, giving a probability of type 1 error P = $1e^{-6}$. Linkage between the

adjacent physical markers was determined by a LOD (Log of the Odds) score of 5.5 or greater.

2.6.10 Cloning and sequencing

2.6.10.1 DNA ligation

Cloning of PCR products was carried out by ligation into the Strataclone PCR cloning vector pSC-A-amp/kan using the Strataclone PCR cloning kit, following the manufacturer's protocol. Briefly, 3 µl Strataclone cloning buffer, 2 µl PCR product and 1 µl Strataclone pSC-A-amp/kan cloning vector were gently mixed and incubated at room temperature for five minutes before placed on ice for transformation.

2.6.10.2 Transformation of *E.coli* competent cells

Transformation of the ligation reaction was carried out in Strataclone Solopack Competent cells according to the manufacturer's protocol. The cells were thawed on ice, 1 µl of cloning reaction mixture added and the mixture incubated on ice for 20 minutes. After incubation, the transfection mixture was heat shocked at 42^{o}C for 45 seconds then returned to ice for two minutes. Two hundred and fifty microlitres of LB media, prewarmed to 37^{o}C, was added to cells and the competent cells allowed to recover by incubating at 37^{o}C, with shaking at 200 rpm, for one hour. During this period, LB plates containing 100 µg/ml ampicillin were prewarmed to room temperature and prepared for blue/white colour screening by spreading 40 µl of 2% (w/v) X-Gal on each plate. Following incubation, the transformations were spread onto the agar plates at a range of dilutions and incubated at 37^{o}C overnight to allow colony growth to occur.

2.6.10.3 Analysis of transformants

Following transformation, single white colonies were inoculated into 5 ml LB media containing 100 µg/ml ampicillin and cultured at 37^{o}C with 200 rpm shaking overnight. *E.coli* cultures aliquots of 1.5 ml were pelleted by centrifugation at 5000 *g* for five minutes and the supernatant discarded. Plasmid purifications were carried out on the cell pellets using the QIAprep Spin Miniprep Kit in accordance with the manufacturer's instructions. The purified plasmid was eluted with 50 µl H_20 and stored at -20^{o}C for further analysis. Plasmid preps were analysed for

successful insertion of the correct sized product by restriction digestion with the restriction endonuclease, EcoRI.

2.6.10.4 Digestion of DNA by restriction enzyme

Digestion of purified plasmid and PCR products with restriction endonucleases was carried out in a 10 µl reaction volume. Typically 1 µl of PCR product or plasmid was added to 1 µl of the supplied restriction digest buffer (optionally contain BSA), 0.5 µl of restriction enzyme and 7.5 µl H_20. The reaction was overlaid with mineral oil to reduce evaporation and incubated at the temperature recommended by the manufacturers for between 2-24 hours. After the specified incubation period the reaction was separated by gel electrophoresis, alongside the appropriate DNA ladder to estimate the fragment sizes.

2.6.10.5 DNA quantification

The DNA concentration of Genomic or Plasmid DNA, when required, was determined from a 1 µl sample, using the nucleic acid measurement program of a nanodrop ND-1000 spectrophotometer, calibrated with the appropriate diluents.

2.6.11 High throughput whole genome sequencing of *T.brucei* strains

2.6.11.1 *T.brucei* strain DNA

Genomic DNA was extracted from procyclic form cell culture of parental lines STIB 247, STIB 386 and TREU 927 LOH line 1 using the genomic DNA extraction protocol (Section 2.6.2.1), and sent to the Pathogen Sequencing Unit, WTSI.

2.6.11.2 Illumina sequencing

All library preparation and Illumina sequencing was performed at WTSI using the Illumina Genome Analyzer System in accordance with the manufacturer's protocols and specifications. One Illumina Genome Analyzer run was performed on each sample to generate paired-end, 36 bp reads. Following standard quality control filtering of the sequence reads, 6,168,680 x 36 bp reads were produced for STIB 247, 7,880,638 reads for STIB 386 and 8,132,452 reads for TREU 927 and made available for data analysis.

2.6.11.3 Data analysis

Jacqueline McQuillan of WTSI, performed data analysis of the Illumina sequence output during a collaborative visit by Anneli Cooper to WTSI Pathogen Sequencing Unit from the 21 April – 2 May, 2008.

The paired 36 bp sequences from each strain were converted to FASTQ format and aligned to the TREU 927 *T.brucei* reference genome (version 4) (GeneDB, 2005) using the Mapping and Assembly with Qualities (MAQ) software v0.6.6 (Li *et al.*, 2008). MAQ aligns paired-end reads to a reference sequence and then calls a consensus based on a statistical model that assigns a quality score to each base position (Li *et al.*, 2008). Only uniquely mapped reads with fewer than two mismatches are retained. After mapping and consensus base calling, MAQ was then used to evaluate each aligned base and its base quality value (phred score) to indicate putative SNPs which were filtered according to a quality criteria > 30. The resulting data files from this analysis are too substantial to be included in this thesis, however relevant data from this output is reported in Chapter 5 and the complete data files are available from the Turner Tait Macleod Group upon request.

2.6.11.4 SNP-RFLPs

To experimentally confirm a number of the SNPs that had been identified by the MAQ analyses of Illumina data, a series of SNP-RFLPs were designed using the VectorNTI software (Invitrogen). The appropriate region of TREU 927 *T.brucei* reference sequence was first imported from GeneDB into VectorNTI (Invitrogen). Then the restriction enzyme function in VectorNTI (Invitrogen) was used to screen the heterozygous SNPs identified by MAQ and detect those SNPs for which one base was within a restriction digest site that was not present in the other base. In addition, screening was also used to ensure that, where possible, the local sequence did not contain another site for the same restriction enzyme within 100 bp of the SNP site. Oligonucleotide primers were then designed approximately 500 bp apart to amplify the SNP site from genomic DNA. PCR products were digested by the appropriate restriction endonuclease to identify whether the site was heterozygous or homozygous in the specific sample line.

2.6.11.5 DNA Sequencing

DNA sequencing of PCR products was used to confirm the MAQ generated SNP data and confirm the LOH boundary of the TREU 927 LOH line 1, as described in Chapter 5. The Primers used for the PCR amplification and sequencing of this region are provided in Appendices 1 and 3. PCR products were cloned using the Strataclone PCR cloning kit (protocol 2.6.10), Colonies analysed by SNP-RFLP (protocol 2.6.11.4) and the plasmid DNA sequenced. Sequencing of DNA was carried out by the Dundee sequencing service (University of Dundee). DNA was prepared in accordance with the service instructions.

2.6.11.6 Sequence analysis

The alignment and assembly of sequence data was performed on imported sequence files from the University of Dundee server using the contigexpress program within the VectorNTI software (Invitrogen).

2.6.12 Pulsed Field Gel Electrophoresis (PFGE)

2.6.12.1 Preparation of chromosome blocks for PFGE

Chromosomal DNA agarose blocks were prepared for strains of normal *T.brucei* and strains exhibiting loss of heterozygosity from procyclic culture based on a published protocol (Leech *et al.*, 2004).

Briefly, a 100 ml culture of log phase procyclic form trypanosomes was grown to 10^7 cells/ml in SDM-79 supplemented with 15% (v/v) FCS. Cells were harvested, centrifuged at 4000 *g* for seven minutes, washed twice in one volume of PBS and resuspended in PBS at 10^9 cells/ml. Cells were mixed with an equal volume of 1.4% (w/v) agarose at 50°C to make a final concentration of 0.5 x 10^9 cells/ml and the mixture quickly transferred to plug moulds using a Pasteur pipette. Plug mould were set at 4°C for 10 minutes and then lysed *in situ* by releasing plugs into a 15 ml falcon tube containing NDS pH 9.0 lysis buffer supplemented with 1 mg/ml proteinase K. After a 24 hour incubation at 50°C the buffer was exchanged for NDS pH 8.0 supplemented with 1 mg/ml proteinase K, and incubated for a further 24 hours at 50°C. Finally the buffer was exchanged for fresh NDS pH 8.0 and the chromosome blocks stored at 4°C until required.

Before use, chromosome blocks were removed from NDS storage buffer and dialysed against TE buffer, exchanging buffer several times over a 24 hour period, to remove excess proteinase K, detergent and EDTA before PFGE.

2.6.12.2 PFGE

PFGE was performed using a CHEF DR-III system (BioRad). Agarose gels were prepared at a percentage (either 0.8 or 1%) and in electrophoresis buffer (either 0.5x TBE or 0.5x TB(0.1)E) appropriate for the program conditions recommended for separating different sized chromosomes (Leech *et al.*, 2004). The dialyzed chromosomes blocks were cut to the required size using a sterile scalpel blade, fixed onto a gel comb using molten agarose and then the gel cast around the comb. Chromosome blocks of the organisms *S.pombe, S.cerevisiae* or *H.wingei* were used as PFGE standards and run alongside the samples blocks, as appropriate, for each size separation program to allow an estimation of *T.brucei* chromosome sizes. Each gel was run in the appropriate electrophoresis buffer and with an appropriate electrophoresis program for optimal separation of the chromosomes of interest (Specific details about each PFGE are given in the relevant figure legends). Electrophoresis buffer was recirculated using a cooling system to maintain a constant buffer temperature of 14^{o}C.

To prevent microorganism contamination and potential degradation of the chromosome blocks the PFGE experiments were performed in a darkened room and the equipment thoroughly cleaned before and after each run with 0.1% (w/v) SDS and dH_20 washes.

After completion of the PFGE program, chromosome bands were visualized by staining the gel with 0.5 mg/ml ethidium bromide overnight followed by washing with dH_20 for one hour, then visualised by UV illumination using a UVIpro platinum 1.1 Gel documentation system.

2.6.13 Southern blotting

Standard procedures were used for the preparation of PFGE agarose gels for Southern blotting. Briefly, the gel was immersed in 0.25 M HCl for 2 x 7 minutes, followed by denaturing solution for 2 x 15 minutes and then finally neutralising solution for 2 x 15 minutes. The gel was transferred to a platform which was covered with a layer of 3MM filter paper of the same width of the gel, and the

ends of which were immersed in 20x SSC solution. The gel was overlaid with a Hybond N+ nylon membrane the exact same size as the gel, and several layers of 3MM filter paper pre-soaked in 3x SSC, gently rolling each layer with a pipette to remove air bubbles. A stack of paper towels, glass plate and weight was then placed on top of the blot and by this method the DNA was transferred to the nylon membrane by capillary action. After a 48 hour transfer the blot was dismantled, the position of the wells was marked on the nylon membrane and the DNA was cross-linked to the membrane using a UV cross linker (Spectronics).

2.6.13.1 Preparation of DNA probes

DNA probes were prepared to three chromosome 10 genes (Ch10/CRAM, Ch10/0620 and Ch10/2920) to investigate the loss of heterozygosity in *T.brucei* culture lines. Probes were generated by amplifying loci from genomic TREU 927 DNA by PCR, using primers Ch10/CRAM A and B, Ch10/0620 A and B and Ch10/2920 A and B respectively (the primers and PCR conditions used are given in Table 5.4). PCR products were separated by gel electrophoresis and then gel slices containing the DNA fragments were excised from the gel. The gel slices were processed to purify DNA using the Qiaquick Gel Extraction Kit according to the manufacturer instructions.

DNA was quantified, then 100 ng of DNA labelled for hybridization with alkaline phosphatase, using the AlkPhos Direct Labelling kit according to the manufacturer's instructions. Probes were kept on ice and used immediately or mixed with one volume of glycerol and stored at –20^{o}C for up to six months.

2.6.13.2 Hybridization

Hybridisation was performed using the Amersham Gene Images AlkPhos Direct Labelling and Detection System following the manufactures instructions. Briefly, the blot was prehybridised with 0.125 ml/cm^2 preheated AlkPhos direct hybridisation buffer at 65^{o}C in a rotating bottle hybridization oven (Hybaid) for at least 15 minutes. Then 5-10 ng labelled probe/ml of hybridisation buffer was added to the reaction and allowed to hybridise at 65^{o}C overnight. After incubation with the probe, the blot was washed twice with preheated primary wash buffer for 10 minutes at 65^{o}C at a volume of 2-5 ml/cm^2, followed by twice with secondary wash buffer at room temperature for five minutes at a volume of 2-5 ml/cm^2.

For detection, blots were placed, DNA facing upwards, onto saran wrap and the AlkPhos CDP-Star detection reagent from the kit applied at a volume of 30-40 µl/cm^2. After five minute incubation at room temperature, excess reagent was shaken off and the blot wrapped in saran wrap and exposed to autoradiography film for a period of five minutes to 24 hours, depending on the band intensity required.

2.6.13.3 Stripping and reprobing blots

Blots could be reprobed by first stripping the blot of the old probe according to the manufacturer's instructions. To do this the blot was incubated in an excess volume of 0.5% (w/v) SDS solution at 60°C for 1 hour and then washed in 100 mM Tris-HCl pH 8 for five minutes at room temperature. Blots were either reprobed immediately or stored wrapped in saran wrap at 4 °C until required.

2.7 *In vitro T.brucei* culture techniques

2.7.1 *T.brucei in vitro* cell culture

T.brucei procyclic form cell cultures were cultivated in SDM-79 medium supplemented with 15% (v/v) FCS, 100 units/ml penicillin and 100 µg/ml streptomycin. Cultures were maintained in phenolic lidded flasks in a 27°C incubator and subpassaged every 2-3 days to maintain a cell density of ~10^6-10^7 trypanosomes/ml.

T.brucei bloodstream form cell cultures were cultivated in HMI-9 medium supplemented with 20% (v/v) Serum plus. Cultures were maintained in vented lid flasks in a 37°C, 5% CO_2 incubator and subpassaged every 2-3 days to maintain a cell density of ~10^5-10^6 trypanosomes per ml.

2.7.2 Adaptation of trypanosomes from blood to procyclic form culture

Blood was collected from ICR mice infected with *T.brucei* at the first peak of parasitaemia (approximately 1 x 10^8 parasites per ml), into CBSS/heparin, under sterile conditions by cardiac puncture. A blood aliquot of 0.1 – 1 ml, was added to 10 ml SDM-79 procyclic culture media supplemented with 20% (v/v) FCS containing 3 mM cis-aconitate and placed at 27 °C overnight. After 24 hours the trypanosome

containing supernatant was carefully decanted into a fresh flask, leaving behind the settled blood layer. Cells were incubated at 27°C to allow transformation to procyclic form to occur until a density of 1 x 10^6 - 1x10^7 cells/ml was reached and then the cells were stabilated as described in protocol 2.7.6 and maintained by sub-passage.

2.7.3 Adaptation of trypanosomes from blood to bloodstream form culture

Blood was collected from ICR mice infected with *T.brucei* at the first peak of parasitaemia (approximately 1 x 10^8 parasites per ml), under sterile conditions by cardiac puncture. One ml of blood was added to 9 ml culture media (modified HMI-9 containing 20% (v/v) serum plus) and gently mixed, then centrifuged at 200 *g* for 10 minutes to pellet the red blood cells.

The trypanosome-containing supernatant was aspirated into a T-25 flask and the cells incubated at 37°c in a humidified 5% CO_2 incubator for two hours to recover. The trypanosomes were then counted by haemocytometer and the concentration adjusted to 1 x 10^5 cells/ml in modified HMI-9 containing 20% (v/v) serum plus. Cells were maintained by diluting trypanosomes ~1:1-1:5 with culture media 2-3 times per week until a density of 1 x 10^6 cells/ml could be reached and then the cells were stabilated (Section 2.7.5) and maintained by sub-passage.

2.7.4 Adaptation of trypanosomes from bloodstream form culture to procyclic form culture

To adapt bloodstream culture cells to procyclic form culture, 5 ml of log stage bloodstream culture cells were isolated at a density of ~10^6/ml and centrifuged at 250 *g* for five minutes. The supernatant was then removed and the cell pellet gently resuspended in 5 ml SDM-79 supplemented with 30% (v/v) FCS and 3 mM cis-aconitate. Cells were incubated at 27°C to allow transformation to procyclic form to occur until a density of 10^6-10^7 cells/ml was reached and cells were stabilated and then maintained by sub-passage.

2.7.5 Preparation of bloodstream culture stabilates

Bloodstream culture cells were isolated at a density of 10^5 - 10^6 cells/ml and combined with an equal volume of bloodstream culture stabilate solution

prewarmed to 37°C. This was mixed by gentle inversion, then dispensed in 1 ml aliquots into cryovials with an assigned stabilate number and frozen at –80 °C in a cryobox. After 24 hours the cryovials were transferred to liquid nitrogen for long term storage.

2.7.6 Preparation of procyclic culture stabilates

Procyclic culture cells were isolated at a density of approx 5 x 10^6 cells /ml and mixed with an equal volume of procyclic culture stabilate solution, prewarmed to 27°C. This was mixed by gentle inversion, then dispensed in 1 ml aliquots into cryovials with an assigned stabilate number and frozen at –80°C in a cryobox. After 24 hours the cryovials were transferred to liquid nitrogen for long term storage.

2.7.7 *In vitro* cloning of new STIB 386 x STIB 247 hybrid clones.

Limiting Dilution *in vitro* cloning was carried out on uncloned populations from several of the STIB 386 x STIB 247 crosses in order to identify F1 progeny clones. Bloodstraws containing uncloned populations (described in section 2.5.3) were amplified in mice. Infected blood was then collected and adapted to bloodstream culture (as described in section 2.7.3). *In vitro* cloning was performed at day 0, day 2 or day 7 of adapting uncloned trypanosome populations from blood into *in vitro* bloodstream culture by limiting dilution. Trypanosomes were counted by haemocytometer and the concentration adjusted to a range of dilutions (5000-0.5 cells/ml) in HMI-9 supplemented with 20% (v/v) serum plus. Each dilution was then seeding into a number of 96-well uncoated flat bottom plates at 200 µl/well to produce a final concentration of 1000 to 0.1 cells/well. Plates were incubated at 37°c in a humidified 5% CO_2 incubator without monitoring for ~7 days and then observed daily by microscopy until wells containing trypanosomes were seen.

Based on Poisson distribution predictions, 96 well plates in which trypanosomes populations grew in > 40% wells were considered likely to have an unacceptable number of non-clonal populations and the whole plate was discarded. Positive wells derived from plates in which < 40% well grew up were analysed by PCR microsatellite analysis from quick boil culture lysates (section 2.6.2.3) to identify hybrid clones. Trypanosomes from wells identified as containing progeny clones were amplified by seeding cells 1:50 into a T-25 flask containing 5 ml HMI-9 20% (v/v) serum plus culture media. Flasks were incubated at 37°C in a humidified 5%

CO_2 incubator until culture reached a cell density of 10^5-10^6 cells/ml. One ml of each flask culture was then used to prepare DNA by crude lysate (section 2.6.2.2) for genotype analysis and 500 µl stabilated for long term storage (section 2.7.5).

2.7.8 Procyclic form culture growth assays

The analysis of growth in procyclic form cultures was performed under standardised conditions.

Cell cultures were initialled at a starting cell density of 5 x 10^5 cells/ml in SDM-79 15% (v/v) FCS at a culture volume of 5 ml. Cultures were maintained in T-25 phenolic lidded flasks in a 27°C incubator. A 10 µl aliquot was removed from the culture every 24 hours and cell density determined by haemocytometer counts using an improved Neubrauer haemocytometer from day 0 until cells reached stationary phase (approximately five days). For cell counts above ~ 5 x 10^6 cell/ml, cells were diluted between $^1/_5$ - $^1/_{20}$ in SDM-79 before counting and the cell densities extrapolated accordingly.

All procyclic culture growth assays were repeated at least three times and from cell cultures originating from at least two different stabilates to control for culture and stabilate condition variables. To maintain consistency, the same incubator and batch of SDM-79, FBS, and flasks were used for all assays.

2.7.9 *In vitro* bloodstream form culture drug sensitivity assays

Bloodstream culture trypanosomes were isolated at a log phase of growth, the cell density determined by haemocytometer and diluted in HMI-9, 20% (v/v) Serum plus to a density of 4 x 10^5 cells/ml. A 1 mM stock solution of pentamidine was diluted in culture media to make a 1 µM stock. Two-fold serial dilutions were then made from this stock to generate a series of 11 different dilutions from 1 µM – 1 nM. One hundred microlitres of the trypanosome culture and 100 µl of each drug dilution were then combined in wells of a 96 well microtitre plate, to give a starting density of 0.5 x 10^5 cells/ml and a pentamidine concentration range of 500 µM – 0.5 nM. Control wells were included under the same conditions but supplemented with 100 µl HMI-9, 20% (v/v) Serum Plus in place of the drug. All drug dilutions and controls were set up in triplicate. Cultures were incubated for 48 hours in a 37°C incubator with 5% CO_2. After 48 hours a 10 µl aliquot of culture

media was removed from each well and the number of live cells counted by haemocytometer for each concentration of drug used. The average growth of trypanosomes at each drug concentration, for each strain, was calculated from the triplicate average as a percentage of the no drug control. All assays were repeated at least twice.

2.8 *T.brucei* infections in mice

2.8.1 Routine growth of *T.brucei* in mice

Routine growth of *T.brucei* bloodstream form trypanosomes were carried out in adult female ICR mice. Immunosuppression of mice was conducted 24 hours prior to inoculation using cyclophosphamide at a dose of 150 mg/kg by intraperitoneal (i.p) inoculation. Blood stage trypanosomes were prepared by placing either a drop of fresh infected blood or a cryopreserved bloodstraw into a 1.5 ml microtube containing 200 µl CBSS and carefully extraction into a syringe. Bloodstream culture samples were prepared by isolating 200 µl log phase bloodstream culture trypanosomes in HMI-9 with 20% (v/v) serum plus.

The tail blood was monitored three times weekly using the rapid matching technique (Herbert and Lumsden, 1976) for the presence of trypanosomes and estimation of parasitemia. Parasites were harvested at high parasitemia by cardiac puncture for DNA extraction, cryopreservation and/or adaptation to *in vitro* culture.

2.8.2 Preparation of blood stabilates

To prepare stabilates of *T.brucei* parasites cultured in mice, 100 µl of infected blood at > 10^5 cells/ml was mixed with an equal volume of bloodstraw cryopreservative solution. This was then carefully injected into a length of 2 mm Portex tubing using a syringe fitted with a 23 gauge blunt needle. The tubing was cut into 1 cm length pieces, placed into a cryotube with an assigned stabilate number and frozen horizontally at –80°C in a cottonwool insulated box. After 24 hours the cryovial was transferred to liquid nitrogen for long term storage.

2.8.3 *In vivo* Minimum Curative Dose (MCD) assays

To assess the *in vivo* drug sensitivity of wild type parental stocks STIB 247, STIB 386 and TREU 927, an *in vivo* minimum curative dose experiment was performed for four trypanocidal drugs; suramin, pentamidine, Berenil and cymelarsen. Drugs were prepared in sterile ddH_20 from fresh or frozen stocks as detailed in section 2.3.

Female ICR mice, weighing ~20-30 g were used for the experiment adapted from a method previously described (Turner *et al.*, 1988). Groups of three mice were inoculated with 10^5 bloodstream tryps by i.p inoculation and drug treatment administered 24 hours later. Mice were weighed and a range of drug concentrations administered by i.p injection at the appropriate mg/kg of mouse weight ratio in a total volume of 200 µl. A further group of two mice were included in each experiment for which trypanosomes were inoculated but no drug therapy administered to serve as positive controls for trypanosome infectivity. The tail blood of the mice was examined three times per week for a total of 21 days after the administration of drug, using the rapid matching technique. Mice were considered insensitive to the administered drug concentration if a patent infection ($> 2.5 \times 10^5$ trypanosomes/ml) was detected in tail blood using the rapid matching technique. Single bloodstream trypanosomes have been demonstrated to bring about patent infections within 14 days (Turner *et al.*, 1988), therefore if no trypanosomes were detected in tail blood after 21 days the mice were considered cured. The lowest concentration of each drug that prevented a patent infection from occurring was considered the curative dose and use to define the sensitivity of each trypanosome strain.

In vivo Minimum Curative Dose assays were performed with the assistance of Dr Caroline Clucas.

Chapter 3

The construction of a genetic map for *Trypanosoma brucei gambiense* and evaluation of drug sensitivity phenotypes for linkage analysis

3.1 Introduction

Genetic maps, as outlined in Chapter 1, can be used to establish the order, location, and relative distances of genetic markers in organisms that undergo sexual recombination, as well as defining some of the basic features of recombination. Their most important application however, is in combination with inheritance data of phenotypes, to identify genetic loci that determine traits that differ between individuals through linkage analysis. The importance of genetic mapping, coupled with positional cloning, is particularly high when analysing both simple and complex phenotypes for which there are no obvious candidate genes, and provides a complementary tool to reverse genetics for analysing gene function.

Genetic maps have been generated for a number of haploid eukaryotic pathogens including *P.falciparum* (Su *et al.*, 1999), *P.c.chabaudi* (Martinelli *et al.*, 2005a), *T.gondii* (Khan *et al.*, 2005) and *E.tenella* (Shirley and Harvey, 2000). The genetic linkage approach, using such maps, has been an important tool for mapping genes which are responsible for traits such as drug resistance (Su *et al.*, 1997; Ferdig *et al.*, 2004), virulence (Saeij *et al.*, 2006; Taylor *et al.*, 2006) and strain-specific immunity (Martinelli *et al.*, 2005b). An important feature of the maps of all these organisms is that the physical size of the recombination unit is relatively small, ranging from 17 Kb per centiMorgan in the case of *P.falciparum* (Su *et al.*, 1999) to 100–215 Kb in the case of *E.tenella* and *T.gondii* (Sibley *et al.*, 1992; Shirley and Harvey, 2000; Khan *et al.*, 2005). This means that the analysis of relatively few progeny can provide high mapping resolution, in contrast to higher eukaryotes where the physical size of the recombination unit is usually considerably larger (NCBI, 1999).

Use of this approach for identifying loci linked to traits of interest in diploid pathogens has been more limited. This is because there is either no evidence for a system of genetic exchange, a crucial requirement for the application of this approach, or the basic parameters of genetic exchange have not been fully defined. In *T.brucei*, genetic exchange has been demonstrated (Jenni *et al.*, 1986) and several experimental crosses have been performed both between and within subspecies over the last 20 years (Gibson and Stevens, 1999). This includes the crossing of two *T.b.brucei* and a *T.b.gambiense* strain in all pair-wise combinations (Turner *et al.*, 1990) from which the products of mating have been defined as the equivalent of F1 progeny (MacLeod *et al.*, 2005a).

The strains used in these crosses (STIB 247, STIB 386 and TREU 927), were isolated from different regions of Africa and different hosts species (Chapter 1, Table 1.2). They also differ in a range of important phenotypes (Tait *et al.*, 2002), allowing the genetic basis of these differences to be analysed. A project to sequence the megabase chromosomes of *T.brucei* has resulted in the availability of the genome sequence for one of the *T.b.brucei* isolates used in these crosses, TREU 927 (Berriman *et al.*, 2005). This has been utilised by our laboratory to generate a genetic map for TREU 927 (MacLeod *et al.*, 2005b) which is now being employed to identify candidate loci that contribute to heritable phenotypes of interest in this strain, such as pathogenicity (Morrison *et al.*, 2009b).

It is the *T.b.gambiense* sub-species however, which is responsible for the majority of current HAT infections in sub-Saharan Africa (WHO, 2006a; Simarro *et al.*, 2008), and although related to *T.b.brucei*, it differs in its ability to infect human hosts as well as several other important characteristics. A separate *T.b.gambiense* genetic map is therefore desirable for the study of specific mechanisms of disease in this pathogenic sub-species. For this reason, the strain STIB 386 is of particular interest, as it was isolated from a human in West Africa and is defined as a group two *T.b.gambiense* (Gibson, 1986). A genetic cross has been performed between STIB 386 and STIB 247, a *T.b.brucei* strain, resulted in 38 unique progeny (Tait *et al.*, 2002; MacLeod *et al.*, 2005a).

The parental stocks from this cross have been shown to differ phenotypically in a number of important traits related to the transmission, pathogenicity and host range of the parasite, thus allowing their genetic basis to be investigated using the progeny from this cross. These key phenotypic differences are outlined in Table 3.1 and include; the ability to resist lysis by human serum, a measure of the virulence of the strains defined by the parasitemia in a murine disease model and two traits related to the transmission of the parasite, the midgut infectivity and transmission index.

Preliminary data had also indicated a fifth potential phenotypic difference between the two strains, relevant to the treatment of trypanosomiasis that required further investigation. Experimental observations of the curative dose for melarsamine hydrochloride (Cymelarsan), a water soluble analogue of the arsenical melarsoprol, in a murine disease model had suggested an innate difference in sensitivity of approximately 6-7 fold between the two wild type strains (Scott *et al.*, 1996).

Phenotype	STIB 247 *T.b.brucei*	STIB 386 *T.b.gambiense*	Reference
Human Serum Resistance	Sensitive	Resistant	(Lindegard, 1999; Turner *et al.*, 2004)
Virulence	+	+++	(Turner *et al.*, 1995)
Midgut infectivity	+++	+	(MacLeod, 1999)
Transmission Index	+++	+	(Welburn *et al.*, 1995)
Cymelarsan sensitivity	MCD 0.3 mg/Kg	MCD 2 mg/Kg	(Scott *et al.*, 1996)

Table 3.1 Differences in phenotype between the two parental stocks used to generate the STIB 386 genetic map.

Relative differences in phenotype between STIB 247 and STIB 386 are indicated by + and +++. MCD, Minimum Curative Dose. Table based on Tait *et al* (2002).

Without the prospect of a vaccine or comprehensive vector control strategy, the control of both human and livestock trypanosome infection relies predominantly on chemotherapy and chemoprophylaxis. However, in addition to problems of toxicity and administration, the emergence of resistance and even cross-resistance to the small number of trypanocidal drugs currently available continues to threaten the effective management of both human and animal disease.

Resistance to all classes of trypanocides including the melaminophenyl arsenicals (Pepin *et al.*, 1994; Brun *et al.*, 2001) and diamidines (Berger *et al.*, 1995; Bridges *et al.*, 2007) as well as suramin (Scott *et al.*, 1996), DFMO (Phillips and Wang, 1987), isometamidium (Geerts *et al.*, 2001) and homidium (Na'Isa, 1967) has been reported either in the field or in laboratory generated lines. More worryingly, cross-resistance between the melaminophenyl arsenicals, which includes melarsoprol and Cymelarsan and the diamidines, which includes both the HAT drug pentamidine and veterinary drug Berenil has been observed in several of these drug resistance lines (Scott *et al.*, 1996; Bridges *et al.*, 2007).

As detailed in Chapter 1, some progress has been made in identifying possible mechanisms of resistance in *T.brucei*, including examples of both reduced drug uptake (Maser *et al.*, 1999) and increased drug efflux (Shahi *et al.*, 2002). In particular, recent advances have been made in our understanding of the role played by the P2/*TbAT1* and HAPT1 transporters of melaminophenyl arsenicals and diamidines, the functional loss of one or both of which can cause varying levels of resistance and cross resistance to drugs of both classes (Matovu *et al.*, 2003; de

Koning *et al.*, 2004; Bridges *et al.*, 2007). However we still lack a full understanding of both the mode of action of these drug and the key determinants of resistance and most of our understanding of drug resistance in trypanosomes so far, has been derived from laboratory-selected lines. The potential difference in drug sensitivity between STIB 247 and STIB 386 is therefore of particular relevance in this context, as a naturally occurring variation which might reflect more realistically what is occurring in the field. This potential innate difference was therefore of great interest as a tool that could be used to help identify additional genes important for drug resistance using a genetic mapping approach.

This chapter describes two of the experimental approaches performed as an important step towards using the STIB 247 x STIB 386 cross to map genes determining traits of importance in the human-infective sub-species of *T.brucei*. Firstly, the construction of a genetic map for the STIB 386 strain of *T.b.gambiense* for linkage analysis of existing well characterised phenotypic differences, alongside a comparative analysis with the genetic map of *T.b.brucei* strain TREU 927. The second approach was an attempt to validate the Cymelarsan phenotype in the parental lines. Analysis of the sensitivity of the parental strains to Cymelarsan and other trypanocidal drugs was performed in order to assess the validity of attempting a linkage analysis approach to identify loci contributing to this phenotype and aid understanding of possible mechanisms of resistance.

3.2 Results

3.2.1 Identification of heterozygous markers and the genotyping of F1 progeny

The *T.brucei* genome sequence of strain TREU 927 had previously been screened using the Tandem Repeat Finder program (Benson, 1999) to identify microsatellites, which were evenly distributed across the genome. Eight hundred and ten pairs of primers were designed to the unique sequence flanking each microsatellite locus (MacLeod *et al.*, 2005b). In this project these primers were used to amplify by PCR the microsatellites from the two parental stocks, STIB 386 and STIB 247, to identify markers that were heterozygous for STIB 386. Heterozygous markers were defined by the amplification of two different sized PCR products in STIB 386, which could be easily separated and visualised by gel electrophoresis. An example of one such microsatellite is given in Figure 3.1, which demonstrates the amplification of alleles from both parental strains alongside the segregation of parental alleles in the progeny clones.

Ninety-nine potentially informative markers were identified by this method that could be used for the construction of a partial genetic map, while the remaining 711 markers either amplified a homozygous band in STIB 386 or failed to amplify any PCR product. Forty-seven of these 99 heterozygous markers had also been previously found to be heterozygous for TREU 927 and so were included in the construction of both the *T.b.brucei* and *T.b.gambiense* genetic maps.

Following this initial microsatellite screen, further markers were sought to fill in regions of the genome that were not covered by a heterozygous marker for STIB 386. An additional 215 primer pairs were designed to screen further microsatellites from these regions, resulting in the identification of an additional 20 heterozygous markers and a total marker coverage of 119 heterozygous markers. Overall the level of heterozygosity for all the markers screened is significantly lower at 12.5%, than the value of 20% reported for the genome strain (χ^2=27.3, d.f=1, P< 0.01) (MacLeod *et al.*, 2005b). Thirty-eight F1 progeny clones from the cross between STIB 386 and STIB 247 (Chapter 2, Table 2.1) were genotyped with the 119 markers and the segregation patterns in the progeny were scored to generate a full genotype of each progeny clone.

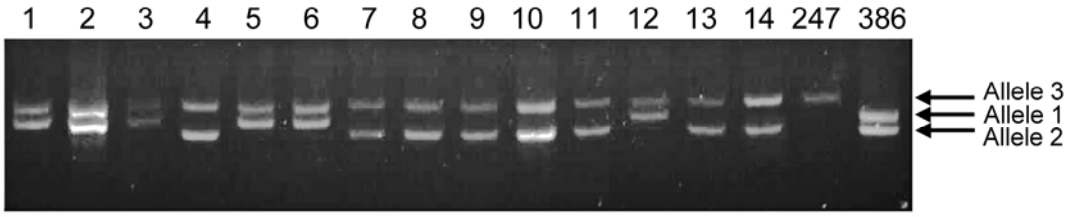


Figure 3.1 A representative microsatellite marker used for mapping analysis.

PCR amplification of the microsatellite Ch6/10 in parental lines STIB 247 and STIB 386 and a selection of progeny clones using primers Ch6/28P18/4-A and B. PCR products were run on a 3% Nusieve agarose gel. The heterozygous alleles in parental lines STIB 386 are identified on the gel as allele 1 and 2 and the homozygous alleles in STIB 247 as allele 3. Progeny clones were analysed for the inheritance of STIB 386 alleles and this segregation data used to generate the genetic linkage map.

Tables detailing the primer sequences and genomic location for all microsatellite markers screened during the development of the genetic map are given in Appendix 2. The segregation data of STIB 386 alleles in the progeny for these markers are given in Appendix 4.

3.2.2 Construction of the STIB 386 genetic linkage map

The inheritance pattern of STIB 386 alleles, at each heterozygous locus, in the 38 F1 progeny was determined, and the segregation data used to construct a genetic map using the Map Manager QTX program (Manly *et al.*, 2001). This linked the 119 markers into 12 linkage groups, which correspond to the 11 housekeeping chromosomes. The genetic linkage map of each chromosome is shown in Figure 3.2, and while ten chromosomes (1, 2, 3, 4, 5, 6, 7, 8, 9, and 11) consist of one linkage group each, chromosome 10 currently comprises two groups. The main characteristics of the linkage groups obtained are summarized in Table 3.2. The genetic distances, based on the number of recombination units between each marker, are expressed in centiMorgans (cM), which added together for all 12 linkage groups gave a total genetic map length of 733.1 cM. The size of each chromosome and the physical distances between markers were based on the TREU 927 *T.b.brucei* sequence (Berriman *et al.*, 2005). Using these figures the genetic map covers 17.9 Mb, which equates to approximate genome coverage of 70%. However, this calculation includes the gene poor subtelomeric regions, which the genetic map does not extend into due to the difficulties in identifying unique sequences in these regions.

On average, the crossover frequency was found to be 0.6 crossovers/per chromosome/per individual progeny clone in the mapped population (Table 3.2) and the average recombination unit size is 24.4 Kb/cM. This provides a 9 cM resolution genetic map with a 90% probability of mapping any locus to within 11 cM (268 Kb). The physical position of each microsatellite marker, based on the genome sequence of *T.b.brucei* (Berriman *et al.*, 2005), allows us to compare the position of markers in the physical map of *T.b.brucei* and the genetic map of *T.b.gambiense*, revealing that synteny is conserved for all markers on all chromosomes.

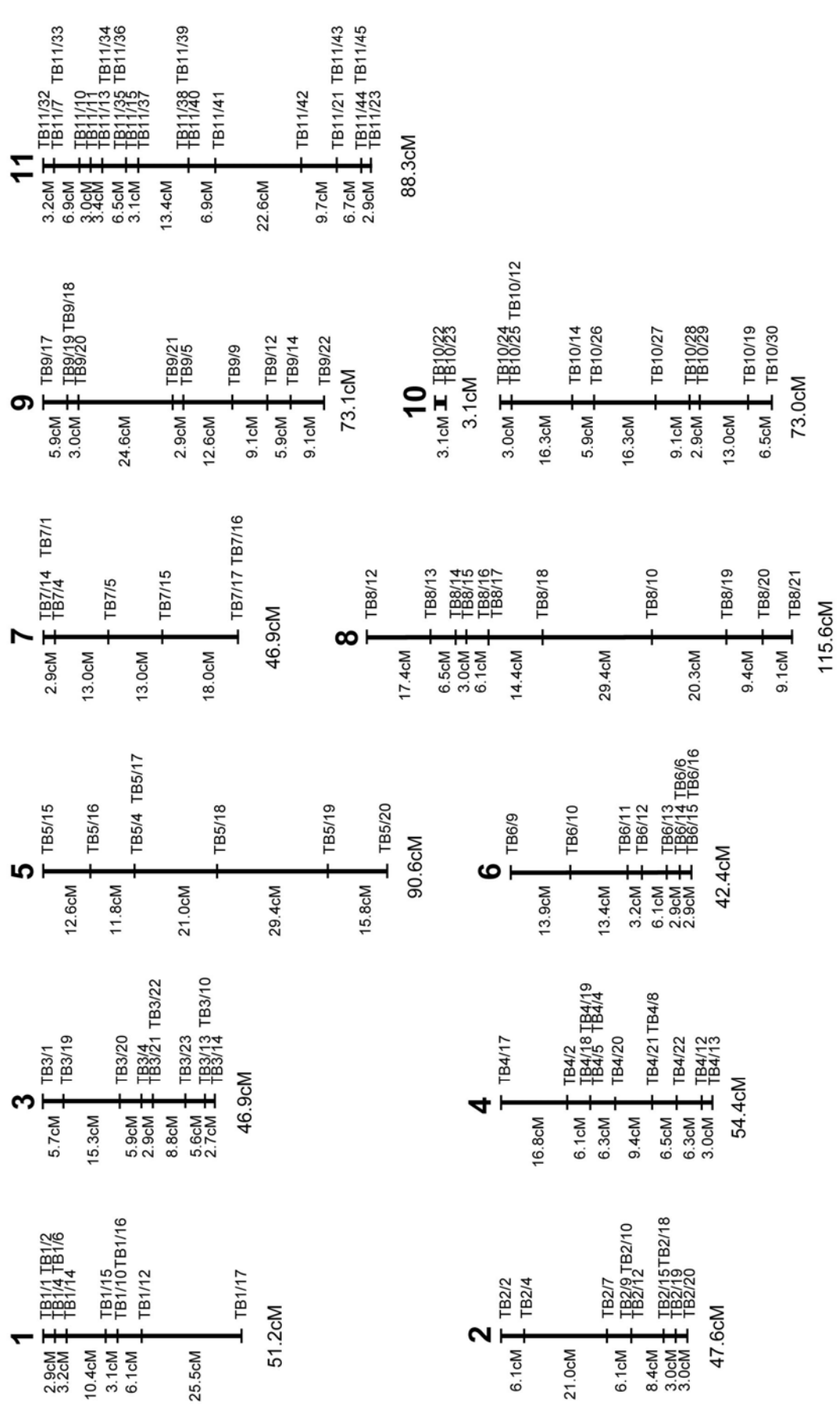

1
2.9cM TB1/1 TB1/2
3.2cM TB1/4 TB1/6
TB1/14
10.4cM TB1/15
3.1cM TB1/10 TB1/16
6.1cM TB1/12
25.5cM TB1/17
51.2cM
3
5.7cM TB3/1
TB3/19
15.3cM TB3/20
5.9cM TB3/4
2.9cM TB3/21 TB3/22
8.8cM TB3/23
5.6cM TB3/13 TB3/10
2.7cM TB3/14
46.9cM
5
TB5/15
12.6cM TB5/16
11.8cM TB5/4 TB5/17
21.0cM TB5/18
29.4cM TB5/19
15.8cM TB5/20
90.6cM
7
2.9cM TB7/14 TB7/1
TB7/4
13.0cM TB7/5
13.0cM TB7/15
18.0cM TB7/17 TB7/16
46.9cM
8
TB8/12
17.4cM TB8/13
6.5cM TB8/14
3.0cM TB8/15
6.1cM TB8/16
TB8/17
14.4cM TB8/18
29.4cM TB8/10
20.3cM TB8/19
9.4cM TB8/20
9.1cM TB8/21
115.6cM
9
5.9cM TB9/17
3.0cM TB9/19 TB9/18
TB9/20
24.6cM TB9/21
2.9cM TB9/5
12.6cM TB9/9
9.1cM TB9/12
5.9cM TB9/14
9.1cM TB9/22
73.1cM
10
3.1cM TB10/22
TB10/23
3.1cM
3.0cM TB10/24
TB10/25 TB10/12
16.3cM TB10/14
5.9cM TB10/26
16.3cM TB10/27
9.1cM TB10/28
2.9cM TB10/29
13.0cM TB10/19
6.5cM TB10/30
73.0cM
11
3.2cM TB11/32
TB11/7 TB11/33
6.9cM TB11/10
3.0cM TB11/11
3.4cM TB11/13 TB11/34
6.5cM TB11/35 TB11/36
3.1cM TB11/15
TB11/37
13.4cM TB11/38 TB11/39
TB11/40
6.9cM TB11/41
22.6cM TB11/42
9.7cM TB11/21 TB11/43
6.7cM TB11/44 TB11/45
2.9cM TB11/23
88.3cM
2
6.1cM TB2/2
TB2/4
21.0cM TB2/7
6.1cM TB2/9 TB2/10
TB2/12
8.4cM TB2/15 TB2/18
3.0cM TB2/19
3.0cM TB2/20
47.6cM
4
TB4/17
16.8cM TB4/2
6.1cM TB4/18 TB4/19
6.3cM TB4/5 TB4/4
9.4cM TB4/20
6.5cM TB4/21 TB4/8
6.3cM TB4/22
3.0cM TB4/12
TB4/13
54.4cM
6
TB6/9
13.9cM TB6/10
13.4cM TB6/11
3.2cM TB6/12
6.1cM TB6/13
2.9cM TB6/14 TB6/6
2.9cM TB6/15 TB6/16
42.4cM

Figure 3.2 Genetic linkage maps corresponding to the 11 megabase chromosomes of *T.b.gambiense*.

Every microsatellite marker, shown to the right of each linkage group, has been anchored to the physical map and the physical location, derived from the *T.b.brucei* genome sequence (Berriman *et al.*, 2005) is identified in the supplementary data (Appendix 2). The corresponding genetic distance between intervals is shown in centiMorgans on the left of each map and the total genetic size of each linkage group given below. Maps were constructed from the segregation analysis of 38 progeny with 119 microsatellite markers using the program MapManager, with a Haldane map function, at the highest level of significance for linkage criteria, with a probability of type 1 error P=1e-6. All markers were mapped with a minimum LOD score of 5.5.

Chromosome	Number of markers	Genetic length (cM)[a]	Physical size (Mb)[b]	Recombination Frequency (Kb/cM)	Average number of crossover events/meiosis[d]
1	10	51.20	0.74	14.53	0.46
2	10	47.60	0.74	15.46	0.42
3	10	46.90	1.25	26.74	0.42
4	12	54.40	1.05	19.30	0.50
5	7	90.60	1.20	13.29	0.74
6	9	42.40	0.94	22.13	0.35
7	7	46.90	1.65	35.08	0.40
8	11	115.60	2.30	19.88	0.95
9	10	73.10	2.10	28.67	0.65
10[c]	12	76.10	2.50	32.85	1.08
11	21	88.30	3.42	38.76	0.71
Average				24.40	0.61
Total	119	733.10	17.89		

Table 3.2 Characteristics of the genetic linkage maps of *T.b.gambiense.*

[a] Total genetic length was calculated by the addition of recombination units between each marker as determined by mapmanager. [b] Physical distances were calculated from the *T.b.brucei* genome sequence (Berriman *et al.*, 2005). [c] Chromosome 10 is a combination of two linkage groups. [d] The statistics regarding the numbers of crossovers of all chromosomes in all hybrids has been compiled and in reported in Appendix 5.

3.2.3 Marker segregation proportions

The availability of segregation data across the length of each chromosome allows a full analysis of the inheritance of the STIB 386 parental chromosome homologues. The ratio of segregation of alleles for each heterozygous marker was calculated along each chromosome with the 95% confidence limits of a 1:1 segregation with 38 F1 progeny. This analysis had been previously conducted for the STIB 386 map of one of the smallest chromosomes, chromosome 1, and detected a region of significant distortion across the left arm of the chromosome (MacLeod *et al.*, 2005a). Segregation analysis has now been performed on the remaining 10 chromosomes (Figure 3.3) and this shows no evidence of distortion from a 1:1 segregation ratio across the length of chromosomes 4, 8, 9, or 10. On chromosomes 2, 5, 6, 7 and 11 there is one marker per chromosome and on chromosome 3 there are two markers that have been inherited at proportions just outside the 95% confidence limits. However, it should be considered that this totals only 7/109 markers analysed (6%), which is close to the 5% of outliers that would be expected with 95% confidence intervals and thus are unlikely to signify regions of true segregation distortion. Therefore, the previously reported region of chromosome 1 (MacLeod *et al.*, 2005a)remains the only region of the STIB 386 genetic map with evidence of any significant segregation distortion. The origin of this distortion is not known but one possibility could be the result of post meiotic selection acting on the uncloned progeny during growth in mice prior to isolation of clones.

3.2.4 Variation in recombination between chromosomes

Although the average rate of recombination in the *T.b.gambiense* map was found to be 24.4 Kb/cM, there is variation both between and within the chromosomes, as is common in many other eukaryotic organisms (Petes, 2001). A correlation of the physical and genetic sizes of every chromosome in the map is shown in Figure 3.4 and the average physical size of a recombination unit ranges from a high of 39 Kb/cM on chromosome 11 to a low of 13 Kb/cM on chromosome 5 (Table 3.2). Variation is also evident between specific intervals across chromosomes where a map unit can vary from < 1 Kb/cM up to 170 Kb/cM on the same chromosome (chromosome 11, Figure 3.5k) representing extremes in recombination frequency. If we define hot and cold spots of recombination as three times less (cold) or three times more (hot) than the average recombination rate, the boundaries for

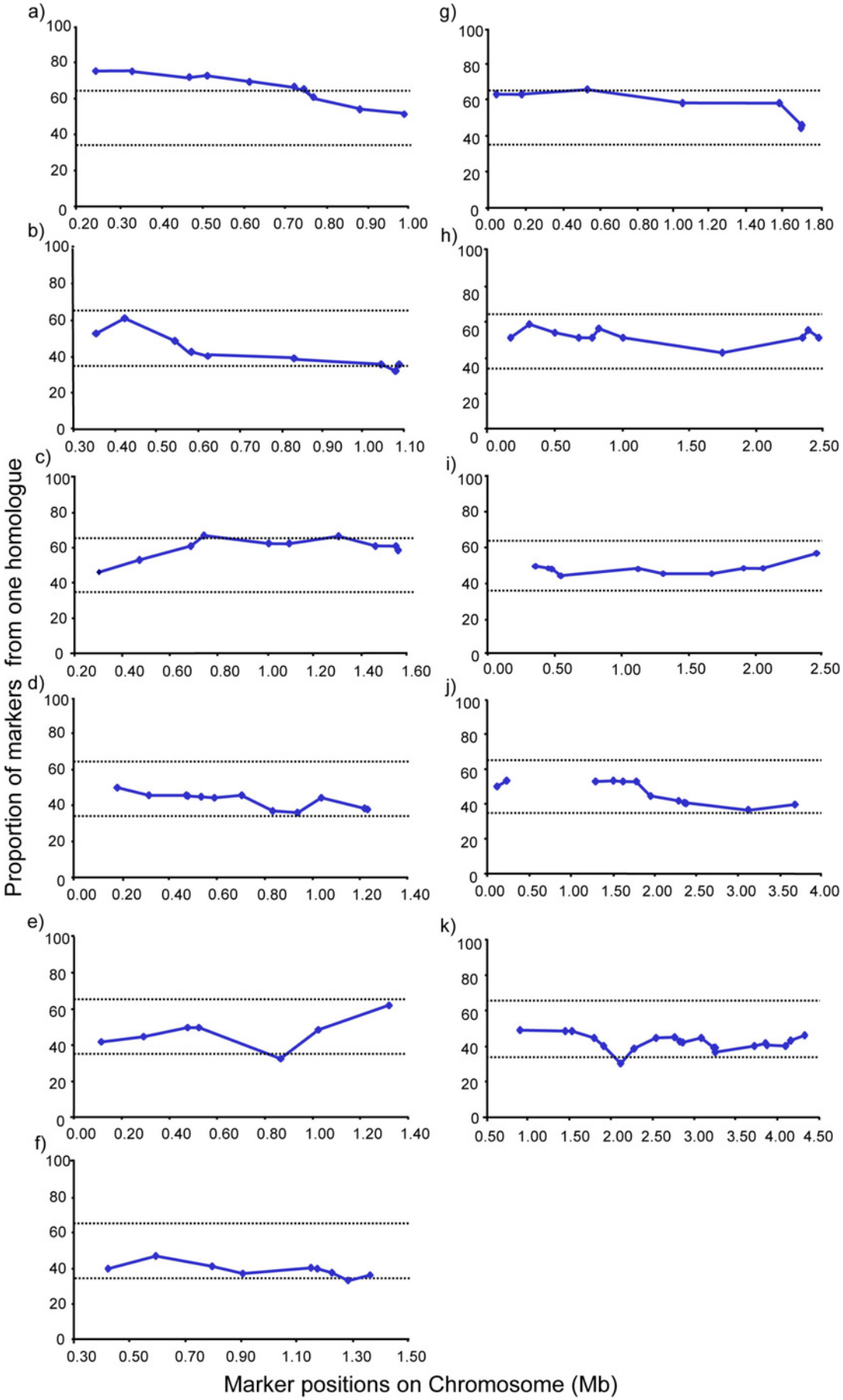

a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
Proportion of markers from one homologue
Marker positions on Chromosome (Mb)

Figure 3.3 Genotype segregation proportions.

Genotype segregation proportions for all microsatellite markers present on chromosomes: 1 (a), 2 (b), 3 (c), 4 (d), 5 (e), 6 (f) 7 (g), 8 (h), 9 (i), 10 (j) and 11 (k). Dashed horizontal lines indicate the approximate 95% probability range for equal segregation of alleles.

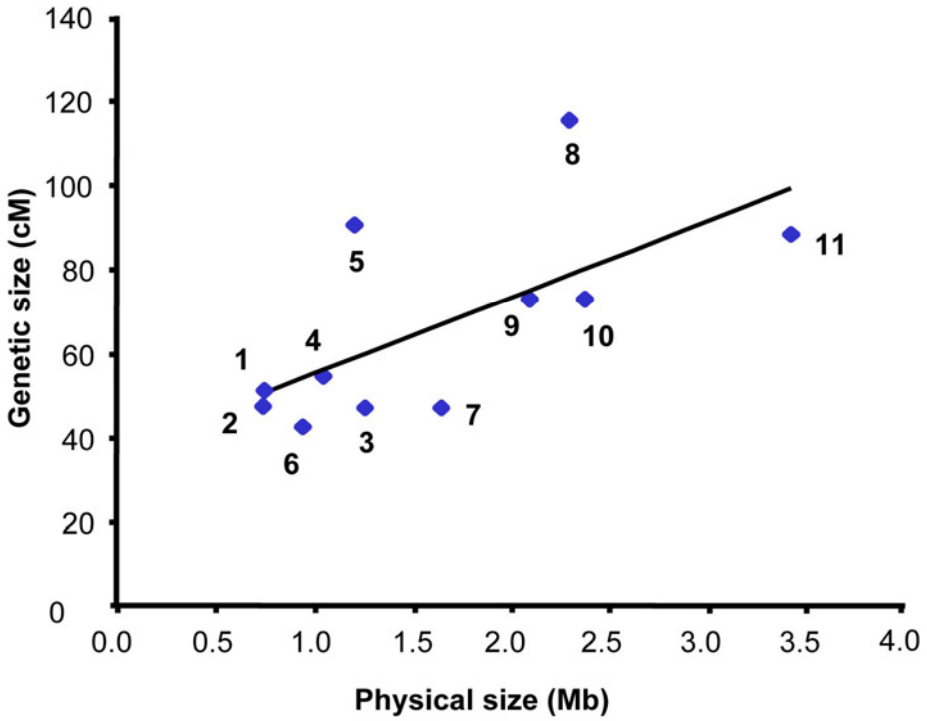


Figure 3.4 The genetic size of each linkage group (cM) relative to its physical size (Mb).

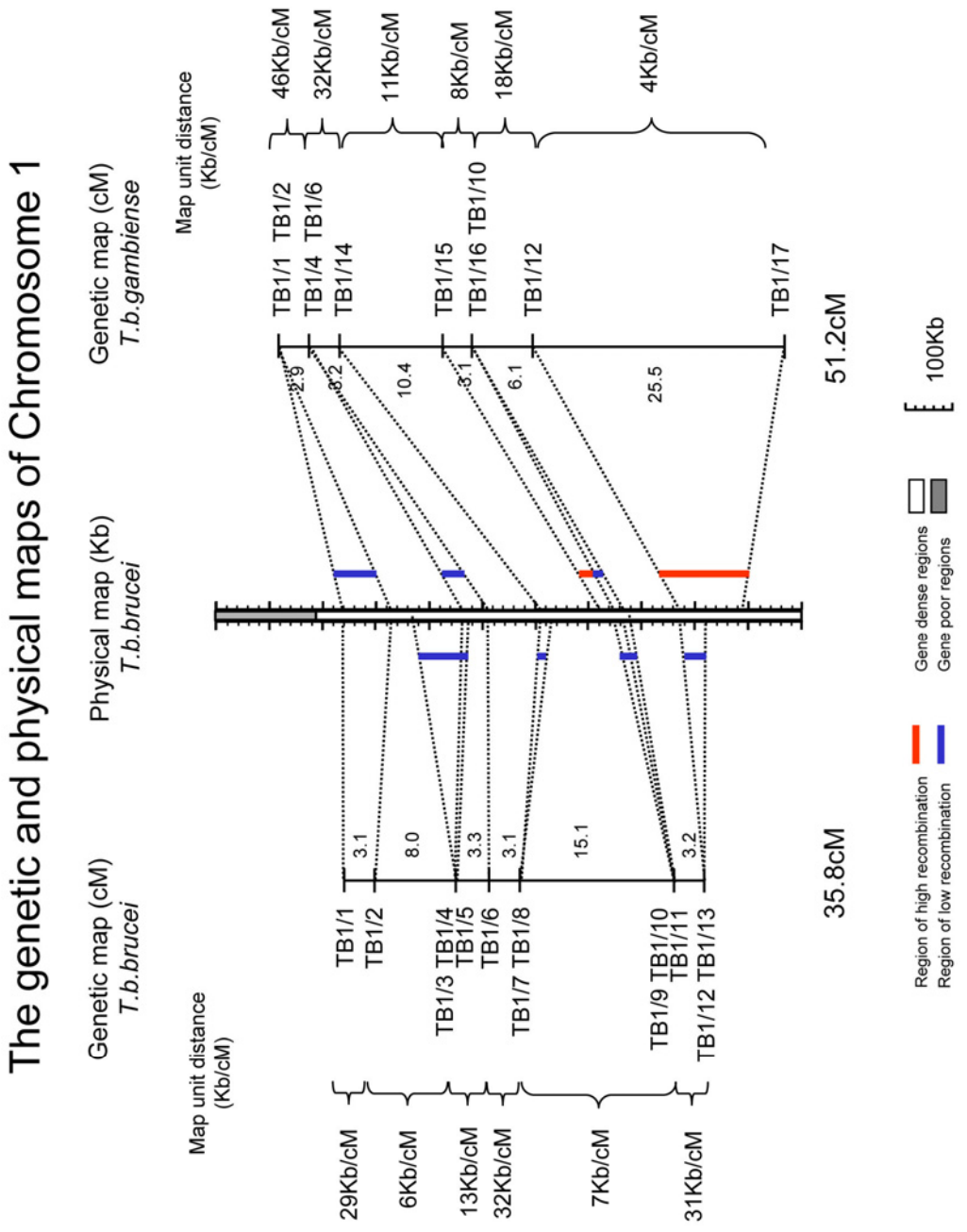
A comparison of the total genetic size of each chromosome genetic map against the predicted physical distance, calculated from the *T.b.brucei* genome sequence (Berriman *et al.*, 2005). The line shown was determined by linear least squares regression analysis (R^2=0.6).

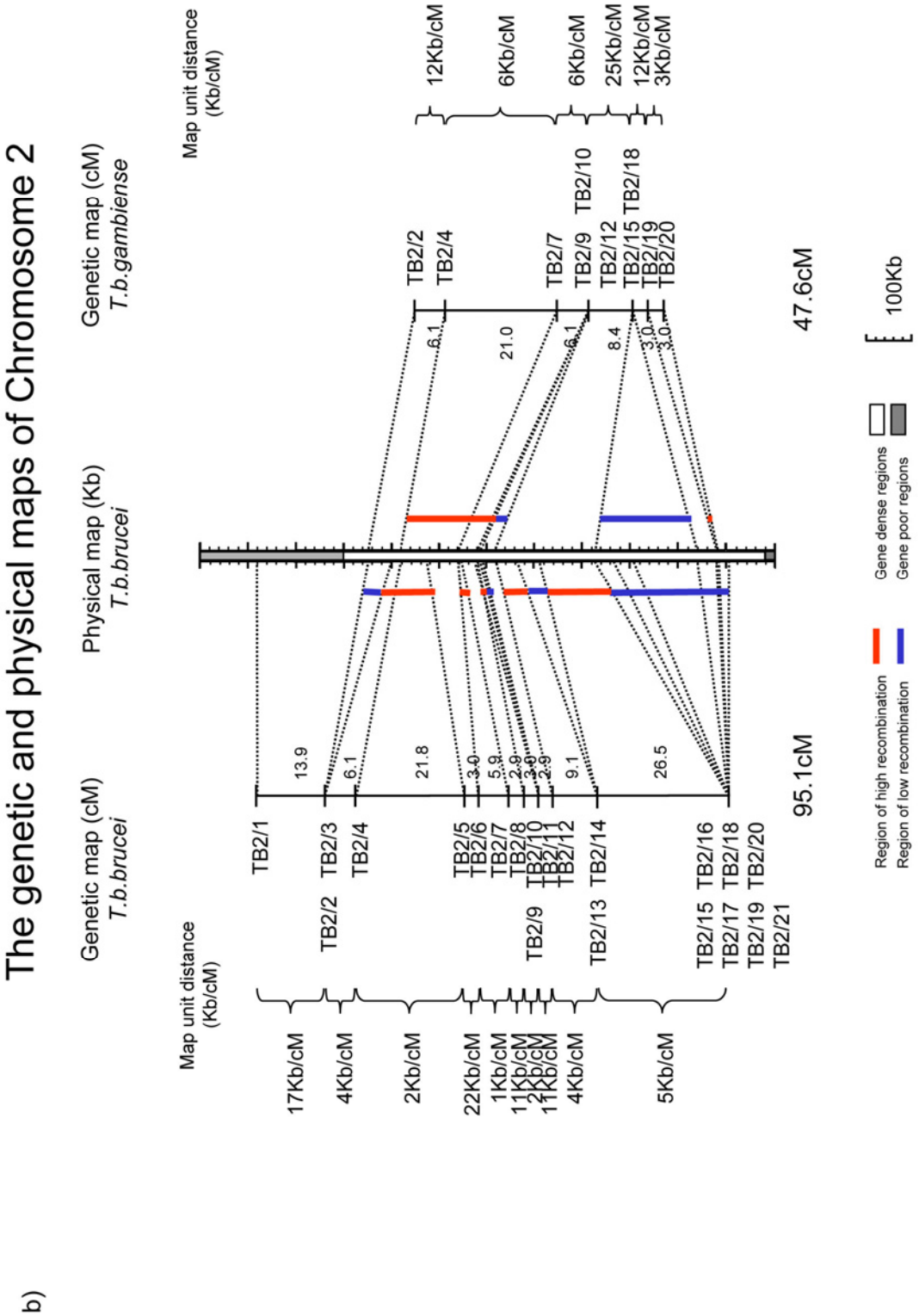
defining hot and cold regions can be set at <8 Kb/cM and >73 Kb/cM respectively, based on an average physical size of a recombination unit of 24 Kb/cM. Analysis of crossovers in the STIB 386 x STIB 247 progeny found that variation in recombination frequency between markers is common, producing a least one hot or cold region on every chromosomes and a total of 15 hot and 27 cold spots overall (Figure 3.5a-k).

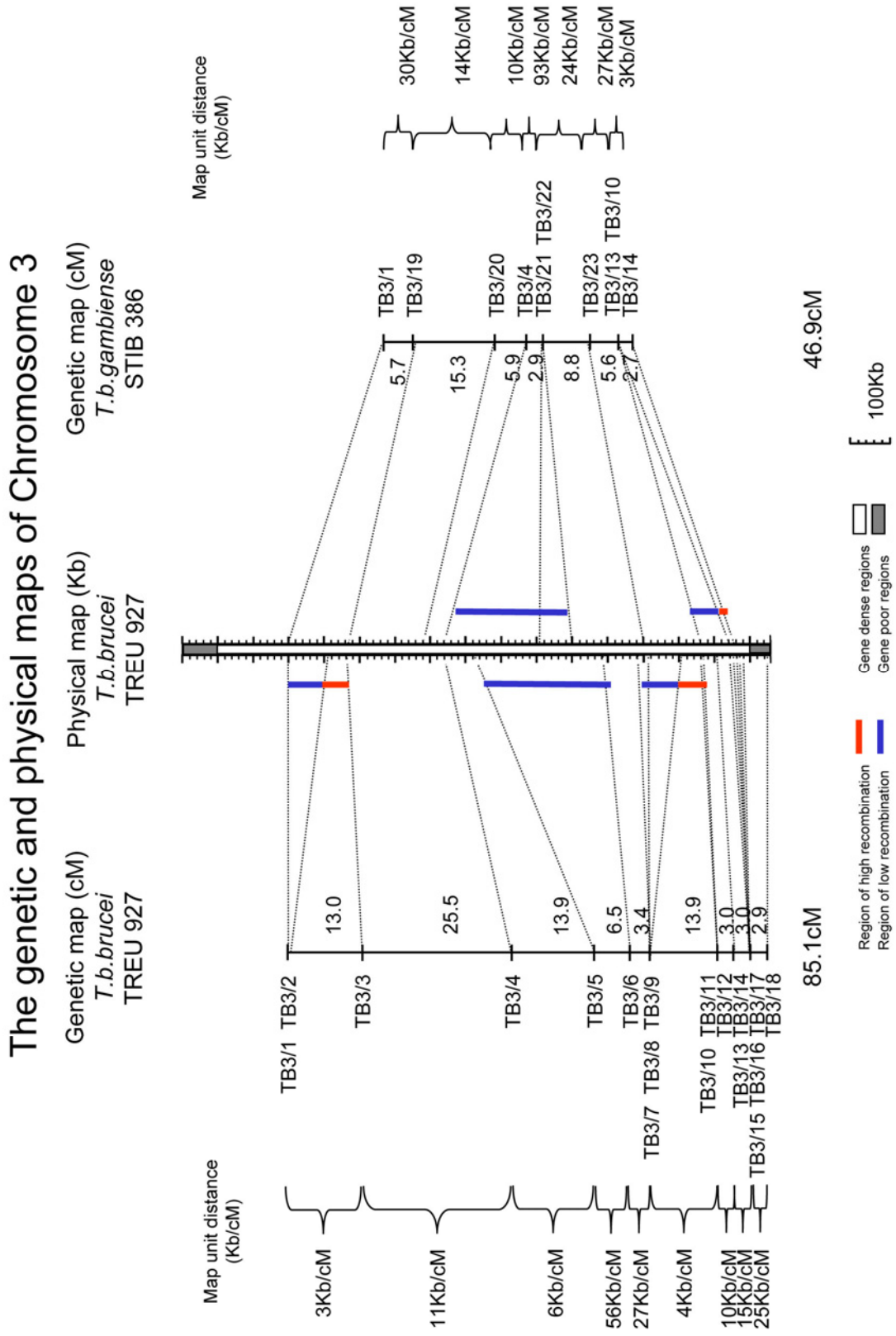
Variation in recombination was also noted as a common feature in the *T.b.brucei* TREU 927 map (MacLeod *et al.*, 2005b). Data from the *T.b.brucei* genetic map was reanalysed alongside the *T.b.gambiense* map to identify regions of high and low recombination using the same definition of boundaries. Based on an average physical recombination unit size of 15.6 Kb/cM for TREU 927, hot and cold spot boundaries could therefore be defined as < 5.2 Kb/cM and > 46.8 Kb/cM respectively. As a result of this analysis, a similar number of hot and cold regions were identified on the TREU 927 map with a total of 20 hot and 32 cold spots overall (Figure 3.5a-k).

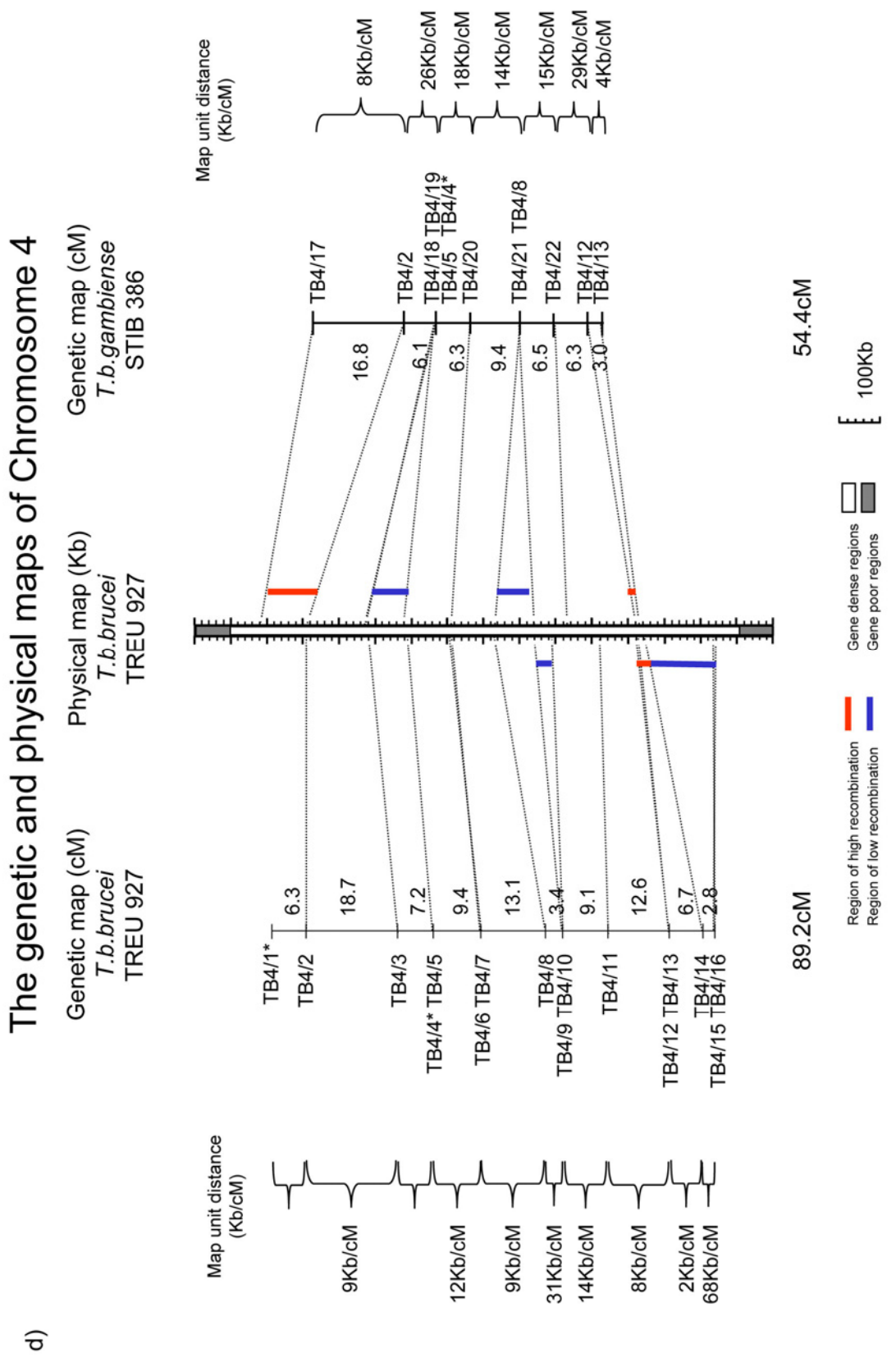
A more detailed comparison of these regions to those identified on STIB 386 was then performed and four areas of high recombination (hot) and 10 of low recombination (cold) were found to overlap the same physical location on both genetic maps. Chromosome 2 for example (Figure 3.5b) has a region of higher recombination towards the centre of the chromosome, (denoted in red), which contains two of the STIB 386 hot spots and four of the TREU 927 hot spots, as well as a large shared cold spot, (denoted in blue), toward the end of the chromosome with no evidence of recombination over a distance of more than 200 Kb on either map. In contrast, there are also several regions, where a STIB 386 hot spot corresponds to a cold spot on TREU 927, as illustrated at the end of Chromosome 1 (Figure 3.5a) and vice versa (for example chromosome 8, Figure 3.5h). Although local variation in crossover frequency appears to be a common feature of both the *T.b.brucei* and *T.b.gambiense* maps, this balances out over the full length of each chromosome with the net result that the total genetic distance of linkage groups is correlated with their physical size (Figure 3.4).

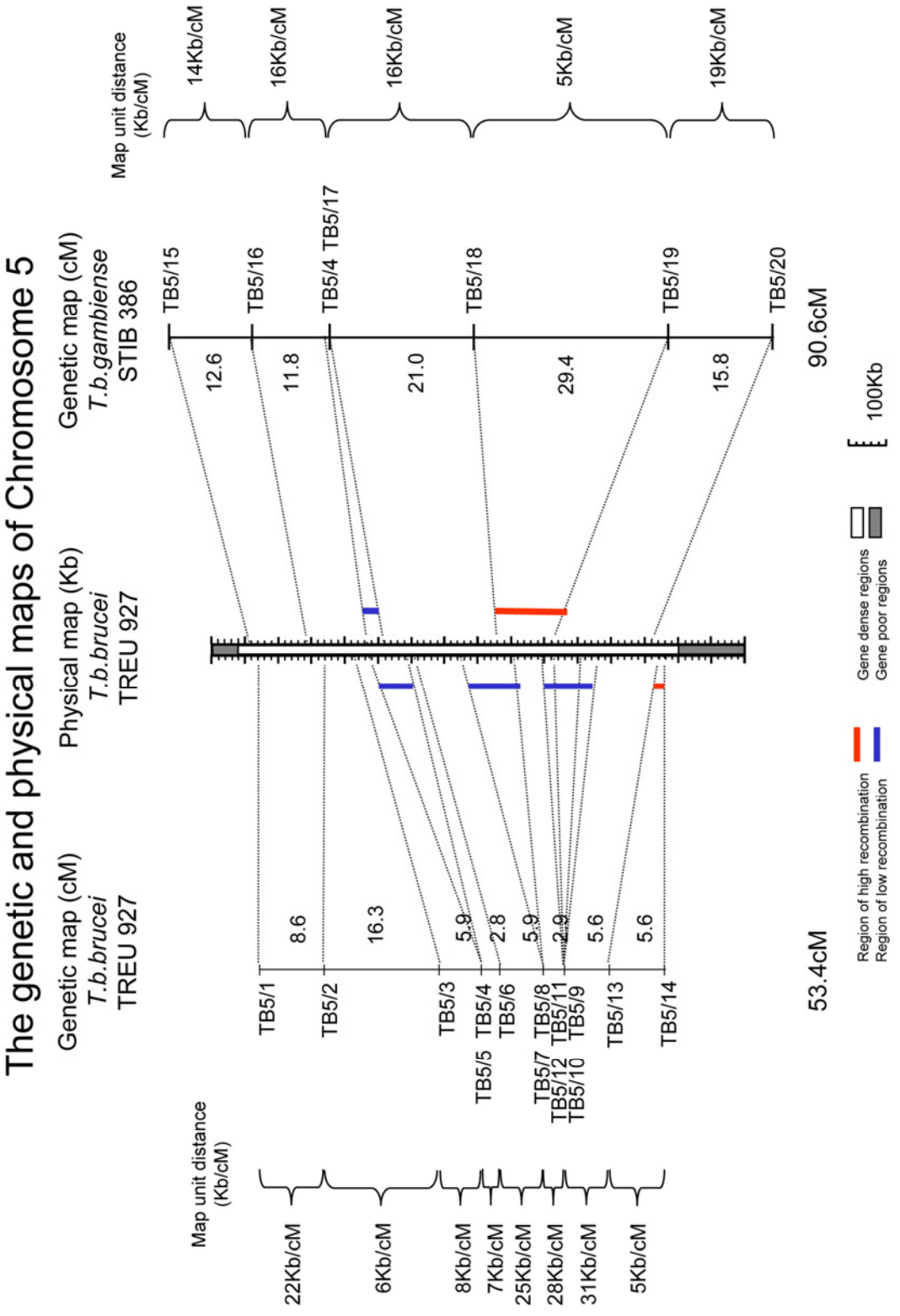
a)

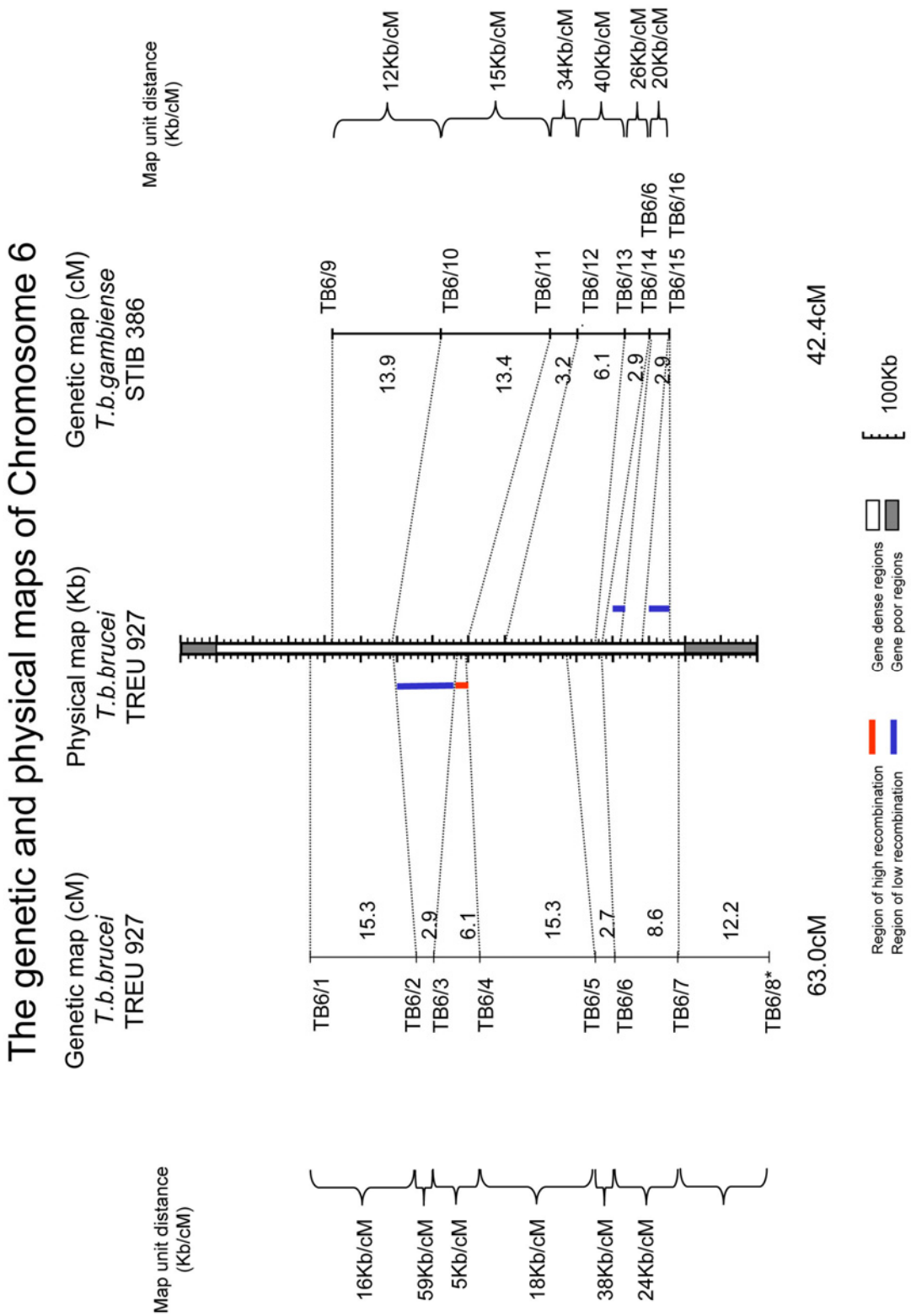

The genetic and physical maps of Chromosome 1
Genetic map (cM)
T.b.brucei
Map unit distance
(Kb/cM)
Physical map (Kb)
T.b.brucei
Genetic map (cM)
T.b.gambiense
Map unit distance
(Kb/cM)
TB1/1
TB1/2
TB1/3 TB1/4
TB1/5
TB1/6
TB1/7 TB1/8
TB1/9 TB1/10
TB1/11
TB1/12 TB1/13
29Kb/cM
6Kb/cM
13Kb/cM
32Kb/cM
7Kb/cM
31Kb/cM
3.1
8.0
3.3
3.1
15.1
3.2
35.8cM
TB1/1 TB1/2
TB1/4 TB1/6
TB1/14
TB1/15
TB1/16 TB1/10
TB1/12
TB1/17
46Kb/cM
32Kb/cM
11Kb/cM
8Kb/cM
18Kb/cM
4Kb/cM
2.9
3.2
10.4
3.1
6.1
25.5
51.2cM
Region of high recombination
Region of low recombination
Gene dense regions
Gene poor regions
100Kb


b)
The genetic and physical maps of Chromosome 2
Genetic map (cM)
T.b.brucei
Physical map (Kb)
T.b.brucei
Genetic map (cM)
T.b.gambiense
Map unit distance
(Kb/cM)
Map unit distance
(Kb/cM)
17Kb/cM
4Kb/cM
2Kb/cM
22Kb/cM
1Kb/cM
11Kb/cM
2Kb/cM
11Kb/cM
4Kb/cM
5Kb/cM
12Kb/cM
6Kb/cM
6Kb/cM
25Kb/cM
12Kb/cM
3Kb/cM
TB2/1
TB2/2
TB2/3
TB2/4
TB2/5
TB2/6
TB2/7
TB2/8
TB2/9
TB2/10
TB2/11
TB2/12
TB2/13
TB2/14
TB2/15
TB2/16
TB2/17
TB2/18
TB2/19
TB2/20
TB2/21
TB2/2
TB2/4
TB2/7
TB2/9
TB2/10
TB2/12
TB2/15
TB2/18
TB2/19
TB2/20
13.9
6.1
21.8
3.0
5.9
2.9
3.0
2.9
9.1
26.5
6.1
21.0
6.1
8.4
3.0
3.0
95.1cM
47.6cM
Region of high recombination
Region of low recombination
Gene dense regions
Gene poor regions
100Kb

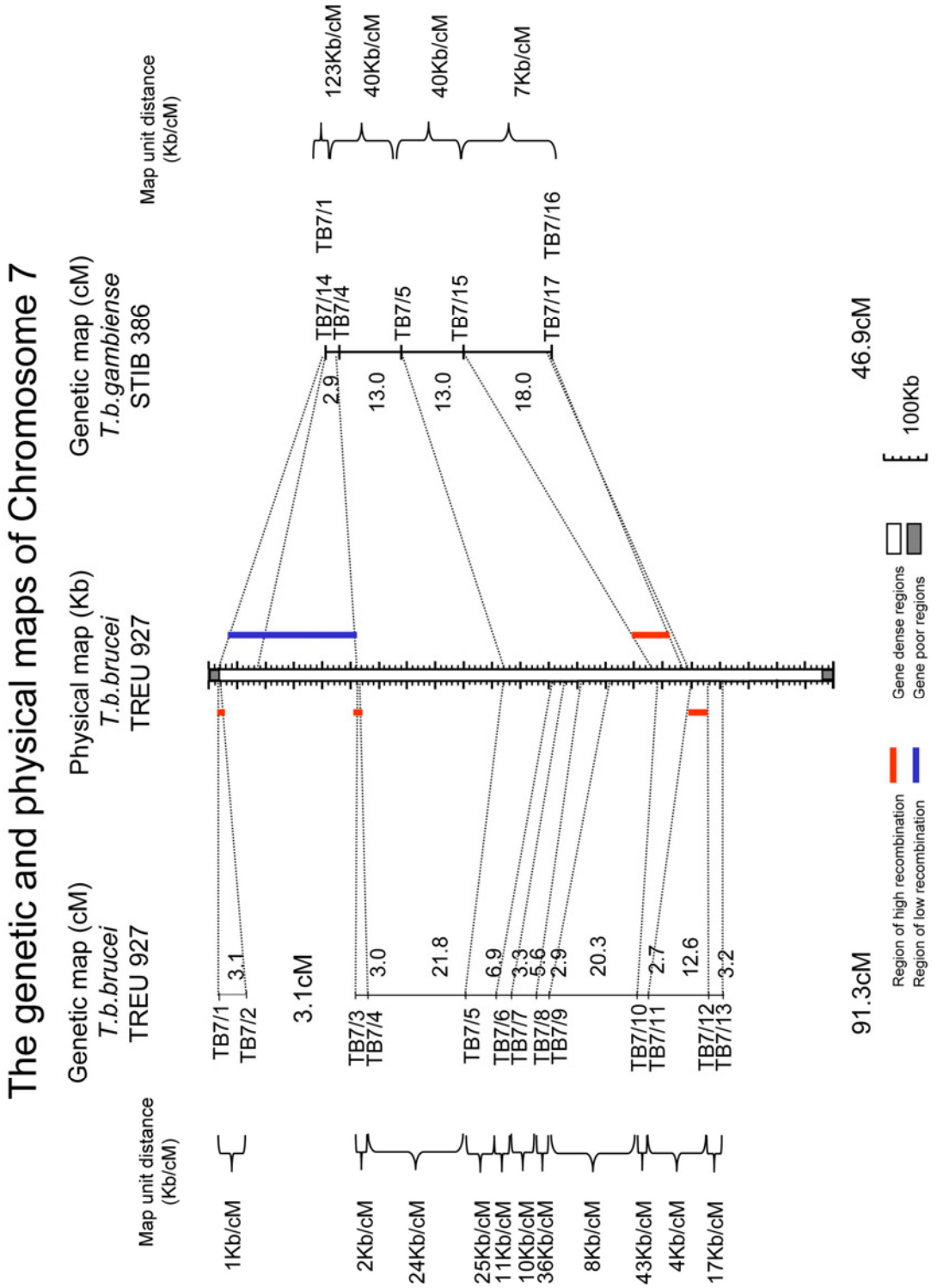

c)
The genetic and physical maps of Chromosome 3
Genetic map (cM)
T.b.brucei
TREU 927
Physical map (Kb)
T.b.brucei
TREU 927
Genetic map (cM)
T.b.gambiense
STIB 386
Map unit distance (Kb/cM)
Map unit distance (Kb/cM)
3Kb/cM
11Kb/cM
6Kb/cM
56Kb/cM
27Kb/cM
4Kb/cM
10Kb/cM
15Kb/cM
25Kb/cM
30Kb/cM
14Kb/cM
10Kb/cM
93Kb/cM
24Kb/cM
27Kb/cM
3Kb/cM
TB3/1 TB3/2
TB3/3
TB3/4
TB3/5
TB3/6 TB3/9
TB3/7 TB3/8
TB3/10 TB3/11
TB3/12
TB3/13 TB3/14
TB3/15 TB3/16 TB3/17
TB3/18
13.0
25.5
13.9
6.5
3.4
13.9
3.0
3.0
2.9
85.1cM
TB3/1 TB3/19
TB3/20
TB3/4 TB3/21 TB3/22
TB3/23
TB3/13 TB3/14 TB3/10
5.7
15.3
5.9
2.9
8.8
5.6
2.7
46.9cM
Region of high recombination
Region of low recombination
Gene dense regions
Gene poor regions
100Kb

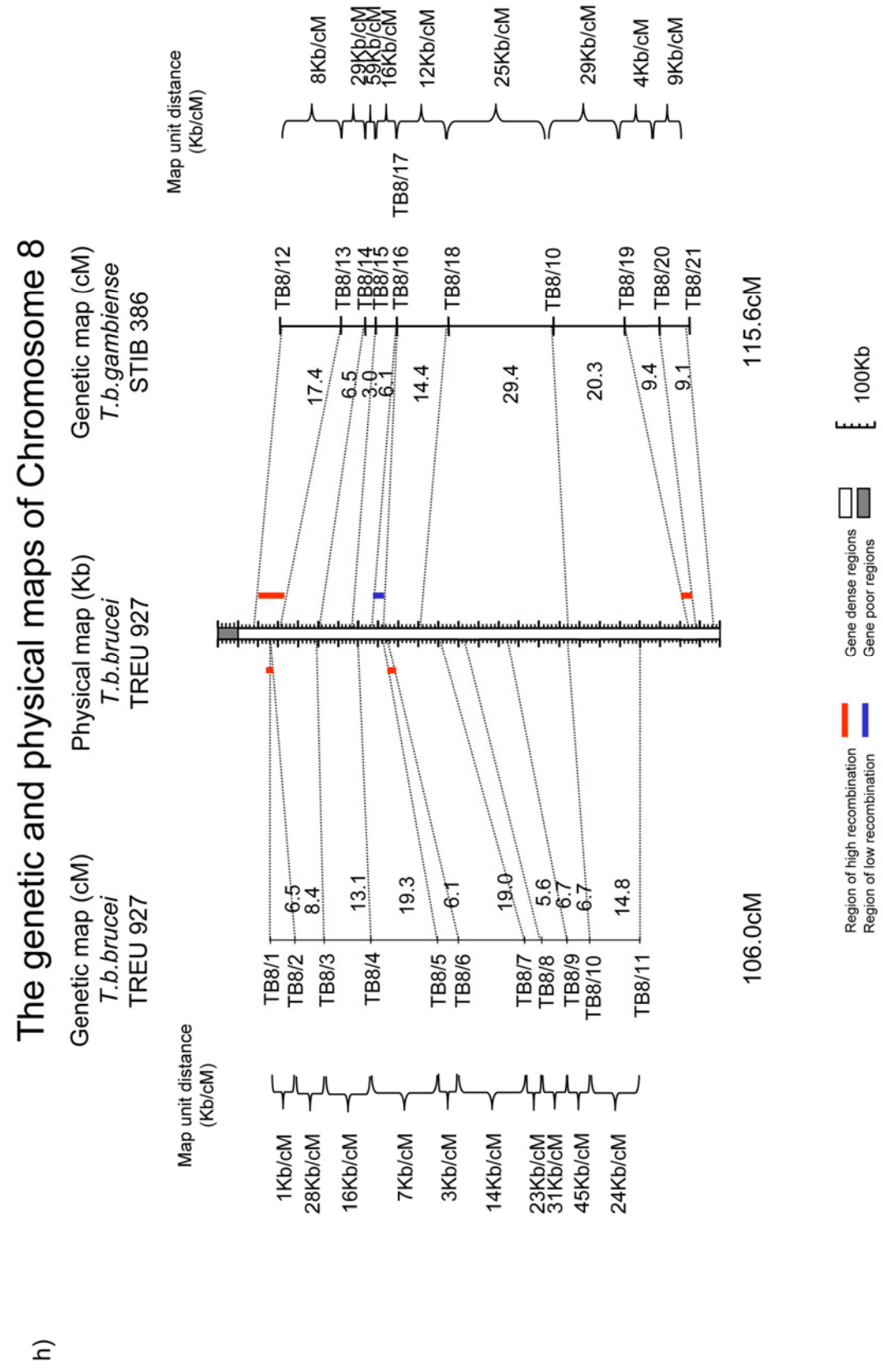

d)
The genetic and physical maps of Chromosome 4
Genetic map (cM)
T.b.brucei
TREU 927
Physical map (Kb)
T.b.brucei
TREU 927
Genetic map (cM)
T.b.gambiense
STIB 386
Map unit distance (Kb/cM)
8Kb/cM
26Kb/cM
18Kb/cM
14Kb/cM
15Kb/cM
29Kb/cM
4Kb/cM
TB4/17
TB4/2
TB4/18 TB4/19
TB4/5 TB4/4*
TB4/20
TB4/21 TB4/8
TB4/22
TB4/12
TB4/13
16.8
6.1
6.3
9.4
6.5
6.3
3.0
54.4cM
TB4/1*
TB4/2
TB4/3
TB4/4* TB4/5
TB4/6 TB4/7
TB4/8
TB4/9 TB4/10
TB4/11
TB4/12 TB4/13
TB4/14
TB4/15 TB4/16
6.3
18.7
7.2
9.4
13.1
3.4
9.1
12.6
6.7
2.8
89.2cM
Region of high recombination
Region of low recombination
Gene dense regions
Gene poor regions
100Kb
Map unit distance (Kb/cM)
9Kb/cM
12Kb/cM
9Kb/cM
31Kb/cM
14Kb/cM
8Kb/cM
2Kb/cM
68Kb/cM

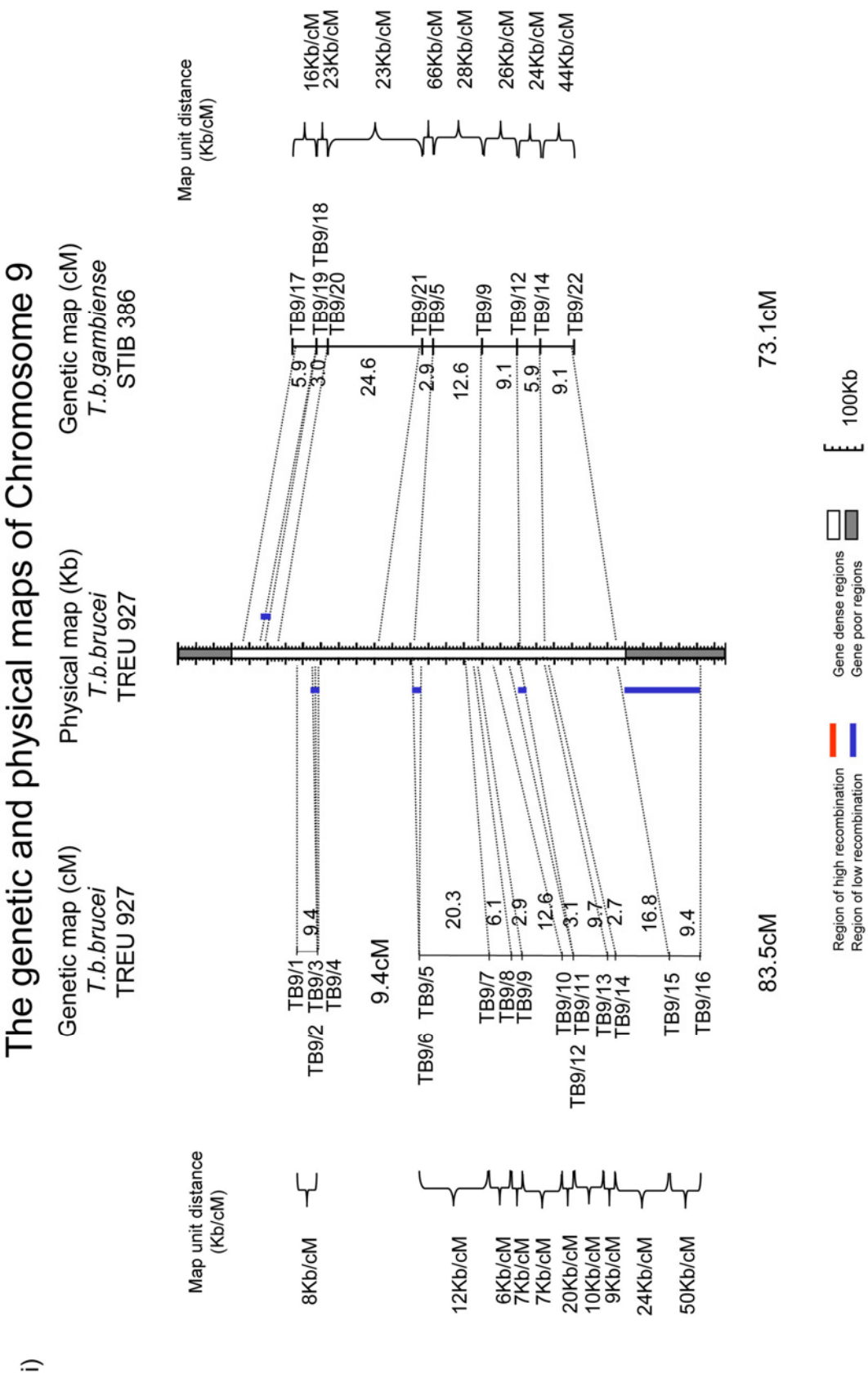

e)
The genetic and physical maps of Chromosome 5
Genetic map (cM)
T.b.brucei
TREU 927
Physical map (Kb)
T.b.brucei
TREU 927
Genetic map (cM)
T.b.gambiense
STIB 386
Map unit distance (Kb/cM)
14Kb/cM
16Kb/cM
16Kb/cM
5Kb/cM
19Kb/cM
TB5/15
TB5/16
TB5/4 TB5/17
TB5/18
TB5/19
TB5/20
12.6
11.8
21.0
29.4
15.8
90.6cM
8.6
16.3
5.9
2.8
5.9
2.9
5.6
5.6
53.4cM
TB5/1
TB5/2
TB5/3
TB5/4
TB5/5
TB5/6
TB5/8
TB5/11
TB5/9
TB5/7
TB5/12
TB5/10
TB5/13
TB5/14
Map unit distance (Kb/cM)
22Kb/cM
6Kb/cM
8Kb/cM
7Kb/cM
25Kb/cM
28Kb/cM
31Kb/cM
5Kb/cM
Region of high recombination
Region of low recombination
Gene dense regions
Gene poor regions
100Kb

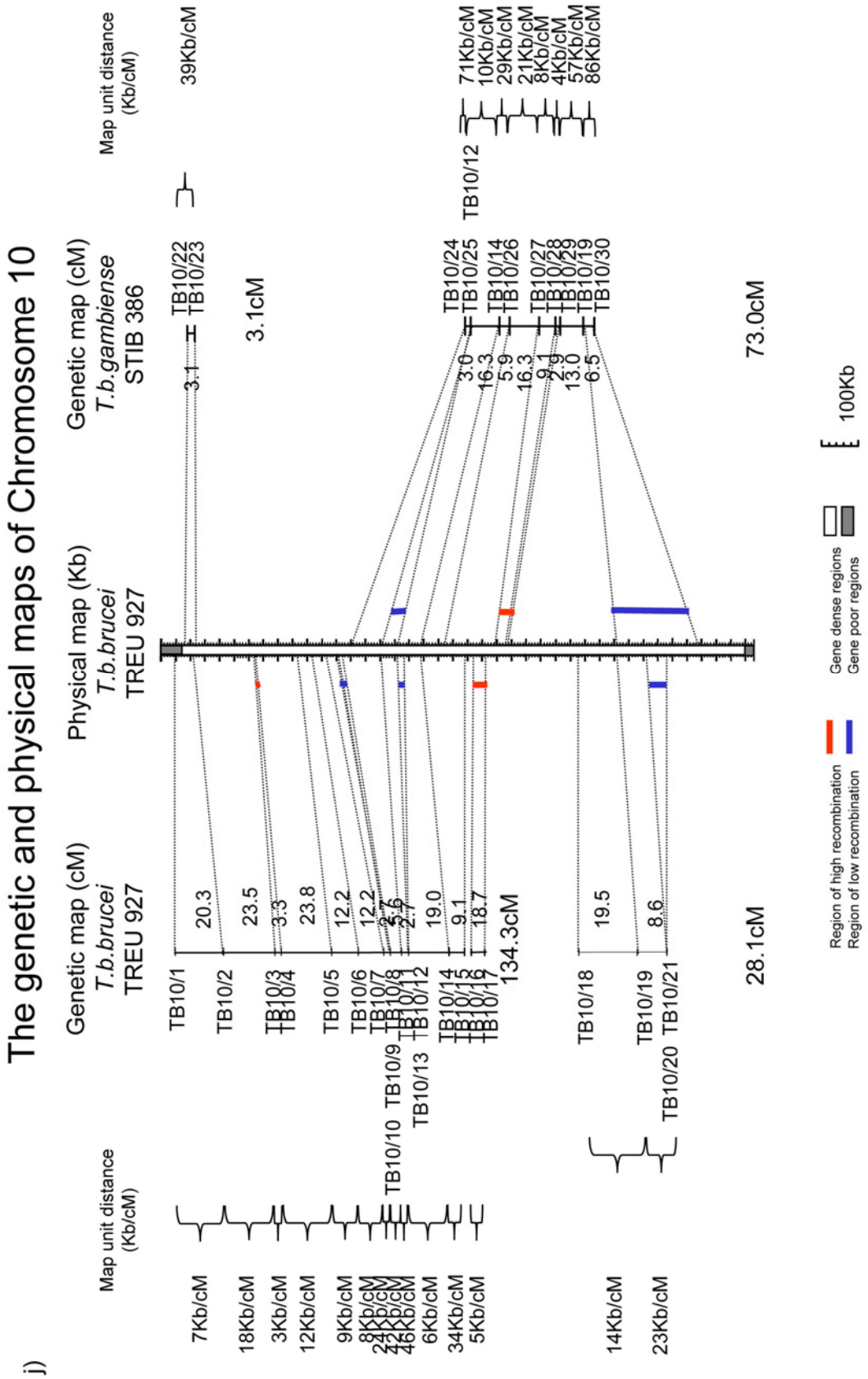

f)
The genetic and physical maps of Chromosome 6
Genetic map (cM)
T.b.brucei
TREU 927
Physical map (Kb)
T.b.brucei
TREU 927
Genetic map (cM)
T.b.gambiense
STIB 386
Map unit distance (Kb/cM)
Map unit distance (Kb/cM)
16Kb/cM
59Kb/cM
5Kb/cM
18Kb/cM
38Kb/cM
24Kb/cM
12Kb/cM
15Kb/cM
34Kb/cM
40Kb/cM
26Kb/cM
20Kb/cM
TB6/1
TB6/2
TB6/3
TB6/4
TB6/5
TB6/6
TB6/7
TB6/8*
TB6/9
TB6/10
TB6/11
TB6/12
TB6/13
TB6/14
TB6/15
TB6/6
TB6/16
15.3
2.9
6.1
15.3
2.7
8.6
12.2
13.9
13.4
3.2
6.1
2.9
2.9
63.0cM
42.4cM
Region of high recombination
Region of low recombination
Gene dense regions
Gene poor regions
100Kb

g)


The genetic and physical maps of Chromosome 7
Genetic map (cM)
T.b.brucei
TREU 927
Physical map (Kb)
T.b.brucei
TREU 927
Genetic map (cM)
T.b.gambiense
STIB 386
Map unit distance (Kb/cM)
Map unit distance (Kb/cM)
TB7/1
TB7/2
3.1
3.1cM
TB7/3
TB7/4
3.0
21.8
TB7/5
TB7/6
TB7/7
TB7/8
TB7/9
6.9
3.3
5.6
2.9
20.3
TB7/10
TB7/11
2.7
12.6
TB7/12
TB7/13
3.2
91.3cM
TB7/1
TB7/14
TB7/4
2.9
13.0
TB7/5
13.0
TB7/15
18.0
TB7/17
TB7/16
46.9cM
123Kb/cM
40Kb/cM
40Kb/cM
7Kb/cM
1Kb/cM
2Kb/cM
24Kb/cM
25Kb/cM
11Kb/cM
10Kb/cM
36Kb/cM
8Kb/cM
43Kb/cM
4Kb/cM
17Kb/cM
Region of high recombination
Region of low recombination
Gene dense regions
Gene poor regions
100Kb


h)
The genetic and physical maps of Chromosome 8
Genetic map (cM)
T.b.brucei
TREU 927
Physical map (Kb)
T.b.brucei
TREU 927
Genetic map (cM)
T.b.gambiense
STIB 386
Map unit distance (Kb/cM)
1Kb/cM
28Kb/cM
16Kb/cM
7Kb/cM
3Kb/cM
14Kb/cM
23Kb/cM
31Kb/cM
45Kb/cM
24Kb/cM
TB8/1
TB8/2
TB8/3
TB8/4
TB8/5
TB8/6
TB8/7
TB8/8
TB8/9
TB8/10
TB8/11
6.5
8.4
13.1
19.3
6.1
19.0
5.6
6.7
6.7
14.8
106.0cM
TB8/12
TB8/13
TB8/14
TB8/15
TB8/16
TB8/17
TB8/18
TB8/10
TB8/19
TB8/20
TB8/21
17.4
6.5
3.0
6.1
14.4
29.4
20.3
9.4
9.1
115.6cM
Map unit distance (Kb/cM)
8Kb/cM
29Kb/cM
59Kb/cM
16Kb/cM
12Kb/cM
25Kb/cM
29Kb/cM
4Kb/cM
9Kb/cM
Region of high recombination
Region of low recombination
Gene dense regions
Gene poor regions
100Kb


i)
The genetic and physical maps of Chromosome 9
Genetic map (cM)
T.b.brucei
TREU 927
Physical map (Kb)
T.b.brucei
TREU 927
Genetic map (cM)
T.b.gambiense
STIB 386
Map unit distance (Kb/cM)
8Kb/cM
12Kb/cM
6Kb/cM
7Kb/cM
7Kb/cM
20Kb/cM
10Kb/cM
9Kb/cM
24Kb/cM
50Kb/cM
16Kb/cM
23Kb/cM
23Kb/cM
66Kb/cM
28Kb/cM
26Kb/cM
24Kb/cM
44Kb/cM
TB9/1
TB9/2
TB9/3
TB9/4
9.4
9.4cM
TB9/5
TB9/6
20.3
TB9/7
TB9/8
TB9/9
6.1
2.9
12.6
TB9/10
TB9/11
TB9/12
3.1
9.7
TB9/13
TB9/14
2.7
16.8
TB9/15
TB9/16
9.4
83.5cM
5.9
TB9/17
3.0
TB9/19
TB9/18
TB9/20
24.6
2.9
TB9/21
TB9/5
12.6
TB9/9
9.1
TB9/12
5.9
TB9/14
9.1
TB9/22
73.1cM
Region of high recombination
Region of low recombination
Gene dense regions
Gene poor regions
100Kb


The genetic and physical maps of Chromosome 10
j)
Genetic map (cM)
T.b.brucei
TREU 927
Physical map (Kb)
T.b.brucei
TREU 927
Genetic map (cM)
T.b.gambiense
STIB 386
Map unit distance
(Kb/cM)
Map unit distance
(Kb/cM)
TB10/1
TB10/2
TB10/3
TB10/4
TB10/5
TB10/6
TB10/7
TB10/8
TB10/9
TB10/10
TB10/11
TB10/12
TB10/13
TB10/14
TB10/15
TB10/16
TB10/17
TB10/18
TB10/19
TB10/20
TB10/21
TB10/22
TB10/23
TB10/24
TB10/25
TB10/26
TB10/27
TB10/28
TB10/29
TB10/30
TB10/12
20.3
23.5
3.3
23.8
12.2
12.2
2.7
5.6
19.0
9.1
18.7
134.3cM
19.5
8.6
28.1cM
3.1
3.1cM
3.0
16.3
5.9
16.3
9.1
7.9
13.0
6.5
73.0cM
7Kb/cM
18Kb/cM
3Kb/cM
12Kb/cM
9Kb/cM
6Kb/cM
34Kb/cM
5Kb/cM
14Kb/cM
23Kb/cM
39Kb/cM
71Kb/cM
10Kb/cM
29Kb/cM
21Kb/cM
8Kb/cM
4Kb/cM
57Kb/cM
86Kb/cM
Region of high recombination
Region of low recombination
Gene dense regions
Gene poor regions
100Kb

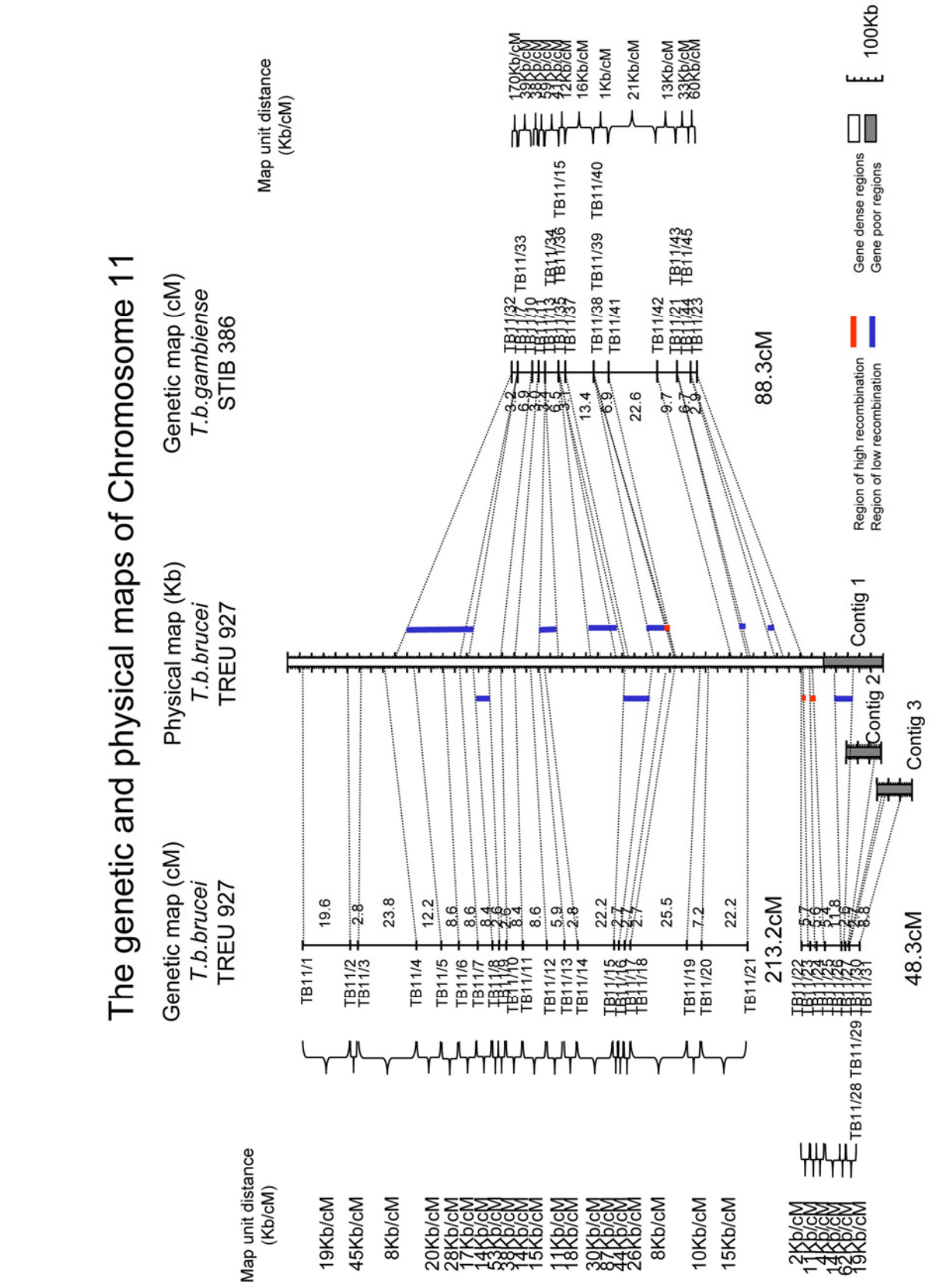

k)
The genetic and physical maps of Chromosome 11
Genetic map (cM)
T.b.brucei
TREU 927
Physical map (Kb)
T.b.brucei
TREU 927
Genetic map (cM)
T.b.gambiense
STIB 386
Map unit distance
(Kb/cM)
Map unit distance
(Kb/cM)
Region of high recombination
Region of low recombination
Gene dense regions
Gene poor regions
100Kb
Contig 1
Contig 2
Contig 3
88.3cM
213.2cM
48.3cM

Figure 3.5 Comparison of the *T.b.gambiense* genetic maps with the physical and genetic maps of *T.b.brucei.*

The genetic maps of *T.b.brucei* isolate TREU 927 and *T.b.gambiense* isolate STIB 386 are shown alongside the TREU 927 physical map of the same chromosome for chromosome 1 (a), 2 (b), 3 (c), 4 (d), 5 (e), 6 (f), 7 (g), 8 (h), 9 (i), 10 (j) and 11 (k). The average physical size of a recombination unit between each marker is given in Kb/cM and the genetic distance given in cM. Dashed lines link the position of all markers on the physical map to their relative position on the genetic maps. Hot and cold spots are defined as three-fold more or less recombination than average for each genetic map and indicated against the physical map by red and blue bars respectively.

3.2.5 Comparison of the genetic maps of *T.b.gambiense* and *T.b.brucei* and the physical map of *T.b.brucei*

The linkage groups of the STIB 386 genetic map comprise a total genetic distance of 733.1 cM covering a physical distance of 17.9 Mb, compared to a genetic map of 1157 cM covering 18.06 Mb for the *T.b.brucei* TREU 927 map (MacLeod *et al.*, 2005b). Although the genetic distance covered by the STIB 386 map is smaller, there is no significant difference in frequency of recombination (Kb/cM) between the two sub species (χ^2=1.936, d.f=1, P= 0.164) and they contain very similar marker densities (average cMs between intervals) of 9.0 cM for STIB 386 and 9.5 cM for TREU 927.

As 47 markers are informative in both the *T.b.brucei* and *T.b.gambiense* maps, this allows a direct evaluation of genetic distances between the maps, and comparison with the physical *T.b.brucei* map. For six chromosomes, where there are four or more shared markers, (chromosomes 1, 2, 3, 4, 9 and 11), synteny in terms of marker order is conserved (Figure 3.5a-d,i,k). The rest of the chromosomes have fewer shared markers making comparisons less informative, however no inconsistencies between the genetic map and the physical map of TREU 927 were detected. The karyotype of both strains has been determined by PFGE (Melville *et al.*, 1998) and, in terms of chromosome size, seven of the chromosome pairs of STIB 386 are found to be considerably larger than those of TREU 927 (chromosomes 1, 4, 6, 7, 8, 9 and 10). If these physical size differences occurred in regions of each chromosome covered by the genetic map, then I would predict that the recombination frequency of the STIB 386 chromosomes would be correspondingly higher and result in larger genetic distances between markers, however this does not appear to be the case.

To illustrate these similarities and differences between chromosomes the data for chromosomes 1 and 2 are considered (Figure 3.5a,b). For chromosome 2, the physical size of the chromosome is similar in both isolates based on PFGE (Melville *et al.*, 1998), but the size of the genetic maps differ significantly. Comparing only the region of the chromosome represented by both genetic maps, from marker TB2/2 to TB2/20, the genetic distances for *T.b.brucei* and *T.b.gambiense* are 81.2 cM and 47.6 cM, respectively (Figure 3.5b), which is significantly different (χ^2 =8.765, d.f=1, P < 0.01). The difference in genetic distance between the chromosome 2 maps is largely due to a hotspot of recombination in the interval between markers TB2/20 and TB2/12 in *T.b.brucei* (35.6 cM), which in not present

in *T.b.gambiense* (14.4 cM) at the same marker interval. However for chromosome 1 (Figure 3.5a), comparing the distance represented by the two genetic maps (35.8 cM and 25.1 cM), the difference is not significant (χ^2=1.88, d.f-1, P= 0.17), despite the physical size of chromosome 1 in the *T.b. gambiense* strain, STIB 386, being estimated to be almost twice that of TREU 927 (Melville *et al.*, 1998).

3.2.6 Mutation frequency

A single spontaneous mutation event, generating a novel sized allele product, distinct from the parental alleles, was detected when genotyping the progeny clones (Figure 3.6). This mutation occurred at marker TB6/15 resulting in a mutation frequency at this locus of 0.028 mutants/alleles genotyped. Combined with all other markers this produces an overall mutation frequency of 2.4 x 10^{-4} mutants/alleles genotyped which is consistent with the mutation frequency of 3.0 x 10^{-4} mutants/alleles genotyped reported for the *T.b.brucei* strain TREU 927 (MacLeod *et al.*, 2005b). In contrast to the TREU 927 mutant loci, the allele in question had lost repeats resulting in an allele smaller than either of the STIB 386 parental alleles. The origin of the mutation has not been determined, but as the original parental allele is not detected in addition to the mutant, the mutation is unlikely to have arisen during vegetative growth of the progeny clone, but prior to the cloning process, probably at meiosis.

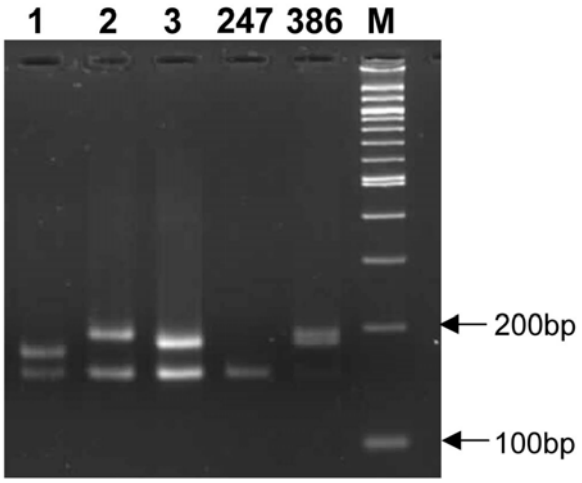


Figure 3.6 Evidence for a mutation in the STIB 386 allele of marker TB6/15 in progeny clone F28/46 bscl 4.

Separation of PCR products amplified for microsatellite TB6/15 demonstrates the expected size of alleles in parental lines STIB 386 and STIB 247. Progeny clones F28/46 bscl 1 (lane 2) and F28/46 bscl 7 (lane 3) display the inheritance of correctly sized alleles from the parental lines, which each inheriting a different STIB 386 allele. Progeny clone F28/46 bscl 4 (lane 1) contains the correctly sized STIB 247 allele but a novel sized STIB 386 allele smaller than either of the parental alleles. M is a 100 bp DNA ladder.

3.2.7 In vivo drug sensitivity assay

Following map construction, an *in vivo* Minimum Curative Dose (MCD) assay was used to determine the relative drug sensitivities of the two parental lines, STIB 247 and STIB 386 to the trypanocidal drugs; Cymelarsan, pentamidine, Berenil and suramin. This assay, based on a method previously described (Scott *et al.*, 1996), uses a murine host to model *T.brucei* infection and the pharmacological activity of the drug in the natural human and animal hosts. Groups of three mice were infected (i.p) with 1 x 10^5 trypanosomes, and 24 hours later a single dose of drug administered, at a different concentration for each group. The tail blood of each mouse was monitored for parasitemia, using the rapid "matching" method (Herbert and Lumsden, 1976), from 48 hours after drug inoculation, for 21 days. If no trypanosomes were observed in wet blood films from tail blood during that time, the drug treatment was considered to be effective and the mouse was considered cured. A drug treatment failure was judged to have occurred if trypanosomes were directly observed by microscopy in tail blood smears within the 21 day period.

The data was used to determine the MCD; the lowest dose required to prevent a patent infection in all mice in a group. Tables 3.3a-d contain the results of the treatment groups for each of the four trypanocides tested. The rapid "matching" method used to estimate parasitemia is capable of detecting a parasitemia greater than ~2.5 x 10^5 trypanosomes/ml (Herbert and Lumsden, 1976), therefore a patent infection would only be detected if trypanosomes had multiplied in the presence of the drug. All drug-free positive control mice developed patent infections between 2-9 days following inoculation, confirming the viability of the parental trypanosome stocks in each experiment.

Preliminary data had indicated a seven-fold difference in the MCD between the two parental lines in respect to the water soluble melaminophenyl arsenical drug, Cymelarsan (Scott *et al.*, 1996). STIB 386 had previously been assigned a MCD of 2 mg/kg and STIB 247 a MCD of 0.3 mg/kg. However, in my dataset I found that there was no difference in the MCD of STIB 247 and STIB 386, with a single administration of 0.5 mg/kg Cymelarsan yielding a 100% cure rate in both lines (Table 3.3a). This was the closest dose to the 0.3 mg/kg curative dose tested for STIB 247 (Scott *et al.*, 1996) confirming the previously observed MCD for this line, but indicating, in contrast, that the STIB 386 strain is equally sensitive to Cymelarsan.

a)

Cymelarsan Concentration mg/kg	No. of Mice with patent infection		
	STIB 247	STIB 386	STIB 386 repeat*
0	2/2	2/2	2/2
0.25	1/3	2/3	2/3
0.5	0/3	0/3	1/3
1.0	0/3	0/3	0/3
2.0	0/3	0/3	0/3

b)

Suramin Concentration mg/kg	No. of Mice with patent infection	
	STIB 247	STIB 386
0	2/2	2/2
0.25	3/3	3/3
0.5	2/3	3/3
1.0	2/3	3/3
2.0	2/3	2/3
4.0	1/3	0/3
8.0	0/3	-

c)

Berenil Concentration mg/kg	No. of Mice with patent infection	
	STIB 247	STIB 386
0	2/2	2/2
0.25	3/3	3/3
0.5	3/3	3/3
1.0	1/3	1/3
2.0	0/3	0/3

d)

Pentamidine Concentration mg/kg	No. of Mice with patent infection	
	STIB 247	STIB 386
0	2/2	2/2
0.25	3/3	3/3
0.5	3/3	3/3
1.0	3/3	3/3
2.0	3/3	3/3
4.0	0/3	3/3
8.0	0/3	0/3

Table 3.3 Susceptibility of STIB 247 and STIB 386 to trypanocidal drugs.

Susceptibility of STIB 247 and STIB 386 to a) Cymelarsan, b) suramin, c) Berenil, d) pentamidine. Groups of three mice (two for controls) were infected on day 0 with 10^5 trypanosomes (i.p) and drug administered 24 hours after inoculation. The number of mice in each group developing parasitemia within a 21 day time period is indicated. Animals surviving 21 days without observed parasitemia by the rapid "matching" method (Herbert and Lumsden, 1976) were considered cured. The lowest dose curing all mice in a group was considered the minimum curative dose (MCD) for each parent. A difference in MCD between the parents for Pentamidine is highlighted in grey. * Indicates the repeat MCD experiment for STIB 386.

Initially it was considered that a technical error might have occurred with either the trypanosome stabilate used in the analysis or the preparation of the drug. To explore this possibility, a PCR genotyping analysis was performed on genomic DNA derived from the trypanosome-positive drug-free control mice. Analysis of five core microsatellite markers however confirmed that the correct STIB 386 strain had been used in the experiment (data not shown).

An additional consideration was that an error might have occurred during the preparation of the drug by the experimenter. To clarify the MCD therefore, the experiment was repeated for the STIB 386 strain with a fresh preparation of drug. The experiment was replicated using an identical protocol and produced almost identical results (shown in column four of Table 3.3a). In contrast to the previous dataset therefore, no significant difference in Cymelarsan sensitivity could be demonstrated between the STIB 386 and STIB 247 strain and this phenotype could not be considered for linkage mapping studies.

In addition to Cymelarsan, three other drugs were also assessed for their sensitivity for which the parental MCD was not known. The data for sensitivity to the early stage HAT drug, suramin (Table 4b), signalled that in both parents, one mouse was cured at the 2 mg/kg dose, but while 100% of STIB 386 mice were cured at 4 mg/kg the MCD for STIB 247 was 8 mg/kg. However, this difference in MCD was the result of only a single mouse in the higher dose group for STIB 247 and thus cannot be considered significant. No difference was also observed between the two parental strains for the veterinary diamidine, Berenil, with a cure rate of 2/3 mice at 1 mg/kg and a 100% cure rate at 2 mg/kg in both lines (Table 3.3c).

In contrast, a difference was observed in the MCD of the parents with regards to the other diamidine tested, pentamidine (Table 3.3d). A 100% drug failure rate was observed at 2 mg/kg in both strains, but STIB 247 was completely cleared at the 4 mg/kg dose while STIB 386 remained 100% resistant only reaching a curative dose at 8 mg/kg.

The results of the *in vivo* MCD experiments therefore indicate that although no difference was observed with regards to Cymelarsan, suramin or Berenil sensitivity between the two parental lines, there was an increased resistance in STIB 386 to pentamidine compared to STIB 247, of approximately two-fold. This innate difference in pentamidine sensitivity, although subtle, is interesting because of an apparent lack of the cross-resistance to Cymelarsan and Berenil that is frequently,

although not always observed in diamidine and melaminophenyl arsenicals resistant lines (Scott, 1995; Matovu *et al.*, 2003; Bridges *et al.*, 2007).

3.2.8 In vitro drug sensitivity assay

Following the observation of a two-fold difference in pentamidine sensitivity between the parental lines, consideration was given as to whether this phenotype would be suitable for linkage analysis. One concern was that the MCD assay would not be an ideal assay for phenotype screening of the progeny because of the time and resources required for this *in vivo* experimental protocol. An alternative *in vitro* assay would be desirable if the difference between the parental lines was preserved, to reduce the number of experimental animals required and offer a system for rapid and repeatable testing of progeny clone phenotype.

Recent refinements to the *in vitro* culture protocols in our laboratory has resulted in the adaptation of the parental strains, STIB 247 and STIB 386, to a long term bloodstream form culture system. This system offers the advantage of a more direct parallel to the *in vivo* assay over the pre-existing procyclic form culture lines for which substantial biochemical differences are known to exist. In addition to well characterised differences in energy metabolism and surface coat (Matthews, 2005), the mammalian and insect forms differ in their expression of purine transporters (De Koning, 2001b), which have frequently been associated with drug resistance in trypanosomes.

An *in vitro* pentamidine sensitivity assay was tested using continuously growing bloodstream cultures of the STIB 247 and STIB 386 cell lines. Assays were set up with 2 x 10^5 trypanosomes/ml with doubling dilutions of pentamidine, at a concentration range of 0.5 nM-500 nM. Haemocytometer counts of cells were taken after a 48 hour incubation period and growth calculated at each drug concentration as a percentage of a drug free control. The resulting growth inhibition profiles of bloodstream cultures of STIB 247 and STIB 386 are shown in Figure 3.7. The concentration of drug required to inhibit growth by 50% (IC_{50}) was calculated from an average of four growth inhibition experiments. An IC_{50} of 2.2 nM was obtained for STIB 247 compared to 8.3 nM for STIB 386, an approximate 3.5 fold difference.

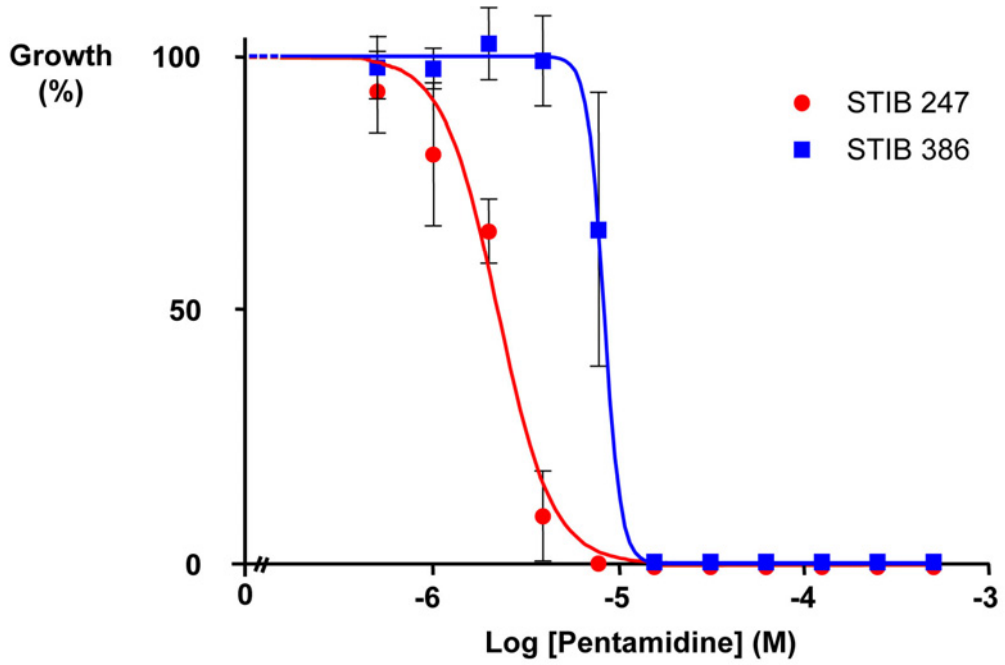


Figure 3.7 48 hour growth inhibition assay with pentamidine.

The growth of *T.brucei* strains STIB 386 and STIB 247 bloodstream culture adapted cells after incubation for 48 hours in the presence of varying concentrations of pentamidine is shown. A dose response curve was fitted to each data set using a non-linear regression, least squares fitting algorithm in the Graphpad prism 5.1 software package. IC_{50} values calculated from these curves were 2.2 nM for STIB 247 and 8.3 nM for STIB 386. The error bars indicate the mean ± one standard deviation of four replicate experiments.

The data obtained from the *in vitro* assays demonstrate than a difference in sensitivity to pentamidine between the parents is retained in the *in vitro* assay system with STIB 386 less sensitive to the trypanocidal action of the drug than STIB 247. The level of resistance is of a relatively comparable level between the two different systems, however in addition to the ethical and technical advantages of replacing an *in vivo* system, the assay also allows a quantitative measurement to be recorded in the form of IC_{50} levels which would allow for the analysis of more complex phenotypes through QTL analysis.

3.3 Discussion

Genetic linkage maps have been determined for a number of parasites including the haploid apicomplexa species *P.falciparum* (Su *et al.*, 1999), *P.c.chabaudi* (Martinelli *et al.*, 2005a), *E.tenella* (Shirley and Harvey, 2000) and *T.gondii* (Khan *et al.*, 2005) and recently the first map for the diploid trypanosomatid *T.b.brucei* has been reported (MacLeod *et al.*, 2005b). Here, knowledge of this parasite is advanced by reporting the construction of the first linkage map of a human-infective strain of the *T.b.gambiense* subspecies to provide a basis for expanding studies on important biological traits in this line such as human infectivity and virulence.

The average recombination rate in this genetic map (24.4 Kb/cM) is close to the values reported for *T.b.brucei* (MacLeod *et al.*, 2005b), *P.falciparum* (Su *et al.*, 1999), and other organisms with a similar size genome (NCBI, 1999). However, as observed for a variety of other eukaryotes, there is considerable variation in the physical size of a centiMorgan. Similar hot and cold spots of meiotic recombination have been reported for a wide variety of eukaryotic species (Petes, 2001) and were also identified on the *T.b.brucei* TREU 927 map (El-Sayed *et al.*, 2003; Hall *et al.*, 2003; MacLeod *et al.*, 2005b). Although local variation in crossover frequency appears to be a common feature of both the *T.b.brucei* and *T.b.gambiense* maps, this balances out over the full length of each chromosome with the total genetic distance of chromosomes correlated with their physical sizes for the *T.b.brucei* map (MacLeod *et al.*, 2005b) and to a lesser degree with the *T.b.gambiense* map, with the caveat that the sequence data of *T.b.brucei* was used to as a basis for estimating the physical size for *T.b.gambiense.*

Size polymorphism in the megabase chromosomes of *T.brucei* has been documented both between isolates and between homologues within a single parasite genome (Gottesdiener *et al.*, 1990; Tait *et al.*, 1996; Melville *et al.*, 1998). PFGE resolution of the molecular karyotype for the genetic map isolate, STIB 386, showed that at least 7 out of 11 chromosome pairs were larger in size than those in the *T.b.brucei* genome reference strain TREU 927 (Melville *et al.*, 1998). On this basis we might therefore anticipate the genetic size of these chromosomes to reflect this physical size difference, with larger genetic distances in those chromosomes that are larger in the *T.b.gambiense* subspecies. Interestingly though, we found no significant difference in recombination, measured in terms of average map unit size, between the two strains. Indeed,

where distance between markers present on both genetic maps were examined, STIB 386 was frequently found to have the smaller genetic map distance, despite the predicted size of homologues being up to twice that of TREU 927 (Gottesdiener *et al.*, 1990; Melville *et al.*, 1998).

Considerable chromosome size variation between isolates has been reported in many protozoan parasites with little or no effect on gene content. Variations in chromosome size between strains of 10-50% in *P.falciparum* (Janse, 1993; Lanzer *et al.*, 1993; Lanzer *et al.*, 1994), *Leishmania spp.* (Blaineau *et al.*, 1991; Ravel *et al.*, 1996; Wincker *et al.*, 1996) and *T.cruzi* (Henriksson *et al.*, 1995) have been attributed primarily to changes in repeat regions in the sub-telomeric sequence. This polymorphism is even more extreme in *T.brucei* isolates where chromosome plasticity results in homologues varying up to four-fold between isolates (Melville *et al.*, 2000) and even two-fold within a single genome (Gottesdiener *et al.*, 1990; Melville *et al.*, 1998; Melville *et al.*, 2000), without an apparent loss of linkage in coding regions.

Comparisons of the Trypanosomatid genome sequence data, comprising the *T.brucei*, *T.cruzi* and *L.major* species, has uncovered a common chromosomal arrangement with a central core exhibiting extensive synteny (El-Sayed *et al.*, 2005). Within *T.brucei* isolates, comparative studies of homologous chromosomes have as yet failed to identify any associated loss of synteny or translocation in coding regions, even between very size divergence chromosomes. In one such study, DNA microarray analysis of the genome content variation of chromosome 1, one of the most size variable chromosomes, was used to identify regions of copy number polymorphism between strains (Callejas *et al.*, 2006). As observed with related protozoan pathogens, the majority of the extensive size variation between isolates appeared to be concentrated in the subtelomerically located genes including the VSGs, VSG expression site associated genes and highly polymorphic gene families such as the retrotransposon hot spot and leucine-rich repeat protein genes. Variation in copy number of these repeat elements was found to compose as much as 75% of the length of a homologue. In contrast, 90% of the diploid core showed little evidence of significant copy number variation with polymorphisms mainly limited to tandemly repeated gene arrays such as tubulin, histone H3 and the pteridine transporters.

Our comparison of the *T.b.brucei* strain TREU 927 and *T.b.gambiense* strain STIB 386 genetic maps is in agreement with these findings. We report no inconsistency in the marker order or average map unit size between the STIB 386 genetic map

and that of *T.b.brucei*. Some strain-specific local variation in the recombination rate between shared marker pairs was identified, which may be attributed to local physical size differences or variation in tandemly repeated gene arrays within the coding regions. Overall though, our data appears to be in agreement with a conservation of synteny between the two subspecies, with the majority of the variation accounting for chromosome size difference between the two strains focused outside the gene-rich coding region (in the sub-telomeres) and therefore not covered by the genetic map.

The genetic distances in the map reflect the number of recombination events that have occurred in the population during meiosis. At least one reciprocal crossover per chromosome is considered essential for the successful disjunction of homologous chromosomes during meiosis (Baker *et al.*, 1976). It is therefore surprising that 48% of all STIB 386 chromosomes analysed in this cross failed to show evidence of any recombination events (The complete crossover data in the progeny is detailed in Appendix 5). Progeny averaged only 0.6 crossovers/chromosome compared to the 1.02 calculated for the TREU 927 map, despite comparable coverage of the genome. Indeed, in several progeny clones, evidence of recombination was extremely rare or, in the case of hybrid F492/50 bscl 23, entirely absent on all 11 chromosomes. The reasons for this low crossover frequency are unknown but may also be a consequence of the larger predicted genome size of the STIB 386 strain. Physical estimates of marker locations were established from the available TREU 927 sequence to produce a total predicted coverage of the genome of 70%. However if the larger physical size of STIB 386 was due to extended sub-telomeric regions, this would leave an increased percentage of the genome outside of the gene-dense centre, uncovered by the map. If the obligate crossover necessary to ensure faithful meiotic segregation of chromosomes is occurring outside the central core on some STIB 386 chromosomes and towards the sub-telomeric regions at the ends of chromosomes, it would not be detected by our analysis.

Estimations of the frequency at which spontaneous microsatellite mutations occur may aid the understanding of the evolution and stability of such markers and their usefulness in genetic analysis of *T.brucei* populations. Only one estimate exists for *T.brucei*; an approximate mutation rate of 3×10^{-4} mutants/allele genotyped was reported in the *T.b.brucei* genetic map from the identification of two spontaneous mutation events in a dataset of 6797 microsatellite alleles. In this *T.b.gambiense* genetic map the identification of a single spontaneous mutation event in a microsatellite marker appears to substantiate this (2.4×10^{-4} mutants/allele

genotyped). These estimates are based on only a small number of mutation events and thus can only be considered an approximation, but are comparable to a similar mutation rate reported in the malaria parasite, *P.falciparum* of 1.6 x 10^{-4} mutants/allele genotyped (Anderson *et al.*, 2000). Given that we have screened an additional 118 markers and found no mutations (~ 4,500 events), we can be confident that the value we have obtained is a maximum. Although the screening of a significantly larger dataset of marker alleles would allow a more accurate mutation rate to be obtained, we consider that our high coverage of the genome sequence in the screen for informative microsatellite markers coupled with the relatively low level of heterozygosity, make it unlikely we would find enough additional microsatellite markers from further screening to detect more mutations.

T.b.gambiense is related to *T.b.brucei*, but differs significantly in important phenotypic characteristics, most notably in its ability to infect humans. Indeed the *T.b.gambiense* and *T.b.brucei* strains examined here not only differ in terms of human infectivity but also in the virulence of infection, their ability to establish midgut infections in the tsetse vector and to progress from the midgut to the salivary glands (transmission index). The availability of a genetic linkage map for *T.b.gambiense* now opens up the possibility of identifying genes that determine these traits. An additional fifth phenotype, that of the ability to resist killing by drugs used in the treatment of trypanosomiasis, was examined in this study to determine if a difference existed between the parents that could also be amenable to genetic linkage analysis studies.

An *in vivo* MCD assay was employed to determine drug sensitivity in a murine disease model to more closely model the natural interaction of parasite, host and drug. Although evidence suggests that the resulting MCD may not be directly extrapolated to the field due to the metabolic differences between mice and the natural hosts of *T.brucei* (Sones *et al.*, 1988), it does allow for a comparison to be made between strains tested within the same model (Eisler *et al.*, 2001).

A study by Scott *et al* (1996) had observed a seven-fold difference in Cymelarsan sensitivity between the STIB 247 and STIB 386 parents using this assay. Attempts to confirm the MCD in this analysis however detected no variation in Cymelarsan sensitivity between the strains. At present no definitive explanation has been established as to why the *in vivo* sensitivity of the previously resistant stock STIB 386 was identical to that of the susceptible stock STIB 247, but it could possibility be attributed to experimental differences in the MCD protocol.

As the study by Scott *et al* was performed ~15 years ago, variation will exist in both the Cymelarsan used and parental stabilate stocks to those available in the laboratory at present. In addition it may be of note that although the MCD experiments were executed for both STIB 247 and STIB 386 by Scott *et al*, this was not performed as a direct comparison of the parental strains. Instead Cymelarsan sensitive of each strain was assessed in relation to a highly drug resistant mutant derived from each wild type strain, therefore addition experimental variables e.g. in the batches of mice and drug and drug dilution ranges may not have been controlled for.

A further explanation may be derived from the variation in MCD protocol between the experiments in regards to the immune status of the murine host. The protocol used by Scott *et al* includes the administration of cyclophosphamide to immunosuppress the mice, 24 hours before the inoculation of trypanosomes. As indicated by Scott (Scott, 1995), the use of this immunosuppression step may considerably reduce the efficacy of the trypanocidal drugs, due to the contributing role of the host's immune system in clearing the infection. This effect was reported as far back as the 1930's (von Jancso and von Jancso, 1934; von Jancso and von Jancso, 1935) where it was observed that resistance to suramin developed more readily in mice immunocompromised by splenectomy than in hosts with an intact immune system. With specific reference to the melamino phenyl arsenicals drugs, a similar result has been observed in both *T.evansi* (Osman *et al.*, 1992) and *T.b.rhodesiense* (Frommel, 1988) in immunosuppressed mice treated with Cymelarsan and melarsoprol, respectively. Frommel (1988) observed that doses of melarsoprol which were curative in treated *T.b.rhodesiense* in immunocompetent mice were ineffective in mice immunosuppressed with cyclophosphamide or radiation. Moreover the trypanosomes that had been treated with subcurative doses of melarsoprol had become increasingly resistant when transferred to immunocompetent mice and retested.

These studies indicate that immunosuppression of the host might considerably reduce the efficacy of trypanocidal drug and more quickly lead to the development of resistant lines. We might therefore expect both strains to exhibit higher resistance in the immunocompromised MCD experiment than the immunocompetent MCD experiment and the fact that only the STIB 386 line appears to have increased resistance is puzzling. It is possible that this is another reflection of interstrain differences, such as the increased virulence already noted in STIB 386 (Turner *et al.*, 1995) (Table 3.1), which might lead to the immune

system playing a more crucial role in clearing the parasites or that resistance to the drug develops more rapidly during drug treatment in one strain than in the other. This hypothesis could not be tested in our MCD model however because of restrictions in the project license preventing the use of cyclophosphamide in conjunction with other trypanocides and is thus presently confined to speculation.

Although no difference was observed with regards to Cymelarsan, suramin or Berenil sensitivity between the two parental lines, there was an increased resistance in STIB 386 to pentamidine compared to STIB 247. This two-fold difference in the MCD for pentamidine, although subtle, is similar to the increased resistance reported in knock-out and loss of function mutants in *T.brucei* strain 427 for a transporter of the drug, P2 coded by the *TbAT1* gene (Matovu *et al.*, 2003). The natural substrate of the P2 transporter are aminopurines but due to a common substrate recognition motif (de Koning and Jarvis, 1999) it is also able to transport a range of other substrates including the melaminophenyl arsenical and diamidine class of trypanocidal drugs (Carter and Fairlamb, 1993; Carter *et al.*, 1995). However, while the loss of *TbAT1*/P2 function in *T.brucei* has been demonstrated to be sufficient for high level resistance to Berenil, it produced only low levels of resistance to pentamidine and the melaminophenyl arsenicals (2-3 fold) (Matovu *et al.*, 2003) suggesting other routes of entry for these drugs and other possible mechanisms of drug resistance.

Two additional pentamidine transporter activities have been biochemically characterised in *T.brucei*, although the genes and natural substrates have not yet been identified. These include a high affinity pentamidine transporter (HAPT1) (De Koning, 2001b; de Koning and Jarvis, 2001), which has also been shown capable of transporting melaminophenyl arsenicals (Matovu *et al.*, 2003) and a low affinity transporter (LAPT1) (De Koning, 2001b; de Koning and Jarvis, 2001) without apparent cross reactivity to other trypanocides. Loss of LAPT1 function has been postulated as a possible mechanism of pentamidine resistance but as yet it has not been experimentally demonstrated (Bridges *et al.*, 2007). The combined loss of both P2 and HAPT1 activity has been observed in both a Cymelarsan selected STIB 386 resistance line and a pentamidine selected *TbAT1* -/- resistant line (Bridges *et al.*, 2007). The exact contribution made by each of the transporters to the drug resistance profile both in isolation and in combination remains to be established, however.

In the parental lines STIB 247 and STIB 386 the *TbAT1* gene is present (Stewart *et al.*, 2005) and transport assays have demonstrated both P2, HAPT1 and LAPT1

activity to be intact (Bridges *et al.*, 2007). Furthermore no cross-resistance was observed between pentamidine, Cymelarsan and Berenil in the MCD experiments, indicating that a different mechanism to the functional loss of drug uptake previously characterised in the literature is likely to be responsible for the difference in sensitivity in pentamidine between the two parental lines. Unravelling the genetic basis responsible for this difference could give insight into potential mechanisms of drug resistance or drug action in this parasite. The ability to map this phenotype using the genetic map however will depend on the ease with which the phenotype segregates in the progeny, and can be accurately measured.

An alternative *in vitro* assay for drug sensitivity screening was developed based on growth inhibition of bloodstream culture cells. This demonstrated that the difference in phenotype is retained *in vitro*, with the growth inhibition curves of the two parents repeatably resulted in a difference in IC_{50} of 3-4 fold. Due to the time constraints of this project, attempts have not yet been made to screen the progeny from the STIB 247 x STIB 386 for inheritance of the parental pentamidine sensitivity trait. Nevertheless the drug action and reduced sensitivity does appear to be retained *in vitro*, suggesting that the *in vitro* growth inhibition assay could substitute for the *in vivo* MCD assay if the progeny were to be phenotype screened in the future.

Substitution of the MCD assay for the growth inhibition assay offers several major advantages. Firstly performing the assay with bloodstream form culture cells reduces the need to use large numbers of laboratory animals. Secondly the *in vitro* assay is rapid; whereas the MCD assays takes > 21 days to complete and is difficult to perform in parallel, results from the *in vitro* assay can be obtained in 48 hours and because they require only a few rows of a 96 well plates, can be easily performed in multiples or repeated to obtain replicate values. As the results of the growth inhibition assay can be expressed as an IC_{50} measurement, this form of drug sensitivity testing would also allow the pentamidine sensitivity trait to be handled as a QTL, allowing more subtle genetic contributions to variation to be elucidated.

In all the traits which differ between STIB 247 and STIB 386 the ability to determine their genetic basis will depend on the mode of inheritance of the genes responsible and the relative magnitude of their effect. If the inherited phenotype has high penetrance and is determined by a single (or small number) of genes for which at the STIB 386 parent is heterozygous, the phenotype should co-segregate

in the progeny with molecular markers close to the gene responsible, and linkage analysis should be relatively straightforward. Positional cloning and reverse genetics could then be used to narrow down and test the candidate genes. If, however, the phenotypically different parents are both homozygous for the alleles responsible or STIB 386 is homozygous and STIB 247 is heterozygous, the phenotype will not segregate in the F1 generation but will require a backcross or F2 cross before the trait can be mapped.

Many traits are complex and expression of phenotype may involve the interaction of segregation alleles at multiple loci. Such traits that are multigenic and quantitative in nature will require a QTL mapping approach to resolve the contribution of different genetic determinants. The accuracy of this approach may depend on the number of progeny available and the number of genes contributing to the trait, but it has been used successfully in other protozoan parasites with a comparable genetic mapping resolution and numbers of progeny (Su *et al.*, 2002; Ferdig *et al.*, 2004; Saeij *et al.*, 2006; Taylor *et al.*, 2006). For example a QTL analysis of Quinine (QN) sensitivity in haploid *P.falciparum* displaying a similar variation in QN sensitivity between the parents (3-fold) was able to identify multiple loci associated with the drug resistance phenotype leading to candidate gene analysis, from the IC_{90} values derived from *in vitro* dose response assays in only 35 progeny clones (Ferdig *et al.*, 2004).

With regards to the traits known to differ between STIB 247 and STIB 386, analysis of progeny clones performed in the laboratory for tsetse salivary gland infectivity and human serum resistance has demonstrate that the phenotype does segregate in the progeny and suggest only a few loci major loci are involved making them amenable to the linkage approach. Virulence and midgut infectivity as with pentamidine resistance at present remain at the parent phenotype level and preliminary analysis of the phenotypes will be required to ascertain the feasibility of genetic analysis for elucidating the genetic basis of these traits.

Chapter 4

The isolation of new progeny clones by *in vitro* bloodstream cloning

4.1 Introduction

A classical genetics approach can be used to determine the genetic basis of phenotypic variation between parental stocks through the generation of a genetic cross and segregation analysis of the resulting progeny. Regions of a genome containing loci that contribute to a phenotype can be located by means of the co-segregation of the phenotype in the progeny with genetic markers from an established genetic linkage map.

The parental strains STIB 247 (*T.b.brucei*) and STIB 386 (*T.b.gambiense*) differ for a number of phenotypes of interest and, as described in the previous chapter, a genetic linkage map for STIB 386 has been generated from 38 F1 progeny isolated from a STIB 386 x STIB 247 cross. The genetic map covers 733 cM with an average resolution of 9 cM and a 90% probability of mapping any loci to within 268 Kb. This is comparable to that of many other published parasite genetic maps (*P.falciparum* (Su *et al.*, 1999), *T.gondii* (Khan *et al.*, 2005), *T.b.brucei* (MacLeod *et al.*, 2005b) but may result in linkage to a locus containing > 100 candidate genes for the next stage of analysis.

The potential of linkage mapping to pinpoint loci that mediate phenotypes of interest is limited by the density of molecular markers and, crucially, the number of recombination events between linkage intervals. This is a product of both the frequency of recombination and the number of progeny available. Positional cloning within the STIB 386 map could therefore be improved by increasing the number of available crossovers within the map, through the identification of additional F1 progeny. This chapter explores the isolation of new hybrids by an *in vitro* cloning approach and their addition to the STIB 386 genetic map.

4.1.1 Introduction to *T.brucei* crosses and methods of hybrid isolation

4.1.1.1 The isolation of hybrids by *in vivo* cloning

Progeny clones can be generated in the laboratory through the co-transmission of two different *T.brucei* strains through the tsetse fly vector. This was first demonstrated by Jenni *et al* in 1986, and genetic analysis has been performed on the products of this cross and a number of other subsequent crosses (reviewed in Gibson and Stevens, 1999; Gibson, 2001). Although the exact mechanism by which

genetic exchange occurs has not yet been determined, this analysis has established that genetic exchange takes place within the salivary glands of the tsetse (Gibson *et al.*, 2006), by a non obligatory process (Jenni *et al.*, 1986; Gibson, 1989; Turner *et al.*, 1990; Gibson and Bailey, 1994), involving both cross and self-fertilisation, to generate parental clones (Jenni *et al.*, 1986; Sternberg *et al.*, 1989), F1 progeny (Jenni *et al.*, 1986; Sternberg *et al.*, 1988; Gibson, 1989; Sternberg *et al.*, 1989) and the products of self-fertilisation (Tait *et al.*, 1996). Statistical analysis of marker segregation in 79 progeny derived from two crosses has also formally demonstrated that allelic segregation and independent assortment occur in ratios consistent with a standard Mendelian system (MacLeod *et al.*, 2005a), a prerequisite for linkage analysis.

Several crosses have been reported between the STIB 247 and STIB 386 strains (Table 4.1), resulting in the identification of a total of 38 unique progeny that were used to construct the genetic map. To generate these crosses, a tsetse fly was co-infected with the two strains, and two different techniques were used to isolate the clonal products from mature infections (Figure 4.1). Individual metacyclic clones (indicated by the prefix mcl in the progeny nomenclature) were obtained from dissected tsetse salivary glands, followed by optical cloning and expansion in individual mice as described in (Van Meirvenne *et al.*, 1975; Jenni *et al.*, 1986; Sternberg *et al.*, 1989). In the second method, the tsetse fly was fed on a mouse to transmit the metacyclics, the uncloned population expanded and then individual bloodstream trypanosomes optically cloned and expanded in individual mice (indicated by the prefix bcl in the progeny nomenclature.(Van Meirvenne *et al.*, 1975; Jenni *et al.*, 1986; Turner *et al.*, 1990) The expanded uncloned bloodstream population derived from each fly was also cryopreserved to enable multiple rounds of cloning to be attempted.

The resulting clones were genotyped with a panel of five polymorphic microsatellite markers to identify new and unique hybrid genotypes. Each of these molecular markers amplifies two different alleles for the STIB 386 parent and a homozygous allele (TB1/1, TB2/20, TB3/13, TB10/14) or two alleles (TB5/4) for the STIB 247 parent. As all the markers are located on a different chromosome and therefore unlinked, this would theoretically allow for the differentiation of 64 possible genotypes from each cross.

Clones which amplified alleles from only a single parent were considered parental clones and discarded. Progeny clones that amplified multiple alleles from one or both parents were considered to be either mixes or polyploid and in most cases

were subsequently recloned and reanalysed to distinguish this. Unique progeny clones were defined as having a different genotype for the five molecular markers from all other progeny derived from the same fly. Progeny clones with an identical genotype but derived from different flies were considered to be independent unique progeny as they were the product of independent crosses and thus separate mating events. Multiple progeny clones that had the same genotype for the five molecular markers and were derived from the same cross were considered to have arisen from the same original progenitor clone and only one example of each was retained. A list of the 38 hybrids generated by this *in vivo* cloning procedure is given in Chapter 2, Table 2.1 and information about the crosses from which they were generated is given in Table 4.1.

It is important to consider however, that these *in vivo* methods, while effective in generating sufficient progeny to confirm Mendelian inheritance (MacLeod *et al.*, 2005a), and construct genetic maps for both *T.b.brucei* (MacLeod *et al.*, 2005b) and *T.b.gambiense* (Cooper *et al.*, 2008) subspecies, are not without their drawbacks. They are time consuming, technically demanding, labour intensive and require a large numbers of animals for clonal expansion. The technical challenge of identifying single trypanosomes by optical cloning methods can results in a high proportion of non-clonal products. Conversely, a large number of trypanosome-negative mice can also result from a failure to successfully inoculate viable trypanosomes or for single trypanosomes to establish patent infections (Taylor, 2004).

It was therefore considered that the development of an alternative *in vitro* method for isolating the products of a genetic cross might have several ethical and technical advantages over traditional *in vivo* protocols.

Uncloned Fly Population	Number of clones analysed	Clone composition				Unique diploid progeny	Reference
		Diploid Progeny clones	Parental clones	Triploid progeny or unresolved mixes	Products of self-mating		
723C	5	0	2	3	0	0	(Jenni *et al.*, 1986; Paindavoine *et al.*, 1986a; Wells *et al.*, 1987)
B80	10	4	6	0	0	1	(Sternberg *et al.*, 1988)
F9/34,41,45	21	18	2	1	0	12	(Sternberg *et al.*, 1989) (Taylor, 2004)
F19/31	16	13	2	1	0	5	(MacLeod, 1999; Taylor, 2004)
F28/46	11	11	0	0	0	6	(Taylor, 2004)
F29/42	5	5	0	0	0	4	(Taylor, 2004)
F57/50	4	0	3	1	0	0	(MacLeod, 1999; Taylor, 2004)
F492/50	47	24	18	1	4	10	(Turner *et al.*, 1990; MacLeod, 1999)
Total	121	75	35	7	4	38	

Table 4.1 A summary of reported STIB 247 x STIB 386 crosses.

The uncloned fly populations are designated by the fly number and the day (post-fly infection) on which the trypanosomes were sampled. In some cases, the same fly was sampled more than once (F9/34, F9/41, F9/45). The number of clonal products analyzed in each experiment is reported alongside the resulting clone composition by genotyping analysis. Mixes that were recloned and resolved to be clonal progeny are recorded only in the progeny clone column. The total number of unique progeny genotypes identified from each cross so far is recorded in column seven alongside the appropriate reference.

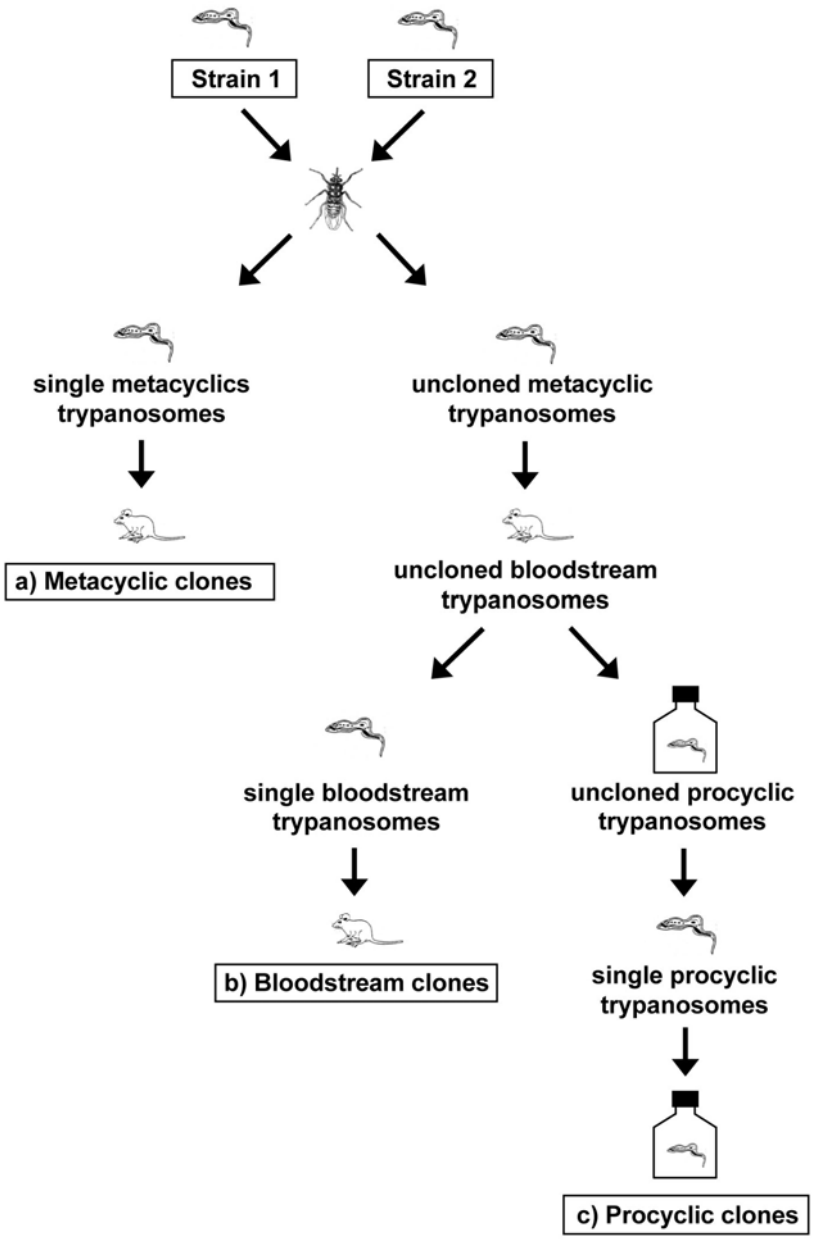


Figure 4.1 Methods used for isolating clonal products from a *T.brucei* genetic cross.

The two *T.brucei* strains are mixed and co-transmitted through a tsetse fly. Single trypanosomes are isolated *in vivo* by optical cloning either (a) from metacyclics taken directly from the salivary glands of the tsetse fly or (b) from the uncloned population resulting from the fly feeding on a mouse. Each single trypanosome is amplified in a mouse. Procyclic clones (c) are isolated by first transferring uncloned bloodstream trypanosomes to *in vitro* procyclic culture by standard methods, followed by limiting dilution cloning in 96 well plates. Clones isolated by all three methods are analysed by genotyping PCR to identify novel progeny genotypes. Adapted from (Tait *et al.*, 2002).

4.1.2 The isolation of clones by *in vitro* procyclic cloning

An attempt to address some of the limitations encountered with the *in vivo* cloning method has been undertaken previously in the laboratory. The implementation of an *in vitro* procyclic cloning approach was proposed for one of the other crosses, a TREU 927 x STIB 247 cross used to generate the *T.b.brucei* genetic map.

In these experiments (Taylor, 2004), procyclic form clones were isolated from the cryopreserved uncloned STIB 247 x TREU 927 cross, fly F532/72 for which a number of progeny had previously been identified. The uncloned bloodstream population, derived from *in vivo* amplification in a mouse, was firstly adapted to *in vitro* procyclic form culture, followed by cloning of culture adapted trypanosomes in 96 well plate cultures. A total of 5000 cloning wells were set up at a one cell/well dilution, of which 9% resulted in trypanosome-positive wells and two-thirds of these were clonal progeny. However of the 309 progeny clones analysed by genotyping analysis only 14 different genotypes were represented and only four of these were new and unique. From these findings it was proposed that the transfer of uncloned bloodstream populations into procyclic form trypanosomes before cloning might have created a bottleneck that substantially reduced the diversity of the population. This potential problem was combined with the limitation that procyclic trypanosomes must be retransmitted through the tsetse fly to generate bloodstream form trypanosomes for *in vivo* growth and phenotype analysis. As a result, *in vitro* procyclic culture cloning was not considered a viable alternative to *in vivo* cloning for generating large number of progeny clones and this approach was not taken forward in the laboratory.

4.1.3 The isolation of clones by *in vitro* bloodstream cloning

Historically all experiments with the bloodstream stage of the parental lines STIB 247, STIB 386 and TREU 927 in the laboratory have been performed *in-vivo* or *ex-vivo.* This is because the long term axenic culture of these trypanosomes could not be established. However, recent experiments by group members A.Tweedie and L.Sweeney has resulted in the successful transfer of parental strains to bloodstream culture, using supplemented culture media and methods adapted from available published protocols (Hirumi and Hirumi, 1989; Carruthers and Cross, 1992). This was a significant technical development for the laboratory and it was considered of interest to determine whether STIB 386 x STIB 247 progeny

might be isolated using an *in vitro* cloning approach combined with this culturing method, as an alternative to *in vivo* cloning in mice. If possible, this would also convey the advantage that, unlike procyclic form culture, bloodstream culture adapted parasites would not require retransmission in tsetse flies to generate bloodstream trypomastigotes.

Uncloned bloodstream populations derived from tsetse flies in which progeny have been previously cloned (and thus mating events are known to have occurred) are available as cryopreserved stabilates in the laboratory. A number of these were chosen to investigate the effectiveness of an *in vitro* bloodstream culture cloning method with the aim of increasing the number of STIB 386 x STIB 247 F1 progeny to assist with linkage mapping of phenotypes of interest in the STIB 386 strain. The following chapter describes the *in vitro* method designed to identify new progeny, microsatellite analysis of the progeny and their addition to the genetic linkage map for STIB 386.

4.2 Results

4.2.1 The isolation of clones by *in vitro* bloodstream cloning

In vitro cloning by limiting dilution (Figure 4.2) was performed on four different uncloned populations, derived from four different tsetse fly crosses, in which STIB 247 x STIB 386 mating events are known to have occurred. Firstly, uncloned parasite populations; F28/46, F29/42, F19/31 and F492/50 were amplified *in vivo*, the infected blood collected and trypanosomes separated from red blood cells by centrifugation. After incubating cells in culture media for up to seven days, they were seeded in 96 well plates, at a range of cell concentrations. The efficiency of *in vitro* cloning in bloodstream culture with these cell lines was unknown so a range of dilutions were prepared (1000, 100, 10, 1, 0.5, 0.25 and 0.1 cells/well), based on haemocytometer counts of the uncloned culture population. Eighty plates (7680 wells) were prepared in total and monitored daily, from day seven onwards, by microscopy for evidence of growing trypanosome cultures.

For all four cross populations, it was observed that trypanosome-positive wells could be detected by microscopy 11 - 21 days after the 96 well plates were prepared. A Poisson distribution calculation was performed at the start of the experiment to predict the polyclonal, monoclonal and empty well distributions expected at different seeding densities (Figure 4.3). Because of the time and labour that would be required to reclone genotypically mixed populations, these predictions were used to formulate a screening policy to determine which cloning plates would proceed to PCR genotyping analysis.

Based on these calculations, it was considered that plates in which < 40% wells were trypanosome-positive (an average of 0.5 cells/well) had an acceptable ratio of empty:clonal:polyclonal wells and that all positive wells from these plates would progress to analysis. Ninety-six wells plates in which > 40% wells were trypanosome-positive were considered likely to have an overabundance of non-clonal wells (>25%) and the entire plate would be discarded.

As predicted by the Poisson distribution, the 28 plates which had been seeded at dilutions of 1000, 100 or 10 cells/well exceeded the 40% trypanosome-positive well cut-off and were therefore discarded from analysis on the basis that they would contain a high proportion of mixed genotypes. Plates containing positive wells, within the 40% cut-off limit, were observed for all plates at the 1 cell/well

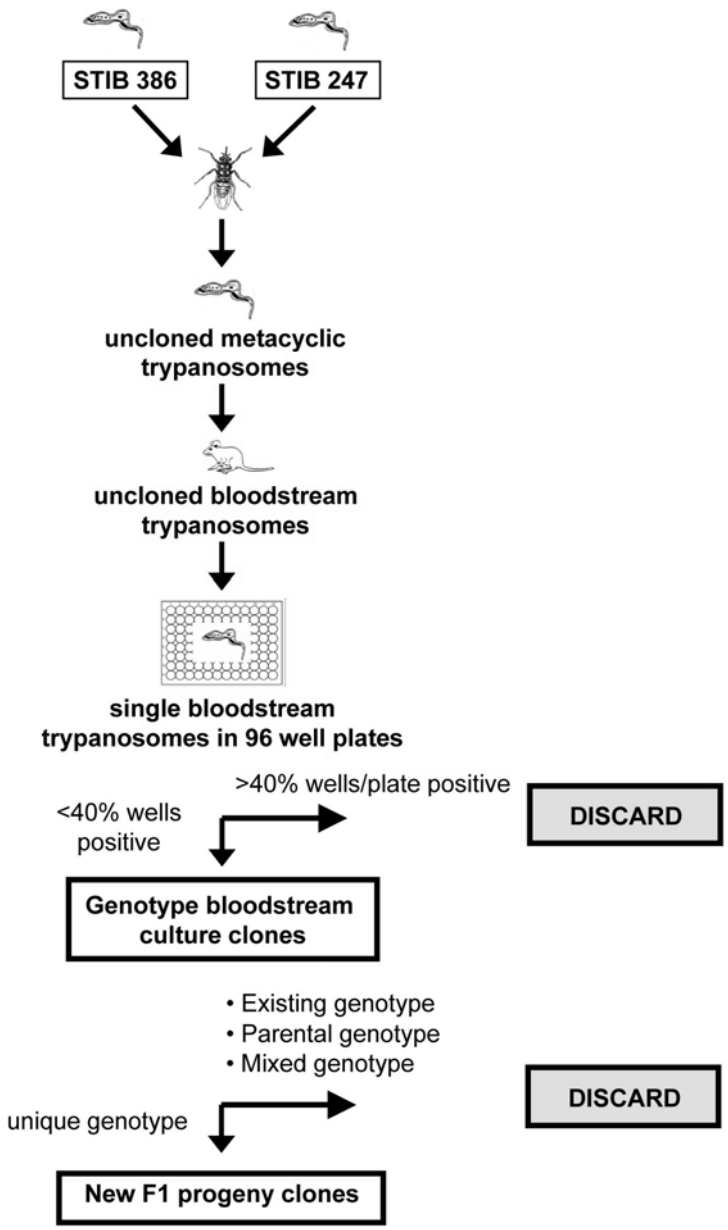


Figure 4.2 The protocol for generating STIB 386 x STIB 247 hybrids by *in vitro* cloning.

Uncloned bloodstream trypanosomes were derived as for the *in vivo* protocol (Figure 4.1), retrieved from cryopreservation and amplified in a mouse. Trypanosomes were then adapted to *in vitro* bloodstream culture and seeded in 96 well plates at a variety of concentrations. Trypanosome-positive wells from plates in which < 40% wells grew up were genotyped by PCR to identify new and unique progeny clones. Clones genotyping as parental, mixed/polyploid or identical to existing F1 progeny were discarded.

dilution and for at least one plate at the 0.5, 0.25 and 0.1 cell/dilutions in each uncloned fly population. In all cases, however, the percentage of positive wells/plate was lower than that predicted by Poisson distribution. This may be due to < 100% efficiency in transfer of bloodstream trypanosomes from *in vivo* growth to *in vitro* bloodstream culture as well as technical inaccuracies in haemocytometer cell counts and dilutions.

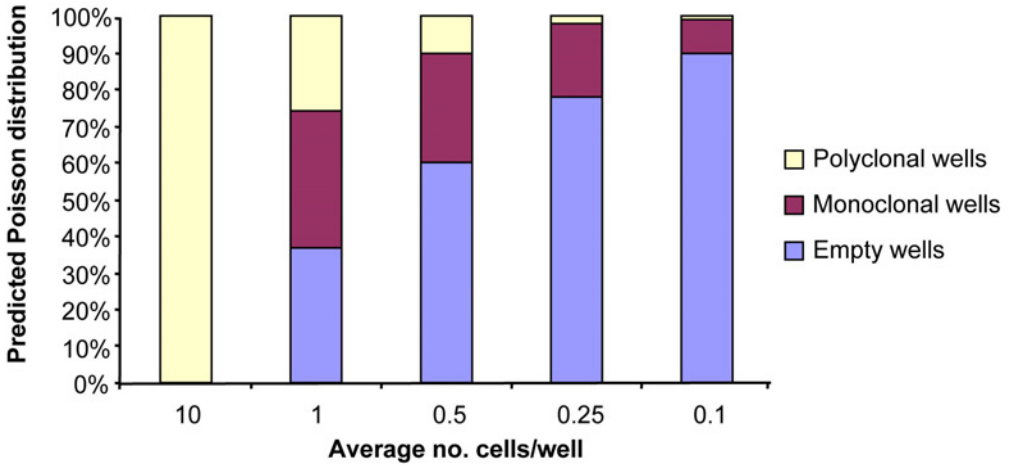


Figure 4.3 The predicted Poisson distribution of *T.brucei* cloning plates prepared at different dilutions in 96 well plates.

The relative percentages of polyclonal, monoclonal and empty wells were calculated at different cell/well dilutions, based on Poisson distribution probability and 100% cell viability. Plate in which < 40% wells were trypanosome positive (equivalent to 0.5 cells/well) were considered to give the optimal ratio of empty:clonal:polyclonal wells.

Once trypanosome-positive well cultures had grown to a sufficient level to be detected by microscopy (~1 x 10^5 cells/ml), 100 µl cultures were harvested from the well and used to prepare a small DNA lysate. This DNA sample was then PCR genotyped with a double heterozygous microsatellite marker (TB5/4) to determine whether the well contained a mixed population, parent or a progeny clone. Although clonality could not be confirmed on the basis of allele inheritance for a single marker, it could be used as a preliminary screening tool before more rigorous genotyping at the next stage of analysis.

An example of screening 13 different well lysates with marker, TB5/4, is shown in Figure 4.4. The genotype of both the parental strains is heterozygous, which results in a total of four different parental alleles (alleles 1 and 2 for STIB 386 and 5 and 6 for STIB 247) that may be inherited in the progeny in four possible

combinations (1-5, 1-6, 2-5 or 2-6). The genotype of 12 lysates in this example are consistent with that of F1 progeny clones, with all four possible genotypes represented. One sample however contains both STIB 386 alleles as well as a STIB 247 allele. This indicates that this well contains either a mixed population or a potentially triploid/trisomic clone.

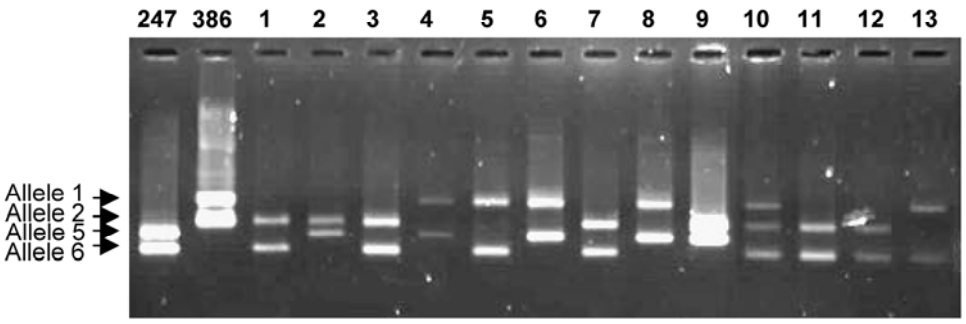


Figure 4.4 PCR microsatellite screening of *in vitro* bloodstream culture clones derived from STIB 247 x STIB 386 genetic crosses.

An example of PCR amplification with the double heterozygous marker TB5/4, used to screen lysates of all trypanosome-positive wells. The parental strains STIB 247 and STIB 386 are shown in the lanes 1 and 2, and 13 *in vitro* derived clones in lanes 1-13. Clones were scored based on the inheritance of parental alleles. Clones exhibiting a parental genotype or amplifying more than two alleles (e.g. lane 10) were discarded from further analysis. Clones exhibiting a progeny genotype (e.g. lane 1-9,11-13) were genotyping with additional microsatellite markers.

Following initial screening, only potential F1 progeny clones were amplified in a greater volume of culture and a larger DNA sample prepared. Further microsatellite screening was performed on these clone samples with the core panel of heterozygous molecular markers used to define all previously isolated F1 progeny clones (TB1/1, TB2/20, TB3/13 and TB10/14). Two additional double heterozygous microsatellite markers (TB10/28 and TB3/1) that had been identified during the construction of the STIB 386 genetic map were also included in the series. This increased the possible number of genotype combinations that could be differentiated from each cross from 64 to 256. The genotype of the parental lines STIB 247 and STIB 386 for these seven markers, with alleles indicated, is shown in Figure 4.5.

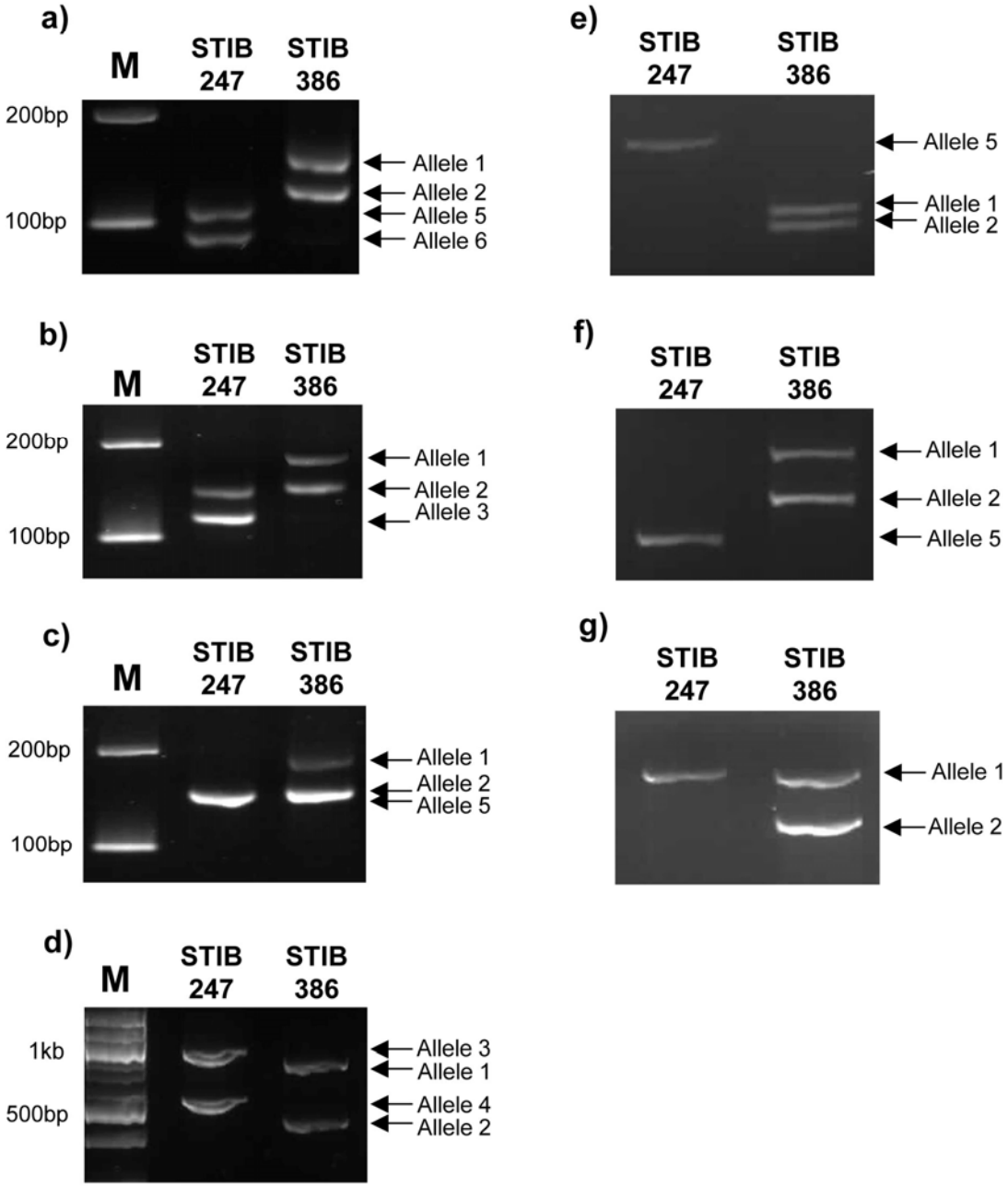


Figure 4.5 PCR amplification of the seven molecular markers for STIB 386 and STIB 247 parental strains used for genotyping *in vitro* clones.

PCR examples of parental stocks with labelled alleles for the microsatellites markers a) TB5/4, b) TB3/1, c) TB2/20, d) TB10/28 and minisatellites e) TB3/13, f) TB1/1, and g) TB10/14. Primer sequences and PCR conditions are given in Appendix 2. M is a 100 bp ladder.

In total, 209 trypanosome-positive wells were analysed from the four different fly populations. The number and composition of these wells is given for each uncloned fly population in Table 4.2 and the complete genotype of each clone screened is given in Appendix 6. One hundred and forty-eight wells (71%) were obtained from the 1 cell/well dilution plates, 44 wells (21%) were obtained from the 0.5 cell/well dilution plates and the remaining 17 wells were obtained from the 0.25 cell/well (5%) and 0.1 cell/well plates (3%).

Twenty-four (11%) wells amplified more than two alleles for at least one of the microsatellite markers and were thus considered "mixed" populations. These wells had all been derived from the more concentrated 1 cell/well dilution (88%) and 0.5 cell/well dilution (12%) and at the approximate level predicted for polyclonal wells by Poisson distribution. While it is possible that these "mixed" wells may have actually contained trisomic or triploid progeny clones, further investigation was considered beyond the scope of this study and all such genotypes were discounted from further analysis.

Only two clones (< 1%) amplified a parental genotype for any of the molecular markers analysed. These two clones were both identical to the STIB 386 parental line and both isolated from the same uncloned cross population (F29/42). Parental clones had not been previously isolated from this cross by *in vivo* cloning. However, genetic exchange has been demonstrated to be a non-obligatory process (Jenni *et al.*, 1986; Gibson, 1989; Turner *et al.*, 1990; Gibson and Bailey, 1994) and as shown in Table 4.1, parental clones have frequently been isolated alongside the products of sexual recombination in seven of the nine STIB 247 x STIB 386 crosses previously analysed.

Four clones previously isolated by optical cloning and amplification *in vivo* from the fly population F492/50 were found to be the products of self-mating (9% clones from this cross). In contrast, none of the 140 clones isolated from this cross (or any of the other crosses) by *in vitro* bloodstream cloning were found to have inherited alleles consistent with self-fertilisation.

The remaining 183 wells (88%) comprised F1 progeny clones. Twenty-two different genotypes were represented, of which 10 were identical to existing progeny isolated in the previous *in vivo* cloning experiments and 12 were new and unique genotypes. The genotype of these 12 new progeny clones for the core panel of microsatellite markers is given in Table 4.3. This increases the number of unique progeny isolated form these crosses in total by 30%, from 38 to 50. Five of these

progeny clones were derived from metacyclic *in vivo* cloning, 33 clones were derived from *in vivo* bloodstream cloning and 12 clones derived from *in vitro* bloodstream culture cloning.

There was some variation in the number of F1 progeny genotypes isolated by the *in vitro* approach from each of the four uncloned fly populations, with the crosses F28/46 and F492/50 producing a much higher number of progeny clones than F29/42 and F19/31. F19/31 in particular appeared to be limited in genotype diversity, with only a single genotype present in the first nine clones isolated from this cross. In part, this appears to be a reflection of the number of trypanosome positive wells analysed in each cross, with F492/50 and F28/46 cloning plates producing a larger number of clones for genotype analysis. However, following the initial genotyping results of F19/31 clones, DNA extracted from flasks containing the uncloned bloodstream population was also PCR genotyped. In this analysis only two alleles were amplified for each microsatellite marker, resulting in a genotype for the uncloned population consistent with that seen in the nine progeny clones isolated from this cross and consequently further genotyping from this cross was abandoned.

Multiple unique progeny had been previously isolated from these uncloned fly populations by *in vivo* bloodstream cloning (four unique genotypes from F29/42 and five from F19/31). However in order to set up the *in vitro* bloodstream cloning plates, the original stabilate had undergone an amplification step, firstly within the mouse and then by *in vitro* incubation of the uncloned culture (2 hr-7 days) to allow the cells to recover before cloning. The potential therefore exists for selection to act on the uncloned population that may have resulted in a population bottleneck, allowing certain progeny genotypes to have become more prevalent in the population and therefore more frequently isolated by cloning.

STIB 247 x STIB 386 uncloned Fly Population	Number of clones analysed from *in vitro* cloning	Composition of *in vitro* cloned genotypes				Total number of unique genotypes identified	Number of new unique genotypes identified
		Progeny clones	Parental clones	Mix or polyploid clone	Products of self-mating		
F28/46	37	34	0	3	0	10	7
F29/42	23	21	2	0	0	2	1
F19/31[1]	9	9	0	0	0	1	0
F492/50	140	119	0	21	0	9	4
Total	209	183	2	24	0	22	12

Table 4.2 Summary of the composition of *in vitro* derived clones for the four different fly populations.

The total number of wells analysed, genotype composition, and number of genotypes is shown for each fly population. The total number of different genotypes is shown in addition to how many of these were new and had not been previously isolated. [1] F19/31 clone analysis was abandoned due to the limited genetic diversity of the uncloned population by genotyping analysis

Lysate	Parents and *in vitro* progeny clones	Genotyping marker scores						
		TB5 /4	TB2 /20	TB1 /1	TB3 /13	TB10 /14	TB3 /1	TB10 /28
	STIB 247 parental	5-6	5-5	5-5	5-5	1-1	2-3	3-4
	STIB 386 parental	1-2	1-2	1-2	1-2	1-2	1-2	1-2
AC NEW 1	F28/46 0.25A C5	2-6	1-5	1-5	1-5	1-2	2-2	1-3
AC NEW 2	F28/46 0.5A E8	1-5	2-5	1-5	2-5	1-2	2-3	1-4
AC NEW 3	F28/46 0.25A C3	1-5	1-5	1-5	2-5	1-1	1-3	2-4
AC NEW 4	F28/46 1A B8	2-5	1-5	1-5	1-5	1-2	1-3	1-3
AC NEW 5	F28/46 0.5B G6	1-6	2-5	1-5	1-5	1-1	2-2	2-4
AC NEW 6	F28/46 1B C8	1-6	2-5	1-5	1-5	1-2	1-2	1-4
AC NEW 7	F28/46 1A G7	1-5	1-5	1-5	2-5	1-2	2-2	1-3
ACNEW8	F29/42 AML DA6	2-6	1-5	2-5	1-5	-	2-3	1-3
ACNEW9	F492/50 15 E3	1-6	2-5	1-5	2-5	1-1	1-2	2-4
ACNEW10	F492/50 0.5 5 F2	2-6	2-5	2-5	1-5	1-1	2-2	2-4
ACNEW11	F492/50 1 8 F7	2-5	1-5	-	-	-	1-3	2-4
ACNEW12	F492/50 1 2 E3	1-6	2-5	1-5	2-5	1-2	2-2	1-3

Table 4.3 Genotype of the 12 new unique hybrids isolated from the STIB 247 x STIB 386 cross by *in vitro* cloning.

Allele inheritance from both parents is shown for the panel of five core genotyping markers (TB1/1, TB2/20, TB3/13, TB5/4 and TB10/14) as well as the two additional double heterozygous markers TB3/1 and TB10/28. Progeny with identical genotypes but isolated from independent experiments (i.e. from different flies) were considered products of separate mating events and thus independent progeny. Dashed lines indicate incomplete genotyping data.

4.2.2 Addition of the new progeny clones to the STIB 386, *T.b.gambiense* genetic map

The *T.b.gambiense* STIB 386 genetic map, as reported in the previous chapter, comprised 119 microsatellite markers in 79 intervals, distributed across the 12 linkage groups of 11 chromosomes. The 38 progeny contain a total of 244 genetic recombination events with an average of 0.6 crossovers/chromosome/individual and a recombination unit size of 24.4 Kb/cM. The twelve new unique progeny isolated by the *in vitro* bloodstream cloning approach, increased the number of progeny available from this cross from 38 to 50. As a step towards the generation of a higher resolution genetic map, the new progeny were genotyped with microsatellite markers for the inheritance of parental alleles, and linkage analysis performed.

To achieve this, bloodstream culture stabilates of each of the progeny clones were revived, grown to high density in culture and genomic DNA extracted. Only one of the new progeny clones stabilates (F492/50 18F7) failed to grow when revived in bloodstream culture. Because of a scarcity of stabilates for this clone, a decision was made not to revive another stabilate for the purposes of this project but to attempt an *in vivo* amplification at a later date. This clone has therefore not been included in any further analysis.

The inheritance of STIB 386 microsatellite alleles, in the remaining 11 F1 progeny was examined for 91 of the molecular markers from the genetic map (at least one microsatellite marker/interval), using the methods previously described. The segregation data for the new progeny clones (Appendix 7) was then combined with the segregation data for the original 38 progeny and the genetic map reconstructed using the Map Manager QTX program (Manly *et al.*, 2001).

The 11 new progeny were integrated into the STIB 386, *T.b.gambiense* genetic map resulting in minor modifications to the original linkage groups. The new genetic linkage maps for each chromosome are shown in Figure 4.6. Synteny and marker order is retained in each new linkage map and the markers remains in 12 linkage groups covering the 11 chromosomes. The genetic distances, based on the number of recombination units between each marker, added together for all 12 linkage groups has increased the total genetic map length from 733.1 cM to 746.3 cM and reduced the average physical size of a recombination unit slightly, from

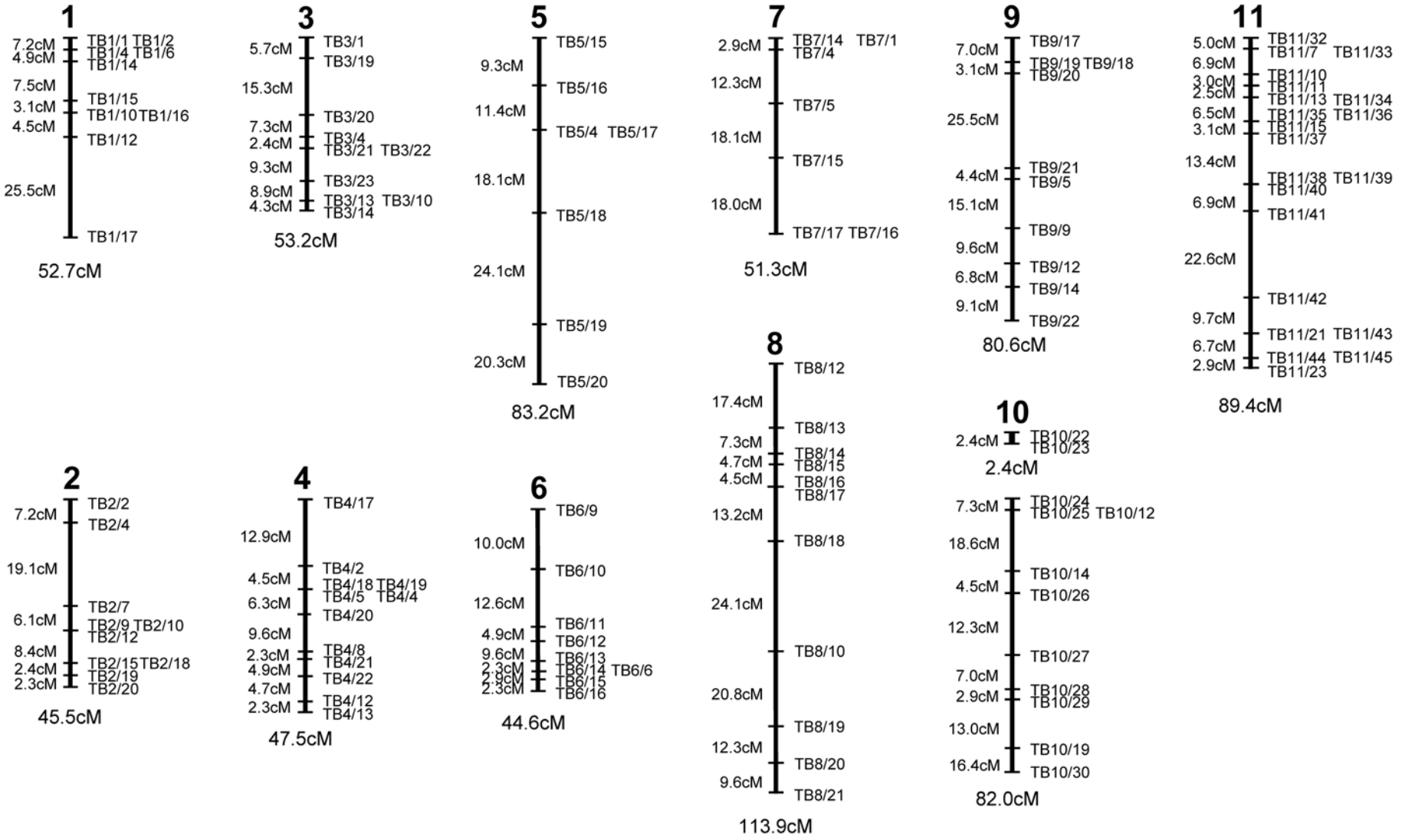

1
7.2cM TB1/1 TB1/2
4.9cM TB1/4 TB1/6
TB1/14
7.5cM
3.1cM TB1/15
TB1/10 TB1/16
4.5cM
TB1/12
25.5cM
TB1/17
52.7cM
3
5.7cM TB3/1
TB3/19
15.3cM
7.3cM TB3/20
TB3/4
2.4cM TB3/21 TB3/22
9.3cM
TB3/23
8.9cM
4.3cM TB3/13 TB3/10
TB3/14
53.2cM
5
TB5/15
9.3cM
TB5/16
11.4cM
TB5/4 TB5/17
18.1cM
TB5/18
24.1cM
TB5/19
20.3cM
TB5/20
83.2cM
7
2.9cM TB7/14 TB7/1
TB7/4
12.3cM
TB7/5
18.1cM
TB7/15
18.0cM
TB7/17 TB7/16
51.3cM
9
7.0cM TB9/17
3.1cM TB9/19 TB9/18
TB9/20
25.5cM
4.4cM TB9/21
TB9/5
15.1cM
TB9/9
9.6cM
6.8cM TB9/12
TB9/14
9.1cM
TB9/22
80.6cM
11
5.0cM TB11/32
6.9cM TB11/7 TB11/33
3.0cM TB11/10
2.5cM TB11/11
TB11/13 TB11/34
6.5cM TB11/35 TB11/36
3.1cM TB11/15
TB11/37
13.4cM
TB11/38 TB11/39
TB11/40
6.9cM
TB11/41
22.6cM
TB11/42
9.7cM
TB11/21 TB11/43
6.7cM
2.9cM TB11/44 TB11/45
TB11/23
89.4cM
2
7.2cM TB2/2
TB2/4
19.1cM
TB2/7
6.1cM TB2/9 TB2/10
TB2/12
8.4cM
2.4cM TB2/15 TB2/18
TB2/19
2.3cM TB2/20
45.5cM
4
TB4/17
12.9cM
4.5cM TB4/2
TB4/18 TB4/19
6.3cM TB4/5 TB4/4
TB4/20
9.6cM
2.3cM TB4/8
4.9cM TB4/21
4.7cM TB4/22
2.3cM TB4/12
TB4/13
47.5cM
6
TB6/9
10.0cM
TB6/10
12.6cM
4.9cM TB6/11
TB6/12
9.6cM TB6/13
2.3cM TB6/14 TB6/6
2.9cM TB6/15
2.3cM TB6/16
44.6cM
8
TB8/12
17.4cM
TB8/13
7.3cM TB8/14
4.7cM TB8/15
4.5cM TB8/16
TB8/17
13.2cM
TB8/18
24.1cM
TB8/10
20.8cM
TB8/19
12.3cM
TB8/20
9.6cM
TB8/21
113.9cM
10
2.4cM TB10/22
TB10/23
2.4cM
7.3cM TB10/24
TB10/25 TB10/12
18.6cM
TB10/14
4.5cM TB10/26
12.3cM
TB10/27
7.0cM
2.9cM TB10/28
TB10/29
13.0cM
TB10/19
16.4cM TB10/30
82.0cM

Figure 4.6 The genetic linkage map of *Trypanosoma brucei gambiense* reconfigured for the addition of 11 new progeny clones.

Microsatellite markers are shown to the right of each linkage group and the corresponding genetic distances between intervals are shown in centiMorgans on the left of each map. The total genetic size of each linkage group is given below the map. Maps were constructed from the segregation analysis of 49 progeny using the program MapManager, with a Haldane map function, at the highest level of significance for linkage criteria, with a probability of type 1 error $P=1e^{-6}$. All markers were mapped with a minimum LOD score of 5.5.

24.40 Kb/cM to 23.97 Kb/cM. The main characteristics of the new linkage groups obtained are summarized in Table 4.4.

Examination of the segregation data (Appendix 7) reveals that 3 - 11 crossovers were detected in each new progeny clone, with an average number of crossovers/chromosome/progeny of 0.61, almost identical to that observed in the original map (0.60). In total the 11 new progeny contribute an additional 74 recombination events to the linkage map and increase the total number of detected crossovers in the map from 244 to 318.

Two of these new recombination events occur between markers that are linked at the same interval on the original map (markers TB6/15 and TB6/16 on chromosome 6, and markers TB4/8 and TB4/21 of chromosome 4), increasing the number of intervals in the genetic map from 79 to 81. Furthermore, there are an additional 14 recombination events that have been linked to regions of the genetic map which contain further groups of markers mapped to single intervals. In many cases, at present, only a single marker per interval has so far been genotyped. Therefore it is possible that completing the genotyping of the new progeny for every marker, at every interval, could locate some of these recombination events between linked markers and add further intervals to the genetic map.

The segregation ratio of alleles from each marker was recalculated with 95% confidence limits. Most loci (108) segregated with the expected 1:1 ratio, with only 11 markers showing significant derivation from Mendelian genetics. The previously reported region at one end of chromosome 1 remains the only section of the STIB 386 genetic map with evidence of any significant segregation distortion and the number of markers skewed in favour of one parental homologue has increased from six to seven (Figure 4.7). As mentioned in the previous chapter, the origin of this distortion is not known but could be the result of post meiotic selection on the uncloned progeny during growth in mice prior to isolation either in vivo or *in vitro*.

Chromosome	Original STIB 386 Map Genetic length (cM)[a]	New STIB 386 Map Genetic length (cM)[a]	Physical size (Mb)[b]	Recombination Frequency (Kb/cM)	Average number of crossover events/meiosis
1	51.2	52.7	0.74	14.04	0.417
2	47.6	45.5	0.74	16.26	0.447
3	46.9	53.2	1.25	23.50	0.469
4	54.4	47.5	1.05	22.11	0.447
5	90.6	83.2	1.2	14.42	0.694
6	42.4	44.6	0.94	21.08	0.375
7	46.9	51.3	1.65	32.16	0.391
8	115.6	113.9	2.3	20.19	0.918
9	73.1	80.6	2.1	26.05	0.667
10[c]	76.1	84.4	2.5	30.49	0.979
11	88.3	89.4	3.42	38.26	0.816
Average				23.97	6.45
Total	733.1	746.3	17.89		

Table 4.4 Characteristics of the genetic linkage maps of *T.b.gambiense* with the new *in vivo* culture derived progeny clones added.

[a] Total genetic length was calculated by the addition of recombination units between each marker as determined by map manager.

[b] Physical distances were calculated from the *T.b.brucei* genome sequence (Berriman *et al.*, 2005).

[c] Chromosome 10 is a combination of two linkage groups.

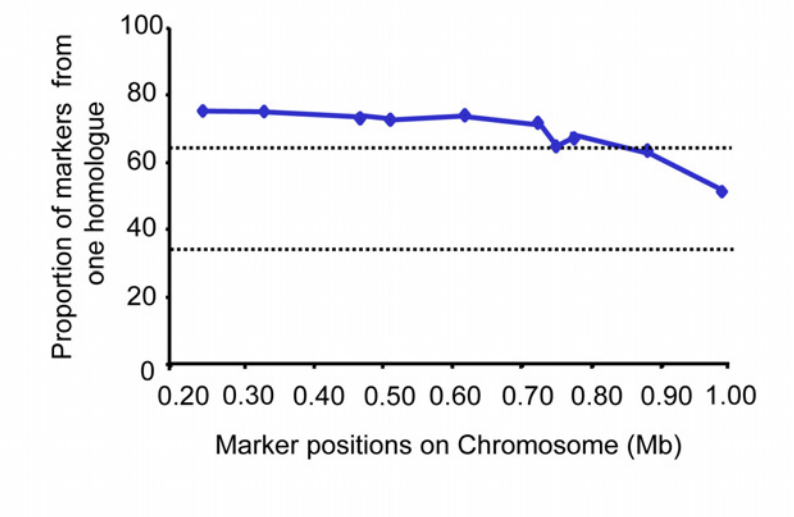


Figure 4.7 Segregation distortion on chromosome 1.

Genotype segregation proportions recalculated for all microsatellite markers present on chromosome 1 incorporating the 11 new progeny clones. Dashed horizontal lines indicate the approximate 95% probability range for equal segregation of alleles.

4.3 Discussion

A 9 cM *T.b.gambiense* genetic map has been constructed, allowing preliminary linkage for mapping phenotypes of interest to within 268 Kb of the gene responsible. Identification of markers that are progressively closer to those genes (fine mapping) will require both additional markers and larger numbers of progeny containing informative crossovers. However, the current *in vivo* methods of isolating progeny are very labour intensive and technically demanding. This chapter investigated an alternative *in vitro* bloodstream cloning method for isolating the products of genetic crosses and resulting in 12 new unique progeny clones. This increased the number of available STIB 247 x STIB 386 progeny from 38 to 50, and they were consequently incorporating into the STIB 386 genetic map.

The number of different genotypes identified per analysed clone by this new *in vitro* bloodstream cloning approach was lower in comparison to that which has been reported for *in vivo* cloning. Although 99% of clones analysed were shown to be the progeny of a cross, only 12% expressed a unique genotype. In comparison, statistics gathered from *in vivo* cloning of crosses of STIB 247, STIB 386 and TREU 927 crossed in all pairwise combinations (Tait *et al.*, 2002), reported that clones derived directly from the salivary glands (metacyclics) contained 27% progeny clones of which 80% were unique and clones derived from bloodstream *in vivo* cloning contained 80% progeny of which 60% were unique.

From these statistics it appears that each additional amplification step of the uncloned cross population; firstly in the salivary glands, then by *in vivo* amplification in mice, and lastly by *in vitro* culture has resulted in a reduced genetic diversity of the population. Correspondingly, the percentage of progeny clones in the uncloned population appears to have increased and the percentage of parental clones appears to have decreased with each additional amplification step. The most likely explanation for these two observations is that selection for particular genotypes is occurring with each expansion, resulting in preferential amplification of certain genotypes and a subsequent reduction in clone diversity. A preferential selection for progeny clones over the two parental lines might likewise be an example of hybrid vigour, in which the progeny exhibit increased fitness over either parent.

In spite of the decrease in unique genotypes identified using the *in vitro* bloodstream culture method, there are also other technical advantages that may make this approach a feasible strategy for the generation of large numbers of progeny clones in the future.

Firstly each successful STIB 247 x STIB 386 cross amplified in a mouse produces $>10^8$ cells which can be cryopreserved for long term storage. Therefore repeated cloning attempts can be made from the same cross to generate an almost limitless number of progeny.

Secondly, because each *in vivo* clone must be amplified in a mouse, there are scale limitations to the number of clones that can be set up at any one time by this approach. The technical challenge of successfully isolating individual trypanosomes by optical cloning, inoculating them into individual mice and developing a patent infection is also labour intensive, time consuming and uses large numbers of animals. *In vivo* cloning experiments in the laboratory have reported as few as 37% of mice developing a patent infection, of which 26% were mixed/trisomic and requiring recloning (Taylor, 2004). The relative ease with which large numbers of cloning plates can be set up at any one time, combined with a screening policy for removal of parental and mixed clones, makes the *in vitro* approach more attractive, despite the decrease in genotype diversity.

Thirdly, the efficiency of *in vitro* bloodstream cloning had not previously been established before starting this experiment and thus a range of dilutions from 1000-0.1 cells/well were analysed. Most of these dilutions proved to be inappropriate (1000, 100 and 10 cells/well) or ineffective (0.5, 0.25 and 0.1 cells/well). The efficiency of the *in vitro* bloodstream cloning technique may therefore be vastly improved in the future by using only the most effective dilution tested (1 cell/well). Based on the figures achieved during this study, setting up the same number of plates at a single dilution would have increased the number of trypanosome positive wells from 209 to 844, and potential number of new hybrids from 12 to 48 (a 4-fold increase).

Fourthly, in the future better use could be made of preliminary PCR microsatellite screening to select only for those clones of interest. Lysates made of individual wells were initially screened with a single marker, TB5/4, to screen out parental lines and mixed wells. This could be extended to screen for either the whole panel of genotyping markers or a particular set of microsatellite markers, flanking a genetic region of interest. This would generate the ability to only grow up those

hybrids likely to be new and unique and/or identify those hybrids with a crossover in the locus of interest saving considerable time and effort.

In chapter three I reported the construction of a genetic map for *T.b.gambiense* from 38 progeny. Here I extend the analysis by reporting a new approach to isolating clones by *in vitro* bloodstream cloning that can be used to identify new progeny clones. Using this approach 12 new progeny clones were successfully identified resulting in the addition of an extra 74 recombination events for mapping analysis. This represented a step towards a second-generation map that in the future, combined with the identification of more molecular markers, will provide a more effective framework for extending studies on the genetic basis of phenotypes such as drug resistance and human infectivity.

Chapter 5

Loss of heterozygosity in laboratory-adapted *T.brucei* strains

5.1 Introduction

Genomic rearrangements, as detailed in previous chapters, can occur in *T.brucei* as a result of sexual recombination when two strains are co-infected in a salivary gland of the tsetse fly vector (Jenni *et al.*, 1986; Gibson *et al.*, 2006). This non-obligatory process adheres to conventional Mendelian principles (MacLeod *et al.*, 2005a) and can result in both F1 progeny clones and the products of self-fertilisation (Tait *et al.*, 1996). However, although the existence of a system of genetic exchange has been proven, the frequency with which it occurs in nature and the role it plays in generating genetic diversity in the field remain under debate (Gibson and Stevens, 1999; MacLeod *et al.*, 2001a).

Primarily, *T.brucei* replicates by mitosis in both the mammalian host and the tsetse fly vector stages of the life cycle. Research into the karyotype of a limited number of laboratory stocks, after prolonged passage in mice, procyclic form culture and transmission through the tsetse fly vector found the karyotype to be mitotically stable (Tait *et al.*, 1996). However, phenotypic changes can be observed in *T.brucei* strains adapted and maintained in laboratory conditions (Lamont *et al.*, 1986; Turner, 1990; Hesse *et al.*, 1995). *T.brucei* also possesses a highly efficient system of homologous recombination for DNA repair and recombination, which is responsible for subtelomeric rearrangements of chromosomes, associated with antigenic variation via VSG switching (described in Chapter 1, Section 1.3).

Much of our current understanding of the processes of antigenic variation, DNA repair and recombination in *T.brucei* has been determined using the laboratory adapted *T.b.brucei* strain, Lister 427. This strain, originally isolated from a sheep in Uganda in 1960 (Cross and Manning, 1973), was the first to be adapted to *in vitro* culture and has been widely disseminated to laboratories around the world, where it continues to be the most commonly utilised strain for research. The 427 line in use in laboratories today is monomorphic, non-tsetse transmissible, antigenically stable and able to grow to a high density in culture and rodents, making it ideal for many experimental uses.

However it has been observed in many organisms that frequent passage, such as by prolonged *in vitro* culture, can cause genomic alterations as cells adapt under the selection pressure of the laboratory environment (Day *et al.*, 1993; Lefort *et al.*, 2008; Spits *et al.*, 2008), and indeed there is some evidence that this might also be the case with laboratory-adapted Lister 427. A genotypic and phenotypic

comparison was recently performed by Wendy Gibson's laboratory between current isolates of this line and an early 427 line cryopreserved in 1964 (Peacock *et al.*, 2008). This confirmed that the historical 427 line was tsetse fly transmissible and genetic exchange compatible, with distinct chromosome size differences by karyotype analysis to its modern day representative. This indicates that significant chromosomal alterations have occurred during its passage history, in correlation with the many phenotypic changes. However, because 427 appear to be highly homozygous and without a full complement of heterozygous genetic markers (Annette MacLeod, personal communication) it is not easy to observe what changes have taken place.

As a result of the genetic maps generated for the megabase genome of *T.b.brucei* line TREU 927 (MacLeod *et al.*, 2005b) and the *T.b.gambiense* line STIB 386 (Chapter 3), an array of heterozygous genetic markers are available for the strains commonly in use in our laboratory. These markers can be used to scrutinise samples of parental and progeny lines for genotype conformation and to screen for contamination or abnormalities. Through this routine strain genotyping over the past few years, several lines have been identified which appeared to have "lost" an allele from a normally heterozygous microsatellite by PCR analysis. These seven lines are composed of a variety of parental and progeny stocks of both *T.b.brucei* and *T.b.gambiense* subspecies, but were mainly isolated after prolonged *in vitro* bloodstream or procyclic culture, and all have experienced an allele loss event on chromosome 10.

The genomic rearrangement that leads to a previously heterozygous genotype becoming homozygous is termed loss of heterozygosity (LOH). LOH events have been identified in a range of eukaryotic cells including the model organisms; *C.albicans* (Diogo *et al.*, 2009), *S.cerevisiae* (Andersen *et al.*, 2008), and *D.melanogaster* (Preston *et al.*, 2006) and multiple mammalian cell lines (Gupta *et al.*, 1997; Blackburn *et al.*, 2004). Such LOH events can be beneficial, as a driving force for evolution in asexually dividing mitotic cells by converting heterozygous beneficial alleles to a homozygous state (Mandegar and Otto, 2007). But also can be detrimental, for example in tumour suppressor genes where LOH is a key event associated with genomic instability and carcinogenesis in somatic cells (Carr and Gottschling, 2008).

LOH is usually the end product of a double strand break (DSB) that can be caused by both endogenous and exogenous sources. This form of DNA damage has the potential to be particularly dangerous if left unchecked, as it compromises both

information-containing strands of the DNA. There are two main mechanisms by which LOH can potentially result from a DSB; the loss of part or all of a chromosome due to an unsuccessfully repaired break (a hemizygous state) or a homologous recombination event that replaces DNA with sequence from a homologous chromosome (a homozygous state, termed true LOH).

Multiple pathways of homologous recombination have been identified in eukaryotes. These include non-reciprocal mechanisms (in which DNA transfer is uni-directional) for example via a gene conversion over short distances, or break-induced replication (BIR) over much longer distances that extend to the telomere (Malkova *et al.*, 1996; Paques and Haber, 1999). Alternatively, recombination can occur though a reciprocal pathway, in which both homologues are exchanged, via a mitotic crossover. These mechanisms have been extensively reviewed (Paques and Haber, 1999; Haber, 2000; Helleday, 2003) and examples of how each could give rise to a homozygous microsatellite pattern are demonstrated in Figure 5.1.

Studies of artificially induced DSBs, VSG switching and targeted transfection in trypanosomes have indicated homologous recombination, utilising short-tract gene conversion to be the dominant pathway in *T.brucei* (McCulloch and Barry, 1999; Conway *et al.*, 2002; Glover *et al.*, 2008). However, much of our current thinking regarding repair of DSBs in *T.brucei* has originated from artificially introducing genetic material by transfection (Conway *et al.*, 2002; Barnes and McCulloch, 2007) and experimentally inducing DSBs in the Lister 427 line (Glover *et al.*, 2008; Boothroyd *et al.*, 2009), which may not fully reflect the natural situation in trypanosomes. The absence of sufficient genetic markers to observe events, and the relative infrequency of the recombination event, means that there has been little or no study of spontaneous genomic rearrangements in trypanosomes, and where undertaken this has been limited to VSG switching events at subtelomeres (Pays *et al.*, 1985; Shea *et al.*, 1986; Rudenko *et al.*, 1996).

The observation of several pleomorphic *T.brucei* lines exhibiting spontaneous LOH in the laboratory and for which genetic markers are available was therefore of interest. The following chapter details the study of seven laboratory-adapted *T.brucei* lines that were observed to exhibit LOH of up to 3 Mb on chromosome 10 following routine microsatellite genotype analysis. A number of these lines were investigated in detail by PFGE analysis of karyotype, Illumina sequencing, and SNP-RFLP analysis to determine the extent and cause of LOH. Finally, as LOH in these lines had emerged in a previously heterozygous population, a selective growth advantage resulting from the chromosome 10 rearrangement was

hypothesized, and tested in two lines exhibiting LOH, compared to their standard equivalents.

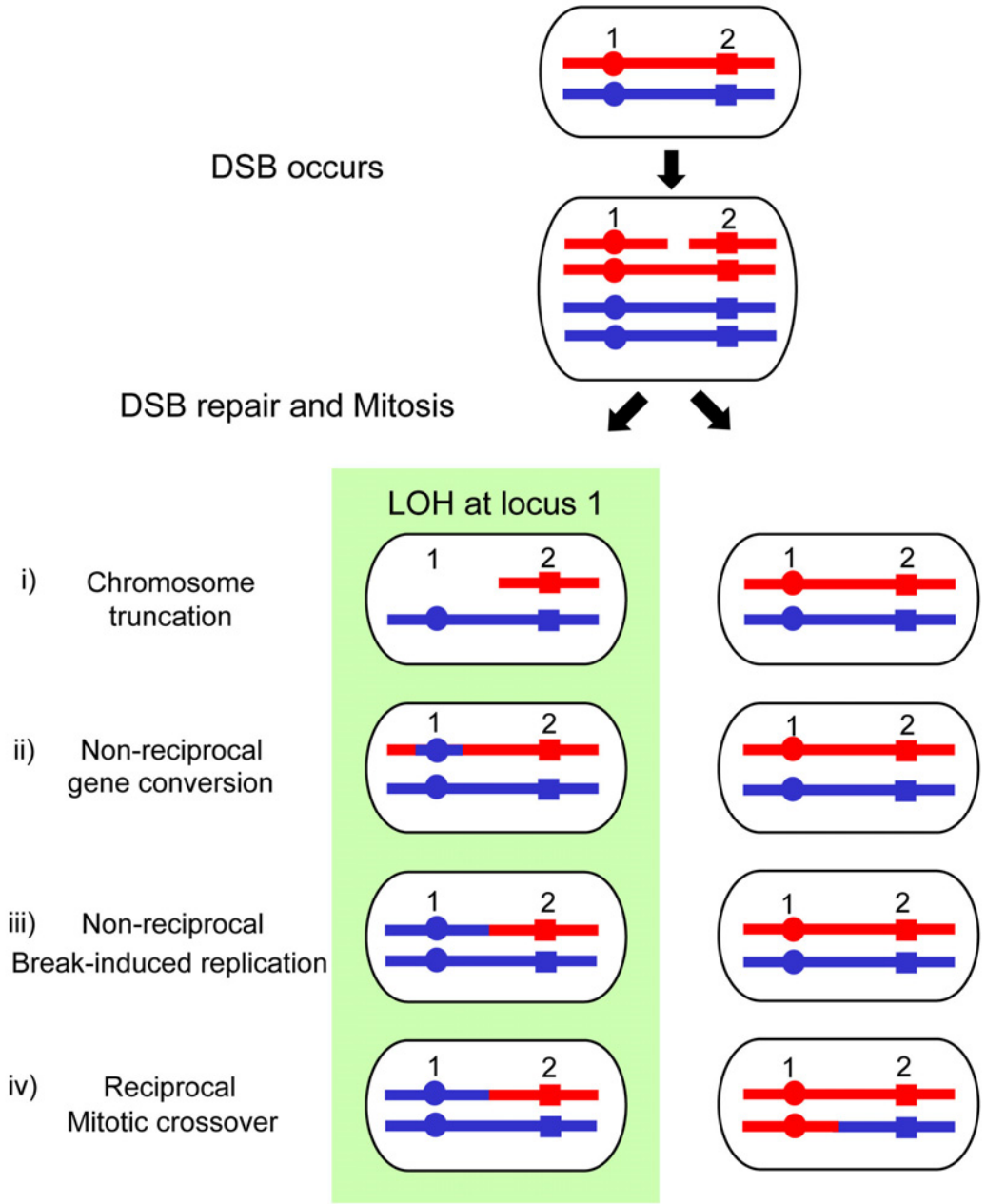

1
2
DSB occurs
1
2
DSB repair and Mitosis
LOH at locus 1
i)
Chromosome
truncation
ii)
Non-reciprocal
gene conversion
iii)
Non-reciprocal
Break-induced replication
iv)
Reciprocal
Mitotic crossover

Figure 5.1 Proposed mechanisms by which a LOH genotype may be generated following a DSB.

Different mechanisms of DSB repair have been observed in diploid eukaryotic organisms that can result in various chromosomal rearrangements. Those mechanisms, which may result in a LOH genotype, are indicated in the schematics i-iv. In these schematics the two chromosome homologues are represented by red and blue lines and the location of two heterozygous microsatellite loci as a filled circle and square on each homologue. Each of the homologues is replicated and one of the red homologues undergoes a DSB. The various outcomes of the subsequent DSB repair and mitotic division in the daughter cells are indicated. i) Chromosome truncation resulting from recovery of a broken chromosome by addition of telomeric repeats, ii) non-reciprocal gene conversion in which a recombination mediated mechanism repairs the break by replacing short tracts of sequence with a copy of sequence from the homologous chromosome, iii) Non-reciprocal BIR repair in which copying of the homologous chromosome proceed all the way to the telomere without reciprocal exchange and iv) A reciprocal homologous recombination mechanism in which a mitotic crossover leads to reciprocal exchange of DNA and a genetic LOH in both daughter cells.

5.2 Results

5.2.1 Microsatellite genotyping of LOH lines

As part of standard laboratory protocol, cryopreserved stabilates derived from *in vivo* passage, bloodstream form culture and procyclic form culture are genotyped with microsatellite markers from the TREU 927 and STIB 386 genetic maps (MacLeod *et al.*, 2005b; Cooper *et al.*, 2008), to confirm that stocks have been correctly catalogued and to ensure strain clonality is maintained.

One-hundred and eighty heterozygous markers are available from the TREU 927 map and 120 from the STIB 386 map. Approximately 10 of these markers are also heterozygous for the predominately homozygous STIB 247 strain. The STIB 247, TREU 927, and STIB 386 parental lines plus all ~100 progeny clones from the crosses can be uniquely identified using seven of the microsatellite markers. All seven markers are heterozygous at each identified locus for TREU 927 and STIB 386 and heterozygous for 3/7 loci in STIB 247. The PCR primers for these seven markers and their relative position on the genome strain, TREU 927, megabase chromosomes are given in Appendix 2 and Figure 5.2, respectively.

Although seven markers are required to distinguish one line from all others, in practise, as the expected genotype of a sample is already known, less than seven markers are generally used during routine screening, unless a problem is suspected. As a result, approximately 500 different DNA samples derived from blood, procyclic form culture or bloodstream form culture have each been genotyped with 3-7 genotyping markers. This sampling identified seven lines (Table 5.1) which repeatedly failed to amplify one of the expected alleles at either one or both of the commonly used chromosome 10 genotyping loci, TB10/14 and TB10/28, resulting in an apparently homozygous rather than the expected heterozygous PCR product.

The failure of an allele to amplify by PCR (allelic drop-out) can be caused by a number of factors including; a mutation in the PCR primer site, sub-optimal PCR conditions, low quality/quantity of DNA, or the absence of the target locus in the genomic DNA (reviewed Bonin *et al.*, 2004). PCRs were thus repeated with variations in DNA concentration, primer and PCR conditions which confirmed that this failure to amplify was not an artefact of the PCR protocol used and consequently these loci were classed as homozygous.

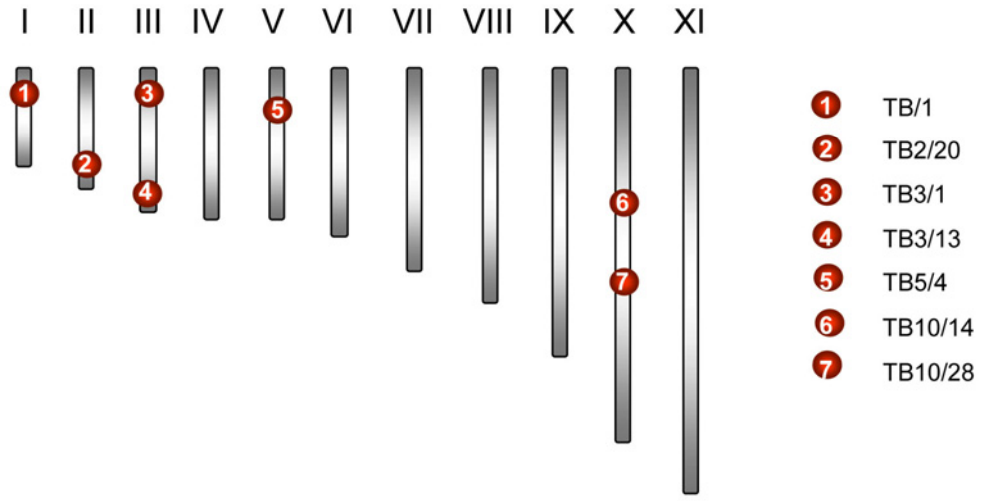


Figure 5.2 The approximate position of seven commonly used genotyping markers on the 11 megabase chromosomes of *T.brucei* TREU 927.

The approximate location of the seven polymorphic genotyping markers; TB1/1, TB2/20, TB3/1, TB3/13, TB5/4, TB10/14 and TB10/28 commonly used for screening samples in the laboratory are indicated as filled red circles. The 11 megabase chromosomes are indicated (in haploid form), appropriately scaled.

Marker	Genotype of standard parental lines			Genotype of lines exhibiting allelic drop-out						
	STIB 247	TREU 927	STIB 386	TREU 927 LOH line 1	TREU 927 LOH line 2	STIB 247 x STIB 386 progeny F29/46 bcl 1 LOH line 1	STIB 247 X STIB 386 progeny F28/46 bcl 6 LOH line 1	STIB 247 LOH line 1	STIB 247 MelR LOH line 1	STIB 386 LOH line 1
TB1/1	5-5 Hom	3-4 Het	1-2 Het	3-4 Het	3-4 Het	1-5 Het	1-5 Het	5-5 Hom	5-5 Hom	ND
TB2/20	5-5 Hom	3-4 Het	1-2 Het	3-4 Het	3-4 Het	1-5 Het	2-5 Het	5-5 Hom	5-5 Hom	ND
TB3/1	3-4 Het	5-6 Het	1-2 Het	5-6 Het	5-6 Het	2-3 Het	2-2 Het	3-4 Het	3-4 Het	1-2 Het
TB3/13	5-5 Hom	3-4 Het	1-2 Het	3-4 Het	3-4 Het	2-5 Het	1-5 Het	5-5 Hom	5-5 Hom	ND
TB5/4	5-6 Het	3-4 Het	1-2 Het	3-4 Het	3-4 Het	1-5 Het	2-5 Het	5-6 Het	5-6 Het	1-2 Het
TB10/14	1-1 Hom	3-4 Het	1-2 Het	3- Hom	3- Hom	1- Hom	2- Hom	1-1 Hom	1-1 Hom	ND
TB10/28	3-4 Het	5-6 Het	1-2 Het	3-4 Het	3- Hom	1- Hom	1- Hom	3- Hom	3- Hom	2- Hom

Table 5.1 Genotype of the standard parent lines and lines exhibiting allele drop-out by PCR.

Genotype of the standard parent lines, STIB 247, TREU 927 and STIB 386 and the seven lines which failed to amplify expected alleles, for the seven common genotyping microsatellite markers. The alleles amplified for each microsatellite marker are indicated by allele number (as shown in Chapter 4, Figure 4.5) and whether the locus is heterozygous (Het) or homozygous (Hom). A failure to amplify both expected alleles of a normally heterozygous marker is identified by shading in grey. ND (not done) indicates microsatellite loci in a recently identified STIB 386 line for which the PCR genotyping has not yet been performed. Details on the PCR primer sequences and locations of all microsatellite markers are in given in Appendix 2. Further information about the seven LOH lines is given in Table 5.2.

Following these initial genotype observations, the next step was to investigate whether the switch from heterozygosity to homozygosity in these seven lines was restricted to individual genetic loci which might indicate a mutation or short gene conversion event, or extended to loci at a considerable distance from the original marker locus. To determine this, it was desirable to genotype markers up and downstream of the standard genotyping markers and on the other megabase chromosomes. Of the seven lines identified exhibiting allele drop-out on chromosome 10, five were derived from either the TREU 927 or STIB 386 parental lines or crosses involving these lines as parents, and thus could be genotyped using the multiple microsatellites identified from the genetic maps. The remaining two lines were derived from the parental STIB 247 line, which is largely homozygous, and thus could not be investigated in further detail.

Of the five strains, four were fully analysed for all chromosome 10 microsatellite loci as well as markers at the distal ends and middle of the other 10 megabase chromosomes. A fifth strain, STIB 386 LOH line 1 was only identified towards the end of the project and thus although LOH has been confirmed for several markers on chromosome 10, a full genotyping analysis of all chromosomes has not yet been undertaken.

An example of the microsatellite genotyping is shown in Figure 5.3, which demonstrates amplification of two chromosome 10 markers in both a standard and LOH line derived from procyclic culture of strain TREU 927. For one marker, TB10/21, situated at the right hand end of chromosome 10, both stabilates amplify heterozygous loci as expected. However for the other marker, TB10/28, situated further towards the centre of chromosome 10, the standard parental line amplifies two alleles, while TREU 927 LOH line 2 amplifies only a single allele.

The genotyping results confirmed that for all four lines investigated, LOH was not restricted to a single microsatellite marker but extended from a centrally positioned marker on chromosome 10 to one of the distal ends of each genetic map. These LOH events occurred on and were confined to Chromosome 10 homologues in all samples, with no other LOH events identified on any of the other megabase chromosomes. Loss of the entire chromosome was also not observed, with all samples retaining heterozygosity for at least two microsatellite loci on chromosome 10. The LOH tracts appear to be continuous in every case with heterozygosity maintained at the extreme right hand end of the chromosome and varying amounts of LOH occurring to the left hand end.

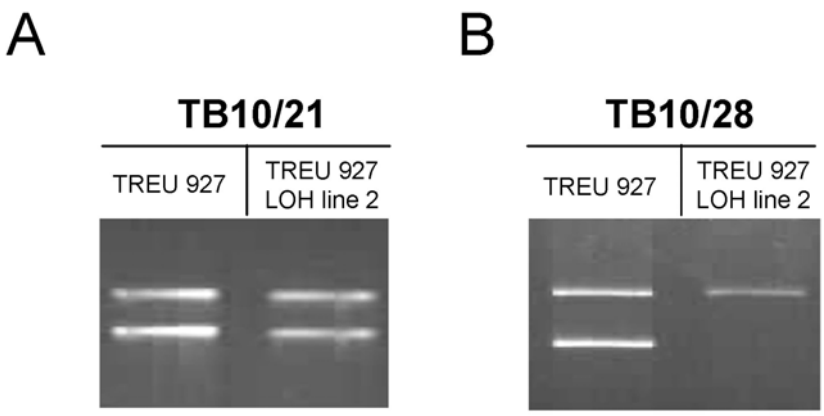


Figure 5.3 LOH in a TREU 927 procyclic form culture line at one microsatellite locus.

Microsatellite PCR products are shown for two genomic loci on chromosome 10, marker TB10/21 (A) and marker TB10/28 (B), the primer sequence and location for which are detailed in Appendix 2. Genomic DNA was amplified from a standard TREU 927 procyclic control (stabilate TGG737), labelled TREU 927 in this figure, and TREU 927 LOH line 2 procyclic culture (stabilate TGG716), labelled TREU 927 LOH 2 in this figure, for each marker. Homozygosity in TREU 927 LOH line 2 is evident at locus TB10/28 by the lack of amplification of the lower allele that is present in the TREU 927 control.

This is demonstrated in detail for two TREU 927 LOH lines and a STIB 247 x STIB 386 progeny clone F29/46 bcl 1 line in Figure 5.4, showing where loss of alleles occurs in microsatellite markers on chromosome 10 compared to wildtype isolates. LOH could be traced to the farthest left hand genetic marker of both the TREU 927 and STIB 386 maps, which are situated at genomic position 124, 019 bp in STIB 386 and 90, 414 bp on the genome strain TREU 927 chromosome homologues.

The amount of heterozygous sequence lost, was estimated from chromosome end to first heterozygous marker position. These regions of LOH are variable but overlap in all lines, with estimated sequence loss of 2-3.5 Mb or 50-75% of the total estimated chromosome size. In addition, three of these stocks possess LOH boundaries (the region between last heterozygous marker and first homozygous marker) that overlap on chromosome 10. The amounts of predicted homozygosity along with further information on the strains and stabilates involved is detailed in Table 5.2 for all strains.

The long, continuous tracts of LOH events identified on chromosome 10 in these lines indicated the LOH was not the result of point mutations, or short tract gene conversion which had previously been cited as one of the main mechanism for DSB repair (Glover *et al.*, 2008). It was therefore considered possible that LOH resulted either due to a chromosome truncation or a mitotic recombination resolved with crossover or break-induced replication. In order to determine whether the LOH strains exhibited major modifications of the karyotype associated with chromosome truncation, PFGE was performed.

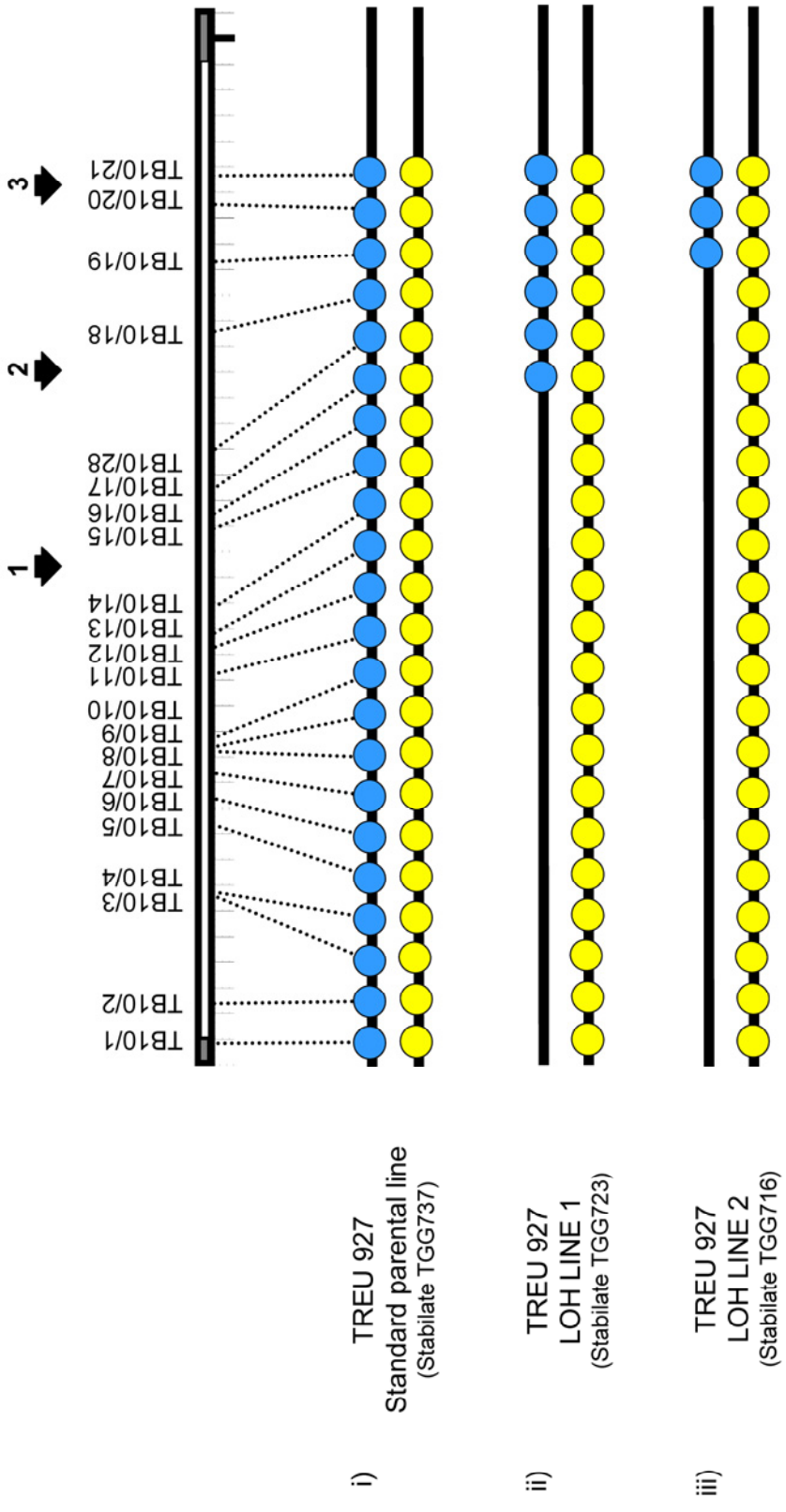

1
2
3
TB10/1
TB10/2
TB10/3
TB10/4
TB10/5
TB10/6
TB10/7
TB10/8
TB10/9
TB10/10
TB10/11
TB10/12
TB10/13
TB10/14
TB10/15
TB10/16
TB10/17
TB10/28
TB10/18
TB10/19
TB10/20
TB10/21
i)
TREU 927
Standard parental line
(Stabilate TGG737)
ii)
TREU 927
LOH LINE 1
(Stabilate TGG723)
iii)
TREU 927
LOH LINE 2
(Stabilate TGG716)

Figure 5.4a Microsatellite evidence for LOH in TREU 927 procyclic culture lines.

Summary of PCR microsatellite results on chromosome 10 for three different TREU 927 procyclic culture lines; standard parental line TREU 927 (stabilate TGG737), TREU 927 LOH line 1 (stabilate TGG723), and TREU 927 LOH line 2 (stabilate TGG716). The two homologues of TREU 927 chromosome 10 were determined by genetic mapping of a TREU 927 x STIB 247 cross (MacLeod *et al.*, 2005b). Yellow circles indicate amplification of homologue B alleles at each PCR microsatellite locus and blue circles indicate homologue A alleles. Amplification of both alleles indicates heterozygosity, amplification of only a single allele indicates LOH. The name and genomic location of each marker on chromosome 10 is indicated at the top of the schematic as described by Macleod *et al*, (2005). PCR primers and genomic locations are given in Appendix 2. The black block arrows labelled 1-3 indicate the relative positions of Southern blot probes used in Southern blot karyotype analysis as described in Table 5.4 and Section 5.2.2.

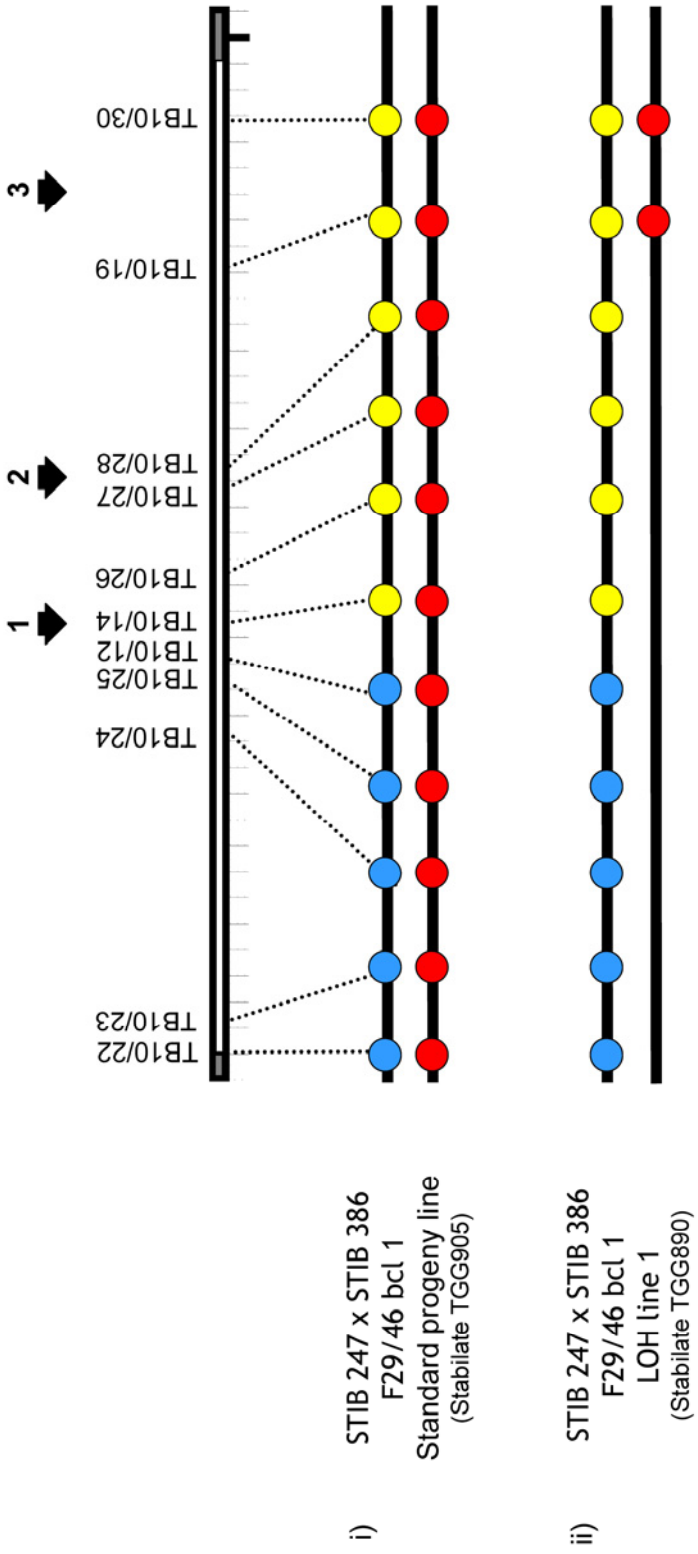

TB10/30
TB10/19
TB10/28
TB10/27
TB10/26
TB10/14
TB10/12
TB10/25
TB10/24
TB10/23
TB10/22
3
2
1
i)
STIB 247 x STIB 386
F29/46 bcl 1
Standard progeny line
(Stabilate TGG905)
ii)
STIB 247 x STIB 386
F29/46 bcl 1
LOH line 1
(Stabilate TGG890)

Figure 5.4b Microsatellite evidence for LOH in a STIB 247 x STIB 386 progeny clone F29/46 bcl 1 procyclic culture line.

Summary of PCR microsatellite results on chromosome 10 for two different progeny clone F29/46 bcl 1 procyclic stabilates: standard progeny clone F29/46 bcl 1 (stabilate TGG905) and LOH progeny clone F29/46 bcl 1 (stabilate TGG890). The genotype at each chromosome 10 marker locus for the control clone was determined during the generation of the STIB 386 genetic map (Chapter 3, Cooper *et al* (2008). Yellow circles indicate inheritance of a STIB 386 homologue B allele and blue circles indicate inheritance of a STIB 386 homologue A allele at each PCR microsatellite locus. Red alleles indicate the inheritance of a STIB 247 allele (homologue A and B indistinguishable for the majority of markers due to the homozygosity of STIB 247 line). Amplification of both a STIB 247 and STIB 386 allele indicates heterozygosity at a locus, amplification of only a single allele indicates LOH. The name and genomic location of each marker on Chromosome 10 is given at the top of the schematic. Further details on each marker including PCR primers and sequence is given in Appendix 2. The black block arrows labelled 1-3 indicate the relative positions of Southern blot probes used in Southern blots to analyse karyotype as described in Table 5.4 and Section 5.2.2.

T.brucei line	Life cycle stage	Sub-species	Estimated LOH*	Stabilate numbers used in analyses	Stabilate numbers of standard line (non-LOH) used in analyses
TREU 927 LOH Line 1	Procyclic culture	*T.b.brucei*	50%	TGG723 TGG821	TGG737 TGG731
TREU 927 LOH Line 2	Procyclic culture	*T.b.brucei*	60-70%	TGG716 TGG805	TGG737 TGG731
STIB 247 x STIB 386 F29/46 bcl 1 LOH line 1	Procyclic culture	*T.b.brucei* x *T.b.gambiense* hybrid	50-70%	TGG890	TGG905
STIB 247 x STIB 386 F28/46 bscl 6 LOH line 1	Procyclic culture	*T.b.brucei* x *T.b.gambiense* hybrid	50-70%	WCMP4463	Not required
STIB 247 LOH line 1	Bloodstream culture	*T.b.brucei*	n/a	TGG0477	Not required
STIB 247 MelR LOH line 1	Bloodstraw selected for melarsoprol resistance in vivo by subcurative dosage	*T.b.brucei*	n/a	TGG732	Not required
STIB 386 LOH line 1	Bloodstream culture	*T.b.gambiense*	50%	TGG044	Not required

Table 5.2 Details of the seven LOH lines identified by routine microsatellite genotyping in the laboratory.

All seven lines are pleomorphic and exhibit LOH events on chromosome 10. Estimates of the amount of the chromosome that have become homozygous, based on microsatellite genotyping results, are given where enough markers are available or labelled n/a where not available. *Estimated LOH is based on the TREU 927 genome sequence.

5.2.2 Karyotype analysis of LOH isolates by PFGE

There are two proposed mechanisms by which the normally heterozygous chromosomes could become genotypically homozygous; the loss of part of one homologue resulting in hemizygosity, or a homologous recombination event that results in replacement of sequence from one homologue with a copy of the other homologue, causing allelic LOH while chromosomes remain diploid.

Chromosome loss or truncation can occur if chromosomes segregate aberrantly during mitosis or a double strand break is not correctly rejoined or repaired. Such losses can be detrimental or even fatal, particularly in haploid organisms in which only one copy of each chromosome is present. However they have been observed to occur in parasitic organisms in laboratory culture conditions without detriment to survival. For example, it has been well documented that *in vitro* mitotic propagation of the haploid apicomplexa parasite, *P.falciparum* is frequently associated with subtelomeric deletions resulting from spontaneous DSBs and their subsequent healing by addition of telomeric repeat sequences (Pologe and Ravetch, 1988). These chromosome truncations often result in the loss of virulence genes (Shirley *et al.*, 1990; Scherf and Mattei, 1992; Scherf *et al.*, 1992; Day *et al.*, 1993), the best characterised of which is *KAHRP* on chromosome 2 (Pologe and Ravetch, 1986; Biggs *et al.*, 1989) which is a component of the knob structure associated with cytoadherance. Parasites with this deletion possess a growth advantage *in vitro* but are non-viable *in vivo* (Barnwell *et al.*, 1983).

The potential extent of chromosome loss predicted from the microsatellite data in the *T.brucei* lines is considerably greater than that seen in *P.falciparum*, extending beyond the subtelomeres to encompass up to 70% of the homologue. However the diploid nature of *T.brucei,* with two copies of each gene, means that such a loss may not have such an extreme effect, providing one copy of essential genes remained. Indeed there are examples of *T.brucei* clones in the literature that have undergone radical chromosomal rearrangements. For example mutants with a disruption of the Mre11 component of the RAD51 pathway have been shown to exhibit a significant loss of chromosomal material, yet remain viable, albeit with a significantly reduced growth rate (Robinson *et al.*, 2002).

In order to assess whether the LOH *T.brucei* lines had undergone a significant truncation of chromosome 10, the megabase karyotype of a number of them were analysed. Classical cytogenetic analysis of the nuclear genome is impossible

because the chromosomes of *T.brucei* do not readily condense and thus cannot be easily visualised. Pulsed field gel electrophoresis is a technique that allows the separation of large molecules of DNA, such as intact chromosomes, on agarose gels. By using a program optimised to target separation of the size of chromosome of interest, it allows significant changes in chromosome size to be detected. To determine if deletion accounts for LOH, PFGE separation of a selection of the LOH lines were undertaken in comparison to standard lines.

Three isolates of interest were chosen for investigation, which exist as procyclic form stabilates (and thus grow to a high density) and can be paired with an equivalent normal isolate for comparison. The two different TREU 927 LOH lines selected have different LOH boundaries and are derived from the genome sequence strain, and thus the exact marker and probe positions can be identified. In addition a non-genome strain was chosen for comparison; one of the *T.b.brucei* x *T.b.gambiense* progeny clones, F29/46 bcl 1 for which a normal procyclic form line also exists.

Chromosome 10 is one of the largest sized chromosomes of *T.brucei* in all parental lines (Melville *et al.*, 1998). The two homologues cannot be separated by PFGE from each other in any of the lines, but can be separated from all the other chromosomes for TREU 927 and from all but chromosome 11 for STIB 247 and STIB 386 (Melville *et al.*, 1998). As standard parental line TREU 927 and progeny clone F29/46 bcl 1 each contain chromosome 10 homologues of equal size, they will appear together as a single band after PFGE separation. LOH due to chromosome truncation, of the amount predicted by the microsatellite data (Table 5.3), should thus be easily visible by PFGE and Southern blotting as a change in size in one of the homologues resulting in two different sized chromosome bands.

Chromosome 10 probes for detection following Southern blotting were designed against target sequence in both the heterozygous and homozygous side of the predicted LOH boundary (Table 5.4 and Figure 5.4a). An additional probe was also designed for the region between the predicted LOH boundaries for the two different TREU 927 LOH examples. Because of the similar regions of predicted LOH in the three lines, the same probes were also appropriate for detection of chromosome 10 homologues in the STIB 247 x STIB 386 progeny clone F29/46 bcl 1 LOH isolate (Figure 5.4b).

LOH strain	Predicted size of normal chromosome 10 homologues	Predicted size of chromosome 10 homologues if chromosome is truncated
TREU 927 LOH line 1 (Stabilate TGG723)	4.4 Mb	1.9-2.3 Mb
TREU 927 LOH line 2 (Stabilate TGG716)	4.4 Mb	1.0-1.6 Mb
Progeny clone F29/46 bcl 1 LOH line (Stabilate TGG890)	> 5.7 Mb	1-2 Mb*

Table 5.3 Predicted sizes of normal and LOH chromosome 10 homologues by PFGE.

Chromosome sizes are predicted, based on microsatellite results, for if LOH had occurred as a result of chromosome truncation. The predicted sizes of normal chromosomes 10 homologues are based on Melville *et al*, (1998). The predicted sizes of LOH chromosomes are based on the genomic position of microsatellite markers in the *T.b.brucei* TREU 927 genome sequence and therefore may differ for STIB 247 x STIB 386 progeny clone F29/46 bcl 1*.

Probe	Probe and primer names	Probe primer sequence	Primer location*	PCR product size (bp)
1	Ch10/CRAM			3017
	Ch10/CRAM-A	CTGCTGATGCCGTACATGATGATTTC	1787688 – 1787713	
	Ch10/CRAM-B	CAAGAACATTCCGTCAGTCCCGTC	1784697 –1784720	
2	Ch10/0620			714
	Ch10/0620-A	TCAACCAACCTCTCGCAA	2364877 – 2364894	
	Ch10/0620-B	GTCGCATTGTCGTGATAGT	2365572 – 2365590	
3	Ch10/2920			814
	Ch10/2920-A	AAGGTTGATGACATGGACTAT	3331623 – 3331643	
	Ch10/2920-B	ACTGACATATCCACGATGAC	3330830 – 3330849	

Table 5.4 PCR primers for probes used in Southern blot analysis of LOH.

Primer information for chromosome 10 probes used in Southern blot analysis of PFGE separated standard and LOH lines. Primers and genomic locations given are based on TREU 927 genome sequence. PCR amplifications conditions were as for the standard microsatellite program for probes 2 and 3 and as for the standard minisatellite program for probe 1.

PFGE plugs of each LOH and standard line from procyclic form culture cells were run on a PFGE program for maximal separation of large chromosome sizes (Melville *et al.*, 1998), Southern blotted and probed. Each experiment was performed at least twice and, where available, with different stabilates of the same line. As demonstrated in Figure 5.5a and b, in all cases the expected chromosome 10 homologue size of 4.4 Mb (TREU 927) and ~6 Mb (247 x 386 progeny clone F29/46 bcl 1) was detected with no evidence of a homologue of the expected truncated size. Karyotype of the standard and LOH isolates appear indistinguishable, indicating that no gross chromosome size changes have occurred.

The only change which could be detected between any of the paired lines was a small additional band of approximately 1 Mb present on the ethidium stained gel in one of the STIB 247 x STIB 386 progeny clone F29/46 bcl 1 LOH samples (indicated by a black arrow in Figure 5.5b). This band was not present in the standard line or the other F29/46 bcl 1 LOH line derived from the same initial stock. This band did not hybridise with any chromosome 10 probes however and is not the same size as any other chromosome in the isolate which might indicate that an additional genomic rearrangement has taken place in this isolate while in culture. However its genotype appears normal with all microsatellite markers tested on every other chromosome, indicating that no other obvious LOH events had occurred on the megabase chromosomes. The origin of this band is unknown and requires further investigation, but similar anomalies have been reported in other PFGE experiments. Melville reported three progeny clones from a STIB 247 x TREU 927 cross, which contain unique bands not present in either of the parents and which did not hybridise to any probes in a panel of 401 probes derived from the 11 megabase chromosomes (Melville *et al.*, 1998).

The karyotype analysis argues against the possibility that LOH has arisen as a result of a chromosome deletion. Instead, it confirms that the chromosomes 10 homologues remain the same size as wildtype. This suggested that the observed LOH genotype must be the result of a homologue recombination event resulting in the replacement of alleles from one homologue with those of the other. The most likely explanation for the LOH genotypes is therefore as the result of a reciprocal exchange of DNA (mitotic crossover) or a non-reciprocal break induced replication. Both of these mechanisms would lead to LOH, replacing existing sequence that could proceed to the very end of the chromosomes.

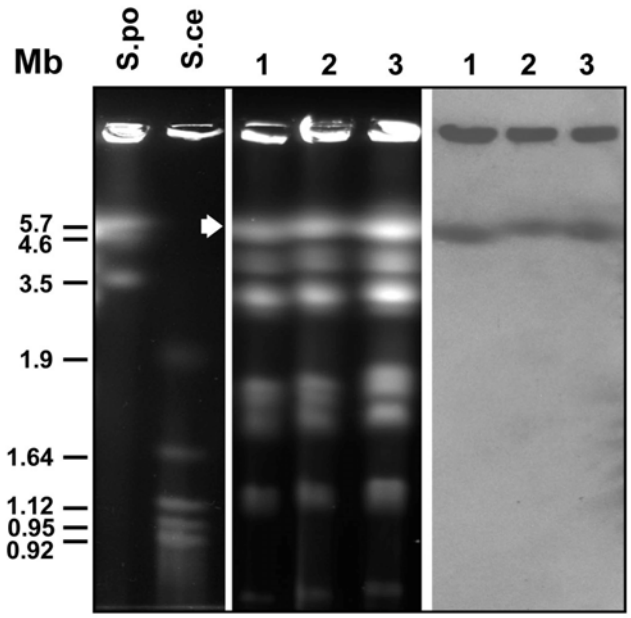


Figure 5.5a The 1-6 Mb molecular karyotype of standard TREU 927 and LOH lines.

The middle panel shows an ethidium bromide stained PFGE separation of intact genomic DNA from procyclic culture of 1. Standard TREU 927 (stabilate TGG737), 2. TREU 927 LOH line 1 (stabilate TGG723) and 3. TREU 927 LOH line 2 (stabilate TGG716). These were separated alongside yeast markers, *S.pombe* (S.po) and *S.cerevisiae* (S.ce), the approximate sizes of which are indicated in megabases (Mb). The PFGE was run out on parameters as set out by (Melville *et al.*, 1998); 3000-2400s switch time, 1.5 v/cm, 106° angle, 72 hr run time followed by 850-450 switch time, 2.4 v/cm, 106° angle, 72 hr run time. The experiment was performed at 14°C using 1% agarose and 0.5x TBE and repeated at least once for each probe. Separated chromosomes were subsequently blotted and hybridised with the chromosome 10 specific probes; Ch10/2920, Ch10/cram, and Ch10/6200 (hybridisation of Ch10/6200 is shown here in the right hand panel) to determine the size of the chromosome 10 homologues in each isolate. The genomic location of each of the probes is indicated in Figure 5.4 and Table 5.4. Hybridization with chromosome 10 probes detected only homologues at the expected 4.4 Mb size in all strains (indicated by a block white arrow).

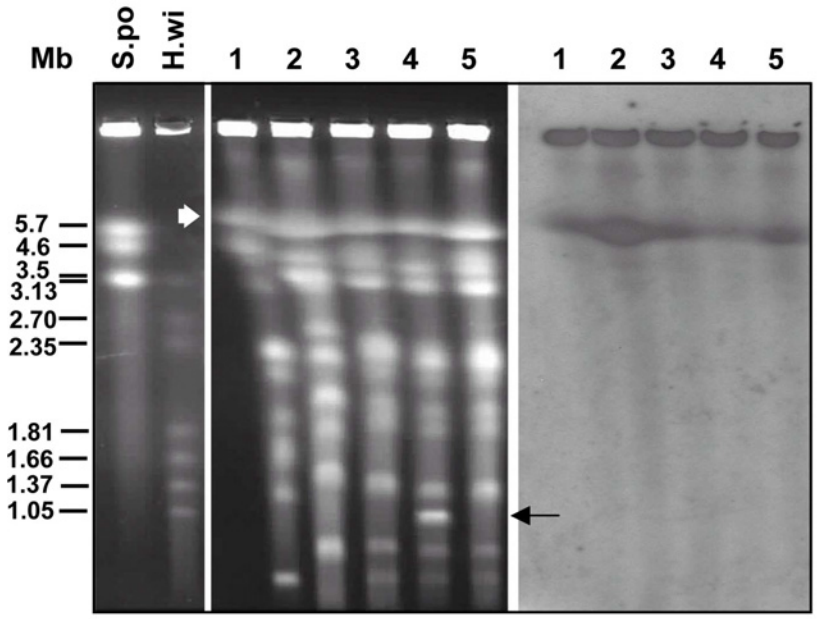


Figure 5.5b The 1-6 Mb molecular karyotype of STIB 247 x STIB 386 progeny clone F29/46 bcl 1 standard and LOH lines.

The middle panel shows an ethidium bromide stained PFGE separation of intact genomic DNA from procyclic culture of 1. STIB 247 wild type (stabilate TGG575), 2. STIB 386 wild type (stabilate TGG667), 3. progeny clone F29/46 bcl 1 wild type (stabilate TGG905), and 4. and 5. progeny clone F29/46 bcl 1 LOH stabilates (TGG890) and (WCMP4463) respectively. These were separated alongside yeast markers, *S.pombe* (S.po) and *H.wingei* (H.wi), the approximate sizes of which are indicated in megabases (Mb). PFGE conditions as defined in Figure 5.5a. Separated chromosomes were subsequently blotted and hybridised with the chromosome 10 specific probes Ch10/2920, Ch10/cram, and Ch10/6200 (hybridisation of Ch10/6200 is shown here in the right hand panel) to determine the size of the Ch10 homologues in each strain. The physical location of the probes is indicated in Figure 5.4 and Table 5.4. Hybridization reveals chromosome 10 homologues are only detected at the expected > 5.7 Mb size (indicated by block white arrow). A black arrow indicates the position of an additional band observed in one stabilate of F29/46 bcl 1 LOH that is not present in the wildtype strain or other stabilate of the same isolate, and does not hybridise to any of the chromosome 10 probes used.

5.2.3 Analysis of the LOH boundaries by Illumina sequencing and SNP-RFLP

Microsatellite screening has identified seven *T.brucei* stocks that have all experienced a LOH event on the same chromosome. In four stocks, that have been analysed in greater detail the LOH appears to run continuously from a centrally placed marker to the last microsatellite marker at the same end of the chromosome's genetic map. In addition, three of these stocks possess LOH boundaries that overlap on chromosome 10.

There is some evidence for the non-random distribution of mitotic recombination along eukaryotic chromosomes, indicating that hotspots for mitotic recombination might exist, as have been observed for meiotic recombination (LaFave and Sekelsky, 2009; Lee *et al.*, 2009). There is also evidence that certain regions of DNA can be more prone to the DSBs which might initiate homologous recombination reactions (Aguilera and Gomez-Gonzalez, 2008) and such "common fragile sites", have so far been identified in both mammalian and yeast cell lines (Glover *et al.*, 2005; Lemoine *et al.*, 2005; Glover, 2006).

Following microsatellite analysis, the region between markers that define boundaries of LOH on Chromosome 10 in each stock remain fairly large. Defined here as the region extending from the last heterozygous microsatellite position to the first homozygous microsatellite position, these boundary regions range from 80 Kb in TREU 927 LOH line 1 to > 800 Kb in STIB 247 x STIB 386 progeny clones F29/46 bscl 1 and F28/46 bscl 6. It was therefore considered of interest to examine the LOH boundary regions in more detail in order to look for potential hotspots for recombination.

To achieve this goal, a further set of high density genotyping markers was required. SNPs are alleles that vary at a single base pair and as the most abundant type of polymorphism in the genome are suitable for this purpose. This data cannot be readily extracted from the haploid *T.brucei* genome database at present, however the recent development in next generation sequencing technology is providing a rapid and cost-effective technique for genome wide polymorphism discovery.

Based on a sequence by synthesis approach, the Illumina Genome Analyser, introduced in 2006, can produce many millions of short, 35 bp reads of sample DNA in a single run (www.illumina.com). By aligning the reads against a reference

sequence, comparisons can be made between genomes allowing the identification of polymorphisms, which can be used as genetic markers. As part of an ongoing collaboration with the Wellcome Trust Sanger Institute, several isolates of interest to this project were paired-end sequenced using Illumina technology. This including one of our LOH lines (TREU 927 LOH line 1) in addition to the standard parents (STIB 247 and STIB 386) of the LOH progeny clone F29/46 bscl 1. Sequence alignments generated from this project could then be used both to identify SNP markers within the LOH boundaries of each strain, as well as exploring the continuity of homozygosity between the microsatellite markers in the TREU 927 LOH line 1.

5.2.3.1 Illumina sequence results

Data analysis using Illumina sequence reads can be challenging to handle, both in terms of the large data file sizes involved and the large number of reads. Short reads below a critical threshold (around 40x coverage) cannot be easily assembled *de novo* and thus data analysis requires a reference genome to "hang" the reads on and specific software for assembly. MAQ (Li *et al.*, 2008), a program designed specifically for handling short read sequences was chosen for the data analysis and performed in collaboration with the Wellcome Trust Sanger Institute.

Wild type parental lines STIB 247 and STIB 386 as well as one of the LOH lines, TREU 927 LOH 1 line were each run on a single lane of the Illumina genome analyser at WTSI. Approximately 8.5-fold, 11-fold and 11.5-fold haploid genome coverage respectively was obtained. The MAQ program was then used to perform an alignment of the paired reads against the haploid TREU 927 reference megabase genome, determine the read consensus and identify SNPs and small indels between the reference and Illumina sequence data.

For the TREU 927 LOH line 1, Illumina sequencing resulted in 8,132,452 x 35 bp paired reads of which 6,904,666 were successfully aligning to the reference sequence (approx 85%), following filtering for poor quality or mismatched sequence. On average this gave an approximate read depth of 9.5 reads per base, along the length of the genome.

The MAQ software was then used to identify SNPs between the aligned TREU 927 LOH line 1 and the haploid TREU 927 reference sequence. These SNPs could be identified as either homozygous or heterozygous in relation to the reference sequence. A heterozygous SNP was called if more than one base was present in the

reads at any one genomic position. SNPs were characterised as homozygous, if only one base was called at a single position in the Illumina sequence that was different to the base present in the haploid *T.brucei* genome reference. Homozygous SNPs were of interest as an indicator of LOH, but are not conclusive as homozygosity could also occur due to mutation or in areas of low coverage where reads from both homologues were not present in sufficient quantity to generate a high quality SNP-call.

Using the MAQ program, a total of 19,534 SNPs were identified across the megabase genome (average of 0.8 SNPs/Kb). Looking specifically at chromosome 10, 3567 SNPs were identified in total (average of 0.88 SNPs/Kb), which does not deviate significantly from the average seen in the genome (Figure 5.6, top). However, when these SNPs are sorted into homozygous and heterozygous SNPs on each chromosome and compared (Figure 5.6, bottom), there is a clearly skewed ratio of homozygous:heterozygous SNPs on chromosome 10. In consequence the percentage of homozygous SNPs in TREU 927 LOH 1 is 16 times greater on chromosome 10 than the genome average. This is as might be predicted if a large region of the chromosome has undergone a LOH event, as a heterozygous SNP would be replaced by a homozygous SNP at that position if the base left was different to the one present in the mosaic haploid genome reference strain.

This can be seen more clearly when the SNPs are plotted at their genomic positions along the length of chromosome 10 (Figure 5.7). It can be seen that the homozygous and heterozygous SNPs are not randomly distributed but clustered so that most heterozygous SNPs appear on the right hand side of the chromosome where the heterozygous microsatellites occur and homozygous SNPs on the left hand end of the chromosome where the homozygous microsatellites occur. Interestingly there are approximately 100 heterozygous SNPs (green) within the LOH region of the chromosome. At present the explanation for these heterozygous SNPs is unknown. It may reflect a technical error in either the MAQ alignment or SNP calling or could indicate real heterozygous SNPs.

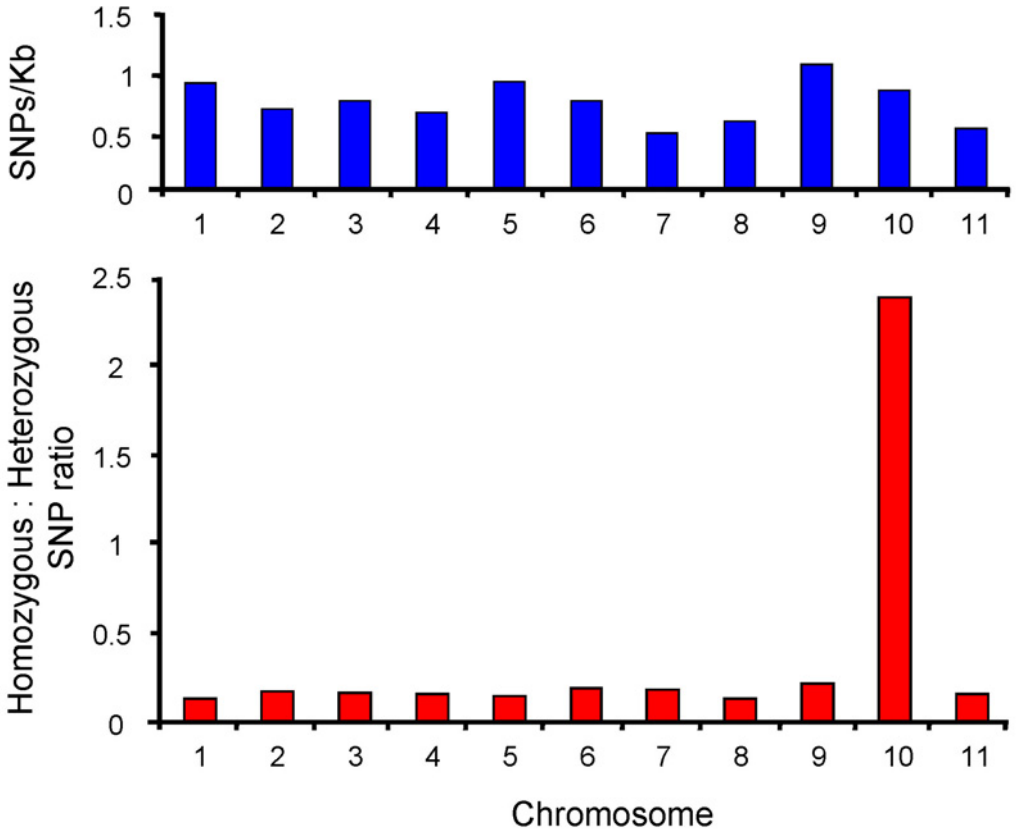


Fig. 5.6 Illumina SNP distribution in TREU 927 LOH Line 1.

SNPs were identified from the TREU 927 LOH 1 Illumina sequence aligned against TREU 927 genome reference sequence using MAQ. The top histogram shows the relative distribution of all SNPs on the 11 megabase chromosomes/per Kb based on *T.b.brucei* TREU 927 genome reference sequence. The bottom histogram shows the average ratio of SNPs characterised as either homozygous or heterozygous on the 11 megabase chromosomes in TREU 927 LOH line 1.

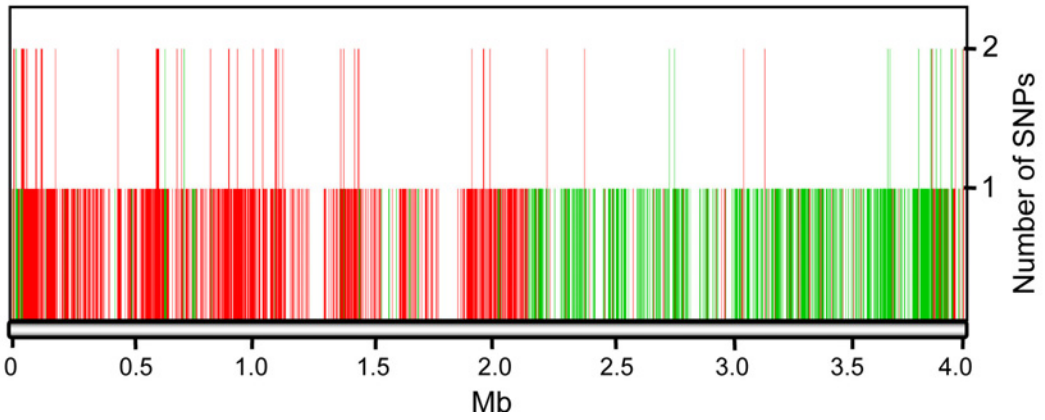


Figure 5.7 TREU 927 LOH line 1 Chromosome 10 SNP plot.

Illumina paired-end reads from the TREU 927 LOH line 1 were aligned against the TREU 927 genome reference sequence using the MAQ program. The heterozygous (green) and homozygous (red) SNPs identified by the MAQ program between the two sequences with a MAQ quality cut-off > 30 have been plotted. Each peak represents a single SNP or 2 adjacent SNPs (double peak) and the X axis indicates their approximate chromosome 10 genomic position. 3567 SNPs were identified, of which 1069 are heterozygous and 2498 homozygous, with a chromosome average of 0.88 SNPs/Kb.

The LOH boundary determined by microsatellite markers, in TREU 927 LOH line 1 extended from marker TB10/16 at genomic position 2,149,291 to marker TB10/17 at position 2,235,898, and this appears to be confirmed by the distribution of heterozygous and homozygous SNPs from Illumina sequencing. There are 46 SNPs present in the 80 Kb region between the two boundary microsatellite markers and based on their distribution, the LOH boundary could be estimated to be between genomic positions 2,200,709 and 2,204,378 (a region of ~3700 bp). To gauge the accuracy of the boundary predicted by Illumina sequence analysis, it was considered prudent to experimentally confirm the SNPs closest to the boundary edge through RFLP analysis.

5.2.4 RFLP analysis

RFLPs are a common laboratory technique that can be used to validate the location of a SNP. In order to discriminate a SNP in a RFLP assay the restriction enzyme has to recognise a digest site involving only one of the SNP bases. To Identify SNP-RFLPs, Vector NTI software (Invitrogen) was used to screen the region of the genome, between the heterozygous and homozygous microsatellite markers, containing each predicted SNP. Each heterozygous SNP base was tried, to look for gain or loss of restriction sites on one homologue relative to the other. For SNPs designated heterozygous by MAQ, the information about both SNPs bases was available directly from the Illumina data. SNPs which were predicted homozygous in TREU 927 LOH line 1 (and thus different to the haploid genome reference) indicated a site which was potentially heterozygous in the wild type sequence but that had undergone LOH, and thus one base could be derived from the Illumina data and the other base from the genome reference sequence.

When a RFLP was identified, primers were designed around this site, amplified by PCR and digested to produce differential digest products in the wild type TREU 927, identifying both SNP bases. In the LOH TREU 927 the expectation was that, as with the microsatellites, where LOH has occurred both chromosome homologues would contain the same base and thus produce same digest pattern.

RFLPs were designed to seven SNPs based on the predicted results of the Illumina sequence data, in order to progressively narrow down the LOH boundary. Six of these RFLPs worked successfully, confirming the Illumina predicted SNP data. The TREU 927 standard line produced heterozygous digest products, while in the LOH TREU 927 LOH line 1, three SNP-RFLPs resulted in homozygous digest products and

three in heterozygous digest products. The details of these SNP-RFLPs are given in Table 5.5. The digest results experimentally confirmed the Illumina SNP results called by the MAQ program in the predicted break point region and experimentally refine the boundary from 80 Kb to 8.7 Kb.

One of the RFLPS, designed to a predicted SNP at position 2,189,549, failed to digest the amplified PCR product. The PCR product was the correct predicted size, however the product was undigested by enzyme following overnight incubation, which had been shown to previously digest control DNA. The reason for this is not known but might indicate either an incorrect alignment of the Illumina reads in this region against the TREU 927 reference, a possible false positive SNP call or a mutation within either the particular stabilate used for Illumina sequencing or the RFLP digest. There are a further two SNPs within the 8.7 Kb LOH boundary identified by MAQ, which do not change any restriction sites and so cannot be experimentally confirmed by RFLP. Based on Illumina sequence alone they are predicted to be homozygous, indicating the boundary between heterozygosity and homozygosity is between position 2,200,709 and 2,204,378 in the genome, a distance of 3669 bp.

In addition to confirming the LOH boundary in TREU 927 LOH line 1, the library of SNPs identified from this isolate can also be used to narrow down the boundary in the other TREU 927 LOH line 2 isolate. Vector NTI was used to screen heterozygous SNPs identified by MAQ within the LOH boundary for TREU 927 LOH 1 to identify RFLPs followed by digesting DNA from both the standard TREU 927 and TREU 927 LOH line 2. By this technique, six RFLPs were tested, all of which worked successfully. Three produced a heterozygous result in TREU 927 LOH line 2 and three a homozygous result, compared to all heterozygous in the TREU 927 reference line (Table 5.6 and Figure 5.8). This allowed the LOH boundary to be progressively reduced from 270 Kb to 15 Kb.

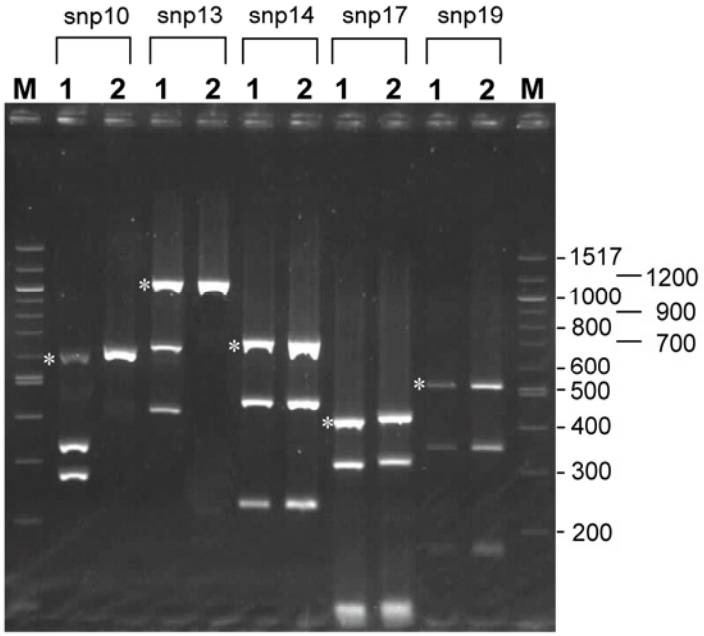


Figure 5.8 Example of five SNP-RFLPs in TREU 927 wildtype and TREU 927 LOH line 2.

Heterozygous SNPs identified from the TREU 927 LOH 1 Illumina sequence in the boundary region between the last heterozygous microsatellite and first homozygous microsatellite of TREU 927 LOH line 2, were screened in Vector NTI (Invitrogen) for polymorphisms in restriction enzyme cut sites (SNP-RFLPs). Sites contains SNP-RFLPs were PCR amplified from the genomic DNA of standard TREU 927 line (TGG737) and TREU 927 LOH line 2 (TGG716) with primers designed to either site of the SNP. This was followed by digestion with the appropriate RFLP enzyme (given in Table 5.6) and gel electrophoresis. For each of the five SNP-RFLPs shown here, the standard TREU 927 line (labelled 1) is heterozygous and thus the PCR product from one allele contains the restriction site and is digested and one does not and remains uncut (uncut size indicated by askeris). For the TREU 927 LOH line 2 (labelled 2) the restriction digest pattern is identical to standard TREU 927 for SNP 927/2 snp14, 927/2 snp17 and 927/2snp19 indicating heterozygosity at these SNP sites however for SNP 927/2 snp10 and 927/2 snp13 the PCR product remains uncut, indicating that the DNA is homozygous at this site. PCR primers, product size, restriction enzyme and fragment sizes are also given in Table 5.6. M indicates 100 bp ladder (NEB).

Finally, the third line in which Illumina data was used to narrow down the LOH boundary was progeny clone F29/46 bscl 1 LOH line, derived from the STIB 247 x STIB 386 cross. In this line potential RFLPs had to be identified through a different technique as neither the normal nor LOH F29/46 bscl 1 lines had been directly Illumina sequenced. Instead we had access to Illumina sequence data of the two normal parental lines STIB 247 and STIB 386. The Illumina reads from each stock were aligned against TREU 927 genome reference and SNPs identified between the reference and Illumina data using the MAQ program as before. In total 4,003,218 (65%) of the STIB 247 reads aligned successfully against TREU 927 sequence, giving six-fold coverage, and 5,621,646 (71%) of the STIB 386 reads aligned giving eight-fold coverage. SNP calling identified 18,000 SNPs on chromosome 10 in STIB 247 and 27,045 in STIB 386 at the quality cut off of >30.

As progeny clone F29/46 bscl 1 is predicted to inherit one homologue from each parent, instead of looking for heterozygous SNPs, homozygous SNPs had to be identified, which were different at the same position in STIB 247 and STIB 386 so that the predicted inherited pattern in standard F29/46 bscl 1 would be a heterozygous SNP. After identifying SNPs, the RFLPs were identified and experimentally characterised as before. The SNP identification process was more time consuming and labour intensive in this strain and consequently only three RFLPs were designed. However all three RFLPs worked successfully, as shown in Table 5.7, resulting is a progressive reduction in the LOH boundary from 700 Kb to 132 Kb.

The result of this RFLP analysis using Illumina SNPs in three of the LOH lines was that the boundaries of TREU 927 LOH 1 could be refined to genome position 2195676-2204378 (8702 bp), TREU 927 LOH 2 to genome position 2968407-2983462 (15052 bp) and F29/46 bcl 1 LOH to genome position 2992541-3124436 (132 Kb). As a result, TREU 927 LOH line 2 and F29/46 bcl 1 LOH lines which previously had overlapping boundaries of LOH as determined for microsatellite analysis, now no longer overlap, although they do remain only 10 Kb apart. The final step in observing the LOH boundary in one of the LOH isolates was to sequence the refined boundary region to confirm that a recombination event had occurred.

SNP name	PCR Primers	PCR product size (bp)	Chromosome 10 SNP position	SNP bases	Restriction Enzyme	Digestion product sizes (bp)		RFLP Results	
						SNP digest site not present	SNP digest site present	Normal TREU 927 line	TREU 927 LOH line 1
927/1snp2	CH10solexa60snp2F CH10solexa60snp2R	809	2187080	A/C	HaeIII	809	190 + 619	Heterozygous	Homozygous
927/1snp16	CH10solexa30snp16.17F CH10solexa30snp16.176	1569	2195667	C/T	EarI	1569	566 + 1003	Heterozygous	Homozygous
927/1snp17	CH10solexa30snp16.17F CH10solexa30snp16.17R	1569	2195676	G/A	EcoNI	1569	574 + 994	Heterozygous	Homozygous
927/1snp19	CH10solexa30snp19F CH10solexa30snp19R	1052	2204378	A/G	NciI	1052	568 + 482	Heterozygous	Heterozygous
927/1snp7	CH10solexa60snp7F CH10solexa60snp7R	1228	2207390	C/T	XmnI	1228	831 + 397	Heterozygous	Heterozygous
927/1snp23	CH10solexa30snp23F CH10solexa30snp23R	1090	2225882	C/T	AluI	47+1043	47 + 475 + 568	Heterozygous	Heterozygous
927/1snp5	CH10solexa60snp5F CH10solexa60snp5R	798	2189549	C/T	KpnI	798	405 + 393	Failed to digest	Failed to digest

Table 5.5 SNP-RFLPs used to refine the LOH boundary in TREU 927 LOH line 1.

SNPs were identified by the MAQ program using TREU 927 LOH line 1 Illumina sequence data aligned against the TREU 927 reference genome. RFLPs were identified using the VECTOR NTI program. PCR primers sequences are detailed in Appendix 1. The PCR primers and product sizes are given along side predicted digest product sizes derived from each allele. Results of the SNP-RFLP reflect whether both digest patterns seen (Heterozygous) or only a single digest pattern seen (Homozygous, shaded grey) for standard TREU 927 control and TREU 927 LOH line 1.

SNP name	PCR Primers	PCR product size (bp)	Chromosome 10 SNP location	SNP bases	Restriction Enzyme	Digestion product sizes (bp)		RFLP Results	
						SNP digest site not present	SNP digest site present	Normal TREU 927 line	TREU 927 LOH line 2
927/2snp10	CH104469snp10F CH104469snp10R	596	2929589	A/G	HincII	596	271+325	Heterozygous	Homozygous
927/2 snp13	CH104469snp13F CH104469snp13R	1058	2955632	C/T	HpaI	1058	417+641	Heterozygous	Homozygous
927/2snp13.7	CH104469snp13.7F CH104469snp13.7R	360	2968407	A/G	EarI	360	138+222	Heterozygous	Homozygous
927/2 snp14	CH104469snp14F CH104469snp14R	680	2983462	A/G	HaeIII	680	228+452	Heterozygous	Heterozygous
927/2snp17	CH104469snp17F CH104469snp17R	514	3002032	C/T	HhaI	105+409	105+305+104	Heterozygous	Heterozygous
927/2snp19	CH104469snp19F CH104469snp19R	509	3021490	T/A	DraI	509	171+338	Heterozygous	Heterozygous

Table 5.6 SNP-RFLPs used to refine the LOH boundary in TREU 927 LOH line 2.

SNPs were identified by the MAQ program using TREU 927 LOH line 1 Illumina sequence data aligned against the TREU 927 reference genome. RFLPs were identified using the VECTOR NTI program. PCR primers sequences are detailed in Appendix 1. The PCR primers and product sizes are given along side predicted digest product sizes derived from each allele. Results of the SNP-RFLP reflect whether both digest patterns seen (Heterozygous) or only a single digest pattern seen (Homozygous, shaded grey) for standard TREU 927 control and TREU 927 LOH line 2.

SNP name	PCR Primers	PCR product size (bp)	Chromosome 10 SNP location	SNP bases			Restriction Enzyme	Predicted digestion product sizes		RFLP Results			
				Solexa STIB 247	Solexa STIB 386	Progeny F29/46 bcl 1*		SNP digest site not present	SNP digest site present	STIB 247	STIB 386	Progeny F29/46 bcl 1 wildtype line	Progeny F29/46 bcl 1 LOH line
414 snp10	CH104469snp10F CH104469snp10R	596	2929383	C/C	T/T	C/T	HhaI	596	65+531	Homozygous (digested)	Homozygous (undigested)	Heterozygous	Homozygous (undigested)
414 snp14	CH104469snp14F CH104469snp14R	680	2983270	A/A	G/G	A/G	AciI	680	36+644	Homozygous (undigested)	Homozygous (digested)	Heterozygous	Homozygous (digested)
414 snp19	CH10414snp19F CH10414snp19R	775	2992541	C/C	T/T	C/T	BbsI	775	199+576	Homozygous (digested)	Homozygous (undigested)	Heterozygous	Homozygous (undigested)

Table 5.7 SNP-RFLPs used to refine the LOH boundary in progeny clone F29/46 bcl 1.

SNPs were identified by the MAQ program using STIB 247 and STIB 386 Illumina sequence data aligned against the TREU 927 reference genome and this used to predict the SNP bases in progeny clone F29/46 bcl 1. RFLPs were identified using the VECTOR NTI program. PCR primers sequence and amplification conditions are detailed in Appendix 1. The PCR primers and product sizes are given alongside the predicted digest product sizes for PCR products derived from each (homozygous) parent. Results of the SNP-RFLP reflect whether both digest patterns seen (Heterozygous) or only a single digest pattern seen (Homozygous) for F29/46 bcl 1 wild type control and F29/46 bcl 1 LOH line as well as the STIB 247 and STIB 386 parental controls. Grey shading indicates homozygosity where heterozygosity would normally be expected.

5.2.5 Sequencing the LOH boundary of TREU 927 LOH line 1

The final step in analysing the LOH boundary was to observe the recombination site in individual homologues in a LOH line. This would definitively confirm that a recombination event had taken place on the chromosome in addition to precisely identifying the point of recombination. To achieve this, the predicted boundary from the Illumina SNP analysis in the TREU 927 LOH line 1 was amplified by PCR, cloned and sequenced. The RFLP results successfully confirmed the LOH boundary in TREU 927 LOH line 1 to within 8 Kb. SNPs within this 8 Kb which did not result in changes to restriction sites, further reduced the predicted boundary to ~ 4 Kb, which is within a suitable range for cloning and sequencing.

A 4 Kb region of the genome covering the LOH boundary was amplified from both the standard TREU 927 line and TREU 927 LOH line 1 at genome position 2200416-2204493. This 4 Kb region included one of the RFLPs previously experimentally validated as heterozygous in both the LOH and standard lines. This digest could therefore be used to distinguish which clones had been derived from the two different chromosome 10 homologues in each line. This allowed 10 clones derived from each homologue to be chosen for sequencing with five primers designed to bridge the 4 Kb region.

Alignment and assembly of the five sequence products derived from each clone to produce complete 4 Kb homologues, identified five polymorphisms between the two homologues in the standard TREU 927 line (Table 5.8). Three of these were SNPs which had been identified previously by the Illumina sequencing, plus an additional two SNPs were identified which had not been picked up previously by the MAQ analysis of Illumina data. The five polymorphisms from the two standard TREU 927 homologues are represented in Table 5.8, with blue shaded bases representing homologue A and yellow shaded bases representing homologue B. In the TREU 927 LOH line 1, SNPs derived from one homologue are identical to homologue B from the standard TREU 927 line (yellow shaded bases), with the same SNPs present at same positions. However the bases present on the PCR product derived from the homologue A contains two SNPs identical to the homologue A of wild type TREU 927 but also three SNPs identical to homologue B. This has narrowed the position of the recombination event down as far as possible to approximate position 2200709-2202089 in the genome where the homologue has switched from containing SNPs from homologue A to SNPs from homologue B.

Overall this has resulted in a successful reduction in the LOH boundary from 80 Kb in the original microsatellite genotyping to just 1380 bp as detailed in Figure 5.9.

Seventeen of the 20 clones produced the sequence results shown in Table 5.8, but there were 3 clones for which the results deviated from the expected SNP base call (Table 5.9). One clone was derived from the TREU 927 LOH line 1 for which an additional gene conversion event appears to have occurred (clone I). Within the standard TREU 927 derived clones, two were identified from the alignment for which SNP bases appeared to be derived from both homologues (Clones 18 and DD). While these aberrant bases may be the result of PCR/sequencing error or mutation they may also indicate additional homologous recombination events within clones of the wildtype population. However, because the sequencing substrate was derived from the PCR of a procyclic culture it cannot be determined if this represents small areas of gene conversion or might extend further along the chromosomes as we have observed in the other LOH isolates. These clones may be examples of other recombination events occurring in the same region of the chromosomes masked by the wildtype population.

Sequencing across the LOH boundary confirms that this LOH line contains one normal homologue but that the second homologue had undergone a recombination event and now contains SNP bases that are the same as the first homologue. No sequence is missing from the recombination site indicating recombination had occurred with high fidelity without any loss of sequence information.

The final position of the LOH boundaries in the four LOH lines for which genotyping has been possible is shown schematically in Figure 5.10.

SNP	Genomic position of SNP	SNP position Within 4 Kb PCR product (from 5' of forward primer)	Haplotype TREU 927 standard		Haplotype TREU 927 LOH line 1	
			Homologue B	Homologue A	Homologue B	Homologue A
1	2200505	90	T	G	T	T
2	2200709	284	T	C	T	T
3	2202089	1674	A	G	A	G
4	2202215	1800	A	G	A	G
5	2204378	3963	G	A	G	A

Table 5.8 SNP Sequence results for both homologues at the predicted chromosome LOH boundary in TREU 927 LOH line 1 compared to standard TREU 927.

Primers were designed around the LOH boundary predicted in TREU 927 LOH line 1 from the Illumina SNP-RFLP results. PCR products were cloned and sequenced to assemble the 4 Kb sequence of the two homologues directly across the LOH site where the recombination event was predicted to occur. Mapping the five SNPs present in 4 Kb of sequence established the haplotype of the two homologues in the standard TREU 927 line, each having a different pattern for the five SNPs. In the LOH line, homologue B is identical to the homologue B in the standard line. PCR products derived from homologue A show a clear switch from displaying homologue A type SNPs at position 2202089 to homologue B type SNPs at position 2200709 indicating the 1.4 Kb region where recombination event occurred. This position is indicated by a switch from blue to yellow in the table.

SNP	Genomic position of SNP	SNP position Within 4 Kb PCR product (from 5' of forward primer)	Haplotype standard TREU 927 Clone 18	Haplotype standard TREU 927 Clone DD	Haplotype TREU 927 LOH line 1 Clone I
1	2200505	90	G	G	T
2	2200709	284	C	C	T
3	2202089	1674	A	G	G
4	2202215	1800	A	G	G
5	2204378	3963	G	G	G

Table 5.9 Clones with SNP haplotypes that deviated from the majority pattern seen in Table 5.8.

Two clones derived from the PCR of the standard TREU 927 line and one from the PCR of the TREU 927 LOH line 1 contained a SNP profile that deviated from the majority. The two wildtype clones contain a mix of SNP bases derived from both homologues A and B which extend to each end of the PCR product and Clone I, derived from the TREU 927 LOH line I appears to contain a SNP pattern consistent with a possible gene conversion event.

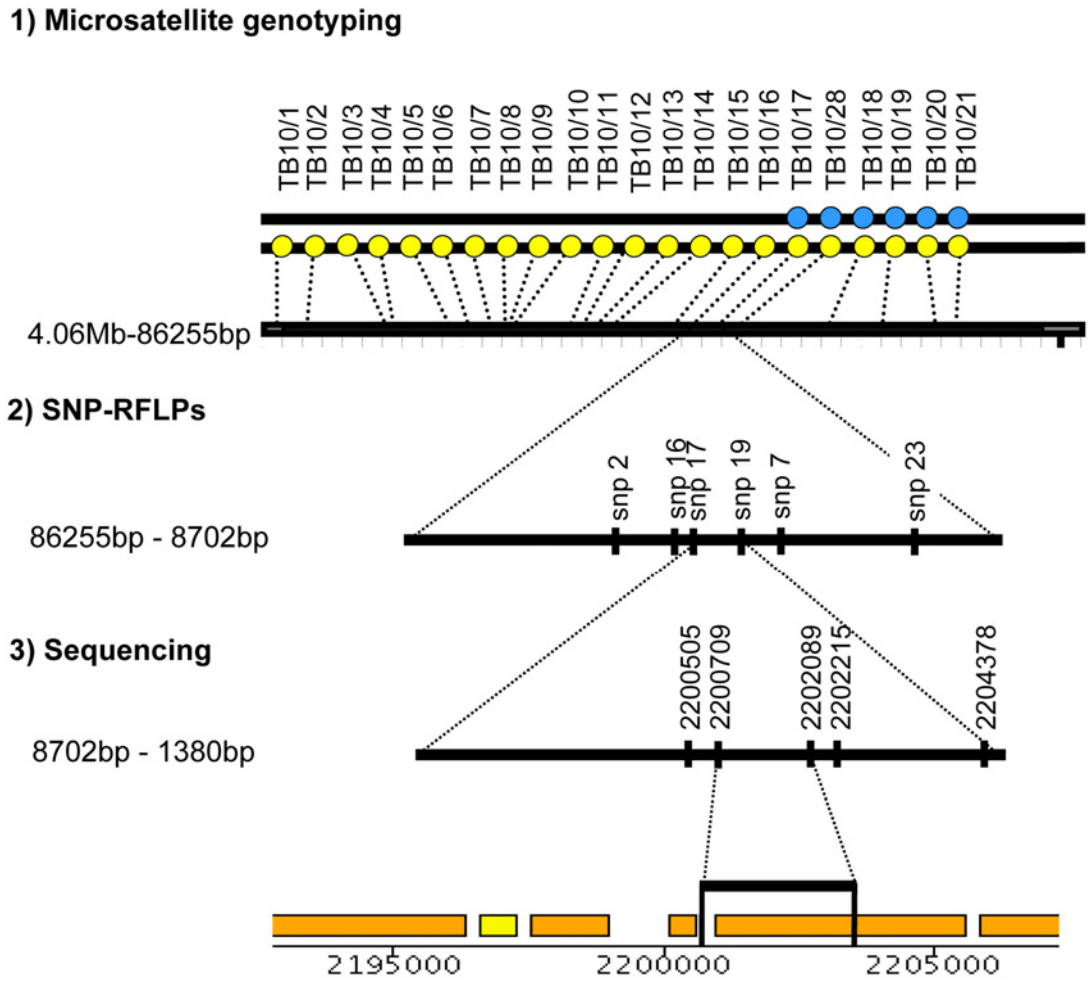


Figure 5.9 Mapping the LOH boundary in TREU 927 LOH line 1.

The LOH boundary in TREU 927 LOH lines was preliminarily mapped by microsatellite genotyping to the region between markers TB10/16 (homozygous) and TB10/17 (heterozygous). SNP-RFLPs predicted by Illumina sequencing the LOH lines were then used to refine the boundary from 86 Kb to 8.7 Kb. Finally the predicted LOH boundary region from both homologues was amplified by PCR, cloned and sequenced to map recombination to within a 1.4 Kb region on chromosome 10. The microsatellite, SNP-RFLP and SNP markers names and approximate position in the genome are indicated on the schematic along with the final position of the LOH boundary in Gene DB.

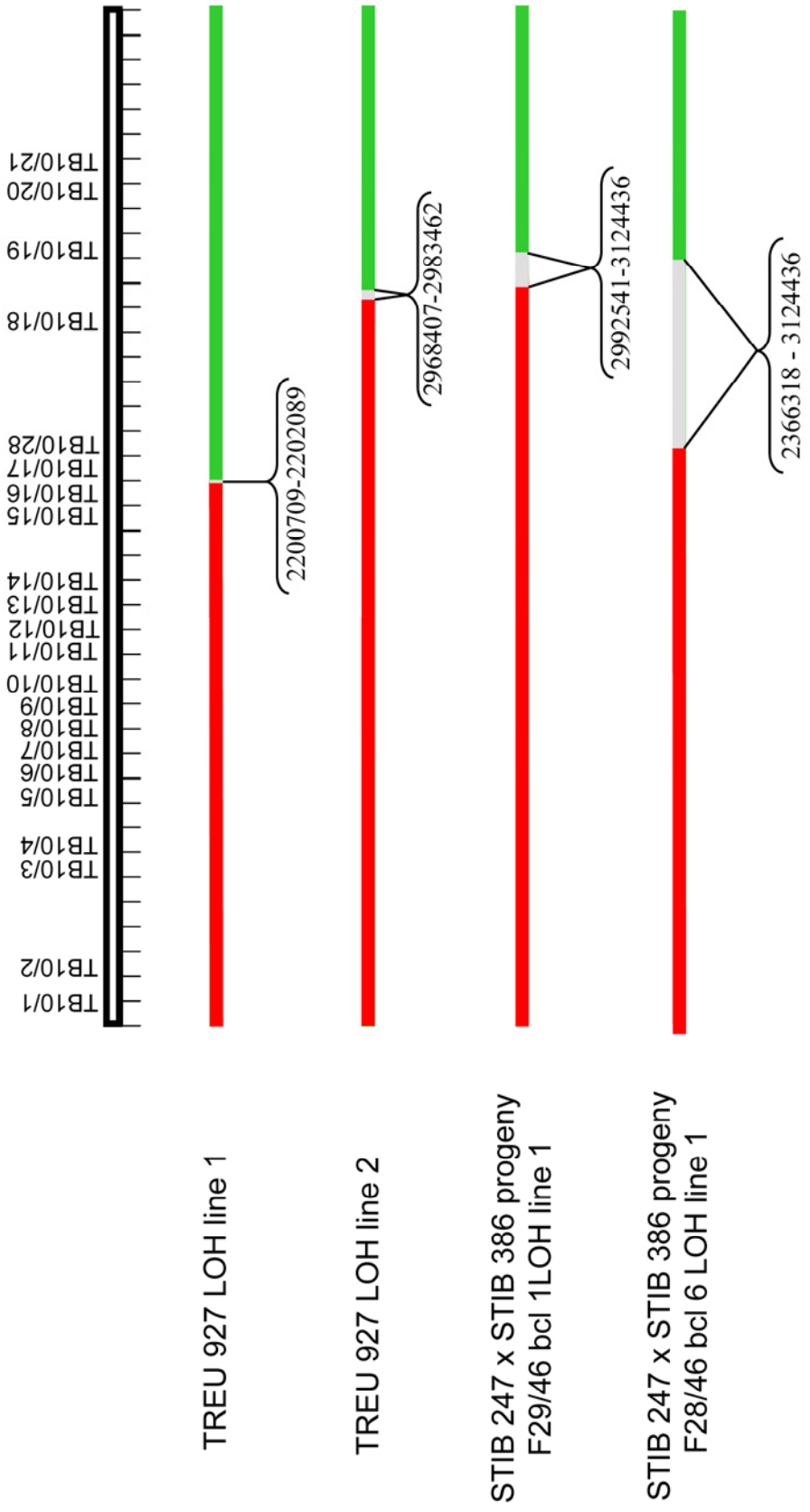

TB10/1
TB10/2
TB10/3
TB10/4
TB10/5
TB10/6
TB10/7
TB10/8
TB10/9
TB10/10
TB10/11
TB10/12
TB10/13
TB10/14
TB10/15
TB10/16
TB10/17
TB10/28
TB10/18
TB10/19
TB10/20
TB10/21
2200709-2202089
2968407-2983462
2992541-3124436
2366318 - 3124436
TREU 927 LOH line 1
TREU 927 LOH line 2
STIB 247 x STIB 386 progeny F29/46 bcl 1LOH line 1
STIB 247 x STIB 386 progeny F28/46 bcl 6 LOH line 1

Figure 5.10 Summary of the LOH boundaries on chromosome 10 in four LOH lines.

The boundary of homozygous and heterozygous sequence on chromosome 10 in four lines of *T.brucei* exhibiting LOH events on chromosome 10 have been refined by microsatellite genotyping in all lines. Additional narrowing-down of the boundary has been performed in three lines using SNP-RFLPs (TREU 927 LOH line 1 and 2, F29/46 bcl 1) and one line by Illumina and capillary sequencing (TREU 927 LOH line 1). The results of this analysis are displayed in the schematic. Green indicates regions of normal heterozygous sequence, red indicated regions in which LOH have occurred and grey indicated the recombination boundary between last heterozygous and first homozygous markers. The exact genomic location of each boundary given below each schematic. The location of microsatellite markers on the reference genome is indicated at the top of the picture.

5.2.6 Growth assays

It is of interest that all lines exhibiting LOH so far identified in the laboratory have experienced recombination events on chromosome 10, while the remaining chromosomes, based on microsatellite genotyping, appear to remain heterozygous. Of the seven lines identified, three were isolated as clones and thus have not been in competition with the heterozygous population, which might outcompete cells with a disadvantageous genomic rearrangement. However, 4/7 (TREU 927 LOH lines 1 and 2, progeny clone F29/46 bcl 1 and STIB 247 LOH line 1) have emerged from a population that was historically cloned and genotyped heterozygous for a set of microsatellite markers.

I would predict the original recombination event that resulted in LOH to have occurred in a single cell from a large population. This would not be detected by microsatellite genotyping as the allele loss would be masked by the presence of heterozygous microsatellite alleles in normal cells. I would therefore predict that any line exhibiting LOH that was not isolated as a clone must have become dominant in a wildtype population and therefore must have a strong selective advantage. Anecdotally it was noticed that procyclic form cultures of several LOH isolates did appear to grow faster then their wildtype equivalents and this was experimentally tested in a series of growth assays.

TREU 927 LOH lines 1 and 2 were compared to standard TREU 927 lines in an *in vitro* procyclic culture growth assay. Growth rates were determined by daily haemocytometer cell counts, from a starting cell concentration of 5×10^5/ml. Two independent stabilates of each line were tested, under identical conditions, and each assay repeated three times. Figure 5.11 shows the linear regression growth curves for the two stabilates of all three lines tested. Statistical analysis was performed on the data by Prof. Mike Turner (University of Glasgow). Firstly the 0-72 hr results for each pair of stabilates of the same line were compared using a one-way ANOVA. In all cases there was no statistically significant difference between each set of paired stabilate data. For standard TREU 927 $F_{1,95} = 0.20$, $P = 0.64$, for TREU 927 LOH line 1 $F_{1,95} = 0.38$, $P = 0.54$ and for TREU 927 LOH line 2 $F_{1,95} = 0.08$, $P = 0.78$. Therefore data from both stabilates could be pooled for subsequent analysis. To determine growth rates a regression analysis of $\log_{10}$ transformed value produced a population doubling time (PDT) of 21.5 hours in standard TREU 927, 14.1 hours in TREU 927 LOH line 1 and 13.0 hours in TREU 927 LOH line 2. Finally the growth rates of the LOH lines and standard TREU 927 were

compared in a 2-way ANOVA. This demonstrated that there were significant differences in growth rates between the strains ($F_{2,287}$ =385.3, P< 0.001). Tukey's post hoc test then showed that the two TREU LOH lines grew at similar rates, but significantly faster than the growth of standard TREU 927.

Although the two TREU 927 LOH lines have different predicted LOH boundaries on chromosome 10 (genome position 2200709-2202089 for TREU 927 LOH line 1 and genome position 2968407-2983462 for TREU 927 LOH line 2), in both cases the same homologue is affected (homologue A) and loss occurs at the same end of the chromosome, with an overlap in homozygosity of > 2 Mb. This strongly indicates that the same genomic rearrangement is responsible for this phenotypic change in both isolates, and supports the hypothesis that a growth advantage is associated with LOH on chromosomes 10 that allows them to outgrow the wildtype population.

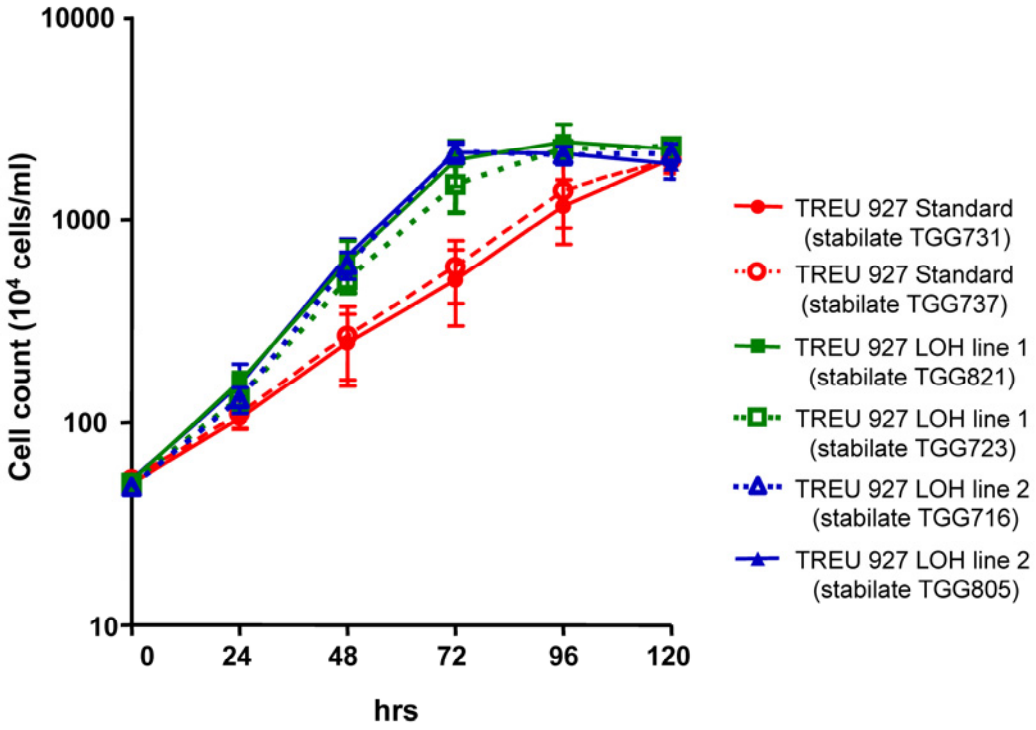


Figure 5.11 Growth comparison of standard TREU 927 and 927 LOH procyclic form culture lines *in vitro*.

The growth curves of standard TREU 927, TREU 927 LOH line 1, and TREU 927 LOH line 2 procyclic form cultures, fitted by linear regression (Graphpad prism 5.1). Two stabilates of each line were measured daily by haemocytometer counts, from a starting concentration of 5 x 10^5/ml. The mean of n=3 experiments is plotted for each time point with standard deviations.

5.3 Discussion

Microsatellite markers derived from genetic maps of *T.brucei* subspecies are useful tools for the monitoring and investigation of genetic stability within *T.brucei* strains. From an initial observation of allele drop-out in microsatellite PCRs, we have identified seven independent lines of *T.brucei*, from a variety of strains that exhibit a LOH genotype on chromosome 10. LOH may reflect the occurrence of chromosomal rearrangements events such as gene conversions, deletions and mitotic crossovers. The results based on a panel of genotyping markers, PFGE karyotype and sequence analysis reveals extensive LOH has occurred from each chromosome 10 homologue incompatible with gene conversion or chromosomes deletion. The retention of partial heterozygosity and no resulting change in chromosome size is most consistent with an explanation of homologous recombination either through a reciprocal mitotic crossover or non-reciprocal break induced replication.

Our analysis of LOH in the seven isolates of *T.brucei*, demonstrate that in every case the same chromosome has been implicated. Much of the previous analysis that has been performed on homologous recombination in *T.brucei*, has examined events in which a DSB has been artificially induced or recombination targeted to a known point in the genome (Conway *et al.*, 2002; Barnes and McCulloch, 2007; Glover *et al.*, 2008). As the events in this study represent spontaneous recombination events that have been observed in a more natural context, it is of interest that all involve chromosome 10 homologues. There are three main hypotheses that may offer an explanation as to why this might have occurred.

The first possibility is that LOH occurs more often on chromosome 10 because of a non-random distribution of mitotic recombination events in the *T.brucei* genome. It has been well established that meiotic recombination sites are unevenly distributed in all but a few eukaryotic genomes (Petes, 2001), and from our analysis this appears to also be the cases in *T.b.brucei* and *T.b.gambiense* (MacLeod *et al.*, 2005b; Cooper *et al.*, 2008). However it has not yet been clearly demonstrated if such hotspots exist for mitotic recombination events.

There are some chromosomal regions in mammalian and yeast genomes that are more sensitive to forming the DSBs which act as a trigger for recombination repair. What makes these fragile sites prone to breakage has not been fully established but such regions of DNA tend to be associated with certain genome

features (Aguilera and Gomez-Gonzalez, 2008). These characteristics include; inverted repeat regions capable of forming secondary structures, stalled replication forks and highly transcribed genes, which may result in the collision of transcriptional and DNA replication machinery, indicating that the molecular basis of their fragility is related to their ability to cause reproductive impairment.

A correlation between these features and recombination events resulting in LOH was found in the fungal pathogen *Candida albicans* where tRNA genes where shown to have an association with examined LOH boundaries (Diogo *et al.*, 2009). A significantly non random distribution of mitotic crossovers has also been observed in LOH studies in human retinoblastoma tumours (Hagstrom and Dryja, 1999) and *S.cerevisiae* (Lee *et al.*, 2009) with putative hotspots identified, however there was no obvious association with structural or sequence elements in the LOH sites.

At present we do not have a sufficient number of LOH isolates to undertake such an study in *T.brucei*, however no association was seen between the highly refined LOH boundary of TREU 927 LOH line 1, and any of the fragile site sequence features previously associate with DSB (Aguilera and Gomez-Gonzalez, 2008). The LOH boundary has been refined to ~1.4 Kb, situated between two putative genes and including the coding region of one of them. An examination of this region indicates no obvious structural or sequence features indicative of elevated level of mitotic recombination (RNAs, repeat regions etc), and the putative genes were not highly transcribed in a digital gene expression analysis of another strain of *T.brucei* (A. MacLeod, personal communication). Interestingly this locus is situated in a meiotic recombination hotspot on the *T.b.brucei* genetic map (MacLeod *et al.*, 2005b). Because of the infrequency and difficulty in detecting mitotic recombination events, the relationship in eukaryotic organisms between mitotic and meiotic recombination sites along the same region of a chromosome is unclear. Where data exists for comparison, for example in the model organism *S.cerevisiae*, (Lee *et al.*, 2009) no clear correlation was found between the two therefore it remains to be seem whether the identification of further LOH events will result in any potential correlation being made in *T.brucei*.

The remaining three LOH boundaries are found in close proximately to each other between markers Ch10/28 and Ch10/17. The TREU 927 LOH line 2 and progeny clone F29/46 bcl 1 LOH line boundaries have been refined by SNPs and found to be separated by only 10 Kb, and the unrefined boundary of progeny clone F28/46 bscl

6 overlaps them both. Further refinement of the boundaries would be needed for a detailed analysis of repeat regions, however GeneDB does indicate arrays of tRNA genes, snoRNA genes, RIME retrotransposon and repeat sequences on chromosome 10 (GeneDB, 2005) within the 700 Kb unrefined boundary of progeny clone F28/46 bscl 6, which need to be fine mapped to see if this close association remains. Although the number of samples exhibiting LOH is at present limited, the close proximity of LOH boundaries on chromosome 10 in three independent isolates may tentatively represent a mitotic recombination hotspot in *T.brucei*. However a greater understanding of the nature of mitotic recombination hotspots in *T.brucei* will likely require the identification and analysis of more LOH boundaries.

An alternative explanation for the identification of mitotic recombination events on chromosome 10 is that recombination is occurring on other chromosomes, but that we only detect particular genomic rearrangements that result in a selective growth advantage. Mathematical modelling of mitotic recombination in asexual cells has been suggested to hasten the spread of beneficial mutations in asexual populations (Mandegar and Otto, 2007). When advantageous mutations arise within an asexual diploid population it takes time for them to fix in the population in a heterozygous state and then for a second mutation to generate homozygosity. Homologous recombination, by generating loss of heterozygosity, could potentially have a large influence on the spread of beneficial alleles and the adaptation of diploid asexual populations.

Most of the lines in use in the laboratory were cloned historically, but have subsequently been through multiple passages and cryopreservation events and thus multiple mitotic divisions. A recombination event resulting in LOH in a single cell in the population would not be detected by standard microsatellite genotyping while it remained at a low level in population as any allele loss would be masked by the relatively large population of heterozygous cells remaining in the population. Recombination events that resulted in no selective advantage or a deleterious growth effect would thus never be detected. However a LOH event which resulted in two copies of a beneficial mutation that gave a selective advantage over the wild type population could, over time, out-compete the wild type population until they become dominant in the population and the allele loss would be detectable by microsatellite genotyping. By this reasoning we could expect that any LOH clone identified that originated from a previously heterozygous population must possess a selective growth advantage. This has been

demonstrated to be the case in the two TREU 927 LOH lines that grew markedly faster than wild type TREU 927 in procyclic form culture and has been anecdotally observed for the hybrid F29/46 bcl 1 and STIB 247 bloodstream form culture, which exist in both wildtype and LOH stabilates.

This growth advantage may also offer an explanation as to why the majority of LOH lines identified to date have originated from procyclic form culture and to a lesser extent bloodstream form culture. Although it is also possible that the selective advantage associated with LOH on chromosome 10 is only advantageous in *in vitro* culture, it is also possible that the selection pressure of continuous passage in *in vitro* culture significantly speeds up the spread of beneficial mutations with a population. *In vivo* bloodstream stabilates in comparison may only be passaged once or twice between cryopreservation, which may not allow sufficient time for homologous recombination events to be detected. It is notable that the one bloodstraw that has been identified with a LOH event is derived from a drug selected clone which was serially passaged in mice > 17 times to generate a drug resistant lines. Therefore homologous recombination resulting in LOH may be occurring in other locations and on other chromosomes but may not be easily detectable because they do not result in a selective growth advantage.

Analysis of multiple LOH events has been a useful technique in the field of cancer research to map the location of tumour suppressor genes contributing to the carcinogenesis of cancer cells (Thiagalingam *et al.*, 2002). Although in this study a locus has been identified by the LOH clones that might contain genes linked to this growth advantage, currently this locus contains too many genes (~900) to be practically mapped. Furthermore the LOH has been detected in several different strains, including both parental and progeny clones of *T.b.gambiense* and *T.b.brucei*. Encouragingly, where > 1 stabilates of a line has been identified the same homologue has been lost in each case, however it remains to be investigated whether the same locus is responsible in each case.

The final explanation that may contribute towards the detection of LOH events on chromosome 10 is that of an unintentional sampling bias caused by the initial microsatellite genotyping that was performed. As demonstrated in Figure 5.2, there are approximately seven markers which can be used for routine genotyping; three of which are on one chromosomes each (chromosome 1, 2, and 5) plus an additional two markers each on a further two chromosomes (chromosome 3 and 10). There are therefore six megabase chromosomes that contain no standard

genotyping markers and thus any LOH events occurring on these chromosomes will not be detected unless they occurred during the original genetic mapping experiments. In addition, as the purpose of genotyping is usually the confirmation of a known genotype, of the seven genotyping markers, only approximately 2-3 are usually utilised at any one time. The most commonly used of these are TB5/4 and TB10/28, both of which are doubly heterozygous markers and thus informative for both parental homologues. It is possible that the identification of so many LOH events on chromosome 10 is a by-product of the frequency with which it is sampled.

Despite this, no LOH have been observed using the equally common TB5/4 marker on chromosomes 5, any of the other common markers, or during the making of the genetic maps in which > 80 hybrids were fully genotyped with all microsatellite markers, thus suggesting that although this is likely to be a contributing factor it is unlikely to be the main influence for the skewed observation of LOH on chromosome 10. It is possible that many other recombination events occurring in *T.brucei* go unseen, either because there is no positive pressure to select for them or that they are not picked up during standard PCR genotyping. At present we do not know the contribution made by each of these factors and it may be that one or all of them play a role.

Previous PFGE analysis carried out on karyotypes suggested *T.brucei* to be mitotically stable in tsetse, mice and culture, outside of the small subtelomeric changes associated with antigenic variation at the telomeres. In contrast, the data presented in this chapter, utilising molecular markers to characterise seven genetic recombination events within a collection of only several hundred clones suggests that there may be implications for the stability of the *T.brucei* genome in long term culture or passaging in rodents.

Passage of mammalian and eukaryotic cells lines has frequently been associated with both karyotypic and phenotypic alterations, as reported in human embryonic cell lines (Draper *et al.*, 2004), murine cell culture (Nelson *et al.*, 1989), *P.falciparum* (Day *et al.*, 1993), *C.albicans* (Diogo *et al.*, 2009), and *S.cerevisiae* (Lee *et al.*, 2009). In this study, no karyotype changes could be observed because the two chromosome 10 homologues can not be distinguished by PFGE separation (Melville *et al.*, 1998). However the molecular markers show a chromosomal alteration to have occurred on chromosomes 10 and a change in the growth phenotype compared to wild type in TREU 927.

This is supported by the comparison of current and historically derived lines of the most common used *T.brucei* cell line Lister 427 (Peacock *et al.*, 2008), where alterations to the karyotype and a loss of tsetse fly transmissibility were correlated with long term passage and adaptation to the laboratory. The genotype of Lister 427 has also been shown to be extremely homozygous with very few heterozygous microsatellites (Annette McLeod personal communication). This may indicate that similar events have occurred during the passage history of Lister 427, resulting in LOH on many chromosomes and possible phenotypic effects. Consequently this might have implications for our interpretation of research derived using Lister 427 and its application to the biology of trypanosomes in the field.

It is not yet known whether the LOH events on chromosomes 10 in our isolates have any further phenotypic effects, such as the ability to be tsetse transmitted. However, taken together this data perhaps indicates that long term passage of *T.brucei* lines should be avoided where possible and abundant stabilates of cell lines made early in adaptation to culture, that can be returned to if subsequent genotyping shows lines to have undergone LOH in culture.

Although LOH also originates from chromosome deletion/non-disjunction, mutation or gene conversion, our results indicate the most likely mechanism in the lines identified here, to be a homologous recombination pathway, either via mitotic crossover or break induced replication. Both of these processes are thought to be triggered by DSB events and utilise a homologous template to repair the interruption in sequence. The basic molecular mechanism of both are also conserved (reviewed in Paques and Haber, 1999), in which homologous recombination is initiated by a 5' to 3' resection of the DSB. This allows the 3' end to interact with recombination proteins, including RAD51, which catalyse the identification and strand invasion of the homologous donor template to prime DNA synthesis from the donor strand to repair the break.

In mitotic recombination, the invasion of both ends of the DSB into the donor sequence leads to the formation of double holiday junctions, which can then be resolved either with or without a crossover event. Alternatively, in BIR only a single DSB end invades the donor template and sets up a repair replication fork in which DNA synthesis proceeds to the end of the chromosome. Although LOH could result from both mechanisms, they do differ in their potential outcomes. In mitotic crossovers, the exchange is reciprocal and following cell division both

daughter cells become homozygous at the affected locus, but for alternate alleles. LOH via a mitotic exchange mechanism can also only occur following DNA synthesis as any crossover event at G1, once duplicated and segregated, would be genotypically neutral. By contrast, a non reciprocal BIR mechanism in which DSB repair of one homologue is primed from the other, results in only one of the two daughter cells becoming homozygous if it occurs at the G2 phase or produces two cells with identical LOH genotypes if it occurs at G1 phase and is then replicated. As these two mechanisms would result in the same outcome in one cell (our observed genotype) but different outcomes in the second, the mechanism responsible could only be distinguished if both daughter cells were available for analysis.

Analysis of the daughter cells of the model organism *S.pombe* (Cullen *et al*., 2007) and the budding yeast *S.cerevisiae*, where micromanipulation of budding daughter cells from a mother cell allows the product of every mitotic event to be analysed, found both mitotic recombination and BIR mechanisms operate in DSB repair (McMurray and Gottschling, 2003; Cullen *et al.*, 2007). The factors determining which method of repair is preferred under which circumstances has not been fully determined, however, in *S.cerevisiae* at least, there appears to be a strong age related preference for mitotic recombination in young cells and BIR as the cells aged (McMurray and Gottschling, 2003). Recent evidence has also implied that the type of DSB may also play a decisive role in the choice of homologous repair mechanism used. BIR is a one-ended DSB recombination event and is thus thought to play a key role in repairing breaks in which there is only a single free DSB end, such as that associated with stalled or collapsed replication forks (Seigneur *et al.*, 1998; Michel, 2000) or erosion of uncapped telomeres (McEachern and Haber, 2006). Where both DSB ends are present, BIR appears to be only a minor pathway in most mammalian and yeast cells, and gene conversion, with, or most often, without crossovers appearing to predominate as the main mechanism of homologous recombination repair (Malkova *et al.*, 1996; Ira *et al.*, 2003; Stark and Jasin, 2003).

Both mitotic crossovers and BIR mechanisms have been reported in *T.brucei* in the context of the most frequently studied aspect of *T.brucei* homologous recombination, the VSG switching, responsible for antigenic variation. Although we do not have both daughter cells for comparison and thus cannot determine which of the processes might be the cause of the LOH identified in our *T.brucei* lines, homologous recombination is the preferred mechanism of DNA damage

repair in trypanosomes. Experimentally, *T.brucei* appear to favour homologous recombination for both DNA repair and antigenic variation with gene conversion reported to be the most common mechanism for repair of induced DSB breaks (Glover *et al.*, 2008), VSG switching (Hoeijmakers *et al.*, 1980; Pays *et al.*, 1983a) and transfection of linear DNA constructs (Barnes and McCulloch, 2007).

Several reports exist of mitotic crossover events causing VSG switching by reciprocal telomere exchange (Pays *et al.*, 1985; Rudenko *et al.*, 1996), although this so far appears to be a relatively minor pathway of VSG switching in *T.brucei*. Interestingly, BIR following induced DSBs in the 70 bp repeats, upstream of an actively expressed VSG was recently shown to be the main mechanism of DSB repair. Heterologous donor templates were favoured in this reaction and resulted in VSG switch events (Boothroyd *et al.*, 2009) suggesting that this mechanism is active in *T.brucei*, at least at the telomeres.

In order for homologous recombination mechanisms to successfully repair a DSB, an intact homologous template is required. Preferentially, sister chromatids are considered the favoured donor template, as exact sequence identity allows DSBs to be repaired without loss of genetic integrity (Kadyk and Hartwell, 1992; Johnson and Jasin, 2000). However the results of our LOH lines presented here indicate that the homologous chromosome has acted as the donor template for repair. It has been observed in many other eukaryotic organisms that allelic and ectopic homologous sequences can also be used (Moynahan and Jasin, 1997; Richardson *et al.*, 1998; Aylon *et al.*, 2003), and indeed may be necessary if a DSB occurs before DNA replication. More specifically in *T.brucei*, in which homologous recombination is utilised as a major mechanism in antigenic variation, the choice of a sister chromatid as donor template would not result in a VSG switch event.

As sister chromatid repair is genetically silent and thus effectively invisible the relative choice of template in *T.brucei* is unknown. When DSBs were induced within the core of one chromosome 11 homologue using an inducible I-SecI cleavage system, and repair using the sister template was compromised (by continuous expression of I-SecI), allelic homology was shown to provide the preferred template for repair, rather than ectopic sequence, with a predominant outcome of gene conversion (Glover *et al.*, 2008). This is in contrast to analysis of an induced DSB by the same mechanism at the 70bp repeats in an active BES close to the telomere in which an ectopic template was preferred and BIR the most commonly used mechanism (Boothroyd *et al.*, 2009).

The preference for allelic recombination in the chromosome cores may reflect the unavailability of the sister chromatid during G1 phase, the result of a homology search or a possibly proximity of the homologous chromosomes at certain stages of the replicative cycle (Barzel and Kupiec, 2008). In contrast to the chromosome core, subtelomeres appear to favour ectopic recombination (Barry *et al.*, 2003). *T.brucei* chromosome homologues exhibit variation in size both within and between strains (Melville *et al.*, 1998; Melville *et al.*, 2000) and thus may be hemizygous outside of the chromosome core, limiting allelic homology at the subtelomeres. Repetitive subtelomeric sequences, VSGs, ESAGs and other polymorphic gene families may therefore provide the homology, that allows subtelomeres to engage in ectopic recombination rather the predominant sister or allelic recombination seen in the housekeeping genes (Horn and Barry, 2005). Telomere positioning, in which the chromosome may be situated such that the telomeres cluster together towards the periphery of the nucleus (Klein *et al.*, 1992; Perez-Morga *et al.*, 2001; Barry *et al.*, 2003) may also provide additional opportunity for ectopic recombination at the subtelomeres that is less accessible for the chromosome core. Interestingly the introduction of a DSB within the 70 bp repeat region close to the VSG in an active ES resulted in a VSG switch rate far higher than when the break was introduced at an upstream site, indicating a strong positional effect to the repair mechanism and confirming that the 70 bp repeats are a likely DSB hotspot (Boothroyd *et al.*, 2009).

Finally we consider the implications of the observed anomalies in the Illumina sequence data for chromosome 10, in which several 'islands' of heterozygous SNPs are observed within the largely homozygous LOH region on the left hand end of the chromosome. While it is formally possible that these represent false positive SNPs or alignment errors within the MAQ software, they may also be indicative of the homologous repair mechanism used.

One intriguing possibility is the proposed mechanism of template switching during BIR outlined by (Smith *et al.*, 2007) in which, following strand invasion, the DNA is replicated for only a limited distance before disengaging from the template. If homology searching identifies the other side of the DSB, the repair can be completed by the synthesis dependant strand annealing pathway, resulting in only a small LOH event by gene conversion. However, if the other side of the DSB break is not detected, further rounds of strand invasion, replication and disassociation ensue until either the second DSB end is recognized or the end of the chromosome is reached as per the tradition BIR model (Malkova *et al.*, 1996). If the potential

exists for homology searching to identify a different homologous template after disengagement this could result in a mosaic pattern of repair with template sequence derived from any combination of sister, allelic or ectopic sequence. In support of this theory, examples of BIR products consistent with multiple template switches were identified by Smith *et al* (2007) in approx 20% cells recovered from a chromosome fragmentation assay in *S.cerevisiae*. If this mechanism is a conserved feature of BIR and functioning in *T.brucei*, it is possible that these 'islands' may represent substrate switching between allelic and sister (or even ectopic) templates, with what appears to be a bias towards the allelic homologue.

As previously mentioned, BIR has been identified to be an active mechanism of DSB repair in *T.brucei* at the 70 bp repeats upstream of an active *VSG* gene (Boothroyd *et al.*, 2009). At this site, ectopic templates appeared to be the preferred substrate for repair, resulting in frequent VSG switching events. All the 18 VSG switch products of BIR observed by Boothroyd *et al*. appear to use a complete donor VSG template copied from other VSG expression sites or mini chromosome locations (consistent with that usually observed in early infection (Robinson *et al.*, 1999; Morrison *et al.*, 2005) without evidence for template switching. In spite of this, perhaps template switching may offer an additional mechanism by which trypanosomes could generate the mosaic VSGs assembled through recombination of sequence from *VSG* genes and pseudo genes seen late in infection (Kamper and Barbet, 1992). It is clear that further investigation is needed into this observation. This would require RFLP or further sequence confirmation of the heterozygous SNP regions identified here, as well as looking at the other LOH boundaries identified in the other LOH lines, before a clear mechanism to explain these anomalies could be drawn.

Chapter 6

Conclusions and future research directions

African trypanosomiasis is a neglected disease that remains a major public health concern and obstacle to development in rural sub-Saharan Africa. The publication of the genome sequence for *Trypanosoma brucei* in 2005 (Berriman *et al.*, 2005), marked an important step forward in the effort to uncover the genetics underlying phenotypic variation in this pathogen, and an abundance of new tools and technology is now available for exploiting this data in the post-genomic age. In this project the genome sequence was utilised to generate a series of molecular markers which were employed alongside both traditional and cutting edge technologies, in a series of genetic analyses of pleomorphic *T.brucei* laboratory strains, to investigate the products of meiotic and mitotic recombination.

In Chapter 3, linkage mapping, a traditional but largely underexploited approach in trypanosome research, was employed to generate a genetic map of the human infective *T.b.gambiense* strain. This research will enable linkage analysis of important phenotypes, comparison to *T.b.brucei*, and a greater understanding of recombination in *T.brucei*. The potential to improve this strategy with a new method to isolate progeny was detailed in Chapter 4 alongside an improved genetic map. Finally, an observation of *T.brucei* laboratory adapted strains exhibiting loss of heterozygosity was investigated in Chapter 5 using next generation sequencing technologies, with implications for the maintenance of trypanosomes in *in vitro* culture.

A detailed discussion of each of these experiments has already been provided in the relevant chapters. In this final chapter, therefore, I will briefly outline the main conclusions and explore future work for the data presented in this thesis.

Chapters 3 and 4 described the construction of a genetic map for the *T.b.gambiense* subspecies through analysis of the segregation of 119 microsatellite markers, in 50 hybrid progeny from a STIB 386 *T.b.gambiense* x STIB 247 *T.b.brucei* cross. The completed linkage map, composed of 12 linkage groups totalling 746 cM and all microsatellite markers, have been made available to the trypanosome community (Cooper *et al.*, 2008; University of Glasgow, 2008). The map can now serve as a valuable tool in a number of potential applications that include the localisation of genes or QTLs associated with traits of interest, comparative strain and subspecies analysis and studying the population genetics and molecular epidemiology of *T.brucei* in the field.

Microsatellites have been the markers of choice for many linkage mapping and population genetics studies. This is because they are highly polymorphic, co-dominant and abundantly distributed in the genome. They can also be quickly and reproducibly amplified by PCR, detect mixed infections, and require only a small amount of target DNA, properties that make them ideal for genetic analysis of field samples. One application of the microsatellite markers that have been generated from the *T.b.gambiense* genetic map and pre-existing *T.b.brucei* map (MacLeod *et al.*, 2005b), is therefore in analysis of the population structure and levels of genetic exchange occurring in *T.brucei* populations in endemic foci. The availability of polymorphic markers is particularly useful in combination with novel sampling techniques, such as FTA cards, and methods to increase PCR sensitivity such as Whole Genome Amplification, which enables multiple genotyping analyses of field samples from very limited source material (Maclean *et al.*, 2007; Morrison *et al.*, 2007; Morrison *et al.*, 2008; Pinchbeck *et al.*, 2008).

Genetic linkage maps have been determined for a number of other, mainly haploid, protozoa parasites (Su *et al.*, 1999; Shirley and Harvey, 2000; Khan *et al.*, 2005; Martinelli *et al.*, 2005a) and the *T.b.gambiense* microsatellite map compares favourably to those of a similar genome size both in terms of genetic size and recombination frequency. With regards to the *T.b.brucei* genetic map, with which several microsatellites are shared, a more detailed comparison confirmed that the marker order and synteny is retained between the two subspecies as well as the average recombination frequency. However as in many other eukaryotic genomes (Petes, 2001), recombination does not occur at a uniform frequency but varies both between and within chromosomes. This results in hot and cold spots of recombination in both maps, some of which are shared and some of which appear to be strain-specific. An improvement to the maps in the future to increase their resolution would allow for these spots to be more accurately defined. As meiotic recombination remains a poorly understood process in trypanosomes an improved knowledge of these hot and cold spots would be relevant both for future genetic mapping experiments and for increases our understanding of recombination at a mechanistic level.

The most important purpose of a genetic map however is to facilitate the localisation of genes and QTL loci underlying heritable traits of interest. The parental stocks from this cross have been shown to vary phenotypically in a number of important traits related to the transmission, pathogenicity and host range of the parasite for which the genetic basis has not yet been determined (Tait *et al.*, 2002). The availability of a genetic linkage map for *T.b.gambiense*

thus opens up the possibility of identifying loci that contribute to these traits. Linkage analysis using the genetic map and hybrids from the cross is currently being performed in the group for one of the most significant of these traits, that of human serum resistance in *T.b.gambiense* (Annette MacLeod, personal communication). Mapping of the other traits previously described, including midgut infectivity and virulence in addition to the difference in pentamidine sensitivity determined in Chapter 3 of this thesis, however remains to be explored.

Without the prospect of a *T.brucei* vaccine and with chemotherapy as the mainstay of control, drug resistance is a trait of particular interest. A higher level of pentamidine resistance was observed in STIB 386 than in STIB 247, both *in vivo* and in an *in vitro* bloodstream culture assay. Analysis of the segregation of this phenotype in the hybrids will be required to ascertain the feasibility of genetic analysis for elucidating the genetic basis of this trait. Because this assay is performed *in vitro* it will require all hybrids from the cross to be adapted to bloodstream culture, an ongoing process in the laboratory, followed by the characterisation of pentamidine sensitivity. If linkage analysis and fine mapping, in combination with the annotated genome sequence, identifies candidate genes associated with this trait, these could be tested in the laboratory by reverse genetics with the eventual downstream aim of association studies of candidate genes with drug resistance in the field.

At its present resolution, there is a 90% probability of mapping any gene to within 11 cM or 263 Kb of its physical location. This will provide a framework for the localisation of phenotypes to particular loci. However loci may still contain >100 genes, requiring further fine mapping to narrow down the number of candidate genes to a suitable number for validation. A higher resolution map is therefore desirable for mapping traits of interest to a smaller physical location and would also facilitate a more accurate measurement of the recombination frequencies in *T.b.gambiense*.

Future work will therefore be directed towards improving the resolution of the genetic map. This can be achieved by two mechanisms, example of both of which were also examined in this thesis; by increasing the number of recombination events in the map through the identification of further hybrids (Chapter 4) and by increasing the density of molecular markers (Chapter 5).

The production of a *T.brucei* genetic cross in the laboratory has traditionally been a laborious and expensive process. A major contributing factor to this is the requirement for the *in vivo* cloning and amplification of individual trypanosomes in a rodent host. The development of an alternative *in vitro* culture cloning technique as described in Chapter 4 of this thesis may therefore be considered a major technical advantage for the ease of developing new genetic crosses or identifying further hybrids from existing cryopreserved uncloned populations for genetic linkage projects in the future.

This new method requires only limited growth *in vivo*; once to amplify the uncloned progeny of a cross from tsetse salivary glands for long term cryopreservation and then (optionally) once to amplify a stabilate each time a cloning project is initiated. The *in vitro* cloning method is easy, effective, and scalable allowing much larger numbers of clones to be isolated at any one time. The price paid for this increased ease however, is a drop in genetic diversity, most probably caused by the increased number of amplification steps required before cloning which selects for faster growing trypanosomes. Future cloning experiments must therefore take care to ensure passage of uncloned cross populations is minimised, both *in vivo* and *in vitro*, and abundant stabilates made at the point of cryopreservation to retain maximal genetic diversity in each cross population.

There are also applications for the polymorphic microsatellites from the *T.b.gambiense* and *T.b.brucei* genetic maps in combination with the new *in vitro* method for future cloning experiments. The ability to perform multiple genotyping PCRs from small culture lysates before cloned cultures are expanded should allow preliminary screening with targeted markers to identify and screen out existing hybrids or select for hybrids exhibiting a crossover event in a regions of interest, for fine mapping, allowing the loci of interest to be more quickly narrowed down.

During the timescale of this project, the long-term axenic culture of bloodstream stage trypanosomes was established in our laboratory for a number of pleomorphic trypanosome strains. This included the three *T.brucei* parental strains (STIB 247, STIB 386, and TREU 927) and several of the hybrids used in the construction of the *T.b.brucei* and *T.b.gambiense* genetic maps, in addition to the new hybrids isolated by the *in vitro* cloning method. This gradual transition from *in vivo* growth to *in vitro* culture of *T.brucei* is in line with the general practice of the trypanosome research community where much of our current understanding of their genetic and biochemistry processes has been derived using *in vitro* adapted lines.

Several of our *in vitro* adapted pleomorphic lines, have been tested and at present appear to maintain their phleomorphism, infectivity to rodents and can be transformed from bloodstream form to procyclic form *in vitro*. In addition, *in vitro* culture is advantageous in avoiding the associated problems of host contamination in genetic analysis of parasite DNA or RNA and for the development of *in vitro* phenotyping assays for several of the traits of interest in *T.b.gambiense*, including HSR and pentamidine resistance. However it has not yet been determined if such *in vitro* adapted lines remain fully transmissible through all lifecycle stages by checking their transmissibility through tsetse flies. This should certainly be determined during future work as it may affect our ability to explore certain key phenotypes such as tsetse midgut infectivity and transmission index or to undertake F1 or back crosses. It may also be considered prudent to exert caution in the interpretation of data derived from *in vitro* adapted lines during this transition phase until a full evaluation has been performed. This is particularly relevant in light of the genomic rearrangement events that have been observed in several *in vitro* derived lines, as reported in Chapter 5 of this thesis.

It has long been observed that *T.brucei* lines adapted to laboratory conditions can undergo a number of phenotypic alterations. These can range from a loss of tsetse transmissibility, commonly observed in parasites maintained in continuous *in vitro* culture (Roditi and Lehane, 2008), to the accumulation of a large number of phenotypic changes affecting virulence, differentiation (Ashcroft, 1959) and antigenic variation switch rate (Lamont *et al.*, 1986; Turner and Barry, 1989), characteristic of serial syringe passaged monomorphic lines. The underlying molecular mechanisms responsible for these changes in phenotype and whether or how they might be related has not been determined. However a recent study comparing a current isolate of the most commonly used monomorphic line Lister 427, to an older pleomorphic and fly transmissible version, observed that a number of karyotype differences existed between the two lines indicating genomic rearrangements had occurred during its passage history (Peacock *et al.*, 2008).

In the TTM laboratory, more than 80 different pleomorphic parental and hybrid lines of *T.brucei* are studied, both *in vivo* and more recently, in culture adapted form. In Chapter 5 of this thesis I reported the observation that the adaptation of several of these trypanosome strains to both procyclic form and bloodstream form *in vitro* culture, was associated with the genomic rearrangements of a chromosome 10 homologue and an altered growth phenotype. Using a combination of microsatellite PCR, PFGE and sequencing, the molecular basis of these LOH events was investigated, focused on paired *T.brucei* isolates of the genome strain

TREU 927. These large naturally occurring LOH events, covering several megabases of sequence are the first such events characterised in *T.brucei* and were attributed to mitotic recombination.

In addition to the clues that these rearrangements offer regarding loci on chromosome 10 that might be associated with the genetic control of growth *in vitro*, this work also highlights the concern that genomic rearrangements can and do arise in long term axenic culture of *T.brucei*. A note of caution is therefore sounded to the community, suggesting care should be taken when extrapolating from results obtained using laboratory-adapted lines. Taking this into consideration the recommended procedure for our laboratory is to continue with the adaptation of pleomorphic strains and isolation of hybrids by the new *in vitro* method but to retain cryopreserved stabilates of each line in all available life cycle stages, both *in vivo* and *in vitro* (e.g. bloodstream form culture, procyclic form culture and bloodstraw formats), in order to ensure no valuable phenotype or genetic information is lost.

The work presented in this thesis represents only a preliminary investigation into this area however and there remains many unanswered questions related to how, where, when, and how often the recombination events responsible for the LOH occur. The mechanistic cause of the genomic rearrangements remains unknown but based on the investigations outlined in Chapter 5 could be narrowed down to mitotic recombination either due to a reciprocal exchange or non-reciprocal BIR. Both mechanisms have previously been observed to occur in *T.brucei* as mechanism related to VSG switching in antigenic variation (Pays *et al.*, 1985; Boothroyd *et al.*, 2009). However several questions remain;

1. What is the exact homologous recombination mechanism responsible for the LOH in these strains and is the same mechanism responsible in each case of LOH observed?

2. How is the process induced (DSB, stalled replication fork)?

3. Are similar mitotic recombination events occurring genome-wide but going undetected?

4. Are these recombination events associated with particular genomic locations/features that might result in hotspots for mitotic recombination?

5. Are such mitotic recombinations, resulting in large regions of LOH, a rare laboratory artefact or a frequent event?

6. Does mitotic recombination contribute significantly to genetic diversity or the spread of beneficial mutations such as drug resistance in the field?

Although I have attributed the LOH events to mitotic recombination, the limited number of isolates identified so far and the method by which they were detected, does not allow us to draw any conclusions regarding the frequency of these mechanisms. It is thus hard to assess how common a problem these events might be in *T.brucei*. A more in-depth analysis of these genomic rearrangements will require the identification of further examples of LOH. These may accumulate naturally through the standard microsatellite genotyping of lines employed by our laboratory that brought the initial seven cases to our attention, or may require an active screen for homozygosity or induction of DSB and repair events in the laboratory. If a significant number of examples are identified these would allow a more detailed characterisation of mutants exhibiting LOH to investigate sequence features located in the vicinity of LOH boundaries that might be associated with recombination hotspots.

The investigation of loss of heterozygosity in this thesis also made use of next generation sequence technology, in order to confirm homozygosity on chromosome 10 and determine the recombination boundary more accurately. Next generation sequencing has emerged over the last couple of years, amid much excitement, allowing the rapid and inexpensive acquisition of gigabases of sequence data in as little as a couple of days. These high throughput, massively parallel technologies are transforming post-genomic research with applications above and beyond basic genome sequencing that include polymorphism detection, transcriptomics, metagenomics, comparative genomics and chromatin immunoprecipitation.

Currently there are three major next generation platforms on the market; Roche 454 Genome Sequencer FLX introduced in 2005 (http://www.454.com), Illumina Genome Analyzer introduced in 2006 (http://www.illumina.com) and Applied Biosystems SOLiD System introduced in 2007 (http://www.appliedbiosystems.com). Each of these platforms utilises a slightly different underlying chemistry, but all work on the basic principle of parallel amplification and sequence of millions of individual DNA fragments, which permits deep coverage and multiple base calls for a single genomic position. Compared to

traditional capillary sequencing the sequence reads are considerably shorter (35-250 bp) and have a higher base call error rate (Shendure and Ji, 2008), which has downstream consequences for the assembly of the sequence reads. *De novo* assembly of short reads is challenging, but possible with a sufficiently high read depth and the appropriate capacity for analysis. However the major application for next generation sequence reads at present relies on their alignment with an established reference sequence.

In this project, whole genome sequencing was performed, using Illumina paired end technology, of a TREU 927 LOH mutant isolate and normal parental isolates of STIB 247 and STIB 386. Strain to reference sequence comparisons were performed by specialist software to align and identify polymorphisms between the TREU 927 genome sequence and the Illumina sequenced reads. The alignments generated by this process were utilised in two ways. Firstly, an examination of the distribution of heterozygous SNPs on chromosome 10 in the TREU 927 LOH mutant line was able to successfully localise the LOH recombination boundary within this mutant isolate to within 4 Kb which was later confirmed by targeted capillary sequencing of the boundary region.

The second application of the strain to reference comparison was for genome wide sequence variant discovery. The identification of polymorphisms between the TREU 927 reference strain and the Illumina sequenced isolates has resulted in a genome wide database of heterozygous and homozygous SNPs for each strain of around 120 000 SNPs for STIB 247, 165 000 SNPs for 386 and 20 000 for TREU 927. Only a small number of the SNPs in these databases have so far been experimentally validated (~20 in Chapter 5), but with that caveat, they can be utilised as a resource of polymorphisms for applications such as comparative strain analysis, mutant mapping and in combination with the genetic maps generated for *T.b.brucei* and *T.b.gambiense* for linkage mapping.

As detailed in Chapter 3, the *T.b.brucei* and *T.b.gambiense* genetic maps currently consist of a density of 180 and 119 markers respectively, with a probability of mapping any loci to within a few hundred kilobases. In both maps we also observe at least one chromosome for which multiple linkage groups are present and remain unlinked, and also regions of the genome that presently lie outside the range of the genetic map. These regions tend to be located towards the subtelomeric regions at the ends of the chromosomes, in which polymorphic microsatellites are scarce. One future aim using SNPs identified from the Illumina sequence project, would therefore be to build on the foundation of the

microsatellite genetic map, to create a high density SNP map that, at their current distribution, might allow linkage of phenotypes of interest to loci containing as little as a few kilobases of sequence. A high density SNP map of *T.brucei* would also allow for a more detailed analysis of the distribution of meiotic recombination in *T.brucei* by more precisely identifying recombination hotspots to examine if they exhibit any unique genetic features.

The frequency of meiotic recombination in the *T.b.gambiense* map is also a particular mystery for which further analysis might be aided by a higher density of markers. At least one reciprocal crossover per chromosome is usually considered essential for the successful disjunction of homologous chromosomes during meiosis (Baker *et al.*, 1976). It was therefore surprising that 48% of all STIB 386 chromosomes analysed in the cross failed to show evidence of any recombination events, and one hybrid clone exhibited no crossover events at all. As the present coverage of the genome by the genetic map is incomplete and many of the *T.b.gambiense* chromosomes are larger than that of the genome reference strain (Melville *et al.*, 1998), it is possible that the obligate crossover necessary to ensure faithful meiotic segregation is occurring outside of the central core covered by the genetic map on some STIB 386 chromosomes and towards the sub-telomeric regions where it would not be detected by our microsatellite analysis.

Alternatively it may be that double crossovers are present in the intervals between the existing microsatellite markers, and that their current spacing does not allow them to be detected. Illumina whole genome resequencing of a rice cultivar cross was recently demonstrated to improve the resolution of an existing microsatellite based map by approximately 35-fold and resulted in the identification of several double crossovers that had previously been missed by the limited marker density of the original map (Huang *et al.*, 2009). The addition of SNP markers to increase the resolution of the *T.b.gambiense* genetic map would allow us to determine if either of these possibilities were responsible for the relatively low average recombination frequency (0.6 crossovers/chromosome) compared to that of *T.b.brucei* (1.02 crossovers/chromosome). This is of interest as the mechanism of meiosis in trypanosomes remains little understood. Mendelian inheritance has been proven in *T.brucei* (MacLeod *et al.*, 2005a), which is consistent with a standard model of meiosis. However no gametes have as yet been identified and a high frequency of trisomy/triploidy has been observed in some genetic crosses (Gibson and Garside, 1991; Gibson and Bailey, 1994; Gibson *et al.*, 1995) which has resulted in the proposal of alternative models of mating (Gibson *et al.*, 1995). The mapping of recombination events, their frequency and

their physical location may provide some insight into how the mechanism of genetic exchange occur in *T.brucei*.

The publication of the *T.brucei* genome sequence, and its associated resources, has revolutionised our understanding of the African trypanosomes and enabled a whole new range of research possibilities. Next generation sequencing technologies are now taking this one step further with their ability to build on a reference sequence by sequencing the whole genome or transcriptomes of individual parasite strains. With the relatively small genome size of *T.brucei*, this capacity would now theoretically allow close to 1000-fold genome coverage in a single run. These advances in throughput are complemented by the multiplex capacity now possible on the three main platforms, which use a bar-coding system to enable multiple samples to be sequenced in parallel on the same run. In combination with such high coverage this would make the sequencing of a whole genetic cross much more economically feasible, greatly improving our ability to undertake high resolution linkage analysis of *T.brucei* in the future.

Appendices

Primers	Primer sequence	Chromosome location as of 12/08/08	PCR product size (bp)	PCR amplification conditions
Primers for amplification and cloning of TREU 927 LOH 1 for sequencing (Section 5.2.5)				
CH10 CLONE F2 CH10 CLONE R	TACCTATATCTTCGCACTTC ACCTCCTAGTAATTGCCTCA	2200416 – 2200435 2204474 -2204493	4077	Denature - 95°C for 50 seconds Anneal - 55°C for 50 seconds Extend - 65°C for 5 minutes x 30 cycles
Primers for SNP-RFLPS (Section 5.2.4)				
CH10solexa60snp2F CH10solexa60snp2R	GGACGGAATGAGATGAGAG AGTGGCGGTGGTCCTATT	2186461 –2186479 2187252 - 2187269	809	Denature - 95°C for 50 seconds Anneal - 60°C for 50 seconds Extend - 65°C for 2 minutes x 30 cycles
CH10solexa30snp16.17F CH10solexa30snp16.17R	CAGAGCGAGGTAGAAGGT CGTCATATTGCCGAGAACA	2195101 –2195118 2196651 -2196669	1569	Denature - 95°C for 50 seconds Anneal - 60°C for 50 seconds Extend - 65°C for 2 minutes x 30 cycles
CH10solexa30snp19F CH10solexa30snp19R	TTGATGATAATTCGATATGAGAC GCCTTCATCCTCTCTACCA	2203810 – 2203832 2204843 - 2204861	1052	Denature - 95°C for 50 seconds Anneal - 60°C for 50 seconds Extend - 65°C for 2 minutes x 30 cycles
CH10solexa60snp7F CH10solexa60snp7R	GTGGCTTCGTACCTCCAT CCACCACAATCGCCTTAC	2206559 –2206576 2207769 -2207786	1228	Denature - 95°C for 50 seconds Anneal - 60°C for 50 seconds Extend - 65°C for 2 minutes x 30 cycles
CH10solexa30snp23F CH10solexa30snp23R	CTGGTGCGAATGGAAGGT ATGGAATGAATGTAAGGATGAA	2225267 – 2225284 2226335 - 2226356	1090	Denature - 95°C for 50 seconds Anneal - 60°C for 50 seconds Extend - 65°C for 2 minutes x 30 cycles
CH10solexa60snp5F CH10solexa60snp5R	GTGAGTGTGAATGAACATATCA TACGCCTCACGGAATCA	2189144 – 2189165 2189925 -2189941	798	Denature - 95°C for 50 seconds Anneal - 60°C for 50 seconds Extend - 65°C for 2 minutes x 30 cycles
CH104469snp10F CH104469snp10R	TGTGTGTGGTCGTCTATGT AGTAACGAATATGACACATCTT	2929318 – 2929336 2929893 - 2929914	996	Denature - 95°C for 50 seconds Anneal - 60°C for 50 seconds Extend - 65°C for 2 minutes x 30 cycles

CH104469snp13F CH104469snp13R	CGTAGCAGAGGACACAAG CTGCTTATGTACTTGGTTGTG	2955215 –2955232 2956253- 2956273	1058	Denature - 95°C for 50 seconds Anneal - 60°C for 50 seconds Extend - 65°C for 2 minutes x 30 cycles
CH104469SNP13.7F CH104469SNP13.7R	TTGTTCCTCAGTACACTTGG CGATTCTGTAACCAATGTGA	2968269 – 2968288 2968610 - 2968629	360	Denature - 95°C for 50 seconds Anneal - 60°C for 50 seconds Extend - 65°C for 2 minutes x 30 cycles
CH104469snp14F CH104469snp14R	AGAATCCACGCTAGGACA ATAGAAGCGTATCATCCAC	2983234 –2983251 2983895 - 2983914	680	Denature - 95°C for 50 seconds Anneal - 60°C for 50 seconds Extend - 65°C for 2 minutes x 30 cycles
CH104469snp17F CH104469snp17R	GAGCGTTGAGGTGGAGA CCGCACCTCCACACCA	3001823 –3001839 3002322 - 3002337	514	Denature - 95°C for 50 seconds Anneal - 60°C for 50 seconds Extend - 65°C for 2 minutes x 30 cycles
CH104469snp19F CH104469snp19R	AGTAATAACAACTGTACCACAC GAAGTGAGAAGAATGATAATACA	3021319 – 3021340 3021806 - 3021828	509	Denature - 95°C for 50 seconds Anneal - 60°C for 50 seconds Extend - 65°C for 2 minutes x 30 cycles
CH10414snp19F CH10414snp19R	CTTACGCTCTGCTAACCT TCAGCAGTAGCATGTTCAC	2992342 –2992359 2993099 - 2993117	775	Denature - 95°C for 50 seconds Anneal - 60°C for 50 seconds Extend - 65°C for 2 minutes x 30 cycles

Appendix 1 Primer information and amplification conditions for PCR reactions in Chapter 5.

Marker	Alias	Primer-A	Primer-B	Chromosome location as of 26/10/06		PCR product size (bp)
				Primer A	Primer B	
Chromosome 1						
TB1/1	Ch1/MS42	GGTGATTCATCGGCTCCCTTGCCA	TTGTGCGGTCGTTAACGCGCGTTCAA	244160 - 244183	245193 - 245218	1058
TB1/2	Ch1/3E	AAGACTGTGAGATTAACATTCC	TTCTCCTCCTCCTCTTTATTT	333311 - 333332	333421 - 333441	130
TB1/4	Ch1/D2(7)	CGGCAGGGGAAGGGAGAA	GGAAGTGAGGGGAGACGGAAGAC	467634 - 467651	467736 - 467758	124
TB1/6	Ch1/9B	CGCTGTTATCATTATGGTCC	CAACAGCAGCATCAGTAGT	510612 - 510631	510783 - 510801	189
TB1/14	Ch1/Ch10/671	GTAAGTCGTGTTCGGGTTGTTCTT	GCATGTAATCTAGTCGCTAAATAG	613770 - 613793	613949 - 613972	202
TB1/15	Ch1/12H	ATGGACTCATGCGAACTTT	GTCTTCACACTTTCCTTCCC	723608 - 723626	723732 - 723751	143
TB1/16	Ch1/14	AGAAAGAGAATATAGTCAGAAAC	AATGCACCACTAATCTCTCT	747047 - 747069	747208 - 747227	180
TB1/10	Ch1/15	TGGCTAGTTACACTGTAGTTCTCC	CCACAACCACTCCTTGATATTCAC	768972 - 768995	769152 - 769175	203
TB1/12	Ch1/15C	GCTTCCTCTTCCTCTTCTT	CCCTACACCAACATAAATAAATAA	879354 - 879372	879449 - 879472	118
TB1/17	Ch1/23	CCCACTTCCCTCGCATTC	GACATAAATGGCAATTAATAATAA	988105 - 988122	988215 - 988238	133
Chromosome 2						
TB2/2	Ch2/A5	TTGAAATTGTGTATCGGTCTG	GGGAACAGATGTACGATCAT	353729 - 353749	353891 - 353910	1058
TB2/4	Ch2/A8	TCCTCTAATCACCGTACCT	TAACATGAAGGTAACTACTCT	423783 - 423801	423933 - 423953	130
TB2/7	Ch2/A11	AAGTGAAGAACACATATCGTAT	GAAACCCTAAAGCTGCTGT	546436 - 546457	546555 - 546573	124
TB2/9	Ch2/A34	TACAACATAAATTACAGCAATATAGT	CACATATTATAACCATTTGTGGA	583868 - 583893	583988 - 584010	189
TB2/10	Ch2/5	ATGGCGTGTATCACATTCGTGATG	CCGTTGGCATTAGGCACAAGTA	584890 - 584913	585067 - 585088	202
TB2/12	Ch2/A36	AAGTTACACCAGCCATAGAA	ATCGCTCCACCGCAACC	625282 - 625301	625433 - 625449	143
TB2/15	Ch2/A22	CTCGTCTCTCATCATTGCT	TGAGAGAATGGGATTGAGG	833171 - 833189	833290 - 833308	180
TB2/18	Ch2/A26	AAGGAAGAAAGTGTAGATATGA	GGGTCGCTTCTTTAACATAA	1044946 - 1044967	1045076 - 1045095	203
TB2/19	Ch2/A27	CTGGTGCGTGTAACTGTG	GAAGTGAGGACATGCACG	1082077 - 1082094	1082157 - 1082174	118
TB2/20	Ch2/PLC	CAACGACGTTGGAAGAGTGTGAAC	CCACTGACCTTTCATTTGATCGCTTTC	1089578 - 1089601	1089705 - 1089731	133
Chromosome 3						
TB3/1	Ch3/1J15/1	GTTAGGTTACGCAAGTCAGT	GAAACACTCAGTTCCACACC	304141 - 304160	304277 - 304296	155
TB3/19	Ch3/30P12/5	CATTCGACCGCTGTCCCTTCC	CCATGAATGGTGGCGGTGGCGG	472852 - 472872	473015 - 473036	184

TB3/20	Ch3/AC21	AGTCGCACAGTACCAACA	TACCGGCAATATTCCTTCC	688682 - 688699	688904 - 688922	240
TB3/4	Ch3/27C5/3	TGCTTCGACATATCTACCTGTTCAG	GCCATTATGTCATTTGCAAGAAAAGGG	745245 - 745269	745475 - 745501	256
TB3/21	Ch3/AC30	AAGTGAAGGTAAATGGATGAG	CAGAAGAGGGTGAAGGAAAG	1014473 - 1014493	1014619 - 1014638	165
TB3/22	Ch3/AC33	GCAGGGAAGGAACAAGGA	CTTCTCCACACATACACACT	1100053 - 1100070	1100344 - 1100363	310
TB3/23	Ch3/AC40	CACGGTTCTTGTTGGTATTG	AAGGTGGTGGGCATGATTT	1307993 - 1308012	1308232 - 1308250	257
TB3/10	Ch3/B3	AGGGAAGTGTTGACGCACATCCTC	ACACTAACCTTGCTCTACACTGTT	1460952- 1460975	1461039 - 1461062	110
TB3/13	Ch3/292	ACACCCCCTCTCCACTTCAGATAC	GCTGAACCTGTGGGCCCCTCAATTG	1547630 - 1547653	1549076 - 1549100	1470
TB3/14	Ch3/A1	ATGCGTTTGTATGCTTCCAGTC	CACGGAGTCTGGTACCAACAGTCG	1557995 - 1558016	1558093 - 1558116	121
Chromosome 4						
TB4/17	Ch4/AC03	AGCGTTTGGGTAGAAGTGA	ATGATCCAGCGGCTTGG	183484 - 183502	183650 - 183666	182
TB4/2	Ch4/2L9/2	GCCGCTTGATCATTAGGTAACCAC	CCGCCTCACTTTAAGGATGGTGCC	314026 - 314049	314095 - 314118	92
TB4/18	Ch4/29M18/2	CGCACACGCATCCAAATGTATATG	GATCATCTGTCGAGGCATCATCC	472621 - 472644	472793 - 472815	194
TB4/19	Ch4/29M18/5	GAATCTTCATGCATCTCCCTTCAG	CCATAGAAGTACATGACAAGAG	479883 - 479906	480128 - 480149	194
TB4/4	Ch4/7B7	GTAGTCGCTTCCTTTACGGCACG	GCACCGCATGTTATTACACG	LOCATION	UNDETERMINED	
TB4/5	Ch4/1H19/1	GTGGTAACGTAAGTTATTGTCACC	CAGTGACGATAGAACTGACACCGC	589893 - 589916	590095 - 590118	225
TB4/20	Ch4/1H19/2	TACCGCCCACGATGGTATCGTCCG	CCACCAATTTAGTCAGCGCAACTG	705155 - 705178	705559 - 705582	427
TB4/8	Ch4/C2 (F/R2)	CCGTTCAGTGTGCATGTTTCAC	CCGTCACACGCCATGCACGATATG	833811 - 833832	833998 - 834021	210
TB4/21	Ch4/26G5/1	CGAGTAGCCGCGCGCTAGTATCCTC	CTGCAGAACGAGCGCTTGGCGGAG	940469 - 940493	942192 - 942215	1746
TB4/22	Ch4/30K5-2	GACATGTACTGCTACACTCCCGCC	GCTTGTGGTCAAGTAGTCGTATAC	1037050 - 1037073	1037225 - 1037248	198
TB4/12	Ch4/3I12/7	GCGGGCATAACCGTCTATCATAC	GAATACCTGATACGTTGTACCTAC	1221550 - 1221572	1221646 - 1221669	74
TB4/13	Ch4/2A13	AACATTGGCGCGAAGCACTTC	GGTGTCAGAAGTTATCACTGCG	1233288 - 1233308	1233537 - 1233558	270
Chromosome 5						
TB5/15	Ch5/29K2/1	AAGCCTCCACCACGCAC	TTCACTGCGTGGCCTGC	116336 - 116352	116483 - 116499	163
TB5/16	Ch5/AC06	AGAGGTGTACGATATGCCA	GCTCTATCTGAATACACGG	292891 - 292909	293181 - 293199	308
TB5/4	Ch5/JS2	GATTGGCGCAACAACTTTCACATACG	CTTTCTTCCTTGGCCATTGTTTTACTAT	479279 - 479304	479432 - 479461	182
TB5/17	Ch5/AC14	ACCTTTCTTTCATCGCACT	TTAGACGACGCAACACACA	523320 - 523338	523545 - 523563	243
TB5/18	Ch5/26K5/3	AATCATGACCACGATGGTGAG	ATTTCATCATGGGCTCGCTGAG	864884 - 864904	865021- 865042	158

TB5/19	Ch5/AC25	TTGAAGAAGTTAAAGGGGTG	AGTAAGCGATGGTGAATGG	1023256 - 1023275	1023673 - 1023691	278
TB5/20	Ch5/26c7/4	CAGTCGCATTATCTATACTCA	TGAGCAGCCAACGTATGTA	1319903 - 1319923	1320009 - 1320027	124
Chromosome 6						
TB6/9	Ch6/AC06	GTGGTAGCGTTGTTGATGT	TGAGAGATGCGTTGGTTC	424311 - 424329	424432 - 424449	138
TB6/10	Ch6/28P18/4	AATAACGATTTGATAACAGAAGA	AAGCAGAGATTGCAGATTTC	594194 - 594216	594457 - 594476	282
TB6/11	Ch6/4M18/4	GAAAGGCGAAACCTTGCGATCC	AACCATTCATGGATCACTCCAGC	797425 - 797446	797569 - 797591	166
TB6/12	Ch6/5f5/4	AATGAGGTTGGAGGAGTGT	CGCGCTAATTATGGTGACA	906562 -906580	906662 - 906680	118
TB6/13	Ch6/4F7/7	GGTCGCACACACAACTC	TGGAGGAGTCGGTGTG	1152020 - 1152036	1152113 - 1152128	108
TB6/6	Ch6/4F7/6	AATGATCCCTCCCGCATAC	CTGCGTGTGGAAGTGGA	1175946 - 1175964	1176044 - 1176060	114
TB6/14	Ch6/26g9/3	CTAACTGCACAACATAGAGG	AAGCACTTGCGCATTCATC	1228275 - 1228294	1228137 - 1228155	157
TB6/15	Ch6/26g9/1	ATGTCGTTTGCGGTGAATG	TGTGCTGTAATTAAGCATACC	1284550 - 1284568	1284686 - 1284706	156
TB6/16	Ch6/30P15/7	GAGTAACTGCTCGCTTCC	AGACGCAAGTACGAAACC	1362684 - 1362701	1362926 - 1362943	259
Chromosome 7						
TB7/1	Ch7/8P12/2A	TGTGGCACCATGTTATTGTT	GCCAATATAGTCAACCGTAG	38910 - 38929	39028 - 39047	147
TB7/14	Ch7/29K4/A2	GGGAGAGATCGTTTGATTCC	AGGTCTAAGCAATATCTATGC	174364 - 174383	174478 - 174498	134
TB7/4	Ch7/43M14/3	GGGAGTATGCACATATATCAACTG	CATACGTTAAGATGTTGGTGCACG	529806 - 529829	530079 - 530102	296
TB7/5	Ch7/5F10/A4	AACTAGTGATTAAGCGAATACAC	GTCTCCGCCATCACTTG	1044318 - 1044340	1044496 - 1044512	194
TB7/15	Ch7/10C21/1	GCATAATGATGGAAGAATAGTAGAC	CTCACTGATCGCTGCATCCACAAC	1565317 - 1565341	1565575 - 1565598	281
TB7/17	Ch7/33N13/4	GGAGGAATGAGCGCAGTGACACC	GTTCCTGTGTGACTGAGCGACACC	1683957 - 1683979	1684130 - 1684153	196
TB7/16	Ch7/33N13/2	GCACAGACACAACCTCGCTCACGC	GCGCAGCGTTTGTTGTGTCAGGTG	1687722 - 1687745	1688015 - 1688038	316
Chromosome 8						
TB8/12	Ch8/B3	GTTGATATTCGGTATTGTTCCT	CTTCTGTTGGGCGACTG	171047 - 171068	171302 - 171318	271
TB8/13	Ch8/AC02	CGTGCCAATTGTACGGAA	AGAAGGAAGGAAGCAACAC	313150 - 313167	313359 - 313377	227
TB8/14	Ch8/AC09	CACACCTGCCTCACTACA	CACCTCCACACTCATTCAC	502421 - 502438	502535 - 502553	132
TB8/15	Ch8/AC13	AACACGCATTGGTTGGTCT	AGATAAGAGTATAGCCGACA	678469 - 678487	678989 - 679008	539
TB8/16	Ch8/AC17	GGTAATGATCCACAACGAAG	CATCAACATCAACACCTAACA	778291 - 778310	778507 - 778527	236
TB8/17	Ch8/AC18	AGGGAAAGAGGAAAGGAGAA	TCACATCACAATCTCAACCA	830756 - 830775	831162 - 831181	425
TB8/18	Ch8/28L1/2	GAAAAGCGGATGCAGAGGATATAC	GTGATACCATGATGCGATTTCTGTC	1009601 - 1009624	1009728 - 1009752	151

TB8/10	Ch8/11J15/2	TGTTGCCCTCACGGTTC	ACAGACCGCTACATATAACA	1746679 - 1746695	1746834 -1746853	174
TB8/19	Ch8/AC32	TAGTTATTGATGAAGTTGCCG	CTTCCATCCTCGTCCTCA	2344450 - 2344470	2344663 - 2344680	230
TB8/20	Ch8/AC35	GCAACGGGATGGAGAGAA	TTCACACCTACACTCCACTT	2386552 - 2386569	2386860 - 2386879	327
TB8/21	Ch8/AC40	ATAGTAAGTGGTTGTGGTATATTA	AGGATGATATGAAGGAATAACC	2469169 - 2469192	2469564 - 2469585	416
Chromosome 9						
TB9/17	Ch9/43	AGGCATCTGTTGAGCACC	CATCTAACGCCACAGTTCT	361746 - 361763	361919 - 361937	191
TB9/18	Ch9/40	ACACTAATAAACATCGTAAATCA	ATTTCGTTTGTTCACTCCG	457858 - 457880	457976 - 457994	136
TB9/19	Ch9/160/5	TTCTCTCTTCTTCCTTTCACT	GGAGAAGTAGTAGTAGTCGAAT	482643 - 482663	482746 - 482767	124
TB9/20	Ch9/160/6	ATTCGTATAAACGGACGGG	GGTGCTACTTGGCGTAAA	550172 - 550190	550437 - 550454	282
TB9/21	Ch9/1	TGCTGTACTGGTTGGTCTC	GATGAGCAATTTGTAGTGCC	1123491 - 1123509	1123828 - 1123847	356
TB9/5	Ch9/21	TCTGATCCCATCCATTTACAC	AGGAGCGAAGGAGAGGAC	1313966 - 1313986	1314188 - 1314205	239
TB9/9	Ch9/8	AATATGTTCCATACCCTTATGTTTT	CGCAATGGGCATACAACT	1671088 - 1671112	1671242 - 1671259	171
TB9/12	Ch9/19	AAGTGTGAGGAGTTGTTGT	CACCCCTTTCATCAACATCAT	1911962 - 1911980	1912086 - 1912106	144
TB9/14	Ch9/68	AATGTAAATATAGCGAGCAGT	AGATTCTCACATCGTGTGTA	2055901 - 2055921	2056000 - 2056019	118
TB9/22	Ch9/132F04/2	CTTCCTACGCATACGTGT	TGTGCTCTCTGATCTTTGTG	2457464 - 2457480	2457632 - 2457651	187
Chromosome 10						
TB10/22	Ch10/9.1	CTGTGGATACACATGTACATGTAC	CCTTCGCATGAGCCTGCATTC	124019 - 124042	124201 - 124221	202
TB10/23	Ch10/32	CGAGCGTGGTTGTAGTATCGCAAC	GCATCAAGTCTGTTTGATGGCCGC	242572 - 242595	243752 - 243775	1203
TB10/24	Ch10/AC10	GCACATTCATATCACTAACTGA	ACACTAGAGAACAGCACAAG	1304153 - 1304174	1304309 - 1304328	175
TB10/25	Ch10/Ch11/10	TAGATTGTGAAAGTTCTTAAATAAT	TCATTATATCACCTGTCACTCT	1515608 - 1515632	1515751 - 1515772	164
TB10/12	Ch10/2A	CTTGAATGTAAGTACGGGTC	TGATGACGGCTACGATTC	1626747 - 1626766	1626849 - 1626866	119
TB10/14	Ch10/CRAM	AACTCCCTCCCGATCGATCACAAC	CTGCTGATGCCGTACATGATGATTTC	1784294 - 1784317	1787688 - 1787713	3319
TB10/26	Ch10/AC14	CCTCCTTCTGTGCAACAC	CAAGAGAGGACATTGGTGG	1959573 - 1959590	1959672 - 1959690	117
TB10/27	Ch10/AC25	ACAGAACGAGTAGGTGAAG	AGCGGCCATCAACCTCT	2294341 - 2294359	2294630 - 2294646	305
TB10/28	Ch10 /3778	CAACAGTGCCCGGAGGAGGTACG	GCGGTGACTCGTCCGTCTTCTTGC	2365809 - 2365831	2366295 - 2366318	509
TB10/29	Ch10/12.1	CACGGAAACGAAGTCAACGGCA	CCAGTCACTGATATCACTTCCTTCC	2378431 - 2378452	2378618 - 2378642	211
TB10/19	Ch10 41.1	CTGTTCGTTCTGAATTGTGTGCG	GTGCACTTCCTTCTCTCATCCTTTTC	3124436 - 3124458	3124574 - 3124599	163

TB10/30	Ch10/909	CTTGTTCGTGGCTGGTGTATGGAC	CGTGTGGATACTTCCATCCAATG	3681514 - 3681537	3681725 - 3681747	233
Chromosome 11						
TB11/32	Ch11/10/490	GCTTATTACTGGTGTGAATATTGTC	CAGTGTGGGCAGTGACAAGAGG	904119 - 904143	904236 - 904257	138
TB11/7	Ch11/51	TGAGATGGTACTTGAAGAAAG	AACCGATCATTCCTGTTCC	1447665 - 1447685	1447747 - 1447765	100
TB11/33	Ch11/53	CGTGTGTCTTGTATATCTTCT	TGAATAAACAAAACATGAAACGAC	1524474 - 1524494	1524539 - 1524562	88
TB11/10	Ch11/2	TCAATTACGGGTATACACCTG	CAGTAAATGACAAGATACGTGG	1796615 - 1796635	1796692 - 1796713	98
TB11/11	Ch11/65	TCGGTCGATCTTTCTCTC	ATGACACATGTGAGTGAGTGAA	1911199 - 1911216	1911255 - 1911276	77
TB11/13	Ch11/49	CAAGAACTCTGCATTGAGC	ATCTGTTGGCGATGGTGA	2110237 - 2110255	2110377 - 2110394	157
TB11/34	Ch11/11	ATTTCTTGGAGTGGATCTTAG	GATGTAACGGAGTGCTTGA	2271895 - 2271915	2272013 - 2272031	136
TB11/35	Ch11/99	AACCGCCTTGTGAGATG	TCTTTGCTCTGCGATTACT	2540103 - 2540119	2540305 - 2540323	220
TB11/36	Ch11/10/753	CGAGAACCAGAACAAGTCGTTATC	GCTGGTGTCGTTGTTGTTGGG	2759249 - 2759272	2759397 - 2759414	165
TB11/15	Ch11/54	TCCACATGATATACCCACCT	CCTCTTTCTTCACTCACCA	2824278 - 2824297	2824384 - 2824402	124
TB11/37	Ch11/124	AGGCGTAGCGATCAGCACATGCCT	TTTGTGCAGTGCAGTATTGTAAAT	2862043 - 2862066	2862318 - 2862341	298
TB11/38	Ch11/127	GTCTCTCACATGGTCGGAGACGAA	AAGGTGCATACCACACGAGTTTGG	3078260 - 3078283	3078589 - 3078612	352
TB11/39	Ch11/132	ATGGACGATACAATGCTGGTCTCC	ACACCGATGCAATATACCGTGGTG	3231439 - 3231462	3231642 - 3231665	226
TB11/40	Ch11/135	GTAGTAAGGTTAAACGCCTTCACG	GTCACACCTTGGCACAGCAGTAAT	3247074 - 3247097	3247243 - 3247266	192
TB11/41	Ch11/33	CGGATATAAGGATGGAGGAA	TGAAATCGGGATCAGTAGAA	3250722 - 3250741	3250853 - 3250872	150
TB11/42	Ch11/102	ATTACTGCTCGCATATTGC	AGCGACGTAAATTTGCATAA	3723749 - 3723767	3723931 - 3723950	201
TB11/21	Ch11/46	GGGGAAATGAATACACAACAG	TACTCACCATGTTTGAATTTCC	3851812 - 3851832	3851939 - 3851960	148
TB11/43	Ch11/30	GTTAAGACAGATGAATACAACAAAA	CTTCCATTCATCCTCTGTGT	3872014 - 3872038	3872150 - 3872169	155
TB11/44	Ch11/20	AAAGGAGCGTGAAGATTGA	AATACCTCTACGAATAAGCAC	4093825 - 4093843	4093915 - 4093935	110
TB11/45	Ch11/106	GTCACGGCTCTCACCTC	ACTTGTCGCTGCTTCTG	4152033 - 4152049	4152195 - 4152211	178
TB11/23	Ch11/24	ATCACGCTCATTCTTCACT	TACTATTACAGAGTGGTGATAAA	4326401 - 4326419	4326551 - 4326573	172

Appendix 2 - Primer information for all the microsatellite markers used in the generation of the STIB 386 genetic map

The assigned marker names are given (column 1), consistent with those present in the *T.b.brucei* genetic map, alongside the laboratory used alias (column 2). Columns 3 and 4 contain the forward and reverse primer pair sequences for each marker and Columns 5 & 6 the genomic location of these primer sequence based on the available TREU 927 sequence. Finally column 7 indicates the approximate size of the amplified PCR product according to the TREU 927 sequence.

Primer	Primer sequence	Primer Location as of 28/8/08
CH10 CLONE F2	TACCTATATCTTCGCACTTC	2200416 - 2200435
2ND CH10 SEQ F2	AACACCCAATTTCACCGTG	2201196 - 2201214
3RD CH10 SEQ F2	ACGGATGTACTTACCGATG	2202001 -2202019
4TH CH10 SEQ R	CAATACAAGAAGAGTGAGAAGT	2203707 -2203728
CH10 CLONE R	ACCTCCTAGTAATTGCCTCA	2204474 - 2204493

Appendix 3 Primers used to sequence the LOH boundary of TREU 927 LOH line 1 (Section 5.2.5).

Marker	Progeny																		
	F9/45 mcl 2	F9/45 mcl10	F9/45 mcl11	F9/45 mcl 12	F9/34 mcl 1	B80 cl 2	F492/50 bscl 1	F492/50 bscl 6	F492/50 bscl 8	F492/50 bscl 9	F492/50 bscl 12	F492/50 bscl 14	F492/50 bscl 21	F492/50 bscl 23	F492/50 bscl 5/1b	F9/41 bscl 5	F9/41 bscl 7	F9/41 bscl 9	F29/46 bscl 3
	30	34	35	36	37	162	174	179	181	182	185	187	194	196	332	383	384	385	386
TB1/1	B	A	A	trisomic	-	B	A	trisomic	A	B	B	A	A	A	A	B	B	A	A
TB1/2	B	A	A	trisomic	-	B	A	trisomic	A	B	B	A	A	A	A	B	B	A	A
TB1/4	B	A	A	trisomic	-	B	A	trisomic	A	B	B	A	A	A	A	B	B	A	A
TB1/6	B	A	A	trisomic	-	B	A	trisomic	A	B	B	A	A	A	A	B	B	A	A
TB1/14	-	A	A	trisomic	A	B	B	trisomic	A	B	B	A	A	A	A	B	B	A	A
TB1/15	B	A	A	trisomic	-	B	B	trisomic	A	B	B	A	A	A	A	A	B	A	A
TB1/16	B	A	A	trisomic	-	B	B	trisomic	A	B	B	A	A	A	A	A	B	A	A
TB1/10	B	A	A	trisomic	-	B	B	trisomic	A	B	B	A	A	A	A	A	B	A	A
TB1/12	B	A	A	trisomic	-	B	B	trisomic	A	B	B	B	A	A	A	A	B	A	A
TB1/17	B	A	B	trisomic	-	B	B	trisomic	A	B	A	B	B	A	A	A	B	A	A

Marker	Progeny																		
	F29/46 bscl 4	F19/31 clone1	F19/31 bscl 11	F28/46 bscl 6	F28/46 bscl 11	F29/46 bscl 2	F28/46 bscl 1	F28/46 bscl 4	F28/46 bscl 7	F28/46 bscl 8	F29/46 bscl 1	F9/41 bscl 1	F9/41 bscl 2	F9/41 bscl 8	F9/41 bscl11	F19/31 bscl 5	F19/31 bscl 10	F19/31 bscl 8	F492/50 bscl 7
	387	389	391	394	405	406	407	409	410	411	414	415	416	420	422	425	430	428	180G12
TB1/1	A	A	A	A	B	A	A	A	A	B	A	B	A	A	A	A	A	A	-
TB1/2	A	A	A	A	B	A	A	A	A	B	A	B	A	A	A	A	A	A	A
TB1/4	A	A	A	A	B	A	B	A	A	B	A	B	A	A	A	A	A	A	A
TB1/6	A	A	A	A	B	A	B	A	A	B	A	B	A	A	A	A	A	A	A
TB1/14	-	A	A	A	B	A	B	A	A	B	A	B	A	A	A	A	A	A	-
TB1/15	A	A	A	B	B	A	B	A	B	B	A	B	A	A	A	A	A	A	A
TB1/16	A	A	A	B	B	A	B	A	B	-	-	B	A	B	A	A	A	A	A
TB1/10	A	A	A	B	B	A	B	A	B	B	B	B	A	B	A	A	A	A	A
TB1/12	A	B	A	B	B	A	B	A	B	B	B	B	A	B	A	A	A	A	A
TB1/17	A	B	A	A	A	A	B	A	B	B	B	B	B	B	B	A	A	A	A

Appendix 4a The segregation data for all of the microsatellite markers on chromosome 1 in the STIB 386 genetic map.

For each table; Column 1 lists the marker name and columns 2-20 the inheritance pattern of the STIB 386 alleles (either A or B) in the progeny clones for each marker. The name of each progeny clone is given, along with laboratory alias along the top of each table. Trisomy is indicated where appropriate and novel sized alleles are marked as mutant. Progeny for which allele inheritance has not been determined are left blank.

Marker	Progeny																		
	F29/46 bscl 4	F19/31 clone 1	F19/31 bscl11	F28/46 bscl 6	F28/46 bscl 11	F29/46 bscl 2	F28/46 bscl 1	F28/46 bscl 4	F28/46 bscl 7	F28/46 bscl 8	F29/46 bscl 1	F9/41 bscl 1	F9/41 bscl 2	F9/41 bscl 8	F9/41 bscl11	F19/31 bscl 5	F19/31 bscl 10	F19/31 bscl 8	F492/50 bscl 7
	387.1	389	391	394.1	405	406	407	409.1	410.1	411	414	415	416	420	422	425	430	428	180G12
TB2/2	B	A	B	B	B	A	B	A	A	B	A	A	A	A	A	A	A	-	B
TB2/4	B	A	B	B	B	A	B	A	A	B	A	A	A	A	A	A	A	A	-
TB2/7	B	A	B	B	B	B	A	A	A	B	A	A	A	A	B	A	B	-	-
TB2/9	B	A	B	B	B	B	A	A	A	B	A	A	A	A	B	B	B	B	-
TB2/10	B	A	B	B	-	B	A	A	A	B	A	A	A	A	B	B	B	B	-
TB2/12	B	A	B	B	B	B	A	A	A	B	A	A	A	A	B	B	B	B	-
TB2/15	B	A	B	B	-	B	A	B	A	B	A	-	-	A	-	-	-	-	-
TB2/18	B	A	B	B	B	B	A	B	A	B	A	A	A	A	B	B	B	B	-
TB2/19	B	A	B	B	-	B	A	B	A	B	A	A	-	A	B	B	B	B	-
TB2/20	A	A	B	B	B	B	A	B	A	B	A	A	A	A	B	B	B	B	-

Marker	Progeny																		
	F9/45 mcl 2	F9/45 mcl 10	F9/45 mcl 11	F9/45 mcl 12	F9/34 mcl 1	B80 cl 2	F492/50 bscl 1	F492/50 bscl 6	F492/50 bscl 8	F492/50 bscl 9	F492/50 bscl 12	F492/50 bscl 14	F492/50 bscl 21	F492/50 bscl 23	F492/50 bscl 5/1 b	F9/41 bscl 5	F9/41 bscl 7	F9/41 bscl 9	F29/46 bscl 3
	30	34	35	36	37	162	174	179	181	182	185	187	194	196	332	383	384	385	386.1
TB2/2	A	A	A	B	-	B	A	B	B	B	B	A	A	B	B	A	A	B	B
TB2/4	A	A	A	B	-	B	A	A	B	B	A	A	A	B	B	A	A	B	B
TB2/7	A	A	B	B	-	B	A	A	B	B	A	A	B	B	B	A	A	B	B
TB2/9	A	A	B	B	-	B	B	A	B	B	A	A	B	B	B	A	A	B	B
TB2/10	A	A	B	B	-	B	B	A	B	B	A	A	B	B	B	A	A	B	B
TB2/12	A	A	B	B	B	B	B	A	B	B	A	A	B	B	B	A	A	B	B
TB2/15	A	A	B	-	-	B	B	-	-	B	A	A	B	B	B	B	A	B	B
TB2/18	A	A	B	B	-	B	B	A	B	B	A	A	B	B	B	B	A	B	B
TB2/19	A	B	B	B	-	B	B	A	B	B	A	A	B	B	B	B	A	B	B
TB2/20	A	B	B	B	-	B	B	A	B	B	A	A	B	B	B	B	A	B	B

Appendix 4b The segregation data for all of the microsatellite markers on chromosome 2 in the STIB 386 genetic map. See Appendix 4a legend for further details.

Marker	Progeny																		
	F9/45 mcl 2	F9/45 mcl 10	F9/45 mcl 11	F9/45 mcl 12	F9/34 mcl 1	B80 cl 2	F492/50 bscl 1	F492/50 bscl 6	F492/50 bscl 8	F492/50 bscl 9	F492/50 bscl 12	F492/50 bscl 14	F492/50 bscl 21	F492/50 bscl 23	F492/50 bscl 5/1b	F9/41 bscl 5	F9/41 bscl 7	F9/41 bscl 9	F29/46 bscl 3
	30	34	35	36	37	162	174	179	181	182	185	187	194	196	332	383	384	385	386.1
TB3/1	B	B	A	A	B	B	A	A	B	A	B	-	B	A	A	B	A	B	B
TB3/19	B	B	A	A	B	B	A	A	B	A	A	A	B	A	A	B	A	B	B
TB3/20	A	B	A	A	B	B	A	A	B	A	B	A	A	A	A	B	A	B	B
TB3/4	A	B	A	A	B	-	A	A	B	A	B	A	A	A	A	B	A	A	A
TB3/21	-	B	A	A	B	B	B	A	B	A	B	A	A	A	A	B	A	A	A
TB3/22	-	B	A	A	B	B	B	A	B	A	B	A	A	A	A	B	A	A	A
TB3/23	A	B	A	B	B	B	B	A	B	A	B	A	A	A	A	B	A	A	A
TB3/10	A	B	A	B	B	B	B	A	B	A	B	A	A	A	A	B	B	A	A
TB3/13	A	B	A	B	B	B	B	A	B	A	B	A	A	A	A	B	B	A	A
TB3/14	B	B	A	B	B	B	B	A	B	A	B	A	A	A	A	B	B	A	A

Marker	Progeny																		
	F29/46 bscl 4	F19/31 clone1	F19/31 bscl 11	F28/46 bscl 6	F28/46 bscl 11	F29/46 bscl 2	F28/46 bscl 1	F28/46 bscl 4	F28/46 bscl 7	F28/46 bscl 8	F29/46 bscl 1	F9/41 bscl 1	F9/41 bscl 2	F9/41 bscl 8	F9/41 bscl11	F19/31 bscl 5	F19/31 bscl 10	F19/31 bscl 8	F492/50 bscl 7
	387.1	389	391	394.1	405	406	407	409.1	410.1	411	414	415	416	420	422	425	430	428	180G12
TB3/1	A	A	A	B	B	B	A	A	B	A	B	B	A	B	A	B	B	B	A
TB3/19	A	A	A	A	B	B	A	A	B	A	B	B	A	B	A	B	B	B	A
TB3/20	A	A	A	A	B	A	A	A	B	A	B	B	A	B	A	B	A	B	A
TB3/4	A	A	A	A	B	A	-	A	B	A	B	B	A	B	A	B	A	B	A
TB3/21	A	A	A	A	B	A	A	A	B	A	B	B	A	B	A	B	A	B	A
TB3/22	A	A	A	A	B	A	A	A	B	A	B	B	A	B	A	B	A	B	A
TB3/23	A	A	A	A	B	A	A	A	B	A	B	B	A	B	A	A	A	A	A
TB3/10	A	B	A	A	B	A	A	A	B	A	B	B	A	B	A	A	A	A	A
TB3/13	A	B	A	A	B	A	A	A	B	A	B	B	A	B	A	A	A	A	A
TB3/14	A	B	A	A	B	A	A	A	B	A	B	B	A	B	A	A	A	A	A

Appendix 4c The segregation data for all of the microsatellite markers on chromosome 3 in the STIB 386 genetic map. See Appendix 4a legend for further details.

Marker	Progeny																		
	F9/45 mcl 2	F9/45 mcl 10	F9/45 mcl 11	F9/45 mcl 12	F9/34 mcl 1	B80 cl 2	F492/50 bscl 1	F492/50 bscl 6	F492/50 bscl 8	F492/50 bscl 9	F492/50 bscl 12	F492/50 bscl 14	F492/50 bscl 21	F492/50 bscl 23	F492/50 bscl 5/1 b	F9/41 bscl 5	F9/41 bscl 7	F9/41 bscl 9	F29/46 bscl 3
	30	34	35	36	37	162	174	179	181	182	185	187	194	196	332	383	384	385	386
TB4/17	A	A	A	A	A	A	B	B	B	B	-	-	B	B	B	A	A	B	B
TB4/2	A	A	A	-	A	A	B	B	B	B	-	-	B	B	B	A	A	B	B
TB4/18	A	A	A	-	A	A	B	B	B	B	-	-	B	B	B	A	A	B	B
TB4/19	A	A	A	-	A	A	B	-	B	B	-	-	B	B	B	A	A	B	B
TB4/4	A	A	A	B	A	A	B	B	B	B	A	B	B	B	B	A	A	B	B
TB4/5	A	A	A	-	A	A	B	B	B	B	-	-	B	B	B	A	A	B	B
TB4/20	A	A	A	-	A	A	A	B	B	B	-	-	B	B	B	A	A	B	B
TB4/8	B	A	A	-	A	A	A	B	B	B	-	-	B	B	B	A	A	B	B
TB4/21	B	A	A	B	A	A	A	B	B	B	-	B	B	B	B	A	A	B	B
TB4/22	B	A	A	-	A	A	A	B	-	B	-	-	B	B	B	A	A	A	A
TB4/12	B	A	A	-	A	A	A	B	-	B	-	-	B	B	B	A	A	A	A
TB4/13	A	A	A	B	A	A	A	B	B	B	-	B	B	B	B	A	A	A	A

Marker	Progeny																		
	F29/46 bscl 4	F19/31 clone 1	F19/31 bscl 11	F28/46 bscl 6	F28/46 bscl 11	F29/46 bscl 2	F28/46 bscl 1	F28/46 bscl 4	F28/46 bscl 7	F28/46 bscl 8	F29/46 bscl 1	F9/41 bscl 1	F9/41 bscl 2	F9/41 bscl 8	F9/41 bscl11	F19/31 bscl 5	F19/31 bscl 10	F19/31 bscl 8	F492/50 bscl 7
	387	389	391	394	405	406	407	409	410	411	414	415	416	420	422	425	430	428	180G12
TB4/17	A	B	A	A	B	A	A	A	B	B	B	B	A	B	A	A	A	B	B
TB4/2	A	B	A	B	B	A	B	A	B	B	B	A	A	B	A	A	B	A	B
TB4/18	B	B	A	B	A	A	B	A	B	B	B	A	A	B	A	A	B	A	B
TB4/19	B	B	A	B	-	A	B	A	B	B	B	-	A	-	A	A	B	A	B
TB4/4	B	B	A	B	A	A	B	A	B	B	B	A	A	B	A	A	B	A	B
TB4/5	B	B	A	B	A	A	B	A	B	B	B	-	A	B	A	A	B	A	B
TB4/20	B	B	B	B	A	A	B	A	B	B	B	A	A	B	A	A	B	A	B
TB4/8	B	B	B	B	B	B	B	A	B	B	B	A	A	B	A	A	B	A	B
TB4/21	B	B	-	B	B	B	B	A	B	B	B	A	A	B	A	A	B	A	B
TB4/22	B	B	B	B	B	B	B	A	B	B	B	A	A	B	A	A	B	A	B
TB4/12	B	B	B	B	B	B	B	A	B	B	B	A	A	B	A	B	B	B	B
TB4/13	B	B	B	B	B	B	B	A	B	B	B	A	A	B	A	B	B	B	B

Appendix 4d The segregation data for all of the microsatellite markers on chromosome 4 in the STIB 386 genetic map. See Appendix 4a legend for further details.

Marker	Progeny																		
	F9/45 mcl 2	F9/45 mcl 10	F9/45 mcl 11	F9/45 mcl 12	F9/34 mcl 1	B80 cl 2	F492/50 bscl 1	F492/50 bscl 6	F492/50 bscl 8	F492/50 bscl 9	F492/50 bscl 12	F492/50 bscl 14	F492/50 bscl 21	F492/50 bscl 23	F492/50 bscl 5/1 b	F9/41 bscl 5	F9/41 bscl 7	F9/41 bscl 9	F29/46 bscl 3
	30	34	35	36	37	162	174	179	181	182	185	187	194	196	332	383	384	385	386
TB5/15	A	A	A	B	B	B	B	A	A	B	-	A	B	B	B	B	A	B	B
TB5/16	A	A	A	A	B	A	B	A	A	B	A	A	B	B	B	B	A	B	B
TB5/4	A	A	A	A	B	A	B	A	A	B	B	A	B	B	B	A	A	B	B
TB5/17	A	A	A	A	B	A	B	A	A	B	-	A	B	B	B	A	A	B	B
TB5/18	A	B	B	A	B	A	B	A	B	-	B	A	B	B	B	A	B	B	B
TB5/19	A	B	A	B	B	A	B	A	A	B	-	A	A	B	B	A	B	A	A
TB5/20	A	B	A	B	B	A	B	A	A	A	-	A	A	B	B	A	B	A	A

Marker	Progeny																		
	F29/46 bscl 4	F19/31 clone1	F19/31 bscl 11	F28/46 bscl 6	F28/46 bscl 11	F29/46 bscl 2	F28/46 bscl 1	F28/46 bscl 4	F28/46 bscl 7	F28/46 bscl 8	F29/46 bscl 1	F9/41 bscl 1	F9/41 bscl 2	F9/41 bscl 8	F9/41 bscl11	F19/31 bscl 5	F19/31 bscl 10	F19/31 bscl 8	F492/50 bscl 7
	387	389	391	394	405	406	407	409	410	411	414	415	416	420	422	425	430	428	180G12
TB5/15	B	B	B	B	B	A	B	B	B	B	A	A	A	A	A	A	A	-	B
TB5/16	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	B	A	B
TB5/4	B	A	A	B	B	B	B	B	B	B	A	A	A	A	A	A	B	A	B
TB5/17	B	A	A	B	B	B	B	B	B	B	A	A	A	A	A	A	B	-	B
TB5/18	B	A	A	B	B	B	B	B	B	B	A	B	A	A	A	B	B	B	B
TB5/19	B	A	A	A	B	B	B	B	B	B	A	B	A	A	A	B	A	B	B
TB5/20	B	A	A	A	A	B	A	B	A	B	A	A	A	A	A	B	A	B	B

Appendix 4e The segregation data for all of the microsatellite markers on chromosome 5 in the STIB 386 genetic map. See Appendix 4a legend for further details.

Marker	Progeny																		
	F9/45 mcl 2	F9/45 mcl 10	F9/45 mcl 11	F9/45 mcl 12	F9/34 mcl 1	B80 cl 2	F492/50 bscl 1	F492/50 bscl 6	F492/50 bscl 8	F492/50 bscl 9	F492/50 bscl 12	F492/50 bscl 14	F492/50 bscl 21	F492/50 bscl 23	F492/50 bscl 5/1 b	F9/41 bscl 5	F9/41 bscl 7	F9/41 bscl 9	F29/46 bscl 3
	30	34	35	36	37	162	174	179	181	182	185	187	194	196	332	383	384	385	386
TB6/9	B	-	B	B	A	A	B	B	A	A	-	-	A	A	B	A	B	B	B
TB6/10	B	B	B	B	A	A	B	A	A	A	-	-	B	A	B	A	B	B	B
TB6/11	B	B	B	B	A	A	B	A	A	B	-	-	B	A	B	A	B	B	B
TB6/12	-	-	B	B	A	A	B	A	A	B	-	B	B	A	B	A	B	B	B
TB6/13	B	B	B	B	A	A	B	A	A	B	-	B	B	A	B	A	B	B	B
TB6/6	B	B	B	B	A	A	B	A	A	B	-	-	B	A	B	A	B	B	B
TB6/14	B	B	B	B	A	A	B	A	A	B	-	B	B	A	B	A	B	B	B
TB6/15	B	B	B	B	A	B	B	A	A	B	-	B	B	A	B	A	B	B	B
TB6/16	B	-	B	B	A	B	B	A	A	B	-	B	B	A	B	A	B	B	B

Marker	Progeny																		
	F29/46 bscl 4	F19/31 clone 1	F19/31 bscl11	F28/46 bscl 6	F28/46 bscl 11	F29/46 bscl 2	F28/46 bscl 1	F28/46 bscl 4	F28/46 bscl 7	F28/46 bscl 8	F29/46 bscl 1	F9/41 bscl 1	F9/41 bscl 2	F9/41 bscl 8	F9/41 bscl11	F19/31 bscl 5	F19/31 bscl 10	F19/31 bscl 8	F492/50 bscl 7
	387	389	391	394	405	406	407	409	410	411	414	415	416	420	422	425	430	428	180G12
TB6/9	B	B	B	B	B	A	B	A	A	B	B	B	B	B	B	A	A	A	A
TB6/10	B	B	A	B	-	A	B	A	A	B	B	B	-	A	B	A	A	A	A
TB6/11	B	B	A	B	-	B	A	A	A	B	B	B	-	A	B	A	A	A	B
TB6/12	B	B	A	B	B	B	A	A	A	B	B	B	B	B	B	A	A	A	B
TB6/13	A	A	A	B	B	B	A	A	A	B	B	B	B	B	B	A	A	A	B
TB6/6	A	A	A	B	B	B	A	A	B	B	B	-	B	B	B	A	A	A	B
TB6/14	A	A	A	B	B	B	A	A	B	B	B	B	B	B	B	A	A	A	B
TB6/15	A	A	A	B	B	B	A	mutant	B	B	B	B	B	B	B	A	A	A	B
TB6/16	A	A	A	B	B	B	A	A	B	B	B	B	B	B	B	A	A	A	B

Appendix 4f The segregation data for all of the microsatellite markers on chromosome 6 in the STIB 386 genetic map. See Appendix 4a legend for further details.

Marker	Progeny																		
	F9/45 mcl 2	F9/45 mcl 10	F9/45 mcl 11	F9/45 mcl 12	F9/34 mcl 1	B80 cl 2	F492/50 bscl 1	F492/50 bscl 6	F492/50 bscl 8	F492/50 bscl 9	F492/50 bscl 12	F492/50 bscl 14	F492/50 bscl 21	F492/50 bscl 23	F492/50 bscl 5/1 b	F9/41 bscl 5	F9/41 bscl 7	F9/41 bscl 9	F29/46 bscl 3
	30	34	35	36	37	162	174	179	181	182	185	187	194	196	332	383	384	385	386
TB7/1	A	B	B	A	A	A	B	A	A	A	-	-	A	B	B	A	A	A	A
TB7/14	A	B	B	A	A	A	B	A	A	A	-	-	A	B	B	A	A	A	A
TB7/4	A	B	B	A	A	A	B	A	A	A	-	-	A	B	B	A	A	A	A
TB7/5	B	B	B	A	A	A	A	A	A	A	-	-	A	B	B	A	A	B	B
TB7/15	-	B	A	A	A	A	A	A	A	A	-	B	A	B	B	B	A	B	B
TB7/17	B	B	A	-	A	A	A	A	A	B	-	-	A	B	B	B	A	B	B
TB7/16	B	B	A	A	A	A	A	A	A	B	-	-	A	B	B	B	A	B	B

Marker	Progeny																		
	F29/46 bscl 4	F19/31 clone 1	F19/31 bscl11	F28/46 bscl 6	F28/46 bscl 11	F29/46 bscl 2	F28/46 bscl 1	F28/46 bscl 4	F28/46 bscl 7	F28/46 bscl 8	F29/46 bscl 1	F9/41 bscl 1	F9/41 bscl 2	F9/41 bscl 8	F9/41 bscl11	F19/31 bscl 5	F19/31 bscl 10	F19/31 bscl 8	F492/50 bscl 7
	387	389	391	394	405	406	407	409	410	411	414	415	416	420	422	425	430	428	180G12
TB7/1	B	A	B	A	B	B	B	A	A	A	B	-	A	B	B	A	A	A	A
TB7/14	B	A	B	A	B	B	B	A	A	A	B	-	A	B	B	A	A	A	A
TB7/4	A	A	B	A	B	B	B	A	A	A	B	-	A	B	B	A	A	A	A
TB7/5	A	A	B	A	B	B	B	A	A	A	B	B	A	B	B	A	A	A	A
TB7/15	A	A	B	A	B	B	B	A	B	A	B	B	A	B	A	A	A	A	A
TB7/17	A	A	B	B	B	B	B	A	B	B	B	-	A	B	A	A	B	A	B
TB7/16	A	A	B	B	B	B	B	A	B	B	B	-	A	B	A	A	B	A	B

Appendix 4g The segregation data for all of the microsatellite markers on chromosome 7 in the STIB 386 genetic map. See Appendix 4a legend for further details.

Marker	Progeny																		
	F9/45 mcl 2	F9/45 mcl 10	F9/45 mcl 11	F9/45 mcl 12	F9/34 mcl 1	B80 cl 2	F492/50 bscl 1	F492/50 bscl 6	F492/50 bscl 8	F492/50 bscl 9	F492/50 bscl 12	F492/50 bscl 14	F492/50 bscl 21	F492/50 bscl 23	F492/50 bscl 5/1 b	F9/41 bscl 5	F9/41 bscl 7	F9/41 bscl 9	F29/46 bscl 3
	30	34	35	36	37	162	174	179	181	182	185	187	194	196	332	383	384	385	386
TB8/12	B	B	A	A	B	B	A	A	B	B	-	A	B	A	B	A	B	B	B
TB8/13	A	-	A	A	-	B	A	A	-	B	-	A	B	A	B	B	A	B	B
TB8/14	A	B	A	A	B	B	A	A	B	B	B	B	B	A	B	B	A	B	B
TB8/15	A	B	A	A	B	B	A	A	-	B	-	B	B	A	B	B	A	B	B
TB8/16	A	B	A	A	B	B	A	A	B	B	-	B	A	A	B	B	A	B	B
TB8/17	A	B	A	A	B	-	A	-	B	B	-	B	A	A	B	B	A	-	B
TB8/18	A	B	B	A	B	B	A	A	B	A	-	B	A	A	B	B	A	B	B
TB8/10	A	B	B	A	B	B	B	A	B	A	A	-	A	A	B	B	B	A	A
TB8/19	A	A	A	A	B	B	B	B	B	A	A	B	A	A	B	B	B	A	A
TB8/20	A	A	A	A	A	B	B	B	B	-	A	-	A	A	B	B	A	A	A
TB8/21	A	A	B	A	A	B	B	A	B	B	A	-	B	A	B	B	A	A	A

Marker	Progeny																		
	F29/46 bscl 4	F19/31 clone 1	F19/31 bscl 11	F28/46 bscl 6	F28/46 bscl 11	F29/46 bscl 2	F28/46 bscl 1	F28/46 bscl 4	F28/46 bscl 7	F28/46 bscl 8	F29/46 bscl 1	F9/41 bscl 1	F9/41 bscl 2	F9/41 bscl 8	F9/41 bscl 1	F19/31 bscl 5	F19/31 bscl 10	F19/31 bscl 8	F492/50 bscl 7
	387	389	391	394	405	406	407	409	410	411	414	415	416	420	422	425	430	428	180G12
TB8/12	A	B	B	B	A	A	A	B	B	A	A	A	B	A	A	A	A	A	B
TB8/13	A	B	B	B	A	A	B	B	A	A	A	A	B	A	A	A	A	A	B
TB8/14	A	B	B	B	A	A	B	A	A	A	A	A	-	A	A	A	A	A	B
TB8/15	B	B	B	B	A	A	B	A	A	A	A	A	B	A	A	A	-	A	B
TB8/16	B	B	B	B	B	A	B	A	A	A	A	A	B	A	A	A	A	A	B
TB8/17	-	B	-	B	B	A	B	A	A	A	A	A	B	A	A	A	A	A	B
TB8/18	B	B	B	B	B	A	B	A	A	A	A	A	B	A	B	A	A	A	A
TB8/10	B	B	B	B	B	B	B	A	A	A	B	B	B	B	B	A	A	A	A
TB8/19	A	B	B	B	A	B	A	A	A	A	B	B	B	B	B	A	-	A	A
TB8/20	A	B	B	B	A	B	A	A	A	A	B	B	B	B	B	A	A	A	B
TB8/21	A	B	B	B	A	B	A	A	A	A	B	B	B	B	B	A	A	A	B

Appendix 4h The segregation data for all of the microsatellite markers on chromosome 8 in the STIB 386 genetic map. See Appendix 4a legend for further details.

Marker	Progeny																		
	F9/45 mcl 2	F9/45 mcl 10	F9/45 mcl 11	F9/45 mcl 12	F9/34 mcl 1	B80 cl 2	F492/50 bscl 1	F492/50 bscl 6	F492/50 bscl 8	F492/50 bscl 9	F492/50 bscl 12	F492/50 bscl 14	F492/50 bscl 21	F492/50 bscl 23	F492/50 bscl 5/1 b	F9/41 bscl 5	F9/41 bscl 7	F9/41 bscl 9	F29/46 bscl 3
	30	34	35	36	37	162	174	179	181	182	185	187	194	196	332	383	384	385	386
TB9/17	A	B	B	A	A	A	A	B	B	B	-	A	B	B	A	A	A	B	B
TB9/18	A	B	B	A	A	B	A	B	B	B	-	-	B	B	A	A	A	B	B
TB9/19	A	B	-	-	A	B	A	B	B	B	-	-	B	B	A	A	A	B	B
TB9/20	A	B	B	A	A	B	A	B	B	B	B	B		B	A	A	A	B	B
TB9/21	A	A	B	B	A	B	A	A	A	B	-	-	B	B	A	A	B	B	B
TB9/5	B	A	B	B	A	B	A	A	A	B	-	-	B	B	A	A	B	B	B
TB9/9	B	B	A	B	A	B	A	A	A	B	-	-	B	B	A	A	B	B	B
TB9/12	B	B	A	B	A	B	B	A	A	B	-	-	A	B	A	A	B	B	B
TB9/14	A	B	A	B	A	B	B	A	A	B	-	-	A	B	A	A	B	B	B
TB9/22	A	B	A	B	A	B	B	A	A	B	-	-	A	B	A	A	B	B	B

Marker	Progeny																		
	F29/46 bscl 4	F19/31clone 1	F19/31 bscl 11	F28/46 bscl 6	F28/46 bscl 11	F29/46 bscl 2	F28/46 bscl 1	F28/46 bscl 4	F28/46 bscl 7	F28/46 bscl 8	F29/46 bscl 1	F9/41 bscl 1	F9/41 bscl 2	F9/41 bscl 8	F9/41 bscl11	F19/31 bscl 5	F19/31 bscl 10	F19/31 bscl 8	F492/50 bscl 7
	387	389	391	394	405	406	407	409	410	411	414	415	416	420	422	425	430	428	180g12
TB9/17	B	B	B	A	B	A	A	A	B	B	A	A	A	A	B	A	B	B	B
TB9/18	B	B	B	A	B	A	A	A	B	B	A	A	A	A	B	A	B	A	B
TB9/19	B	B	B	A	B	A	A	A	B	B	A	A	A	A	B	A	B	A	B
TB9/20	B	B	B	A	B	A	A	A	B	B	A	B	A	A	B	A	B	A	B
TB9/21	A	B	B	A	B	A	A	B	B	B	A	B	A	A	B	A	B	A	B
TB9/5	A	B	B	A	B	A	A	B	B	B	A	B	A	A	B	A	B	A	B
TB9/9	B	B	B	A	B	A	A	B	B	B	A	B	A	A	A	A	B	A	B
TB9/12	B	B	B	A	A	A	A	B	B	B	A	B	A	A	A	A	B	A	B
TB9/14	B	B	B	A	A	A	B	B	B	B	A	B	A	A	A	A	B	A	B
TB9/22	B	A	B	A	A	A	B	B	B	B	A	A	A	A	A	A	A	A	B

Appendix 4i The segregation data for all of the microsatellite markers on chromosome 9 in the STIB 386 genetic map. See Appendix 4a legend for further details.

Marker	Progeny																		
	F9/45 mcl 2	F9/45 mcl 10	F9/45 mcl 11	F9/45 mcl 12	F9/34 mcl 1	B80 cl 2	F492/50 bscl 1	F492/50 bscl 6	F492/50 bscl 8	F492/50 bscl 9	F492/50 bscl 12	F492/50 bscl 14	F492/50 bscl 21	F492/50 bscl 23	F492/50 bscl 5/1b	F9/41 bscl 5	F9/41 bscl 7	F9/41 bscl 9	F29/46 bscl 3
	30	34	35	36	37	162	174	179	181	182	185	187	194	196	332	383	384	385	386
TB10/22	-	A	B	B	B	B	A	B	A	B	B	-	A	B	B	B	B	A	A
TB10/23	A	A	B	-	B	B	A	B	A	B	-	-	A	B	B	B	B	A	A
TB10/24	B	A	B	B	A	B	A	B	A	A	-	-	A	B	B	B	B	A	A
TB10/25	B	A	-	-	A	B	A	B	A	A	-	-	A	B	B	B	B	A	A
TB10/12	B	A	B	A	A	B	A	B	A	A	-	-	A	B	B	B	B	A	A
TB10/14	B	A	B	A	A	B	A	B	A	B	A	A	A	B	B	B	B	A	A
TB10/26	B	A	B	A	A	B	A	B	A	B	-	-	A	B	B	B	B	A	A
TB10/27	B	B	A	A	A	B	B	B	A	B	-	-	B	B	B	B	B	A	A
TB10/28	B	B	A	A	A	B	B	B	A	B	-	A	B	B	B	B	B	B	B
TB10/29	B	B	A	-	A	B	B	B	A	B	-	-	B	B	B	A	B	B	B
TB10/19	B	B	A	A	A	B	B	B	A	B	-	-	B	B	B	A	B	B	B
TB10/30	B	B	A	-	A	A	B	B	-	-	-	-	B	B	B	A	B	B	B

Marker	Progeny																		
	F29/46 bscl 4	F19/31 clone 1	F19/31 bscl 11	F28/46 bscl 6	F28/46 bscl 11	F29/46 bscl 2	F28/46 bscl 1	F28/46 bscl 4	F28/46 bscl 7	F28/46 bscl 8	F29/46 bscl 1	F9/41 bscl 1	F9/41 bscl 2	F9/41 bscl 8	F9/41 bscl 11	F19/31 bscl 5	F19/31 bscl 10	F19/31 bscl 8	F492/50 bscl 7
	387	389	391	394	405	406	407	409	410	411	414	415	416	420	422	425	430	428	180
TB10/22	A	A	A	A	A	A	B	A	B	A	A	A	A	A	B	B	B	B	B
TB10/23	A	B	A	A	A	A	B	A	B	A	A	A	A	A	B	B	-	B	B
TB10/24	A	B	A	A	B	B	A	B	B	B	A	A	A	A	B	A	B	A	A
TB10/25	A	B	A	B	B	B	A	B	B	B	A	A	A	A	B	A	B	A	A
TB10/12	A	B	A	B	B	B	A	B	B	B	A	A	A	A	B	A	B	A	A
TB10/14	A	B	A	B	B	A	A	B	A	B	A	A	A	A	B	A	B	B	B
TB10/26	A	B	A	B	B	A	A	B	A	B	B	A	A	B	B	A	B	B	B
TB10/27	A	B	A	B	B	A	A	B	A	B	B	A	A	B	A	A	B	B	B
TB10/28	A	B	A	B	B	A	A	B	A	B	B	A	A	B	A	A	B	A	B
TB10/29	A	B	A	B	B	A	A	B	A	B	B	A	A	B	A	A	B	A	B
TB10/19	B	B	A	B	B	A	B	A	B	B	B	A	A	B	A	A	B	A	B
TB10/30	B	B	A	B	B	A	B	A	B	B	B	A	A	B	A	A	B	A	A

Appendix 4j The segregation data for all of the microsatellite markers on chromosome 10 in the STIB 386 genetic map. See Appendix 4a legend for further details.

Marker	Progeny																		
	F9/45 mcl 2	F9/45 mcl 10	F9/45 mcl 11	F9/45 mcl 12	F9/34 mcl 1	B80 cl 2	F492/50 bscl 1	F492/50 bscl 6	F492/50 bscl 8	F492/50 bscl 9	F492/50 bscl 12	F492/50 bscl 14	F492/50 bscl 21	F492/50 bscl 23	F492/50 bscl 5/1b	F9/41 bscl 5	F9/41 bscl 7	F9/41 bscl 9	F29/46 bscl 3
	30	34	35	36	37	162	174	179	181	182	185	187	194	196	332	383	384	385	386
TB11/32	A	A	-	B	B	A	A	A	B	A	A	A	A	A	B	A	B	B	B
TB11/7	A	A	B	B	B	A	A	A	B	A	-	-	A	A	B	A	B	B	B
TB11/33	A	A	B	B	B	A	A	A	B	A	-	-	A	A	B	A	B	B	B
TB11/10	A	A	-	-	B	A	A	A	B	A	-	-	B	A	B	A	B	B	B
TB11/11	A	A	B	-	B	A	A	A	B	A	-	-	B	A	B	A	B	B	B
TB11/13	-	-	-	-	B	A	A	B	B	A	-	-	B	A	B	A	B	B	B
TB11/34	A	A	-	-	B	A	A	B	B	A	-	-	B	A	B	A	B	B	B
TB11/35	A	A	-	-	B	A	-	B	A	A	-	-	B	A	A	A	B	B	B
TB11/36	A	A	B	B	B	A	A	B	A	A	A	A	B	A	A	A	B	B	B
TB11/15	A	A	B	-	-	A	A	B	A	A	-	-	B	-	A	A	B	B	B
TB11/37	B	A	B	B	B	A	A	B	A	A	A	A	B	A	A	A	B	B	B
TB11/38	B	B	A	B	B	A	A	B	A	A	-	A	B	A	A	B	B	B	B
TB11/39	B	B	A	B	B	A	A	B	A	-	A	-	B	A	A	B	B	B	B
TB11/40	B	B	A	B	B	A	A	B	A	A	A	B	B	-	A	B	B	B	B
TB11/41	A	B	-	-	B	A	A	B	A	A	-	-	B	A	A	B	B	B	B
TB11/42	A	B	A	-	B	B	B	B	A	B	-	-	B	A	A	B	B	B	B
TB11/21	A	B	-	-	B	B	B	B	A	B	-	-	B	A	A	B	B	B	B
TB11/43	A	B	-	-	B	B	B	B	A	B	-	-	-	A	A	B	B	B	B
TB11/44	A	B	-	B	B	B	B	B	A	B	-	-	B	A	A	B	B	B	B
TB11/45	A	B	A	-	B	B	B	B	A	B	-	-	B	A	A	B	B	B	B
TB11/23	A	B	A	-	B	B	B	B	A	B	-	-	B	A	A	B	A	B	B

Marker	Progeny																		
	F29/46 bscl 4	F19/31 clone1	F19/31 bscl 11	F28/46 bscl 6	F28/46 bscl 11	F29/46 bscl 2	F28/46 bscl 1	F28/46 bscl 4	F28/46 bscl 7	F28/46 bscl 8	F29/46 bscl 1	F9/41 bscl 1	F9/41 bscl 2	F9/41 bscl 8	F9/41 bscl11	F19/31 bscl 5	F19/31 bscl 10	F19/31 bscl 8	F492/50 bscl 7
	387	389	391	394	405	406	407	409	410	411	414	415	416	420	422	425	430	428	180G12
TB11/32	B	B	B	A	A	B	A	A	A	B	B	B	A	B	B	B	B	B	A
TB11/7	B	B	A	A	-	B	A	A	A	-	B	-	A	B	B	B	B	B	A
TB11/33	B	B	A	A	-	-	A	A	A	B	B	-	A	B	B	B	B	B	A
TB11/10	B	B	A	A	A	B	A	B	A	B	B	B	A	B	B	B	B	B	A
TB11/11	B	B	A	A	A	B	A	B	B	B	B	B	A	B	B	B	B	B	A
TB11/13	B	B	-	A	A	B	A	B	B	B	B	B	A	B	B	B	B	B	-
TB11/34	B	B	A	A	A	B	A	B	B	B	B	B	A	B	B	B	B	B	A
TB11/35	B	B	A	A	A	B	A	B	B	B	B	B	A	B	B	B	B	B	A
TB11/36	B	B	A	A	A	B	A	B	B	B	B	B	A	B	B	B	B	B	A
TB11/15	B	B	A	A	A	B	A	B	B	B	B	B	A	B	B	B	B	B	A
TB11/37	B	B	A	A	A	B	A	B	B	B	B	B	A	B	B	B	B	B	A
TB11/38	B	-	A	A	A	-	A	B	B	B	B	B	A	-	A	B	B	B	A
TB11/39	B	B	A	A	A	B	A	B	B	B	B	B	A	B	A	B	B	B	A
TB11/40	B	B	A	A	A	B	A	B	B	B	B	B	A	-	A	B	B	B	A
TB11/41	B	B	A	A	A	B	-	B	B	B	B	B	A	B	A	B	B	B	B
TB11/42	A	A	A	A	A	B	A	A	B	B	B	B	A	B	A	B	B	B	B
TB11/21	A	A	A	B	A	B	A	A	B	B	B	B	A	B	A	A	B	A	B
TB11/43	A	A	A	B	A	B	A	A	B	B	B	B	A	B	A	A	B	-	B
TB11/44	A	A	A	B	A	B	B	A	B	B	B	B	A	B	A	A	A	A	B
TB11/45	A	A	A	B	A	B	B	A	B	B	B	B	A	B	A	A	A	A	B
TB11/23	A	A	A	B	A	B	B	A	B	B	B	B	A	B	A	A	A	A	B

Appendix 4k The segregation data for all of the microsatellite markers on chromosome 11 in the STIB 386 genetic map. See Appendix 4a legend for further details.

CHROMOSOME	PROGENY																			
	F9/45 mcl 2	F9/45 mcl 10	F9/45 mcl 11	F9/45 mcl 12	F9/34 mcl 1	B80 cl 2	F492/50 bscl 1	F492/50 bscl 6	F492/50 bscl 8	F492/50 bscl 9	F492/50 bscl 12	F492/50 bscl 14	F492/50 bscl 18	F492/50 bscl 21	F492/50 bscl 23	F492/50 bscl 5/1 b	F9/41 bscl 5	F9/41 bscl 7	F9/41 bscl 9	F29/46 bscl 3
	30	34	35	36	37	162	174	179	181	182	185	187	194	196	332	383	384	385	386	387
1	0	0	1	1	n/a	0	1	0	0	0	1	1	1	0	0	1	0	0	0	0
2	0	1	1	0	n/a	0	1	1	0	0	1	0	1	0	0	1	0	0	0	1
3	2	0	0	1	0	0	1	0	0	0	2	0	1	0	0	0	1	1	1	0
4	2	0	0	1	0	0	1	0	0	0	n/a	n/a	0	0	0	0	0	1	1	1
5	0	1	2	2	0	1	0	0	2	1	1	0	1	0	0	1	1	1	1	0
6	0	0	0	0	0	1	0	1	0	1	n/a	0	1	0	0	0	0	0	0	1
7	1	0	1	0	0	0	1	0	0	1	n/a	n/a	0	0	0	1	0	1	1	1
8	1	1	3	0	1	0	1	2	0	2	1	1	2	0	0	1	3	1	1	2
9	2	2	1	1	0	1	1	1	1	0	n/a	1	1	0	0	0	1	0	0	2
10	1	1	1	1	1	1	1	0	0	2	n/a	n/a	1	0	0	1	0	1	1	1
11	2	1	1	0	0	1	1	1	1	1	0	1	1	0	1	1	1	0	0	1
Total no. crossovers	11	7	11	7	2	5	9	6	4	8	6	4	10	0	1	7	7	6	6	10
Average no. crossovers	1.0	0.6	1.0	0.6	0.2	0.5	0.8	0.5	0.4	0.7	1.0	0.5	0.9	0.0	0.1	0.6	0.6	0.5	0.5	0.9

Appendix 5 Recombination events in every linkage group of every individual for progeny clones 1-19

CHROMOSOME	PROGENY																		
	F29/46 bscl 4	F19/31clone 1	F 9/31 bscl11	F28/46 bscl 6	F28/46 bscl 11	F29/46 bscl 2	F28/46 bscl 1	F28/46 bscl 4	F28/46 bscl 7	F28/46 bscl 8	F29/46 bscl 1	F9/41 bscl 1	F9/41 bscl 2	F9/41 bscl 8	F9/41 bscl 11	F19/31 bscl 5	F19/31 bscl 10	F19/31 bscl 8	crossovers/ chromosome
	389	391	394	405	406	407	409	410	411	414	415	416	420	422	425	430	428	180G12	
1	1	0	2	1	0	1	0	1	0	1	0	1	1	1	0	0	0	0	0.46
2	0	0	0	0	1	1	1	0	0	0	0	0	0	1	1	1	1	n/a	0.42
3	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0.42
4	0	1	1	2	1	1	0	0	0	0	1	0	0	0	1	1	2	0	0.50
5	1	1	1	1	1	1	0	1	0	0	2	0	0	0	1	2	1	0	0.74
6	1	1	0	0	1	1	0	1	0	0	0	0	2	0	0	0	0	1	0.35
7	0	0	1	0	0	0	0	1	1	0	n/a	0	0	1	0	1	0	1	0.40
8	0	0	0	2	1	2	1	1	0	1	1	0	1	1	0	0	0	2	0.95
9	1	0	0	1	0	1	1	0	0	0	2	0	0	1	0	1	1	0	0.65
10	1	0	1	1	2	2	2	2	1	1	0	0	1	1	1	0	3	3	1.00
11	1	1	1	0	0	1	2	1	0	0	0	0	0	1	1	1	1	1	0.71
Total no. crossovers	7	4	8	8	8	11	7	8	2	3	6	1	5	7	6	8	10	8	
Average no. crossovers	0.6	0.4	0.7	0.7	0.7	1.0	0.6	0.7	0.2	0.3	0.6	0.1	0.5	0.6	0.5	0.7	0.9	0.8	

Appendix 5 Recombination events in every linkage group of every individual for progeny clones 20-38

A breakdown of the number of recombination events in every linkage group of every individual progeny clone is given and the total and average for each individual and linkage group calculated. The name of each progeny clone is given, along with laboratory alias along the top of each table. One progeny clone for which no crossover events were detected on any chromosome is highlighted in grey.

Fly Stock	Clone identity	Culture stabilate	Blood straw stabilate	Markers							Genotype
				JS2 (TB5/4)	PLC (TB2/20)	IJ15/1 (TB3/1)	3778 (TB10/28)	292 (TB3/13)	CRAM (TB10/14)	MS42 (TB1/1)	
STIB 247 parental stock				5-6	5-5	2-3	3-4	5-5	1-1	5-5	
STIB 386 parental stock				1-2	1-2	1-2	1-2	1-2	1-2	1-2	
F19/31	1 cell/well 3 E7	ND	ND	1-6	2-5	2-2	2-3	-	-	-	same as existing F19/31 bscl 5
F19/31	1 cell/well 3 C7	ND	ND	1-6	2-5	2-2	2-3	-	-	-	same as existing F19/31 bscl 5
F19/31	1 cell/well 3 E4	ND	ND	1-6	2-5	2-2	2-3	-	-	-	same as existing F19/31 bscl 5
F19/31	1 cell/well 3 C9	ND	ND	1-6	2-5	2-2	2-3	-	-	-	same as existing F19/31 bscl 5
F19/31	1 cell/well 1 C8	ND	ND	1-6	2-5	2-2	2-3	-	-	-	same as existing F19/31 bscl 5
F19/31	1 cell/well 1 C10	ND	ND	1-6	2-5	-	2-3	-	-	-	same as existing F19/31 bscl 5
F19/31	1 cell/well 1 E6	ND	ND	1-6	-	-	-	-	-	-	same as existing F19/31 bscl 5?
F19/31	1 cell/well 1 C7	ND	ND	1-6	-	-	-	-	-	-	same as existing F19/31 bscl 5?
F19/31	1 cell/well 3 C1	ND	ND	1-6	-	-	-	-	-	-	same as existing F19/31 bscl 5?
F28/46	0.25 cell/well A C5	TGG0152	TGG0162	2-6	1-5	2-2	1-3	1-5	1-2	1-5	New Progeny clone AC NEW 1
F28/46	1 cell/well A C9	TGG0149	TGG0167	2-6	1-5	2-2	1-3	1-5	1-2	1-5	same as AC NEW 1
F28/46	1 cell/well B D11	ND	TGG0172	2-6	1-5	2-2	1-3	1-5	-	-	same as AC NEW 1
F28/46	0.25 cell/well A G2	TGG0153	TGG0174	2-6	1-5	2-2	1-3	1-5	-	-	same as AC NEW 1
F28/46	0.5 cell/well B D4	TGG0142	TGG0175	2-6	1-5	2-2	1-3	1-5	-	-	same as AC NEW 1
F28/46	0.5 cell/well B F9	TGG0134	ND	2-6	1-5	2-2	1-3	-	-	-	same as AC NEW 1
F28/46	0.5 cell/well B G5	TGG0135	ND	2-6	1-5	2-2	1-3	-	-	-	same as AC NEW 1
F28/46	0.5 cell/well A E10	TGG0189	ND	2-6	1-5	2-2	1-3	-	-	-	same as AC NEW 1
F28/46	0.5 cell/well A F7	TGG0140	ND	2-6	?	2-2	1-3	-	-	-	same as AC NEW 1
F28/46	0.5 cell/well A G4	TGG0155	ND	2-6	1-5	2-2	1-3	-	-	-	same as AC NEW 1
F28/46	1 cell/well B B4	TGG0147	ND	2-6	1-5	2-2	1-3	-	-	-	same as AC NEW 1
F28/46	1 cell/well B D1	TGG0136	ND	2-6	1-5	2-2	1-3	-	-	-	same as AC NEW 1
F28/46	1 cell/well B F2	TGG0150	ND	2-6	1-5	2-2	1-3	-	-	-	same as AC NEW 1
F28/46	1 cell/well B F4	TGG0143	ND	2-6	1-5	2-2	1-3	-	-	-	same as AC NEW 1
F28/46	1 cell/well B F5	TGG0141	ND	2-6	1-5	2-2	1-3	-	-	-	same as AC NEW 1
F28/46	1 cell/well B F9	TGG0146	ND	2-6	1-5	2-2	1-3	-	-	-	same as AC NEW 1
F28/46	1 cell/well B G4	TGG0145	ND	2-6	1-5	2-2	1-3	-	-	-	same as AC NEW 1
F28/46	1 cell/well B G5	TGG0144	ND	2-6	1-5	2-2	1-3	-	-	-	same as AC NEW 1

F28/46	1 cell/well A C2	TGG0190	ND	2-6	1-5	-	1-3	-	-	-	same as AC NEW 1
F28/46	1 cell/well A D7	TGG0160	ND	2-6	1-5	2-2	-	-	-	-	same as AC NEW 1
F28/46	1 cell/well A F4	TGG0148	ND	2-6	1-5	2-2	1-3	-	-	-	same as AC NEW 1
F28/46	1 cell/well A G1	TGG0138	ND	2-6	1-5	2-2	1-3	-	-	-	same as AC NEW 1
F28/46	0.5 cell/well A E8	TGG0157	TGG0164	1-5	2-5	2-3	1-4	2-5	1-2	1-5	New Progeny clone AC NEW 2
F28/46	0.25 cell/well A C3	TGG0151	TGG0165	1-5	1-5	1-3	2-4	2-5	1-1	1-5	New Progeny clone AC NEW 3
F28/46	0.25 cell/well B D2	TGG0139	TGG0168	1-5	1-5	1-3	2-4	2-5	-	1-5	same as AC NEW 3
F28/46	0.5 cell/well B D12	TGG0137	TGG0181	1-5	1-5	1-3	2-4	2-5	-	-	same as AC NEW 3
F28/46	1 cell/well B E8	TGG0191	ND	1-5	1-5	?-3	2-4	-	-	-	same as AC NEW 3
F28/46	1 cell/well A B8	TGG0161	TGG0169	2-5	1-5	1-3	1-3	1-5	1-2	1-5	New Progeny clone AC NEW 4
F28/46	0.5 cell/well B G6	TGG0192	TGG0170	1-6	2-5	2-2	2-4	1-5	1-1	1-5	New Progeny clone AC NEW 5
F28/46	1 cell/well B C8	TGG0132	TGG0171	1-6	2-5	1-2	1-4	1-5	1-2	-	New Progeny clone AC NEW 6
F28/46	1 cell/well A G7	TGG0158	TGG0177	1-5	1-5	2-2	1-3	2-5	1-2	-	New Progeny clone AC NEW 7
F28/46	0.1 cell/well B F3	TGG0154	TGG0344	2-6	2-5	1-3	1-3	-	-	-	same as existing F28/46 bscl 4
F28/46	0.25 cell/well A D10	TGG0159	TGG0163	2-5	2-5	2-2	1-4	1-5	1-2	1-5	same as existing F28/46 bscl 6
F28/46	0.5 cell/well B F8	ND	TGG0193	2-5	1-5	2-2	2-4	-	-	-	same as existing F28/46 bscl 7
F28/46	0.5 cell/well B D10	TGG0133	ND	2-5-6 mix	1-5	-	-	-	-	-	mixed or polyploid
F28/46	0.5 cell/well A C1	TGG0156	ND	2-6	1-5	2-3?	1-2-3-4	-	-	-	mixed or polyploid
F28/46	0.5 cell/well A F8	ND	ND	1-2-6	1-5	-	-	-	-	-	mixed or polyploid
F28/46	0.1 cell/well A H3	ND	ND	2-6	1-5	-	-	-	-	-	Culture failed
F28/46	1 cell/well B E5	ND	ND	-	1-5	-	-	-	-	-	Culture failed
F28/46	1 cell/well B E7	ND	ND	2-6	-	-	-	-	-	-	Culture failed
F28/46	1 cell/well B F12	ND	ND	2-6	-	-	-	-	-	-	Culture failed
F28/46	1 cell/well B G7	ND	ND	2-6	-	-	-	-	-	-	Culture failed
F29/42	AML DA6	ND	TGG0198	2-6	1-5	2-3	1-3	1-5	-	-	New Progeny clone AC NEW 8
F29/42	0.5 cell/well 2 C8	TGG320	ND	2-6	-	2-3	1-3	-	-	-	same as AC NEW 8
F29/42	0.5 cell/well 5 E12	TGG322	ND	2-6	-	2-3	1-3	-	-	-	same as AC NEW 8
F29/42	0.5 cell/well 4 E12	TGG324	ND	2-6	-	2-3	1-3	-	-	-	same as AC NEW 8
F29/42	0.5 cell/well 4 F7	TGG325	ND	2-6	-	2-3	1-3	-	-	1-5	same as AC NEW 8
F29/42	0.5 cell/well 4 F7	TGG326	ND	-	-	2-3	1-3	-	-	-	same as AC NEW 8
F29/42	0.5 cell/well 2 G3	TGG327	ND	2-6	-	2-3	1-3	-	-	-	same as AC NEW 8
F29/42	0.1 cell/well 2 H7	TGG328	ND	2-6	-	2-3	1-3	-	-	-	same as AC NEW 8
F29/42	0.5 cell/well 1 C7	TGG330	ND	2-6	-	2-3	1-3	-	-	-	same as AC NEW 8
F29/42	0.5 cell/well 1 E6	TGG331	ND	2-6	-	2-3	1-3	-	-	-	same as AC NEW 8
F29/42	0.5 cell/well 1 D12	TGG332	ND	2-6	-	2-3	1-3	-	-	-	same as AC NEW 8

F29/42	0.5 cell/well 4 E7	TGG333	ND	2-6	-	2-3	1-3	-	-	-	same as AC NEW 8
F29/42	AML AD8	ND	TGG0166	1-5?	1-5	2-3	1-2-3	2-5	-	-	same as existing F29/46 bscl 1
F29/42	AML BH4	ND	TGG0197	1-5	1-5	2-3	1-3	2-5	-	-	same as existing F29/46 bscl 1
F29/42	AML AA3	ND	TGG0176	1-5	1-5	2-3	1-3	2-5	-	-	same as existing F29/46 bscl 1
F29/42	AML CG10	ND	TGG0203	1-5	1-5	2-3	1-3	2-5	-	-	same as existing F29/46 bscl 1
F29/42	AML AD10	ND	TGG0204	1-5	1-5	2-3	1-3	2-5	-	-	same as existing F29/46 bscl 1
F29/42	AML BF9	ND	TGG0206	1-5	1-5	2-3	1-3	2-5	-	-	same as existing F29/46 bscl 1
F29/42	0.1 4 H10	TGG0321	ND	1-5	-	2-3	1-3	-	-	-	same as existing F29/46 bscl 1
F29/42	0.1 3 E10	TGG0323	ND	1-5	-	2-3	1-3	-	-	-	same as existing F29/46 bscl 1
F29/42	0.5 4 D4	TGG0334	ND	1-5	-	-	1-3	-	-	-	same as existing F29/46 bscl 1
F29/42	AML BE4	ND	TGG0205	1-2	1-2	1-2	1-2	-	-	-	parental strain STIB 386
F29/42	AML AC11	ND	TGG0173	1-2	1-2	1-2	1-2	-	-	-	parental strain STIB 386
F492/50	1 cell/well 5 E3	TGG0386	ND	1-6		1-2	2-4	2-5	1-1	1-5	New Progeny clone AC NEW 9
F492/50	1 cell/well 2 B3	TGG0403	ND	1-6	2-5	1-2	2-4	-	-	-	same as AC NEW 9
F492/50	1 cell/well 1 C2	TGG0398	ND	1-6	2-5	1-2	2-4	2-5	1-1	1-5	same as AC NEW 9
F492/50	0.5 cell/well 5 F2	TGG0388	ND	2-6	2-5	2-2	2-4	1-5	1-1	2-5	New Progeny clone AC NEW 10
F492/50	1 cell/well 7 E9	TGG0376	ND	2-6	2-5	2-2	2-4	1-5	1-1	2-5	same as AC NEW 10
F492/50	1 cell/well 8 F7	TGG0400	ND	2-5	1-5	1-3	2-4	-	-	-	New Progeny clone AC NEW 11
F492/50	1 cell/well 2 E3	TGG0406	ND	1-6	2-5	2-2	1-3	2-5	1-2	-	New Progeny clone AC NEW 12
F492/50	0.5 cell/well 4 G3	TGG0363	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 2 E4	TGG0364	ND	2-5	1-5	2-3	2-3	-	-	2-5	same as existing F492/50 bscl 12
F492/50	0.1 cell/well 2 C10	TGG0367	ND	2-5	1-5	2-3	2-3	-	-	2-5	same as existing F492/50 bscl 12
F492/50	0.25 cell/well 6 G6	TGG0368	ND	2-5	1-5	2-3	2-3	-	-	2-5	same as existing F492/50 bscl 12
F492/50	0.1 cell/well 1 B6	TGG0369	ND	2-5	-	2-3	2-3	-	-	2-5	same as existing F492/50 bscl 12
F492/50	1 cell/well 7 G8	TGG0370	ND	2-5	-	2-3	2-3	-	-	2-5	same as existing F492/50 bscl 12
F492/50	1 cell/well 7 G7	TGG0372	ND	2-5	1-5	2-3	2-3	-	-	2-5	same as existing F492/50 bscl 12
F492/50	1 cell/well 1 E2	TGG0373	ND	2-5	-	2-3	2-3	-	-	2-5	same as existing F492/50 bscl 12
F492/50	1 cell/well 6 E7	TGG0378	ND	2-5	1-5	2-3	2-3	-	-	2-5	same as existing F492/50 bscl 12
F492/50	0.5 cell/well 3 G9	TGG0380	ND	2-5	1-5	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	0.5 cell/well 2 D2	TGG0384	ND	2-5	1-5	2-3	2-3	-	-	2-5	same as existing F492/50 bscl 12
F492/50	0.5 cell/well 8 C5	TGG0385	ND	2-5	-	2-3	2-3	-	-	2-5	same as existing F492/50 bscl 12
F492/50	0.5 cell/well 7 G8	TGG0389	ND	-	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	0.5 cell/well 6 F4	TGG0390	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	0.25 cell/well 6 H4	TGG0395	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 1 B3	ND	ND	2-5	-	-	2-3	-	-	-	same as existing F492/50 bscl 12

F492/50	1 cell/well 1 B4	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 1 C7	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 1 F5	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 1 G7	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 2 C6	ND	ND	2-5	1-5	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 2 G8	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 3 D2	ND	ND	2-5	1-5	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 3 F3	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 3 G5	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 3 H4	ND	ND	2-5	1-5	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 4 B6	ND	ND	2-5	-	-	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 4 C3	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 5 B7	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 5 D2	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 6 D12	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 6 E10	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 6 G2	ND	ND	2-5	-	2-3	-	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 6 G3	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 6 G5	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 7 B11	ND	ND	2-5	-	-	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 7 C2	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 7 C10	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 7 D9	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 7 G4	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 8 G1	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	0.5 cell/well 3 B4	ND	ND	2-5	1-5	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	0.5 cell/well 3 C2	ND	ND	2-5	-	-	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	0.5 cell/well 4 F6	ND	ND	2-5	-	-	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	0.5 cell/well 6 E2	ND	ND	2-5	-	-	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	0.5 cell/well 7 E10	ND	ND	2-5	1-5	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	0.25 cell/well 1E11	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	0.25 cell/well 3 F6	ND	ND	2-5	-	-	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	0.25 cell/well 4 E8	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1D cell/well 6 G3	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1D cell/well 7 D4	ND	ND	2-5	-	2-3	2-3	-	-	-	same as existing F492/50 bscl 12
F492/50	1 cell/well 1 E6	ND	ND	2-5	-	1-3	1-3	-	-	-	same as existing F492/50 bscl 9

F492/50	1 cell/well 3 E6	ND	ND	2-5	-	-	1-3	-	-	-	same as existing F492/50 bscl 9
F492/50	1 cell/well 7 F5	ND	ND	2-5	-	-	1-3	-	-	-	same as existing F492/50 bscl 9
F492/50	1 cell/well 1 C8	TGG0397	ND	2-5	-	1-3	1-3	-	-	-	same as existing F492/50 bscl 9
F492/50	1 cell/well 1 F4	TGG0387	ND	2-5	2-5	1-3	1-3	-	-	2-5	same as existing F492/50 bscl 9
F492/50	1 cell/well 3 C4	TGG0362	ND	2-6	-	-	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 2 C10	TGG0365	ND	2-6	2-5	1-2	1-4	-	-	1-5	same as existing F492/50 bscl 1
F492/50	1 cell/well 1 D6	TGG0371	ND	2-6	-	1-2	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 6 B7	TGG0374	ND	2-6	-	1-2	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 1 D10	ND	ND	2-6	2-5	1-2	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 2 B4	ND	ND	2-6	-	1-2	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 2 D9	ND	ND	2-6	-	1-2	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 2 F3	ND	ND	2-6	-	1-2	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 3 D6	ND	ND	2-6	-	1-2	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 4 D10	ND	ND	2-6	-	1-2	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 4 F4	ND	ND	2-6	-	1-2	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 5 D5	ND	ND	2-6	-	-	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 5 E7	ND	ND	2-6	-	1-2	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 5 F6	ND	ND	2-6	-	1-2	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 5 G10	ND	ND	2-6	-	1-2	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 6 C7	ND	ND	2-6	-	1-2	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	0.5 cell/well 6 B10	ND	ND	2-6	-	-	1-4	-	-	-	same as existing F492/50 bscl 1
F492/50	1 cell/well 8 H7	TGG0366	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 4 H6	TGG0375	ND	2-6	-	2-3	1-3	-	-	1-5	same as existing F492/50 bscl 23
F492/50	1 cell/well 6 D11	TGG0377	ND	2-6	2-5	2-3	1-3	-	-	1-5	same as existing F492/50 bscl 23
F492/50	0.5 cell/well 8 G10	TGG0381	ND	-	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 2 D10	TGG0383	ND	2-6	2-5	-	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	0.5 cell/well 2 D10	TGG0391	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	0.5 cell/well 4 H5	TGG0393	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	0.5 cell/well 5 C6	TGG0394	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	0.5 cell/well 5 D9	TGG0396	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 6 B5	TGG0399	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 2 B9	ND	ND	2-6	2-5	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 2 B10	ND	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 3 D8	ND	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 3 F7	ND	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23

F492/50	1 cell/well 3 F11	ND	ND	2-6	2-5	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 3 G12	ND	ND	2-6	2-5	-	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 4 C4	ND	ND	2-6	2-5	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 4 C5	ND	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 4 H9	ND	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 5 C7	ND	ND	2-6	-	-	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 5 F5	ND	ND	2-6	2-5	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 6 D3	ND	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 6 E8	ND	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 7 B6	ND	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 7 B9	ND	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 7 E5	ND	ND	2-6	-	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 8 C7	ND	ND	2-6	-	-	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	0.25 cell/well 4D10	ND	ND	2-6	2-5	2-3	1-3	-	-	-	same as existing F492/50 bscl 23
F492/50	1 cell/well 2 E10	TGG0379	ND	1-6	-	1-3	2-3	-	-	-	same as existing F492/50 bscl 14
F492/50	1 cell/well 8 C2	TGG0382	ND	1-6	1-5	1-3	2-3	-	-	-	same as existing F492/50 bscl 14
F492/50	0.5 cell/well 8 G4	TGG0392	ND	1-6	1-5	1-3	2-3	-	-	-	same as existing F492/50 bscl 14
F492/50	1 cell/well 4 B4	TGG401/402	ND	1-6	1-5	1-3	2-3	-	-	-	same as existing F492/50 bscl 14
F492/50	1 cell/well 3 G10	ND	ND	1-6	1-5	1-3	2-3	-	-	-	same as existing F492/50 bscl 14
F492/50	1 cell/well 4 E5	ND	ND	1-6	1-2-5	1-3	2-3	-	-	-	same as existing F492/50 bscl 14
F492/50	1 cell/well 5 E11	ND	ND	1-6	-	1-3	2-3	-	-	-	same as existing F492/50 bscl 14
F492/50	1 cell/well 5 F7	ND	ND	1-6	-	1-3	2-3	-	-	-	same as existing F492/50 bscl 14
F492/50	1 cell/well 6 D8	ND	ND	1-6	1-5	1-3	2-3	-	-	-	same as existing F492/50 bscl 14
F492/50	1 cell/well 6 G10	ND	ND	1-6	-	1-3	2-3	-	-	-	same as existing F492/50 bscl 14
F492/50	0.5 cell/well 4 C7	ND	ND	1-6	-	-	2-3	-	-	-	same as existing F492/50 bscl 14
F492/50	1 cell/well 1 D8	ND	ND	2-6	-	-	1-3-4	-	-	-	mixed or polyploid
F492/50	1 cell/well 2 F7	ND	ND	2-6	-	1-2-3	1-4	-	-	-	mixed or polyploid
F492/50	1 cell/well 3 D7	ND	ND	wk 2-6	-	1-2-3	1-3	-	-	-	mixed or polyploid
F492/50	1 cell/well 3 D9	ND	ND	2-5-6	-	-	1-2-3-4	-	-	-	mixed or polyploid
F492/50	1 cell/well 3 E3	ND	ND	1-2-6	-	-	3-2-4	-	-	-	mixed or polyploid
F492/50	1 cell/well 3 E7	ND	ND	2-5-6	-	-	2-3	-	-	-	mixed or polyploid
F492/50	1 cell/well 3 E9	ND	ND	2-5-6	-	-	1-2-3-4	-	-	-	mixed or polyploid
F492/50	1 cell/well 3 E11	ND	ND	2-5	-	1-2-3	2-3	-	-	-	mixed or polyploid
F492/50	1 cell/well 5 D10	ND	ND	2-5-6	-	1-2-3	2-3	-	-	-	mixed or polyploid
F492/50	1 cell/well 5 E6	ND	ND	1-6	1-5	1-2-3	2-3	-	-	-	mixed or polyploid
F492/50	1 cell/well 5 G5	ND	ND	2-6	-	-	1-3-4	-	-	-	mixed or polyploid

F492/50	1 cell/well 6 B6	ND	ND	1-6-wk5	-	1-3	2-3	-	-	-	mixed or polyploid
F492/50	1 cell/well 6 C9	ND	ND	2-5-6	-	-	2-3	-	-	-	mixed or polyploid
F492/50	1 cell/well 6 E5	ND	ND	2-5-6	2-5	-	1-2-3	-	-	-	mixed or polyploid
F492/50	1 cell/well 6 E6	ND	ND	2-5-6	-	-	1-4-2	-	-	-	mixed or polyploid
F492/50	1 cell/well 6 F6	ND	ND	2-6	-	-	1-3-4	-	-	-	mixed or polyploid
F492/50	1 cell/well 6 G8	ND	ND	1-6	-	1-3-2	2-3	-	-	-	mixed or polyploid
F492/50	1 cell/well 7 C3	ND	ND	1-6	-	-	2-3-4	-	-	-	mixed or polyploid
F492/50	1 cell/well 7 D8	ND	ND	2-5-6	-	-	1-3	-	-	-	mixed or polyploid
F492/50	1 cell/well 8 B11	ND	ND	2-5-6	-	-	1-2-3-4	-	-	-	mixed or polyploid
F492/50	1 cell/well 8 D1	ND	ND	2-6	-	-	1-2-3	-	-	-	mixed or polyploid
F492/50	1 cell/well 4 F8	ND	ND	wk 2-6	-	-	-	-	-	-	Culture failed
F492/50	0.5 cell/well 5 C4	ND	ND	2-6	-	-	-	-	-	-	Culture failed
F492/50	0.5 cell/well 8 B11	ND	ND	2-5	-	-	-	-	-	-	Culture failed

Appendix 6 Genotype of all clones isolated by *in vitro* limiting dilution cloning of bloodstream culture in Chapter 4.

For each clone the uncloned fly population from which is was derived is given alongside; the clone identity name, number of any cryopreserved stabilates made and genotype for seven microsatellite markers (markers not done are indicated by a dash). ND indicates a stabilate was not made. The final column signifies whether the resulting genotype is unique, the same as an existing hybrid, mixed/polyploid, or failed to grow in culture. Grey shading indicates the first example identified of a new hybrid genotype.

Marker	Progeny										
	F28/46/ AC 0.25A C5	F28/46/ AC 0.5A E8	F28/46/ AC 0.25A C3	F28/46/ AC 1A B8	F28/46/ AC 0.5B G6	F28/46/ AC 1B C8	F28/46/ AC 1A G7	F29/42/ AML DA6	F492/50 / 1 5 E3	F492/50 / 0.5 5 F2	F492/50 / 1 2 E3
	AC NEW 1	AC NEW 2	AC NEW 3	AC NEW 4	AC NEW 5	AC NEW 6	AC NEW 7	AC NEW8	AC NEW9	AC NEW10	AC NEW12
TB1/1	A	A	A	A	A	A	A	B	A	B	A
TB1/4	A	A	B	A	A	A	A	B	A	A	A
TB1/6	A	A	B	A	A	A	A	B	A	A	A
TB1/14	A	A	B	A	A	A	A	A	A	A	A
TB1/15	A	A	B	A	A	A	A	A	A	A	A
TB1/10	A	A	B	A	A	A	A	A	A	A	A
TB1/12	A	A	B	A	A	A	A	A	A	A	A
TB2/2	B	LOH?	A	A	A	A	A	A	A	A	A
TB2/4	B	LOH?	A	A	A	A	A	A	A	B	A
TB2/7	B	LOH?	A	A	A	A	A	A	B	B	A
TB2/12	B	LOH?	A	A	A	B	A	A	B	B	A
TB2/18	A	LOH?	A	A	B	B	A	A	-	B	B
TB2/19	A	LOH?	A	A	B	B	A	A	B	B	B
TB2/20	A	LOH?	A	A	B	B	A	A	B	B	B
TB3/1	B	B	A	A	B	A	B	B	A	B	B
TB3/20	B	B	B	A	A	A	B	B	A	B	B
TB3/4	B	B	B	A	-	A	B	-	A	-	B
TB3/21	B	B	B	A	A	A	B	A	A	B	B
TB3/22	B	B	B	A	A	A	B	-	A	B	B
TB3/23	B	B	B	A	A	A	B	A	B	B	B
TB3/10	A	B	B	A	A	A	B	A	B	A	B
TB3/13	A	B	B	A	A	A	B	A	B	A	B
TB3/14	A	B	B	A	A	A	B	A	B	A	A
TB4/17	B	B	B	A	A	A	A	-	B	B	-
TB4/2	B	B	B	A	A	A	A	B	B	B	B
TB4/18	B	B	B	A	A	A	A	B	B	B	B
TB4/19	B	B	B	A	A	A	A	B	B	B	B
TB4/20	B	B	B	A	A	A	B	B	B	B	B
TB4/8	B	B	B	A	A	A	B	A	B	B	B
TB4/21	A	B	B	A	A	-	B	A	B	B	B
TB4/22	A	B	B	A	A	A	B	A	B	B	B
TB4/12	A	B	B	A	A	A	B	A	B	B	B
TB4/13	A	B	B	A	A	A	B	A	B	B	B
TB5/15	B	A	A	B	B	A	A	B	A	B	A
TB5/16	B	A	A	B	B	A	A	B	A	B	A
TB5/4	B	A	A	B	A	A	A	B	A	B	A
TB5/17	B	A	A	B	A	A	A	B	A	B	A
TB5/18	B	A	A	B	A	A	A	A	A	B	A
TB5/19	B	A	A	A	A	A	A	A	A	B	A
TB5/20	B	B	A	A	B	A	A	A	A	A	A
TB6/9	A	A	B	A	A	B	B	B	B	A	B
TB6/10	A	A	B	A	A	B	B	B	B	A	B
TB6/11	A	A	B	A	A	B	B	B	A	A	B
TB6/12	A	A	B	A	B	B	B	B	A	A	B
TB6/13	A	A	A	A	B	B	B	B	A	A	A
TB6/6	A	A	A	A	B	-	B	B	A	-	A
TB6/15	A	A	A	A	B	B	B	B	A	A	A
TB6/16	A	A	A	A	B	B	B	B	A	B	-
TB7/4	A	A	B	B	A	A	B	B	B	B	B
TB7/5	A	A	B	B	B	A	B	B	B	B	B
TB7/15	A	A	B	A	B	A	A	B	B	B	A
TB7/16	A	A	B	A	B	A	A	B	B	B	A
TB8/13	B	A	B	A	B	B	A	A	B	B	A
TB8/14	B	A	B	A	B	B	A	A	B	A	A
TB8/15	B	A	A	A	B	B	A	A	B	A	A
TB8/16	B	A	A	A	B	B	A	A	B	A	A
TB8/17	B	A	A	A	B	B	A	A	B	A	A
TB8/18	B	A	A	A	B	B	B	A	B	A	A
TB8/10	B	A	A	A	B	A	B	A	B	A	A

TB8/19	A	A	A	A	B	A	B	B	B	A	A
TB8/20	A	A	A	B	A	A	B	B	B	A	A
TB8/21	A	A	A	B	B	-	B	B	B	A	A
TB9/17	B	A	B	A	A	B	A	B	A	A	B
TB9/18	A	A	B	A	A	B	A	-	A	A	B
TB9/19	A	A	B	A	A	B	A	B	A	A	B
TB9/21	A	A	B	A	A	B	A	A	A	A	B
TB9/5	A	B	B	A	A	B	A	A	A	A	B
TB9/9	A	B	B	B	A	B	A	A	B	-	B
TB9/12	A	B	B	B	A	B	A	B	B	B	B
TB9/14	A	B	B	B	B	B	A	B	B	B	B
TB10/22	B	B	B	A	A	B	B	-	B	-	B
TB10/23	B	B	B	A	A	B	B	B	B	A	B
TB10/24	B	B	B	A	A	B	B	B	A	A	B
TB10/25	A	B	B	B	-	B	B	B	A	A	B
TB10/12	A	B	B	B	A	B	B	B	A	-	B
TB10/14	B	B	A	B	A	B	B	-	A	A	B
TB10/26	B	B	A	B	A	B	B	B	A	A	B
TB10/27	B	B	A	B	A	B	B	B	A	-	B
TB10/28	B	B	A	B	A	B	B	B	A	A	B
TB10/19	-	B	B	B	A	B	B	B	A	A	B
TB11/32	B	B	A	B	B	A	A	B	B	-	A
TB11/7	B	A	A	B	B	A	A	B	B	A	A
TB11/11	B	A	A	B	B	B	A	B	B	A	A
TB11/13	B	A	A	B	B	B	A	B	B	A	A
TB11/37	B	A	B	B	A	B	B	B	B	B	A
TB11/40	A	A	B	B	A	B	B	B	B	B	A
TB11/42	B	A	B	B	A	B	B	A	B	B	B
TB11/43	B	A	B	B	A	B	B	A	B	B	B
TB11/45	B	B	A	B	A	B	B	A	B	A	B

Appendix 7 Segregation data for the new progeny clones for all microsatellite markers on each of the 11 megabase chromosomes in the STIB 386 genetic map.

Column 1 lists the marker name and columns 2-12 the inheritance pattern of the STIB 386 alleles (either A or B) in the progeny clones for each marker. The name of each progeny clone is given, along with laboratory alias along the top of each table. Progeny for which allele inheritance has not been determined are left blank. LOH indicates possible loss of heterozygosity.

References

Acosta-Serrano, A., Cole, R.N., Mehlert, A., Lee, M.G., Ferguson, M.A., and Englund, P.T. (1999). The procyclin repertoire of Trypanosoma brucei. Identification and structural characterization of the Glu-Pro-rich polypeptides. J Biol Chem **274,** 29763-29771.

Afewerk, Y., Clausen, P.H., Abebe, G., Tilahun, G., and Mehlitz, D. (2000). Multiple-drug resistant Trypanosoma congolense populations in village cattle of Metekel district, north-west Ethiopia. Acta Trop **76,** 231-238.

Agbo, E.E., Majiwa, P.A., Claassen, H.J., and te Pas, M.F. (2002). Molecular variation of Trypanosoma brucei subspecies as revealed by AFLP fingerprinting. Parasitology **124,** 349-358.

Aguilera, A., and Gomez-Gonzalez, B. (2008). Genome instability: a mechanistic view of its causes and consequences. Nat Rev Genet **9,** 204-217.

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

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

Andersen, M.P., Nelson, Z.W., Hetrick, E.D., and Gottschling, D.E. (2008). A genetic screen for increased loss of heterozygosity in Saccharomyces cerevisiae. Genetics **179,** 1179-1195.

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

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

Ashcroft, M.T. (1959). A comparision between a syringe-passaged and a tsetse-fly-transmitted line of a strain of Trypanosoma rhodesiense. Ann Trop Med Parasitol. **54,** 44-53.

Auffret, C.A., and Turner, M.J. (1981). Variant specific antigens of Trypanosoma brucei exist in solution as glycoprotein dimers. Biochem J **193,** 647-650.

Aylon, Y., Liefshitz, B., Bitan-Banin, G., and Kupiec, M. (2003). Molecular dissection of mitotic recombination in the yeast Saccharomyces cerevisiae. Mol Cell Biol **23,** 1403-1417.

Baker, B.S., Carpenter, A.T., Esposito, M.S., Esposito, R.E., and Sandler, L. (1976). The genetic control of meiosis. Annu Rev Genet **10,** 53-134.

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

Barbet, A.F., and Kamper, S.M. (1993). The importance of mosaic genes to trypanosome survival. Parasitol Today **9,** 63-66.

Barker, R.H., Jr., Liu, H., Hirth, B., Celatka, C.A., Fitzpatrick, R., Xiang, Y., Willert, E.K., Phillips, M.A., Kaiser, M., Bacchi, C.J., Rodriguez, A., Yarlett, N., Klinger, J.D., and Sybertz, E. (2009). Novel S-adenosylmethionine decarboxylase inhibitors for the treatment of human African trypanosomiasis. Antimicrob Agents Chemother **53,** 2052-2058.

Barnes, R.L., and McCulloch, R. (2007). Trypanosoma brucei homologous recombination is dependent on substrate length and homology, though displays a differential dependence on mismatch repair as substrate length decreases. Nucleic Acids Res **35,** 3478-3493.

Barnwell, J.W., Howard, R.J., Coon, H.G., and Miller, L.H. (1983). Splenic requirement for antigenic variation and expression of the variant antigen on the erythrocyte membrane in cloned Plasmodium knowlesi malaria. Infect Immun **40,** 985-994.

Barrett, M.P. (2001). Veterinary link to drug resistance in human African trypanosomiasis? Lancet **358,** 603-604.

Barrett, M.P., and Fairlamb, A.H. (1999). The biochemical basis of arsenical-diamidine crossresistance in African trypanosomes. Parasitol Today **15,** 136-140.

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

Barry, J.D. (1997). The relative significance of mechanisms of antigenic variation in African trypanosomes. Parasitol Today **13,** 212-218.

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

Barry, J.D., Ginger, M.L., Burton, P., and McCulloch, R. (2003). Why are parasite contingency genes often associated with telomeres? Int J Parasitol **33,** 29-45.

Barry, J.D., Marcello, L., Morrison, L.J., Read, A.F., Lythgoe, K., Jones, N., Carrington, M., Blandin, G., Bohme, U., Caler, E., Hertz-Fowler, C., Renauld, H., El-Sayed, N., and Berriman, M. (2005). What the genome sequence is revealing about trypanosome antigenic variation. Biochem Soc Trans **33,** 986-989.

Barzel, A., and Kupiec, M. (2008). Finding a match: how do homologous sequences get together for recombination? Nat Rev Genet **9,** 27-37.

Becker, M., Aitcheson, N., Byles, E., Wickstead, B., Louis, E., and Rudenko, G. (2004). Isolation of the repertoire of VSG expression site containing telomeres of Trypanosoma brucei 427 using transformation-associated recombination in yeast. Genome Res **14,** 2319-2329.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res **27,** 573-580.

Berger, B.J., Carter, N.S., and Fairlamb, A.H. (1995). Characterisation of pentamidine-resistant Trypanosoma brucei brucei. Mol Biochem Parasitol **69,** 289-298.

Bernhard, S.C., Nerima, B., Maser, P., and Brun, R. (2007). Melarsoprol- and pentamidine-resistant Trypanosoma brucei rhodesiense populations and their cross-resistance. Int J Parasitol **37,** 1443-1448.

Berriman, M., Hall, N., Sheader, K., Bringaud, F., Tiwari, B., Isobe, T., Bowman, S., Corton, C., Clark, L., Cross, G.A., Hoek, M., Zanders,

T., Berberof, M., Borst, P., and Rudenko, G. (2002). The architecture of variant surface glycoprotein gene expression sites in Trypanosoma brucei. Mol Biochem Parasitol **122,** 131-140.

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

Biggs, B.A., Kemp, D.J., and Brown, G.V. (1989). Subtelomeric chromosome deletions in field isolates of Plasmodium falciparum and their relationship to loss of cytoadherence in vitro. Proc Natl Acad Sci U S A **86,** 2428-2432.

Bingle, L.E., Eastlake, J.L., Bailey, M., and Gibson, W.C. (2001). A novel GFP approach for the analysis of genetic exchange in trypanosomes allowing the in situ detection of mating events. Microbiology **147,** 3231-3240.

Biteau, N., Bringaud, F., Gibson, W., Truc, P., and Baltz, T. (2000). Characterization of Trypanozoon isolates using a repeated coding sequence and microsatellite markers. Mol Biochem Parasitol **105,** 185-201.

Blackburn, A.C., McLary, S.C., Naeem, R., Luszcz, J., Stockton, D.W., Donehower, L.A., Mohammed, M., Mailhes, J.B., Soferr, T., Naber, S.P., Otis, C.N., and Jerry, D.J. (2004). Loss of heterozygosity occurs via mitotic recombination in Trp53+/- mice and associates with mammary tumor susceptibility of the BALB/c strain. Cancer Res **64,** 5140-5147.

Blaineau, C., Bastien, P., Rioux, J.A., Roizes, G., and Pages, M. (1991). Long-range restriction maps of size-variable homologous chromosomes in *Leishmania infantum*. Mol Biochem Parasitol **46,** 292-302.

Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C., and Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. Mol Ecol **13,** 3261-3273.

Boothroyd, C.E., Dreesen, O., Leonova, T., Ly, K.I., Figueiredo, L.M., Cross, G.A., and Papavasiliou, F.N. (2009). A yeast-endonuclease-generated DNA break induces antigenic switching in Trypanosoma brucei. Nature.

Borst, P., and Cross, G.A. (1982). Molecular basis for trypanosome antigenic variation. Cell **29,** 291-303.

Borst, P., and Ulbert, S. (2001). Control of VSG gene expression sites. Mol Biochem Parasitol **114,** 17-27.

Boyle, J.P., Saeij, J.P., Harada, S.Y., Ajioka, J.W., and Boothroyd, J.C. (2008). Expression quantitative trait locus mapping of toxoplasma genes reveals multiple mechanisms for strain-specific differences in gene expression. Eukaryot Cell **7,** 1403-1414.

Boyt, W.P. (1986). A field guide for diagnosis, treatment and prevention of African animal trypanosomiasis. Rome: FAO).

Bridges, D.J., Gould, M.K., Nerima, B., Maser, P., Burchmore, R.J., and de Koning, H.P. (2007). Loss of the high-affinity pentamidine transporter is responsible for high levels of cross-resistance between arsenical and diamidine drugs in African trypanosomes. Mol Pharmacol **71,** 1098-1108.

Brun, R., and Schonenberger. (1979). Cultivation and in vitro cloning or procyclic culture forms of Trypanosoma brucei in a semi-defined medium. Short communication. Acta Trop **36,** 289-292.

Brun, R., Schumacher, R., Schmid, C., Kunz, C., and Burri, C. (2001). The phenomenon of treatment failures in Human African Trypanosomiasis. Trop Med Int Health **6,** 906-914.

Burri, C., and Brun, R. (2003). Eflornithine for the treatment of human African trypanosomiasis. Parasitol Res **90 Supp 1,** S49-52.

Callejas, S., Leech, V., Reitter, C., and Melville, S. (2006). Hemizygous subtelomeres of an African trypanosome chromosome may account for over 75% of chromosome length. Genome Res **16,** 1109-1118.

Carlton, J., Mackinnon, M., and Walliker, D. (1998). A chloroquine resistance locus in the rodent malaria parasite Plasmodium chabaudi. Mol Biochem Parasitol **93,** 57-72.

Carr, L.L., and Gottschling, D.E. (2008). Does age influence loss of heterozygosity? Exp Gerontol **43,** 123-129.

Carruthers, V.B., and Cross, G.A. (1992). High-efficiency clonal growth of bloodstream- and insect-form Trypanosoma brucei on agarose plates. Proc Natl Acad Sci U S A **89,** 8818-8821.

Carter, N.S., and Fairlamb, A.H. (1993). Arsenical-resistant trypanosomes lack an unusual adenosine transporter. Nature **361,** 173-176.

Carter, N.S., Berger, B.J., and Fairlamb, A.H. (1995). Uptake of diamidine drugs by the P2 nucleoside transporter in melarsen-sensitive and -resistant Trypanosoma brucei brucei. J Biol Chem **270,** 28153-28157.

Chappuis, F., Udayraj, N., Stietenroth, K., Meussen, A., and Bovier, P.A. (2005). Eflornithine is safer than melarsoprol for the treatment of second-stage Trypanosoma brucei gambiense human African trypanosomiasis. Clin Infect Dis **41,** 748-751.

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

Chello, P.L., and Jaffe, J.J. (1972). Comparative properties of trypanosomal and mammalian thymidine kinases. Comp Biochem Physiol B **43,** 543-562.

Clausen, P.H., Adeyemi, I., Bauer, B., Breloeer, M., Salchow, F., and Staak, C. (1998). Host preferences of tsetse (Diptera: Glossinidae) based on bloodmeal identifications. Med Vet Entomol **12,** 169-180.

Codjia, V., Mulatu, W., Majiwa, P.A., Leak, S.G., Rowlands, G.J., Authie, E., d'Ieteren, G.D., and Peregrine, A.S. (1993). Epidemiology of bovine trypanosomiasis in the Ghibe valley, southwest Ethiopia. 3. Occurrence of populations of Trypanosoma congolense resistant to diminazene, isometamidium and homidium. Acta Trop **53,** 151-163.

Conway, C., Proudfoot, C., Burton, P., Barry, J.D., and McCulloch, R. (2002). Two pathways of homologous recombination in Trypanosoma brucei. Mol Microbiol **45,** 1687-1700.

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

Criscione, C.D., Valentim, C.L., Hirai, H., LoVerde, P.T., and Anderson, T.J. (2009). Genomic linkage map of the human blood fluke Schistosoma mansoni. Genome Biol **10,** R71.

Cross, G. (2009). Pedigrees of some Trypanosoma brucei strains in common use and an index of Lister 427 VSGs. http://tryps.rockefeller.edu/trypsru2_pedigrees.html 23rd Nov, 2009.

Cross, G.A. (1975). Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of Trypanosoma brucei. Parasitology **71,** 393-417.

Cross, G.A., and Manning, J.C. (1973). Cultivation of Trypanosoma brucei sspp. in semi-defined and defined media. Parasitology **67,** 315-331.

Cullen, J.K., Hussey, S.P., Walker, C., Prudden, J., Wee, B.Y., Dave, A., Findlay, J.S., Savory, A.P., and Humphrey, T.C. (2007). Break-induced loss of heterozygosity in fission yeast: dual roles for homologous recombination in promoting translocations and preventing de novo telomere addition. Mol Cell Biol **27,** 7745-7757.

Damper, D., and Patton, C.L. (1976). Pentamidine transport in Trypanosoma brucei-kinetics and specificity. Biochem Pharmacol **25,** 271-276.

Davis, M.W., and Hammarlund, M. (2006). Single-nucleotide polymorphism mapping. Methods Mol Biol **351,** 75-92.

Day, K.P., Karamalis, F., Thompson, J., Barnes, D.A., Peterson, C., Brown, H., Brown, G.V., and Kemp, D.J. (1993). Genes necessary for expression of a virulence determinant and for transmission of Plasmodium falciparum are located on a 0.3-megabase region of chromosome 9. Proc Natl Acad Sci U S A **90,** 8292-8296.

de Koning, H.P. (2001a). Transporters in African trypanosomes: role in drug action and resistance. Int J Parasitol **31,** 512-522.

De Koning, H.P. (2001b). Uptake of pentamidine in Trypanosoma brucei brucei is mediated by three distinct transporters: implications for cross-resistance with arsenicals. Mol Pharmacol **59,** 586-592.

de Koning, H.P., and Jarvis, S.M. (1999). Adenosine transporters in bloodstream forms of Trypanosoma brucei brucei: substrate

recognition motifs and affinity for trypanocidal drugs. Mol Pharmacol **56,** 1162-1170.

de Koning, H.P., and Jarvis, S.M. (2001). Uptake of pentamidine in Trypanosoma brucei brucei is mediated by the P2 adenosine transporter and at least one novel, unrelated transporter. Acta Trop **80,** 245-250.

de Koning, H.P., Anderson, L.F., Stewart, M., Burchmore, R.J., Wallace, L.J., and Barrett, M.P. (2004). The trypanocide diminazene aceturate is accumulated predominantly through the TbAT1 purine transporter: additional insights on diamidine resistance in african trypanosomes. Antimicrob Agents Chemother **48,** 1515-1519.

De Lange, T., Kooter, J.M., Michels, P.A., and Borst, P. (1983). Telomere conversion in trypanosomes. Nucleic Acids Res **11,** 8149-8165.

Diogo, D., Bouchier, C., d'Enfert, C., and Bougnoux, M.E. (2009). Loss of heterozygosity in commensal isolates of the asexual diploid yeast Candida albicans. Fungal Genet Biol **46,** 159-168.

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

Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T.P., Bowden, D.W., Smith, D.R., Lander, E.S., and et al. (1987). A genetic linkage map of the human genome. Cell **51,** 319-337.

Draper, J.S., Moore, H.D., Ruban, L.N., Gokhale, P.J., and Andrews, P.W. (2004). Culture and characterization of human embryonic stem cells. Stem Cells Dev **13,** 325-336.

Dukes, P. (1984). Arsenic and old taxa: subspeciation and drug sensitivity in Trypanosoma brucei. Trans R Soc Trop Med Hyg **78,** 711-725.

Eisler, M.C., Brandt, J., Bauer, B., Clausen, P.H., Delespaux, V., Holmes, P.H., Ilemobade, A., Machila, N., Mbwambo, H., McDermott, J., Mehlitz, D., Murilla, G., Ndung'u, J.M., Peregrine, A.S., Sidibe, I., Sinyangwe, L., and Geerts, S. (2001). Standardised tests in mice and cattle for the detection of drug resistance in tsetse-transmitted trypanosomes of African domestic cattle. Vet Parasitol **97,** 171-182.

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

El-Sayed, N.M., Myler, P.J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E.A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D.C., Haas, B.J., Tran, A.N., Wortman, J.R., Alsmark, U.C., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J.M., Cerqueira, G.C., Creasy, T., Delcher, A.L., Djikeng, A., Embley, T.M., Hauser, C., Ivens, A.C., Kummerfeld, S.K., Pereira-Leal, J.B., Nilsson, D., Peterson, J., Salzberg, S.L., Shallom, J., Silva, J.C., Sundaram, J., Westenberger, S., White, O., Melville, S.E.,

Donelson, J.E., Andersson, B., Stuart, K.D., and Hall, N. (2005). Comparative genomics of trypanosomatid parasitic protozoa. Science **309,** 404-409.

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

Fairlamb, A.H., Henderson, G.B., and Cerami, A. (1989). Trypanothione is the primary target for arsenical drugs against African trypanosomes. Proc Natl Acad Sci U S A **86,** 2607-2611.

Fairlamb, A.H., Carter, N.S., Cunningham, M., and Smith, K. (1992). Characterisation of melarsen-resistant Trypanosoma brucei brucei with respect to cross-resistance to other drugs and trypanothione metabolism. Mol Biochem Parasitol **53,** 213-222.

Ferdig, M.T., and Su, X.Z. (2000). Microsatellite markers and genetic mapping in Plasmodium falciparum. Parasitol Today **16,** 307-312.

Ferdig, M.T., Cooper, R.A., Mu, J., Deng, B., Joy, D.A., Su, X.Z., and Wellems, T.E. (2004). Dissecting the loci of low-level quinine resistance in malaria parasites. Mol Microbiol **52,** 985-997.

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

Fidock, D.A., Nomura, T., Talley, A.K., Cooper, R.A., Dzekunov, S.M., Ferdig, M.T., Ursos, L.M., Sidhu, A.B., Naude, B., Deitsch, K.W., Su, X.Z., Wootton, J.C., Roepe, P.D., and Wellems, T.E. (2000). Mutations in the P. falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. Mol Cell **6,** 861-871.

Forche, A., Magee, P.T., Magee, B.B., and May, G. (2004). Genome-wide single-nucleotide polymorphism map for Candida albicans. Eukaryot Cell **3,** 705-714.

Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M.M., Tsui, S.K., Xue, H., Wong, J.T., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallee, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., Mak, W., Song, Y.Q., Tam, P.K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., de Bakker, P.I., Barrett, J., Chretien, Y.R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D.J., Sabeti, P., Saxena, R., Schaffner, S.F.,

Sham, P.C., Varilly, P., Altshuler, D., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M.K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Tsunoda, T., Mullikin, J.C., Sherry, S.T., Feolo, M., Skol, A., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal, C.D., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Gibbs, R.A., Belmont, J.W., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G.M., Wheeler, D.A., Yakub, I., Gabriel, S.B., Onofrio, R.C., Richter, D.J., Ziaugra, L., Birren, B.W., Daly, M.J., Altshuler, D., Wilson, R.K., Fulton, L.L., Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archeveque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., McEwen, J.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R., and Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature **449,** 851-861.

Frommel, T.O. (1988). Trypanosoma brucei rhodesiense: effect of immunosuppression on the efficacy of melarsoprol treatment of infected mice. Exp Parasitol **67,** 364-366.

Gaunt, M.W., Yeo, M., Frame, I.A., Stothard, J.R., Carrasco, H.J., Taylor, M.C., Mena, S.S., Veazey, P., Miles, G.A., Acosta, N., de Arias, A.R., and Miles, M.A. (2003). Mechanism of genetic exchange in American trypanosomes. Nature **421,** 936-939.

Geerts, S., Holmes, P.H., Eisler, M.C., and Diall, O. (2001). African bovine trypanosomiasis: the problem of drug resistance. Trends Parasitol **17,** 25-28.

Geerts, S.H., P.H. (1998). Drug management and parasite resistance in bovine trypanosomiasis. (PAAT Technical and Scientific Series No. 1). (FAO).

Geigy, R., Mwambu, P.M., and Kauffmann, M. (1971). Sleeping sickness survey in Musoma district, Tanzania. IV. Examination of wild mammals as a potential reservoir for T.rhodesiense. Acta Trop **28,** 211-220.

GeneDB. (2005). *Trypanosoma brucei* GeneDB. http://www.genedb.org/genedb/tryp/index.jsp 23rd Nov, 2009.

Ghedin, E., Bringaud, F., Peterson, J., Myler, P., Berriman, M., Ivens, A., Andersson, B., Bontempi, E., Eisen, J., Angiuoli, S., Wanless, D., Von Arx, A., Murphy, L., Lennard, N., Salzberg, S., Adams, M.D., White, O., Hall, N., Stuart, K., Fraser, C.M., and El-Sayed, N.M. (2004). Gene synteny and evolution of genome architecture in trypanosomatids. Mol Biochem Parasitol **134,** 183-191.

Gibson, W. (2001). Sex and evolution in trypanosomes. Int J Parasitol **31,** 643-647.

Gibson, W., and Garside, L. (1991). Genetic exchange in Trypanosoma brucei brucei: variable chromosomal location of housekeeping genes in different trypanosome stocks. Mol Biochem Parasitol **45,** 77-89.

Gibson, W., and Whittington, H. (1993). Genetic exchange in Trypanosoma brucei: selection of hybrid trypanosomes by introduction of genes conferring drug resistance. Mol Biochem Parasitol **60,** 19-26.

Gibson, W., and Bailey, M. (1994). Genetic exchange in Trypanosoma brucei: evidence for meiosis from analysis of a cross between drug-resistant transformants. Mol Biochem Parasitol **64,** 241-252.

Gibson, W., and Stevens, J. (1999). Genetic exchange in the *trypanosomatidae*. Adv Parasitol **43,** 1-46.

Gibson, W., 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. Mol Biochem Parasitol **51,** 189-199.

Gibson, W., Kanmogne, G., and Bailey, M. (1995). A successful backcross in Trypanosoma brucei. Mol Biochem Parasitol **69,** 101-110.

Gibson, W., 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. Tropenmed Parasitol **29,** 335-345.

Gibson, W., Winters, K., Mizen, G., Kearns, J., and Bailey, M. (1997). Intraclonal mating in Trypanosoma brucei is associated with out-crossing. Microbiology **143 (Pt 3),** 909-920.

Gibson, W., Peacock, L., Ferris, V., Williams, K., and Bailey, M. (2006). Analysis of a cross between green and red fluorescent trypanosomes. Biochem Soc Trans **34,** 557-559.

Gibson, W.C. (1986). Will the real Trypanosoma b. gambiense please stand up. Parasitol Today **2,** 255-257.

Gibson, W.C. (1989). Analysis of a genetic cross between Trypanosoma brucei rhodesiense and T. b. brucei. Parasitology **99 Pt 3,** 391-402.

Gibson, W.C., de, C.M.T.F., and Godfrey, D.G. (1980). Numerical analysis of enzyme polymorphism: a new approach to the epidemiology and taxonomy of trypanosomes of the subgenus Trypanozoon. Adv Parasitol **18,** 175-246.

Glover, L., McCulloch, R., and Horn, D. (2008). Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. Nucleic Acids Res **36,** 2608-2618.

Glover, T.W. (2006). Common fragile sites. Cancer Lett **232,** 4-12.

Glover, T.W., Arlt, M.F., Casper, A.M., and Durkin, S.G. (2005). Mechanisms of common fragile site instability. Hum Mol Genet **14 Spec No. 2,** R197-205.

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 (Pt 2),** 337-347.

Gottesdiener, K., Garcia-Anoveros, J., Lee, M.G., and Van der Ploeg, L.H. (1990). Chromosome organization of the protozoan *Trypanosoma brucei*. Mol Cell Biol **10,** 6079-6083.

Grech, K., Martinelli, A., Pathirana, S., Walliker, D., Hunt, P., and Carter, R. (2002). Numerous, robust genetic markers for Plasmodium chabaudi by the method of amplified fragment length polymorphism. Mol Biochem Parasitol **123,** 95-104.

Guo, Z.G., and Johnson, A.M. (1995). Genetic characterization of Toxoplasma gondii strains by random amplified polymorphic DNA polymerase chain reaction. Parasitology **111 (Pt 2),** 127-132.

Gupta, P.K., Shao, C., Zhu, Y., Sahota, A., and Tischfield, J.A. (1997). Loss of heterozygosity analysis in a human fibrosarcoma cell line. Cytogenet Cell Genet **76,** 214-218.

Gut, I.G. (2001). Automation in genotyping of single nucleotide polymorphisms. Hum Mutat **17,** 475-492.

Haber, J.E. (2000). Partners and pathwaysrepairing a double-strand break. Trends Genet **16,** 259-264.

Hagstrom, S.A., and Dryja, T.P. (1999). Mitotic recombination map of 13cen-13q14 derived from an investigation of loss of heterozygosity in retinoblastomas. Proc Natl Acad Sci U S A **96,** 2952-2957.

Haldane, J.B.S. (1919). The mapping function. Journal of Genetics **8,** 299-309.

Hall, N., Berriman, M., Lennard, N.J., Harris, B.R., Hertz-Fowler, C., Bart-Delabesse, E.N., Gerrard, C.S., Atkin, R.J., Barron, A.J., Bowman, S., Bray-Allen, S.P., Bringaud, F., Clark, L.N., Corton, C.H., Cronin, A., Davies, R., Doggett, J., Fraser, A., Gruter, E., Hall, S., Harper, A.D., Kay, M.P., Leech, V., Mayes, R., Price, C., Quail, M.A., Rabbinowitsch, E., Reitter, C., Rutherford, K., Sasse, J., Sharp, S., Shownkeen, R., MacLeod, A., Taylor, S., Tweedie, A., Turner, C.M., Tait, A., Gull, K., Barrell, B., and Melville, S.E. (2003). The DNA sequence of chromosome I of an African trypanosome: gene content, chromosome organisation, recombination and polymorphism. Nucleic Acids Res **31,** 4864-4873.

Hayton, K., Gaur, D., Liu, A., Takahashi, J., Henschen, B., Singh, S., Lambert, L., Furuya, T., Bouttenot, R., Doll, M., Nawaz, F., Mu, J., Jiang, L., Miller, L.H., and Wellems, T.E. (2008). Erythrocyte binding protein PfRH5 polymorphisms determine species-specific pathways of Plasmodium falciparum invasion. Cell Host Microbe **4,** 40-51.

Heisch, R.B., Mc, M.J., and Mansonbahr, P.E. (1958). The isolation of Trypanosoma rhodesiense from a bushbuck. Br Med J **2,** 1203-1204.

Helleday, T. (2003). Pathways for mitotic homologous recombination in mammalian cells. Mutat Res **532,** 103-115.

Henriksson, J., Porcel, B., Rydaker, M., Ruiz, A., Sabaj, V., Galanti, N., Cazzulo, J.J., Frasch, A.C., and Pettersson, U. (1995). Chromosome specific markers reveal conserved linkage groups in spite of extensive chromosomal size variation in *Trypanosoma cruzi*. Mol Biochem Parasitol **73,** 63-74.

Herbert, W.J., and Lumsden, W.H. (1976). Trypanosoma brucei: a rapid "matching" method for estimating the host's parasitemia. Exp Parasitol **40,** 427-431.

Hertz-Fowler, C., Figueiredo, L.M., Quail, M.A., Becker, M., Jackson, A., Bason, N., Brooks, K., Churcher, C., Fahkro, S., Goodhead, I., Heath, P., Kartvelishvili, M., Mungall, K., Harris, D., Hauser, H., Sanders, M., Saunders, D., Seeger, K., Sharp, S., Taylor, J.E., Walker, D., White, B., Young, R., Cross, G.A., Rudenko, G., Barry, J.D., Louis, E.J., and Berriman, M. (2008). Telomeric expression sites are highly conserved in Trypanosoma brucei. PLoS One **3,** e3527.

Hesse, F., Selzer, P.M., Muhlstadt, K., and Duszenko, M. (1995). A novel cultivation technique for long-term maintenance of bloodstream form trypanosomes in vitro. Mol Biochem Parasitol **70,** 157-166.

Hide, G. (1999). History of sleeping sickness in East Africa. Clin Microbiol Rev **12,** 112-125.

Hirumi, H., and Hirumi, K. (1989). Continuous cultivation of Trypanosoma brucei blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. J Parasitol **75,** 985-989.

Hirumi, H., Doyle, J.J., and Hirumi, K. (1977). African trypanosomes: cultivation of animal-infective Trypanosoma brucei in vitro. Science **196,** 992-994.

Hoeijmakers, J.H., Frasch, A.C., Bernards, A., Borst, P., and Cross, G.A. (1980). Novel expression-linked copies of the genes for variant surface antigens in trypanosomes. Nature **284,** 78-80.

Hope, M., MacLeod, A., Leech, V., Melville, S., Sasse, J., Tait, A., and Turner, C.M. (1999). Analysis of ploidy (in megabase chromosomes) in Trypanosoma brucei after genetic exchange. Mol Biochem Parasitol **104,** 1-9.

Horn, D., and Barry, J.D. (2005). The central roles of telomeres and subtelomeres in antigenic variation in African trypanosomes. Chromosome Res **13,** 525-533.

Hoskins, R.A., Phan, A.C., Naeemuddin, M., Mapa, F.A., Ruddy, D.A., Ryan, J.J., Young, L.M., Wells, T., Kopczynski, C., and Ellis, M.C. (2001). Single nucleotide polymorphism markers for genetic mapping in Drosophila melanogaster. Genome Res **11,** 1100-1113.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T., Dong, G., Sang, T., and Han, B. (2009). High-throughput genotyping by whole-genome resequencing. Genome Res **19,** 1068-1076.

ILRAD. (1993). Estimating the costs of animal trypanosomiasis in Africa. ILRAD reports.

ILRI. (1999). Using the Economic Surplus Model to Measure Potential Returns to International Livestock Research. Impact Assessment Series.

Ira, G., Malkova, A., Liberi, G., Foiani, M., and Haber, J.E. (2003). Srs2 and Sgs1-Top3 suppress crossovers during double-strand break repair in yeast. Cell **115,** 401-411.

Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M.A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S.M., Bianchettin, G., Borzym, K., Bothe, G., Bruschi, C.V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R.M., Cronin, A., Cruz, A.K., Davies, R.M., De Gaudenzi, J., Dobson, D.E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A.C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J.C., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'Neil, S., Pentony, M., Pohl, T.M., Price, C., Purnelle, B., Quail, M.A., Rabbinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J.C.,

Rutter, S., Saunders, D., Schafer, M., Schein, J., Schwartz, D.C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D.F., Blackwell, J.M., Stuart, K.D., Barrell, B., and Myler, P.J. (2005). The genome of the kinetoplastid parasite, Leishmania major. Science **309,** 436-442.

Janse, C.J. (1993). Chromosome size polymorphism and DNA rearrangements in plasmodium. Parasitol Today **9,** 19-22.

Janssen, P., Coopman, R., Huys, G., Swings, J., Bleeker, M., Vos, P., Zabeau, M., and Kersters, K. (1996). Evaluation of the DNA fingerprinting method AFLP as an new tool in bacterial taxonomy. Microbiology **142 (Pt 7),** 1881-1893.

Jenni, L., Marti, S., Schweizer, J., Betschart, B., Le Page, R.W., 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.

Johnson, P.J., and Borst, P. (1986). Mapping of VSG genes on large expression-site chromosomes of Trypanosoma brucei separated by pulsed-field gradient electrophoresis. Gene **43,** 213-220.

Johnson, P.J., Kooter, J.M., and Borst, P. (1987). Inactivation of transcription by UV irradiation of T. brucei provides evidence for a multicistronic transcription unit including a VSG gene. Cell **51,** 273-281.

Johnson, R.D., and Jasin, M. (2000). Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. Embo J **19,** 3398-3407.

Kadyk, L.C., and Hartwell, L.H. (1992). Sister chromatids are preferred over homologs as substrates for recombinational repair in Saccharomyces cerevisiae. Genetics **132,** 387-402.

Kamper, S.M., and Barbet, A.F. (1992). Surface epitope variation via mosaic gene formation is potential key to long-term survival of Trypanosoma brucei. Mol Biochem Parasitol **53,** 33-44.

Keita, M., Bouteille, B., Enanga, B., Vallat, J.M., and Dumas, M. (1997). Trypanosoma brucei brucei: a long-term model of human African trypanosomiasis in mice, meningo-encephalitis, astrocytosis, and neurological disorders. Exp Parasitol **85,** 183-192.

Kennedy, P.G. (2004). Human African trypanosomiasis of the CNS: current issues and challenges. J Clin Invest **113,** 496-504.

Kennedy, P.G. (2006). Diagnostic and neuropathogenesis issues in human African trypanosomiasis. Int J Parasitol **36,** 505-512.

Khan, A., Taylor, S., Su, C., Mackey, A.J., Boyle, J., Cole, R., Glover, D., Tang, K., Paulsen, I.T., Berriman, M., Boothroyd, J.C., Pfefferkorn, E.R., Dubey, J.P., Ajioka, J.W., Roos, D.S., Wootton, J.C., and Sibley, L.D. (2005). Composite genome map and recombination parameters derived from three archetypal lineages of *Toxoplasma gondii*. Nucleic Acids Res **33,** 2980-2992.

Kilgour, V. (1980). The electrophoretic mobilities and activities of eleven enzymes of bloodstream and culture forms of Trypanosoma brucei compared. Mol Biochem Parasitol **2,** 51-62.

Klein, F., Laroche, T., Cardenas, M.E., Hofmann, J.F., Schweizer, D., and Gasser, S.M. (1992). Localization of RAP1 and topoisomerase II

in nuclei and meiotic chromosomes of yeast. J Cell Biol **117,** 935-948.

Kodama, H., Denso, Okazaki, F., and Ishida, S. (2008). Protective effect of humus extract against Trypanosoma brucei infection in mice. J Vet Med Sci **70,** 1185-1190.

Koffi, M., Solano, P., Barnabe, C., de Meeus, T., Bucheton, B., Cuny, G., and Jamonneau, V. (2007). Genetic characterisation of Trypanosoma brucei s.l. using microsatellite typing: new perspectives for the molecular epidemiology of human African trypanosomiasis. Infect Genet Evol **7,** 675-684.

Koffi, M., De Meeus, T., Bucheton, B., Solano, P., Camara, M., Kaba, D., Cuny, G., Ayala, F.J., and Jamonneau, V. (2009). Population genetics of Trypanosoma brucei gambiense, the agent of sleeping sickness in Western Africa. Proc Natl Acad Sci U S A **106,** 209-214.

Kooter, J.M., Winter, A.J., de Oliveira, C., Wagter, R., and Borst, P. (1988). Boundaries of telomere conversion in Trypanosoma brucei. Gene **69,** 1-11.

Kooter, J.M., van der Spek, H.J., Wagter, R., d'Oliveira, C.E., van der Hoeven, F., Johnson, P.J., and Borst, P. (1987). The anatomy and transcription of a telomeric expression site for variant-specific surface antigens in T. brucei. Cell **51,** 261-272.

Kosambi, D.D. (1944). The estimation of map distances from recombination values. Annals of Eugenics **12,** 172-175.

Krafsur, E.S. (2009). Tsetse flies: genetics, evolution, and role as vectors. Infect Genet Evol **9,** 124-141.

LaFave, M.C., and Sekelsky, J. (2009). Mitotic recombination: why? when? how? where? PLoS Genet **5,** e1000411.

Lamont, G.S., Tucker, R.S., and Cross, G.A. (1986). Analysis of antigen switching rates in Trypanosoma brucei. Parasitology **92 (Pt 2),** 355-367.

Lanzer, M., de Bruin, D., and Ravetch, J.V. (1993). Transcriptional differences in polymorphic and conserved domains of a complete cloned P. falciparum chromosome. Nature **361,** 654-657.

Lanzer, M., de Bruin, D., Wertheimer, S.P., and Ravetch, J.V. (1994). Organization of chromosomes in *Plasmodium falciparum*: a model for generating karyotypic diversity. Parasitol Today **10,** 114-117.

Lee, P.S., Greenwell, P.W., Dominska, M., Gawel, M., Hamilton, M., and Petes, T.D. (2009). A fine-structure map of spontaneous mitotic crossovers in the yeast Saccharomyces cerevisiae. PLoS Genet **5,** e1000410.

Leech, V., Quail, M.A., and Melville, S.E. (2004). Separation, digestion, and cloning of intact parasite chromosomes embedded in agarose. Methods Mol Biol **270,** 335-352.

Lefort, N., Feyeux, M., Bas, C., Feraud, O., Bennaceur-Griscelli, A., Tachdjian, G., Peschanski, M., and Perrier, A.L. (2008). Human embryonic stem cells reveal recurrent genomic instability at 20q11.21. Nat Biotechnol **26,** 1364-1366.

Legros, D., Fournier, C., Gastellu Etchegorry, M., Maiso, F., and Szumilin, E. (1999). [Therapeutic failure of melarsoprol among patients treated for late stage T.b. gambiense human African trypanosomiasis in Uganda]. Bull Soc Pathol Exot **92,** 171-172.

Lemoine, F.J., Degtyareva, N.P., Lobachev, K., and Petes, T.D. (2005). Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. Cell **120,** 587-598.

Li, F.J., Gasser, R.B., Zheng, J.Y., Claes, F., Zhu, X.Q., and Lun, Z.R. (2005). Application of multiple DNA fingerprinting techniques to study the genetic relationships among three members of the subgenus Trypanozoon (Protozoa: Trypanosomatidae). Mol Cell Probes **19,** 400-407.

Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res **18,** 1851-1858.

Lindegard, L.A.G. (1999). Human serum resistance in *T.brucei* PhD (University of Glasgow).

Liu, A.Y., Van der Ploeg, L.H., Rijsewijk, F.A., and Borst, P. (1983). The transposition unit of variant surface glycoprotein gene 118 of Trypanosoma brucei. Presence of repeated elements at its border and absence of promoter-associated sequences. J Mol Biol **167,** 57-75.

Liu, B., Liu, Y., Motyka, S.A., Agbo, E.E., and Englund, P.T. (2005). Fellowship of the rings: the replication of kinetoplast DNA. Trends Parasitol **21,** 363-369.

Lun, Z.R., Li, A.X., Chen, X.G., Lu, L.X., and Zhu, X.Q. (2004). Molecular profiles of Trypanosoma brucei, T. evansi and T. equiperdum stocks revealed by the random amplified polymorphic DNA method. Parasitol Res **92,** 335-340.

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

MacLeod, A. (1999). Genetic analysis of *Trypanosoma brucei* University of Glasgow).

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

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

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

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

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

MacLeod, A., Tweedie, A., McLellan, S., Taylor, S., Hall, N., Berriman, M., El-Sayed, N.M., Hope, M., Turner, C.M., and Tait, A. (2005b). The genetic map and comparative analysis with the physical map of

Trypanosoma brucei. Nucleic Acids Res **33,** 6688-6693. Oxford University Press.

MacLeod, A., Turner, C.M.R., Tait, A. (2007). The system of genetic exchange in Trypanosoma brucei and other trypanosomatids. In Trypanosomes:After the Genome, M.R. Barry J.D., Mottram J. and Acosta-Serrano A., ed (Norfolk, UK: Horizon Bioscience), pp. 71-90.

Malkova, A., Ivanov, E.L., and Haber, J.E. (1996). Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. Proc Natl Acad Sci U S A **93,** 7131-7136.

Mandegar, M.A., and Otto, S.P. (2007). Mitotic recombination counteracts the benefits of genetic segregation. Proc Biol Sci **274,** 1301-1307.

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

Manly, K.F., Cudmore, R.H., Jr., and Meer, J.M. (2001). Map Manager QTX, cross-platform software for genetic mapping. Mamm Genome **12,** 930-932.

Marcello, L., and Barry, J.D. (2007). Analysis of the VSG gene silent archive in Trypanosoma brucei reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. Genome Res **17,** 1344-1352.

Martinelli, A., Hunt, P., Fawcett, R., Cravo, P.V., Walliker, D., and Carter, R. (2005a). An AFLP-based genetic linkage map of *Plasmodium chabaudi chabaudi*. Malar J **4,** 11.

Martinelli, A., Cheesman, S., Hunt, P., Culleton, R., Raza, A., Mackinnon, M., and Carter, R. (2005b). A genetic approach to the *de novo* identification of targets of strain-specific immunity in malaria parasites. Proc Natl Acad Sci U S A **102,** 814-819.

Martinez, E., Alonso, V., Quispe, A., Thomas, M.C., Alonso, R., Pinero, J.E., Gonzalez, A.C., Ortega, A., and Valladares, B. (2003). RAPD method useful for distinguishing Leishmania species: design of specific primers for L. braziliensis. Parasitology **127,** 513-517.

Maser, P., Sutterlin, C., Kralli, A., and Kaminsky, R. (1999). A nucleoside transporter from Trypanosoma brucei involved in drug resistance. Science **285,** 242-244.

Masiga, D.K., and Turner, C.M. (2004). Amplified (restriction) fragment length polymorphism (AFLP) analysis. Methods Mol Biol **270,** 173-186.

Masiga, D.K., Ndung'u, K., Tweedie, A., Tait, A., and Turner, C.M. (2006). Trypanosoma evansi: genetic variability detected using amplified restriction fragment length polymorphism (AFLP) and random amplified polymorphic DNA (RAPD) analysis of Kenyan isolates. Exp Parasitol **114,** 147-153.

Matovu, E., Seebeck, T., Enyaru, J.C., and Kaminsky, R. (2001a). Drug resistance in Trypanosoma brucei spp., the causative agents of sleeping sickness in man and nagana in cattle. Microbes Infect **3,** 763-770.

Matovu, E., Iten, M., Enyaru, J.C., Schmid, C., Lubega, G.W., Brun, R., and Kaminsky, R. (1997). Susceptibility of Ugandan Trypanosoma brucei rhodesiense isolated from man and animal reservoirs to diminazene, isometamidium and melarsoprol. Trop Med Int Health **2,** 13-18.

Matovu, E., Geiser, F., Schneider, V., Maser, P., Enyaru, J.C., Kaminsky, R., Gallati, S., and Seebeck, T. (2001b). Genetic variants of the TbAT1 adenosine transporter from African trypanosomes in relapse

infections following melarsoprol therapy. Mol Biochem Parasitol **117,** 73-81.

Matovu, E., Stewart, M.L., Geiser, F., Brun, R., Maser, P., Wallace, L.J., Burchmore, R.J., Enyaru, J.C., Barrett, M.P., Kaminsky, R., Seebeck, T., and de Koning, H.P. (2003). Mechanisms of arsenical and diamidine uptake and resistance in Trypanosoma brucei. Eukaryot Cell **2,** 1003-1008.

Matthews, K.R. (2005). The developmental cell biology of Trypanosoma brucei. J Cell Sci **118,** 283-290.

Matthews, K.R., Sherwin, T., and Gull, K. (1995). Mitochondrial genome repositioning during the differentiation of the African trypanosome between life cycle forms is microtubule mediated. J Cell Sci **108 (Pt 6),** 2231-2239.

Maudlin, I., Welburn, S.C., and Milligan, P.J. (1998). Trypanosome infections and survival in tsetse. Parasitology **116 Suppl,** S23-28.

McCulloch, R. (2004). Antigenic variation in African trypanosomes: monitoring progress. Trends Parasitol **20,** 117-121.

McCulloch, R., and Barry, J.D. (1999). A role for RAD51 and homologous recombination in Trypanosoma brucei antigenic variation. Genes Dev **13,** 2875-2888.

McCulloch, R., Vassella, E., Burton, P., Boshart, M., and Barry, J.D. (2004). Transformation of monomorphic and pleomorphic Trypanosoma brucei. Methods Mol Biol **262,** 53-86.

McDermott, J., Woitag, T., Sidibe, I., Bauer, B., Diarra, B., Ouedraogo, D., Kamuanga, M., Peregrine, A., Eisler, M., Zessin, K.H., Mehlitz, D., and Clausen, P.H. (2003). Field studies of drug-resistant cattle trypanosomes in Kenedougou Province, Burkina Faso. Acta Trop **86,** 93-103.

McEachern, M.J., and Haber, J.E. (2006). Break-induced replication and recombinational telomere elongation in yeast. Annu Rev Biochem **75,** 111-135.

McMurray, M.A., and Gottschling, D.E. (2003). An age-induced switch to a hyper-recombinational state. Science **301,** 1908-1911.

Méda, H.A., and Pepin, J. (2001). Report of the Scientific Working Group meeting on African trypanosomiasis.).

Mehlitz, D., Zillmann, 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. Tropenmed Parasitol **33,** 113-118.

Melville, S.E., Gerrard, C.S., and Blackwell, J.M. (1999). Multiple causes of size variation in the diploid megabase chromosomes of African tyrpanosomes. Chromosome Res **7,** 191-203.

Melville, S.E., Leech, V., Navarro, M., and Cross, G.A. (2000). The molecular karyotype of the megabase chromosomes of Trypanosoma brucei stock 427. Mol Biochem Parasitol **111,** 261-273.

Melville, S.E., Leech, V., Gerrard, C.S., Tait, A., and Blackwell, J.M. (1998). The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers. Mol Biochem Parasitol **94,** 155-173.

Meudt, H.M., and Clarke, A.C. (2007). Almost forgotten or latest practice? AFLP applications, analyses and advances. Trends Plant Sci **12,** 106-117.

Michel, B. (2000). Replication fork arrest and DNA recombination. Trends Biochem Sci **25,** 173-178.

Morrison, L.J., Marcello, L., and McCulloch, R. (2009a). Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity. Cell Microbiol.

Morrison, L.J., Majiwa, P., Read, A.F., and Barry, J.D. (2005). Probabilistic order in antigenic variation of Trypanosoma brucei. Int J Parasitol **35,** 961-972.

Morrison, L.J., Tait, A., McLellan, S., Sweeney, L., Turner, C.M., and MacLeod, A. (2009b). A major genetic locus in Trypanosoma brucei is a determinant of host pathology. PLoS Negl Trop Dis **3,** e557.

Morrison, L.J., McCormack, G., Sweeney, L., Likeufack, A.C., Truc, P., Turner, C.M., Tait, A., and MacLeod, A. (2007). Use of multiple displacement amplification to increase the detection and genotyping of trypanosoma species samples immobilized on FTA filters. Am J Trop Med Hyg **76,** 1132-1137.

Morrison, L.J., Tait, A., McCormack, G., Sweeney, L., Black, A., Truc, P., Likeufack, A.C., Turner, C.M., and MacLeod, A. (2008). Trypanosoma brucei gambiense Type 1 populations from human patients are clonal and display geographical genetic differentiation. Infect Genet Evol **8,** 847-854.

Moynahan, M.E., and Jasin, M. (1997). Loss of heterozygosity induced by a chromosomal double-strand break. Proc Natl Acad Sci U S A **94,** 8988-8993.

Mulugeta, W., Wilkes, J., Mulatu, W., Majiwa, P.A., Masake, R., and Peregrine, A.S. (1997). Long-term occurrence of Trypanosoma congolense resistant to diminazene, isometamidium and homidium in cattle at Ghibe, Ethiopia. Acta Trop **64,** 205-217.

Na'Isa, B.K. (1967). Follow-up of a survey on the prevalence of homidium-resistant strains of trypanosomes in cattle in Northern Nigeria and drug cross-resistance tests on the strains with Samorin and Berenil. Bull Epizoot Dis Afr **15,** 231-241.

NCBI. (1999). Table comparing genetic map unit distances of Plasmodium falciparum and other eukaryotes. http://www.ncbi.nlm.nih.gov/projects/Malaria/Mapsmarkers/tabphysdistvsgmu.html 20th Aug, 2008.

Nelson, F.K., Frankel, W., and Rajan, T.V. (1989). Mitotic recombination is responsible for the loss of heterozygosity in cultured murine cell lines. Mol Cell Biol **9,** 1284-1288.

Njiokou, F., Simo, G., Mbida Mbida, A., Truc, P., Cuny, G., and Herder, S. (2004). A study of host preference in tsetse flies using a modified heteroduplex PCR-based method. Acta Trop **91,** 117-120.

Njiokou, F., Laveissiere, C., Simo, G., Nkinin, S., Grebaut, P., Cuny, G., and Herder, S. (2006). Wild fauna as a probable animal reservoir for Trypanosoma brucei gambiense in Cameroon. Infect Genet Evol **6,** 147-153.

Nkinin, S.W., Njiokou, F., Penchenier, L., Grebaut, P., Simo, G., and Herder, S. (2002). Characterization of Trypanosoma brucei s.l. subspecies by isoenzymes in domestic pigs from the Fontem sleeping sickness focus of Cameroon. Acta Trop **81,** 225-232.

Nok, A.J. (2003). Arsenicals (melarsoprol), pentamidine and suramin in the treatment of human African trypanosomiasis. Parasitol Res **90,** 71-79.

Onyango, R.J., Van Hoeve, K., and De Raadt, P. (1966). The epidemiology of Trypanosoma rhodesiense sleeping sickness in Alego location, Central Nyanza, Kenya. I. Evidence that cattle may act as reservoir hosts of trypanosomes infective to man. Trans R Soc Trop Med Hyg **60,** 175-182.

Osman, A.S., Jennings, F.W., and Holmes, P.H. (1992). The rapid development of drug-resistance by Trypanosoma evansi in immunosuppressed mice. Acta Trop **50,** 249-257.

Paindavoine, P., Zampetti-Bosseler, F., Coquelet, H., Pays, E., and Steinert, M. (1989). Different allele frequencies in Trypanosoma brucei brucei and Trypanosoma brucei gambiense populations. Mol Biochem Parasitol **32,** 61-71.

Paindavoine, P., Zampetti-Bosseler, F., Pays, E., Schweizer, J., Guyaux, M., Jenni, L., and Steinert, M. (1986a). Trypanosome hybrids generated in tsetse flies by nuclear fusion. Embo J **5,** 3631-3636.

Paindavoine, P., Pays, E., Laurent, M., Geltmeyer, Y., Le Ray, D., Mehlitz, D., and Steinert, M. (1986b). The use of DNA hybridization and numerical taxonomy in determining relationships between Trypanosoma brucei stocks and subspecies. Parasitology **92 (Pt 1),** 31-50.

Paques, F., and Haber, J.E. (1999). Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev **63,** 349-404.

Payne, R.C., Sukanto, I.P., Partoutomo, S., and Jones, T.W. (1994). Efficacy of Cymelarsan treatment of suramin resistant Trypanosoma evansi in cattle. Trop Anim Health Prod **26,** 92-94.

Pays, E., Guyaux, M., Aerts, D., Van Meirvenne, N., and Steinert, M. (1985). Telomeric reciprocal recombination as a possible mechanism for antigenic variation in trypanosomes. Nature **316,** 562-564.

Pays, E., Van Assel, S., Laurent, M., Darville, M., Vervoort, T., Van Meirvenne, N., and Steinert, M. (1983a). Gene conversion as a mechanism for antigenic variation in trypanosomes. Cell **34,** 371-381.

Pays, E., Delauw, M.F., Van Assel, S., Laurent, M., Vervoort, T., Van Meirvenne, N., and Steinert, M. (1983b). Modifications of a Trypanosoma b. brucei antigen gene repertoire by different DNA recombinational mechanisms. Cell **35,** 721-731.

Pays, E., Van Assel, S., Laurent, M., Dero, B., Michiels, F., Kronenberger, P., Matthyssens, G., Van Meirvenne, N., Le Ray, D., and Steinert, M. (1983c). At least two transposed sequences are associated in the expression site of a surface antigen gene in different trypanosome clones. Cell **34,** 359-369.

Peacock, L., Ferris, V., Bailey, M., and Gibson, W. (2008). Fly transmission and mating of Trypanosoma brucei brucei strain 427. Mol Biochem Parasitol **160,** 100-106.

Pepin, J., and Milord, F. (1994). The treatment of human African trypanosomiasis. Adv Parasitol **33,** 1-47.

Pepin, J., Milord, F., Khonde, A., Niyonsenga, T., Loko, L., and Mpia, B. (1994). Gambiense trypanosomiasis: frequency of, and risk factors for, failure of melarsoprol therapy. Trans R Soc Trop Med Hyg **88,** 447-452.

Peregrine, A.S. (1994). Chemotherapy and delivery systems: haemoparasites. Vet Parasitol **54,** 223-248.

Perez-Morga, D., Amiguet-Vercher, A., Vermijlen, D., and Pays, E. (2001). Organization of telomeres during the cell and life cycles of Trypanosoma brucei. J Eukaryot Microbiol **48,** 221-226.

Petes, T.D. (2001). Meiotic recombination hot spots and cold spots. Nat Rev Genet **2,** 360-369.

Phillips, M.A., and Wang, C.C. (1987). A Trypanosoma brucei mutant resistant to alpha-difluoromethylornithine. Mol Biochem Parasitol **22,** 9-17.

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

Pinchbeck, G.L., Morrison, L.J., Tait, A., Langford, J., Meehan, L., Jallow, S., Jallow, J., Jallow, A., and Christley, R.M. (2008). Trypanosomosis in The Gambia: prevalence in working horses and donkeys detected by whole genome amplification and PCR, and evidence for interactions between trypanosome species. BMC Vet Res **4,** 7.

Pologe, L.G., and Ravetch, J.V. (1986). A chromosomal rearrangement in a P. falciparum histidine-rich protein gene is associated with the knobless phenotype. Nature **322,** 474-477.

Pologe, L.G., and Ravetch, J.V. (1988). Large deletions result from breakage and healing of P. falciparum chromosomes. Cell **55,** 869-874.

Poltera, A.A., Hochmann, A., Rudin, W., and Lambert, P.H. (1980). Trypanosoma brucei brucei: a model for cerebral trypanosomiasis in mice--an immunological, histological and electronmicroscopic study. Clin Exp Immunol **40,** 496-507.

Pospichal, H., Brun, R., Kaminsky, R., and Jenni, L. (1994). Induction of resistance to melarsenoxide cysteamine (Mel Cy) in Trypanosoma brucei brucei. Acta Trop **58,** 187-197.

Preston, C.R., Flores, C., and Engels, W.R. (2006). Age-dependent usage of double-strand-break repair pathways. Curr Biol **16,** 2009-2015.

PRIDE. (1998). PRIDE Primer Design Software (PRIDE 1.2). http://pride.molgen.mpg.de/ 23rd Nov, 2009.

Ravel, C., Wincker, P., Blaineau, C., Britto, C., Bastien, P., and Pages, M. (1996). Medium-range restriction maps of five chromosomes of Leishmania infantum and localization of size-variable regions. Genomics **35,** 509-516.

Richardson, C., Moynahan, M.E., and Jasin, M. (1998). Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. Genes Dev **12,** 3831-3842.

Robinson, N.P., Burman, N., Melville, S.E., and Barry, J.D. (1999). Predominance of duplicative VSG gene conversion in antigenic variation in African trypanosomes. Mol Cell Biol **19,** 5839-5846.

Robinson, N.P., McCulloch, R., Conway, C., Browitt, A., and Barry, J.D. (2002). Inactivation of Mre11 does not affect VSG gene duplication mediated by homologous recombination in Trypanosoma brucei. J Biol Chem **277,** 26185-26193.

Roditi, I., and Lehane, M.J. (2008). Interactions between trypanosomes and tsetse flies. Curr Opin Microbiol **11,** 345-351.

Roth, C., Bringaud, F., Layden, R.E., Baltz, T., and Eisen, H. (1989). Active late-appearing variable surface antigen genes in Trypanosoma

equiperdum are constructed entirely from pseudogenes. Proc Natl Acad Sci U S A **86,** 9375-9379.

Rudenko, G., McCulloch, R., Dirks-Mulder, A., and Borst, P. (1996). Telomere exchange can be an important mechanism of variant surface glycoprotein gene switching in Trypanosoma brucei. Mol Biochem Parasitol **80,** 65-75.

Saeij, J.P., Coller, S., Boyle, J.P., Jerome, M.E., White, M.W., and Boothroyd, J.C. (2007). *Toxoplasma* co-opts host gene expression by injection of a polymorphic kinase homologue. Nature **445,** 324-327.

Saeij, J.P., Boyle, J.P., Coller, S., Taylor, S., Sibley, L.D., Brooke-Powell, E.T., Ajioka, J.W., and Boothroyd, J.C. (2006). Polymorphic secreted kinases are key virulence factors in toxoplasmosis. Science **314,** 1780-1783.

Sasse, J. (1998). The development of genetic markers for the *T. brucei* genome University of Cambridge).

Scherf, A., and Mattei, D. (1992). Cloning and characterization of chromosome breakpoints of Plasmodium falciparum: breakage and new telomere formation occurs frequently and randomly in subtelomeric genes. Nucleic Acids Res **20,** 1491-1496.

Scherf, A., Petersen, C., Carter, R., Alano, P., Nelson, R., Aikawa, M., Mattei, D., da Silva, L.P., and Leech, J. (1992). Characterization of a Plasmodium falciparium mutant that has deleted the majority of the gametocyte-specific Pf11-1 locus. Mem Inst Oswaldo Cruz **87 Suppl 3,** 91-94.

Scott, A. (1995). Drug-resistance and genetic exchange in *Trypanosoma brucei* PhD (University of Glasgow).

Scott, A.G., Tait, A., and Turner, C.M. (1996). Characterisation of cloned lines of Trypanosoma brucei expressing stable resistance to MelCy and suramin. Acta Trop **60,** 251-262.

Seigneur, M., Bidnenko, V., Ehrlich, S.D., and Michel, B. (1998). RuvAB acts at arrested replication forks. Cell **95,** 419-430.

Shahi, S.K., Krauth-Siegel, R.L., and Clayton, C.E. (2002). Overexpression of the putative thiol conjugate transporter TbMRPA causes melarsoprol resistance in Trypanosoma brucei. Mol Microbiol **43,** 1129-1138.

Shapiro, T.A., and Englund, P.T. (1990). Selective cleavage of kinetoplast DNA minicircles promoted by antitrypanosomal drugs. Proc Natl Acad Sci U S A **87,** 950-954.

Shea, C., Glass, D.J., Parangi, S., and Van der Ploeg, L.H. (1986). Variant surface glycoprotein gene expression site switches in Trypanosoma brucei. J Biol Chem **261,** 6056-6063.

Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. Nat Biotechnol **26,** 1135-1145.

Shirley, M.W., and Harvey, D.A. (2000). A genetic linkage map of the apicomplexan protozoan parasite *Eimeria tenella*. Genome Res **10,** 1587-1593.

Shirley, M.W., Biggs, B.A., Forsyth, K.P., Brown, H.J., Thompson, J.K., Brown, G.V., and Kemp, D.J. (1990). Chromosome 9 from independent clones and isolates of Plasmodium falciparum undergoes subtelomeric deletions with similar breakpoints in vitro. Mol Biochem Parasitol **40,** 137-145.

Shlomai, J. (2004). The structure and replication of kinetoplast DNA. Curr Mol Med **4,** 623-647.

Sibley, L.D., LeBlanc, A.J., Pfefferkorn, E.R., and Boothroyd, J.C. (1992). Generation of a restriction fragment length polymorphism linkage map for *Toxoplasma gondii*. Genetics **132,** 1003-1015.

Simarro, P.P., Jannin, J., and Cattand, P. (2008). Eliminating human African trypanosomiasis: where do we stand and what comes next? PLoS Med **5,** e55.

Simo, G., Cuny, G., Demonchy, R., and Herder, S. (2008a). Trypanosoma brucei gambiense: study of population genetic structure of Central African stocks using amplified fragment length polymorphism (AFLP). Exp Parasitol **118,** 172-180.

Simo, G., Asonganyi, T., Nkinin, S.W., Njiokou, F., and Herder, S. (2006). High prevalence of Trypanosoma brucei gambiense group 1 in pigs from the Fontem sleeping sickness focus in Cameroon. Vet Parasitol **139,** 57-66.

Simo, G., Njiokou, F., Mbida Mbida, J.A., Njitchouang, G.R., Herder, S., Asonganyi, T., and Cuny, G. (2008b). Tsetse fly host preference from sleeping sickness foci in Cameroon: epidemiological implications. Infect Genet Evol **8,** 34-39.

Simpson, L. (1972). The kinetoplast of the hemoflagellates. International Review of Cytology **32,** 137-207.

Smith, C.E., Llorente, B., and Symington, L.S. (2007). Template switching during break-induced replication. Nature **447,** 102-105.

Smith, D.H., Pepin, J., and Stich, A.H. (1998). Human African trypanosomiasis: an emerging public health crisis. Br Med Bull **54,** 341-355.

Sones, K.R., Njogu, A.R., and Holmes, P.H. (1988). Assessment of sensitivity of Trypanosoma congolense to isometamidium chloride: a comparison of tests using cattle and mice. Acta Trop **45,** 153-164.

Spits, C., Mateizel, I., Geens, M., Mertzanidou, A., Staessen, C., Vandeskelde, Y., Van der Elst, J., Liebaers, I., and Sermon, K. (2008). Recurrent chromosomal abnormalities in human embryonic stem cells. Nat Biotechnol **26,** 1361-1363.

Stanghellini, A., and Josenando, T. (2001). The situation of sleeping sickness in Angola: a calamity. Trop Med Int Health **6,** 330-334.

Stark, J.M., and Jasin, M. (2003). Extensive loss of heterozygosity is suppressed during homologous repair of chromosomal breaks. Mol Cell Biol **23,** 733-743.

Sternberg, J., and Tait, A. (1990). Genetic exchange in African trypanosomes. Trends Genet **6,** 317-322.

Sternberg, J., Turner, C.M., Wells, J.M., Ranford-Cartwright, L.C., Le Page, R.W., and Tait, A. (1989). Gene exchange in African trypanosomes: frequency and allelic segregation. Mol Biochem Parasitol **34,** 269-279.

Sternberg, J., Tait, A., Haley, S., Wells, J.M., Le Page, R.W., Schweizer, J., and Jenni, L. (1988). Gene exchange in African trypanosomes: characterisation of a new hybrid genotype. Mol Biochem Parasitol **27,** 191-200.

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 (Pt 2),** 181-186.

Stewart, M.L., Krishna, S., Burchmore, R.J., Brun, R., de Koning, H.P., Boykin, D.W., Tidwell, R.R., Hall, J.E., and Barrett, M.P. (2005).

Detection of arsenical drug resistance in Trypanosoma brucei with a simple fluorescence test. Lancet **366,** 486-487.

Stich, A., Abel, P.M., and Krishna, S. (2002). Human African trypanosomiasis. Bmj **325,** 203-206.

Su, C., Howe, D.K., Dubey, J.P., Ajioka, J.W., and Sibley, L.D. (2002). Identification of quantitative trait loci controlling acute virulence in *Toxoplasma gondii*. Proc Natl Acad Sci U S A **99,** 10753-10758.

Su, X., Kirkman, L.A., Fujioka, H., and Wellems, T.E. (1997). Complex polymorphisms in an approximately 330 kDa protein are linked to chloroquine-resistant *P. falciparum* in Southeast Asia and Africa. Cell **91,** 593-603.

Su, X., Ferdig, M.T., Huang, Y., Huynh, C.Q., Liu, A., You, J., Wootton, J.C., and Wellems, T.E. (1999). A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. Science **286,** 1351-1353.

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

Tait, A., Buchanan, N., Hide, G., and Turner, C.M. (1996). Self-fertilisation in Trypanosoma brucei. Mol Biochem Parasitol **76,** 31-42.

Tait, A., Masiga, D., Ouma, J., MacLeod, A., Sasse, J., Melville, S., Lindegard, G., McIntosh, A., and Turner, M. (2002). Genetic analysis of phenotype in Trypanosoma brucei: a classical approach to potentially complex traits. Philos Trans R Soc Lond B Biol Sci **357,** 89-99.

Taylor, S., Barragan, A., Su, C., Fux, B., Fentress, S.J., Tang, K., Beatty, W.L., Hajj, H.E., Jerome, M., Behnke, M.S., White, M., Wootton, J.C., and Sibley, L.D. (2006). A secreted serine-threonine kinase determines virulence in the eukaryotic pathogen Toxoplasma gondii. Science **314,** 1776-1780.

Taylor, S.D.A. (2004). Genetic analysis of drug resistance in *Trypanosoma brucei* PhD thesis (University of Glasgow).

Thiagalingam, S., Foy, R.L., Cheng, K.H., Lee, H.J., Thiagalingam, A., and Ponte, J.F. (2002). Loss of heterozygosity as a predictor to map tumor suppressor genes in cancer: molecular basis of its occurrence. Curr Opin Oncol **14,** 65-72.

Thon, G., Baltz, T., Giroud, C., and Eisen, H. (1990). Trypanosome variable surface glycoproteins: composite genes and order of expression. Genes Dev **4,** 1374-1383.

Thuita, J.K., Karanja, S.M., Wenzler, T., Mdachi, R.E., Ngotho, J.M., Kagira, J.M., Tidwell, R., and Brun, R. (2008). Efficacy of the diamidine DB75 and its prodrug DB289, against murine models of human African trypanosomiasis. Acta Trop **108,** 6-10.

Turner, C.M. (1990). The use of experimental artefacts in African trypanosome research. Parasitol Today **6,** 14-17.

Turner, C.M. (1997). The rate of antigenic variation in fly-transmitted and syringe-passaged infections of Trypanosoma brucei. FEMS Microbiol Lett **153,** 227-231.

Turner, C.M., and Barry, J.D. (1989). High frequency of antigenic variation in Trypanosoma brucei rhodesiense infections. Parasitology **99 Pt 1,** 67-75.

Turner, C.M., Aslam, N., and Dye, C. (1995). Replication, differentiation, growth and the virulence of Trypanosoma brucei infections. Parasitology **111 (Pt 3),** 289-300.

Turner, C.M., Barry, J.D., Maudlin, I., and Vickerman, K. (1988). An estimate of the size of the metacyclic variable antigen repertoire of Trypanosoma brucei rhodesiense. Parasitology **97 (Pt 2),** 269-276.

Turner, C.M., Sternberg, J., Buchanan, N., Smith, E., Hide, G., and Tait, A. (1990). Evidence that the mechanism of gene exchange in *Trypanosoma brucei* involves meiosis and syngamy. Parasitology **101 Pt 3,** 377-386.

Turner, C.M., McLellan, S., Lindergard, L.A., Bisoni, L., Tait, A., and MacLeod, A. (2004). Human infectivity trait in Trypanosoma brucei: stability, heritability and relationship to sra expression. Parasitology **129,** 445-454.

Ulbert, S., Chaves, I., and Borst, P. (2002). Expression site activation in Trypanosoma brucei with three marked variant surface glycoprotein gene expression sites. Mol Biochem Parasitol **120,** 225-235.

University of Glasgow. (2008). Trypanosome Genetic Mapping Database. http://www.gla.ac.uk/centres/wcmp/research/macleod/trypanosomegeneticmappingdatabase/#d.en.38283 1st Dec, 2009.

Vaidya, A.B., Muratova, O., Guinet, F., Keister, D., Wellems, T.E., and Kaslow, D.C. (1995). A genetic locus on Plasmodium falciparum chromosome 12 linked to a defect in mosquito-infectivity and male gametogenesis. Mol Biochem Parasitol **69,** 65-71.

Van der Ploeg, L.H., Cornelissen, A.W., Barry, J.D., and Borst, P. (1984). Chromosomes of kinetoplastida. Embo J **3,** 3109-3115.

Van der Ploeg, L.H., Cornelissen, A.W., Michels, P.A., Borst, P. (1984). Chromosome rearrangements in Trypanosoma brucei. Cell **39,** 213-221.

van Deursen, F.J., Shahi, S.K., Turner, C.M., Hartmann, C., Guerra-Giraldez, C., Matthews, K.R., and Clayton, C.E. (2001). Characterisation of the growth and differentiation in vivo and in vitro-of bloodstream-form Trypanosoma brucei strain TREU 927. Mol Biochem Parasitol **112,** 163-171.

Vansterkenburg, E.L., Coppens, I., Wilting, J., Bos, O.J., Fischer, M.J., Janssen, L.H., and Opperdoes, F.R. (1993). The uptake of the trypanocidal drug suramin in combination with low-density lipoproteins by Trypanosoma brucei and its possible mode of action. Acta Trop **54,** 237-250.

Vickerman, K. (1965). Polymorphism and mitochondrial activity in sleeping sickness trypanosomes. Nature **208,** 762-766.

Vickerman, K. (1969). On the surface coat and flagellar adhesion in trypanosomes. J Cell Sci **5,** 163-193.

Vickerman, K. (1985). Developmental cycles and biology of pathogenic trypanosomes. Br Med Bull **41,** 105-114.

Vickerman, K., Tetley, L., Hendry, K.A., and Turner, C.M. (1988). Biology of African trypanosomes in the tsetse fly. Biol Cell **64,** 109-119.

Vignal, A., Milan, D., SanCristobal, M., and Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. Genet Sel Evol **34,** 275-305.

von Jancso, N., and von Jancso, H. (1934). The role of the natural defence forces in the evolution of the drug-resistance of trypansomes I. A method for the exclusion of the natural defence mechanisms from chemotherapeutic processes. Annals of Tropical Medicine and Parasitology **28,** 419-438.

von Jancso, N., and von Jancso, H. (1935). The role of the natural defence forces in the evolution of the drug-resistance of trypanosomes II. The rapid production of Germanin-fast T. brucei strains in animals with paralysed defence. Annals of Tropical Medicine and Parasitology **29,** 95-109.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and et al. (1995). AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res **23,** 4407-4414.

Walker-Jonah, A., Dolan, S.A., Gwadz, R.W., Panton, L.J., and Wellems, T.E. (1992). An RFLP map of the Plasmodium falciparum genome, recombination rates and favored linkage groups in a genetic cross. Mol Biochem Parasitol **51,** 313-320.

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,** 1661-1666.

Wang, C.C. (1995). Molecular mechanisms and therapeutic approaches to the treatment of African trypanosomiasis. Annu Rev Pharmacol Toxicol **35,** 93-127.

Wang, P., Read, M., Sims, P.F., and Hyde, J.E. (1997). Sulfadoxine resistance in the human malaria parasite Plasmodium falciparum is determined by mutations in dihydropteroate synthetase and an additional factor associated with folate utilization. Mol Microbiol **23,** 979-986.

Welburn, S.C., Maudlin, I., and Milligan, P.J. (1995). Trypanozoon: infectivity to humans is linked to reduced transmissibility in tsetse. I. Comparison of human serum-resistant and human serum-sensitive field isolates. Exp Parasitol **81,** 404-408.

Wellems, T.E., Walker-Jonah, A., and Panton, L.J. (1991). Genetic mapping of the chloroquine-resistance locus on Plasmodium falciparum chromosome 7. Proc Natl Acad Sci U S A **88,** 3382-3386.

Wellems, T.E., Panton, L.J., Gluzman, I.Y., do Rosario, V.E., Gwadz, R.W., Walker-Jonah, A., and Krogstad, D.J. (1990). Chloroquine resistance not linked to mdr-like genes in a Plasmodium falciparum cross. Nature **345,** 253-255.

Wells, J.M., Prospero, T.D., Jenni, L., and Le Page, R.W. (1987). DNA contents and molecular karyotypes of hybrid Trypanosoma brucei. Mol Biochem Parasitol **24,** 103-116.

WHO. (1998). WHO Expert Committee on Control and Surveillance of African Trypanosomiasis. WHO Technical Report Series, No. 881. (Geneva: World Health Organization).

WHO. (2004a). Control of Human African Trypanosomiasis. Fifty-seventh world health assembly (A57/6). (Geneva: World Health Organization).

WHO. (2004b). Changing history. The world health report 2004. (Geneva: World Health Organization).

WHO. (2006a). Human African trypanosomiasis (sleeping sickness): epidemiological update. Weekly Epidemiological Record 81. (Geneva: World Health Organization).

WHO. (2006b). WHO fact sheet no. 259 African trypanosomiasis (sleeping sickness). Geneva: World Health Organization).

Wickstead, B., Ersfeld, K., and Gull, K. (2004). The small chromosomes of Trypanosoma brucei involved in antigenic variation are constructed around repetitive palindromes. Genome Res **14,** 1014-1024.

Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res **18,** 6531-6535.

Williams, R.O., Young, J.R., and Majiwa, P.A. (1982). Genomic environment of T. brucei VSG genes: presence of a minichromosome. Nature **299,** 417-421.

Williamson, J., and Rollo, I.M. (1959). Drug resistance in trapanososmes: cross-resitance analyses. Br J Pharmacol Chemother **14,** 423-430.

Wincker, P., Ravel, C., Blaineau, C., Pages, M., Jauffret, Y., Dedet, J.P., and Bastien, P. (1996). The Leishmania genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. Nucleic Acids Res **24,** 1688-1694.