Mosaic *VSG*s in *Trypanosoma brucei* antigenic variation

James Peter John Hall

B.A. M.Res.

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

Wellcome Trust Centre for Molecular Parasitology
Institute of Infection, Immunity and Inflammation
College of Medical Veterinary and Life Sciences
University of Glasgow

November 2012

Abstract

Many parasites of mammals avoid elimination by varying their exposed antigens. African trypanosomes—deadly parasites of humans and livestock in tropical Africa—possess a comprehensive system of antigenic variation (AV). *Trypanosoma brucei* undergo frequent, stochastic changes to their variant surface glycoprotein (VSG) coats, and therefore a developing immune response will be only partially effective against the trypanosome population as some trypanosomes will have already switched to a different VSG coat. The source of VSG variability is an archive of ~2000, mostly pseudogenic, silent *VSG* genes, of which only one is expressed. *VSG* genes can also be segmentally recombined: 'mosaic' *VSG*s, constructed from more than one silent *VSG* donor, allow both the reparation of pseudogenes and potentially generation of additional *VSG* variability. The aim of this research was to investigate the patterns of segmental *VSG* gene conversion in *T. brucei*, and assess its contribution to AV. Multiple, longitudinal samples were taken from chronic infections to follow the course of AV, and *VSG* cDNA sequences were analysed, building a detailed portrait of *VSG* expression across infection. *VSG* variability during an infection was extensive, and segmental gene conversion was found to be a frequent occurrence from approximately week three. Two main patterns were found: (i) expressed *VSG*s readily acquired a 3' end different from their silent copy, a pattern that probably represents the 3' boundary of gene conversion occurring within the coding sequence; (ii) expressed *VSG*s often appeared in sets of related 'mosaics', whereby more than one donor gene had contributed to the putative epitope-encoding part of the *VSG*. To test whether varying donor contributions represents an additional source of antigenic variability available to trypanosomes, a set of five mosaic *VSG*s retrieved from a single infection was expressed in non-switching trypanosomes and used to raise antibody responses. Indirect immuno-fluorescence, complement-mediated lysis, and agglutination assays using both polyclonal and monoclonal antibodies showed that although 4/5 mosaics were cross-reactive, one variant was completely antigenically distinct. Segmental gene conversion was therefore found to be both prominent in chronic African trypanosome antigenic variation, and capable of bringing antigenic novelty to an infection, with important consequences for the dynamics of AV, and the nature of selection pressure on the silent *VSG* archive.

Table of Contents

List of Tables

List of Figures

Acknowledgements

My thanks go to:

Dave, for his support, advice and guidance throughout, and for giving me space to develop my interests whilst making sure I stayed on course

My family, particularly my parents, for always being there

All the lab members, past and present, for the pleasure that has been working with you. Special thanks go to Lucio for the tahini and the pseudogenes, Richard for the coffee and a critical eye, Lindsey for perspectives from the archive, and Huanhuan for all the questions and helping with the samples. Thanks also to Dan and Erida, for invaluable advice and ideas at various points

The technical and administrative staff for their help and patience, particularly Alex, Ingrid, and the staff at the JRF

Friends and flatmates, especially Storm, for the food and the stories, Lanark, for the walks and unending enthusiasm, Kanye and Ira for company, and Eddie, for the brilliant illustrations and exciting artistic collaborations

Bill, Darren and Olwyn, for organizing an excellent Ph.D. programme and giving me the opportunity to undertake this work, and the other students of the programme for being an interested, interesting, interdisciplinary community

My examiners Mark Carrington (University of Cambridge) and Lisa Ranford-Cartwright (University of Glasgow) for an enjoyable *viva*, constructive feedback, and many new ideas

The Wellcome Trust, for wonderful workshops and meetings, and the financial support that made all this possible.

This thesis is dedicated to my Grandfather, John.

Author’s declaration

I declare that this thesis and the results presented within are entirely my own work except where otherwise stated. No part of this thesis has been previously submitted for a degree at any university.

James Hall

Definitions and abbreviations

aa	amino acid
AnTat	Antwerp Trypanosome antigenic type
AV	antigenic variation
BES	bloodstream ES
bp	base pair
BSA	bovine serum albumin
cDNA	complementary DNA
CTD	C-terminal domain
DAPI	4', 6-diamidino-2-phenylindole
dH_2O	deionized water
DIC	differential interference contrast
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dsDNA	double-stranded DNA
EDTA	ethylenediaminetetraacetic acid
ES	expression site
ESAG	expression site associated gene
ESB	expression site body
FITC	fluorescein isothiocyanate
gDNA	genomic DNA
GPI	glycosylphosphatidylinositol
GPP	guinea pig plasma
GUTat	Glasgow University Trypanosome antigenic type
h	hour
HAT	human African trypanosomiasis
HAT	(medium) hypoxanthine-aminopterin-thymidine
HMI	Hirumi's modified Iscove's medium
HR	homologous recombination
kbp	10^3 bp
kDa	kilodalton
ILTat	International Laboratory (Nairobi, Kenya) Trypanosome antigenic type
mAb	monoclonal antibody
MES	metacyclic ES

MITat	Molteno Institute Trypanosome antigenic type
min	minute
mRNA	messenger RNA
nt	nucleotide
NTD	N-terminal domain
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PE	phycoerythrin
Pol I	RNA polymerase I
RNA	ribonucleic acid
rRNA	ribosomal RNA
RT-PCR	reverse transcriptase polymerase chain reaction
SDS-PAGE	sodium dodecyl sulphate poly-acrylamide gel electrophoresis
SRA	serum resistance associated
ssDNA	single-stranded DNA
TDB	trypanosome dilution buffer
TREU	Trypanosomiasis Research Edinburgh University
UTR	untranslated region
VSG	variant surface glycoprotein
VSG	variant surface glycoprotein gene
WHO	World Health Organization

Chapter 1

Introduction

1 Introduction

1.1 Introduction to antigenic variation

1.1.1 Antigenic variation is a means of coping with a changing immunological environment

To a parasite, a mammal is essentially a stable, mild, nutrient-rich habitat. An abundant supply of chemical energy flows through its blood vessels, dedicated organs degrade and excrete toxic compounds, and numerous homeostatic systems ensure that extremes of pH, temperature, and osmotic pressure are controlled. Over the course of evolutionary history, many different kinds of organism have tapped these resources, benefiting from a close, exploitative relationship with mammals. From miniscule viruses to giant worms, from bacteria to fungi and protozoa: many have succeeded. But success is unstable: while mammals with adaptations allowing them to resist or eliminate parasites often have advantages over susceptible peers, parasites able to withstand or evade such adaptations are likewise selected. This iterative process, likened to an arms race, has shaped the course of evolution, resulting in an extensive and complex mammalian immune system on one hand (Trowsdale & Parham, 2004), and intricate and elegant parasite strategies of outmanoeuvring it on the other (Finlay & McFadden, 2006) .

Mammalian adaptive immune responses have staggering potential for diversity (for example, antibody diversity is estimated at ~10^{10} within a single host (Glanville *et al.*, 2009)), target specific parasite antigens, and, in evolutionary terms, develop very rapidly. Antigenic variation is the name given to a widely used parasite strategy of coping with the host's adaptive immunity (Deitsch *et al.*, 2009). Mechanisms of antigenic variation exploit the fact that an adaptive immune response against a particular newly encountered parasite antigen takes several days to reach its peak. This lag gives time for the antigens of some of the parasites to change. While the bulk of the parasite population are destroyed by that specific immune response, those parasites that have changed antigens are not. The antigenically different parasites can proliferate to replenish the parasite population, and the whole process can repeat again and again as specific immune responses develop against successive antigens. As a survival

strategy, antigenic variation relies on the expression of antigenic diversity by the parasite population to exceed the rate at which the host develops efficacious immune responses against those antigens.

1.1.2 True antigenic variation is a distinct survival mechanism

Where does antigenic variability come from? Compared with general mutation of antigens caused by background cellular processes, true antigenic variation is a system that has evolved under pressure from adaptive immunity to present an ever-changing phenotype from a clonal population. We might infer that such selection has occurred by witnessing rapid change to the expressed antigen, or by identifying distinct resources and machinery specifically associated with changing antigen structure (Caporale, 2003). A clear case is the presence in the genome of multiple different antigen genes: an 'archive' of silent genes (Deitsch *et al.*, 2009). An archive is an in-built source of antigenic diversity: switching exclusive expression between archival copies, by transcriptional or recombinatorial mechanisms, changes the antigen expressed. The potential for variation can be expanded further through combinatorial mechanisms: if segments of different archival genes are able to join in various ways to create antigenically distinct products, the potential antigenic diversity of a small archive could be greatly increased (Zhuang *et al.*, 2007).

Other mechanisms exist for specifically expressing antigenic variability. Accumulating point mutations in an antigen gene is a less efficient process, but viruses, with their smaller genomes, depend on point mutations to generate the variability required for persistence (Frost *et al.*, 2005; Wolinsky *et al.*, 1996). In the case of HIV gp120, functionally-unconstrained 'hypervariable' loops sterically hinder access to invariant parts of the molecule whilst being able themselves to accumulate diversifying point mutations by the activity of an error-prone polymerase (Sodroski *et al.*, 1998). Variability brought about by differential enzyme activation—*Neisseria* lipooligosaccharide structures for example (Kahler & Stephens, 1998)—could also constitute a form of antigenic variation (Moxon *et al.*, 2006; van der Woude & Bäumler, 2004). A completely inclusive definition of antigenic variation might even stretch to include mechanisms like influenzavirus antigenic drift (Carrat & Flahault, 2007)

(essentially antigenic variation across a population of hosts) and the 'controlled chaos' of *Schistosoma* mucin polymorphisms (Roger *et al.*, 2008).

Mechanism		Species	Antigen(s)	Approx. # silent elements	References
A silent archive	Switching by recombination	African trypanosomes	VSG	1500 §	Barry & McCulloch, 2001; Marcello & Barry, 2007a
		Borrelia burgdorferi	VlsE	15 §	Coutte *et al.*, 2009
		Borrelia hermsii	Vsp/Vlp	21 / 38	Dai *et al.*, 2006
		Anaplasma marginale	MSP2	6 §	Palmer *et al.*, 2006
			MSP3	7 §	
		Neisseria gonorrhoeae	Pilin (PilE)	19 §	Hill & Davies, 2009
		Treponema pallidum	TprK	50 §	Centurion-Lara *et al.*, 2004
		Mycoplasma genitalium	MgpB, MgpC	9 §	Iverson-Cabral *et al.*, 2007
		Mycoplasma synoviae	MSPA, MSPB	30 §	Noormohammadi *et al.*, 2000
		Babesia bovis	VESA1	350 (*ves-1α*), 80 (*ves-1β*) §	Al-Khedery & Allred, 2006
	Switching by transcriptional activation	African trypanosomes	VSG	~15*	Young *et al.*, 2008b
		Plasmodium falciparum	PfEMP1	60	Kyes *et al.*, 2007
			RIFIN	>200	Kyes *et al.*, 1999
			STEVOR	30	Niang *et al.*, 2009
		Babesia bovis	VESA1	24*	Brayton *et al.*, 2007
		Giardia lamblia	VSP	200	Prucca & Lujan, 2009
Differential enzyme activation		Numerous bacterial species, including:			van der Woude & Bäumler, 2004
		Neisseria meningitidis	LOS		Kahler & Stephens, 1998
		Haemophilus influenzae	LPS		Moxon *et al.*, 2006
Point mutation		HIV	gp120		Frost *et al.*, 2005
		Influenza A**	HA, NA		Carrat & Flahault, 2007; Smith *et al.*, 2004
		Hepatitis C virus	E2		Bowen & Walker, 2005

Table 1.1 Mechanisms of antigenic variation. Protozoa are coloured green, eubacteria red, viruses blue. This table is intended to give an overview and is not exhaustive. *This number relates to the number of antigen expression sites (see section 1.4.3). **Unlike the other pathogens described here, influenza A undergoes biologically-relevant antigenic drift only on a population level. '§' indicates segmental combination of antigen genes occurs readily during infection.

Table 1.1 summarizes the primary mechanisms underlying antigenic variation, and some of the antigenically variable pathogens employing them. Features common to phylogenetically diverse pathogens have convergently evolved under similar immunological and environmental pressures, most notably the possession of a silent archive and its segmental recombination to increase its total antigenic potential.

Bearing these common strategies in mind, I shall now continue and describe the antigenically variable pathogen that was the primary subject of this study, the African trypanosome *Trypanosoma brucei*.

1.2 African trypanosomes – historical and medical aspects

1.2.1 African trypanosomes have been a focus of medical science for over a century

African trypanosomes became a principal subject of medical science in the early 1900s. For many centuries previously, the relationship between the bite of the tsetse fly and a debilitating disease of humans—'sleeping sickness'—and livestock—'nagana'—had been known (Steverding, 2008). Human African trypanosomiasis (HAT) manifests first as periodic fever, accompanied by headache, anaemia, general malaise and tell-tale swollen lymph glands. This 'first stage' of the disease can last for weeks to years, varying in severity and occasionally undergoing periods of asymptomatic infection, which can be prolonged (Jamonneau *et al.*, 2012; Songa *et al.*, 1991). The 'second stage' of the disease is accompanied by central nervous system involvement, causing severe headache and disturbance in circadian rhythm (thus 'sleeping sickness') and culminating in irreversible coma and death (Barrett *et al.*, 2003). In animals, trypanosomiasis is often a wasting disease associated with profound anaemia, although different species and breeds show varying susceptibility with indigenous species and breeds being more resistant to the effects of the disease (Mulla & Rickman, 1988; Taylor, 1998).

Historically tolerated by socio-ecological factors such as the size and separation of tribes and their management of the local environment (Ford, 1969), sleeping sickness exploded into a transcontinental epidemic in the late 19th century following the huge social, ecological and economic upheavals of European colonial expansion[1] (Lyons, 2008; Tilley, 2008). Governments seeking to control the epidemic established scientific commissions into the causes and control of sleeping sickness (Figure 1.1).

Figure 1.1 Photograph taken during the Third Royal Sleeping Sickness Commission, 1911 (Uganda/East Africa). Sir David Bruce, after whom *T. brucei* is named, is on the left of the photograph. © Wellcome Library, London. Permission to reproduce this image has been granted by the Wellcome Library.

The confluence of germ theory, innovative microbiological techniques, precedents for insect transmission set by malaria and filariasis, and industrial chemistry saw, within four years (de Raadt, 2005):

[1] In his diaries written only a few decades earlier the missionary explorer David Livingstone reported that the bite of the tsetse in (what is now) Zambia was harmless to man (Livingstone, 1857).

- the identification of the causative agent: in humans, *Trypanosoma brucei* subspecies *gambiense*, later, subspecies *rhodesiense* was also identified as human-infective,
- the insect host species responsible for transmission: the tsetse fly, *Glossina* spp.),
- a drug for its treatment, atoxyl.[2]

A recently renewed drive towards eradicating the disease completely saw reported incidence of HAT drop below 10,000 cases in the year 2009 (WHO, 2012), thanks largely to vigilant pro-active surveillance and improved diagnostic techniques (Cattand *et al.*, 2001; Molyneux *et al.*, 2010; Welburn *et al.*, 2009) (Figure 1.2). Patients diagnosed by serological tests and microscopical examination of blood and cerebrospinal fluid, are currently treated with the drugs suramin, pentamidine, melarsoprol or eflornithine, depending on whether parasites have penetrated the blood-brain barrier (Legros *et al.*, 2002). There have been no new drugs since the discovery that eflornithine was an effective trypanocide for *T. b. gambiense* in 1980 (Steverding, 2010). There are currently no vaccines: indeed, the antigenic variability of the trypanosome's dense surface coat and the immunosuppression that trypanosomes can induce make slim the possibility of finding an efficacious vaccine (La Greca & Magez, 2011).

[2] Amid a context of exoticism and imperial pride, the speed of such advances in the face of the epidemic's devastating effects inspired audiences in Europe and America, leading Britain's poet laureate to-be to write a novel on the subject (John Masefield's 'Multitude and Solitude', (1909), in which a playwright abandons the Arts for a career as a parasitologist) and Universal Pictures to release a dramatic feature film ('Nagana', starring Melvyn Douglas (1933)).

Figure 1.2 Sleeping sickness surveillance by Medecins Sans Frontieres (MSF) in Central African Republic. Between 2008 and 2011, prevalence in the Maitikoulou area fell from 5.9% to 0.5% (http://www.msf.org.uk/car.focus). © Carmen Barra/MSF. Permission to reproduce this image has been granted by MSF.

1.2.2 African trypanosome antigenic variation is an integral part of their life cycle and is diagnostically relevant

African trypanosome antigenic variation was recognized in the early 20th century, when, within a strain, some trypanosomes were found to survive in sera that lysed other trypanosomes of that strain (for example Laveran & Mesnil (1902), early 20th century research on trypanosome antigenic variation was reviewed by Soltys (1963)). This differential response, later ascribed to the presence of different 'variable antigen types' (VATs), was eventually taken to explain how trypanosomes survived in a host: the sequential growth of different VATs was consistent with the pattern of undulating parasitaemia, cyclical fever, and chronicity of infection (Barry & McCulloch, 2001). Over a hundred years after the initial observation of antigenic variation, it is clear that the African trypanosomes' is the most comprehensive system of antigenic variation identified to date (Deitsch *et al.*, 2009), an adaptation reflecting their exposed bloodstream habitat and precarious life cycle. Besides revealing biological survival mechanisms that are fascinating and elegant in their own right, studies of African trypanosome antigenic variation have directly led to improved diagnostic tools (such as the cheap, easy to use Card Agglutination Test for

Trypanosomiasis (CATT) (Inojosa *et al.*, 2006)), and have made African trypanosomes an excellent model for understanding this widespread parasite survival strategy in general.

1.3 The natural history of African trypanosomes

1.3.1 African trypanosomes probably evolved from parasites of arthropods

African salivarian trypanosomes and their closest relatives form Trypanosomatida, a phylum of parasitic/symbiotic kinetoplastids.[3] Into Trypanosomatida, rRNA and protein phylogenies have placed parasites of fish, rodents, plants, reptiles, birds, marsupials, leeches and insects alongside the medically important *Trypanosoma cruzi*, African salivarian trypanosomes, and *Leishmania* species (Simpson *et al.*, 2006). Trypanosomatids were most likely parasites of haematophagic insects before they were parasites of vertebrates (Hamilton *et al.*, 2004). During feeding, parasites that accidentally found themselves deposited in the blood were usually destroyed rapidly by the immune system. But parasites that were able to resist immunity, multiply, and build up numbers in vertebrate hosts would have benefited from increased exposure and transmission to other feeding insects. Opportunistic infections of immunosuppressed humans by monoxenous trypanosomatids (Dedet & Pratlong, 2000) and the isolation of a *Herpetomonas* trypanosomatid from dogs and rats in Egypt found to be closely related to parasites of hemipterid bugs (Podlipaev *et al.*, 2004) suggests that attempts at parasitism may occur often, although the development of successful long-term strategies by invaders is probably much rarer (Hamilton *et al.*, 2007).

1.3.2 There are seven major species of African trypanosome

The African salivarian trypanosomes comprise seven 'species', described in Table 1.2 and Figure 1.3, although it should be noted that the division between species and subspecies remains fluid: *T. equiperdum* and *T. evansi* could easily be

[3] Kinetoplastids themselves are a remarkable phylum of unicellular flagellate protists, exhibiting a number of intriguing and peculiar genetic and metabolic traits, including unprecedented modification of nucleotides (Borst & Sabatini, 2008) compartmentalization of glycolysis (Haanstra *et al.*, 2008), and a unique mitochondrial DNA architecture (Lukeš *et al.*, 2005)–the kinetoplast, after which they are named.

considered subspecies of *brucei*, for example (Lai *et al.*, 2008). Historically, African trypanosomes have been confined to the range of their insect host, the tsetse fly,[4] whose distribution extends between the Sahara and Kalahari deserts. Three species of animal-infective 'African salivarian' trypanosome (*T. equiperdum*, *T. evansi*, and American *T. vivax*) have adopted a simplified tsetse-independent means of transmission, as described below (section 1.3.3): this independence has allowed them to spread beyond the range of *Glossina* and leave Africa.

Subgenus	(Sub)species	Transmission	Human infective?	Distribution
Trypanozoon	*T. brucei brucei*	tsetse fly		Tropical Africa
	T. brucei rhodesiense	tsetse fly	acute	East Africa
	T. brucei gambiense	tsetse fly	chronic	West Africa
	T. equiperdum	contact (sexual)		Cosmopolitan
	T. evansi	tabanid fly (mechanical)	~	North Africa, Middle East, Asia, South America
Nannomonas	*T. congolense*	tsetse fly		Tropical Africa
	T. simiae	tsetse fly		Tropical Africa
	T. godfreyi	tsetse fly		Tropical Africa
Duttonella	*T. vivax*	tsetse fly		Tropical Africa
		tabanid fly (mechanical)		South America*

Table 1.2 Species of African trypanosome. Details were compiled from Gibson (2007). Note the details presented here are subject to change: some species divisions indicated are weak (for example, those within *Trypanozoon* may be demoted to subspecies), whereas it is likely that the phylogenetic diversity seen amongst *T. congolense* isolates represents different subspecies (see Figure 1.3). A rare case of human *T. evansi* infection in India reported in 2005 (indicated by a '~') was subsequently found to be associated with a deficiency in the patient's trypanolytic factor (Vanhollebeke *et al.*, 2006). *It is likely that *T. vivax* are transmitted mechanically in regions of Africa where tsetse are sparse (Cortez *et al.*, 2006).

Of the African trypanosomes, *Trypanosoma brucei* has been the most intensively studied, being as it is the only species capable of infecting humans. Three subspecies of *T. brucei* have been identified: the non-human infective *T. b. brucei*, the 'chronic' disease-causing West African *T. b. gambiense*, and the 'acute' disease-causing East African *T. b. rhodesiense*. As typing of *brucei* and

[4] *Glossina* species, of which there are about 20. Different species have distinctly preferred habitats, and tsetse flies can be broadly grouped into 'savannah', 'riverine' and 'forest' demes (Krafsur, 2008).

rhodesiense sspp. is defined by survival in human serum, a phenotype that is conferred by a single gene (Xong *et al.*, 1998), the boundary between these two 'subspecies' should not be considered strict (*gambiense* uses different mechanisms for human serum resistance and thus is more distinct (Capewell *et al.*, 2011)). Two other species have also received significant attention: *T. vivax*, which causes the most virulent animal infection; and *T. congolense*, which is the most geographically widespread trypanosome in Africa.

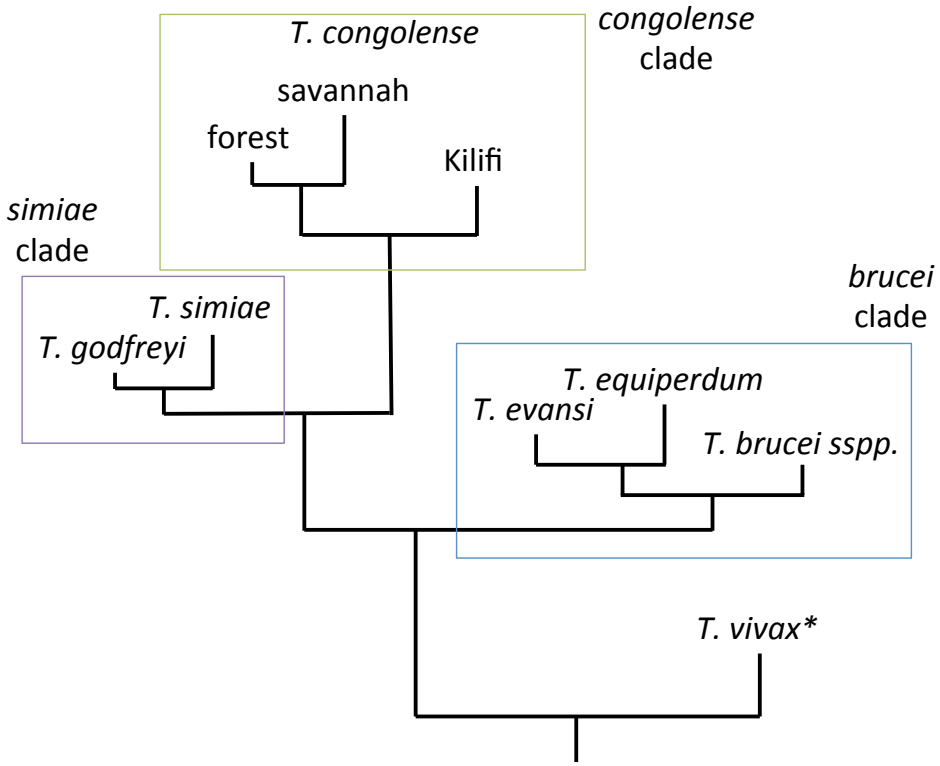


Figure 1.3 The evolutionary relationships of African trypanosomes. Diagram is intended to show key cladistic relationships only and is not to scale. Based on data in Gibson (2007), Hamilton et al. (2007) and Brun et al. (1998). *T. congolense* shows considerable phylogenetic diversity amongst different habitats, as indicated. **T. vivax* South American populations are closely related to West African *T. vivax* (Cortez *et al.*, 2006).

1.3.3 African trypanosomes have a complex life cycle involving two dissimilar hosts and numerous developmental stages

African trypanosomes are usually dixenous, alternately infecting mammal and insect during life-cycle progression.[5] As they travel between each host

[5] The exceptions to this are *T. equiperdum*, *T. evansi* and American *T. vivax*. It appears that *T. evansi* and *T. equiperdum* have diverged from *T. brucei* relatively recently (Lai *et al.*, 2008) and the similarity between West African and South American *T. vivax* isolates indicates that the South American strains have African origins (Cortez *et al.*, 2006); it can thus be assumed that the ancestral African trypanosome was dixenous between mammal and tsetse.

environment, trypanosomes undergo a number of morphological changes. Some of their forms are mitotically active, specialised for colonisation and proliferation of a specific environment, while others are transmission-ready cell-cycle arrested forms, pre-emptively adapted to survive as they move between different environments (Barry & McCulloch, 2001).

For *T. brucei*, infection of the mammal occurs when the tsetse fly takes its blood meal, depositing metacyclic-stage trypanosomes subcutaneously. The trypanosomes migrate to the draining lymph node and enter the vasculature, undergoing a morphological change to the 'long slender' bloodstream form. The long slender form is mitotically active and undergoes repeated cycles of binary fission with a doubling time of about six hours (Turner *et al.*, 1995). Inside the mammal, trypanosomes generally proliferate extracellularly in the blood, lymph and extravascular matrix, although they are capable of invading other tissues, including muscle and the central nervous system (CNS).[6] Besides the dynamic processes of immune killing and population recovery following antigenic variation, parasite numbers are restrained by a process of irreversible differentiation to a cell-cycle arrested 'short stumpy' form (see section 1.7.5). It is this short stumpy form, with a relatively short half-life of 24–72 hours (Seed *et al.*, 2003) that is capable of infecting a biting tsetse fly (Robertson, 1912). In the midgut of the tsetse, stumpy form trypanosomes develop into the proliferative procyclic form trypanosomes. Lumenal migration from the gut to the salivary glands is coupled to a series of developmental changes: a procyclic cell differentiates into the long mesocyclic form, which undergoes asymmetric division to yield a small daughter cell, the progenitor for the salivary gland-colonising proliferative epimastigote form (Vickerman, 1985). Few trypanosomes survive this migration (Oberle *et al.*, 2010). The salivary gland epimastigote population grows as a dense monolayer, and is the point at which trypanosomes can undergo sexual genetic exchange (Gibson *et al.*, 2008). Epimastigotes undergo asymmetric division to produce a trypomastigote daughter cell (Rotureau *et al.*, 2012), which develops to form the mammalian-infective metacyclic stage, and the cycle can begin again. The lifecycle is shown in Figure 1.4.

[6] Tissue invasion is *T. brucei* specific – neither *T. congolense* nor *T. vivax* leave the blood in large numbers, if at all (Ssenyonga & Adam, 1974; Barry, 1986).

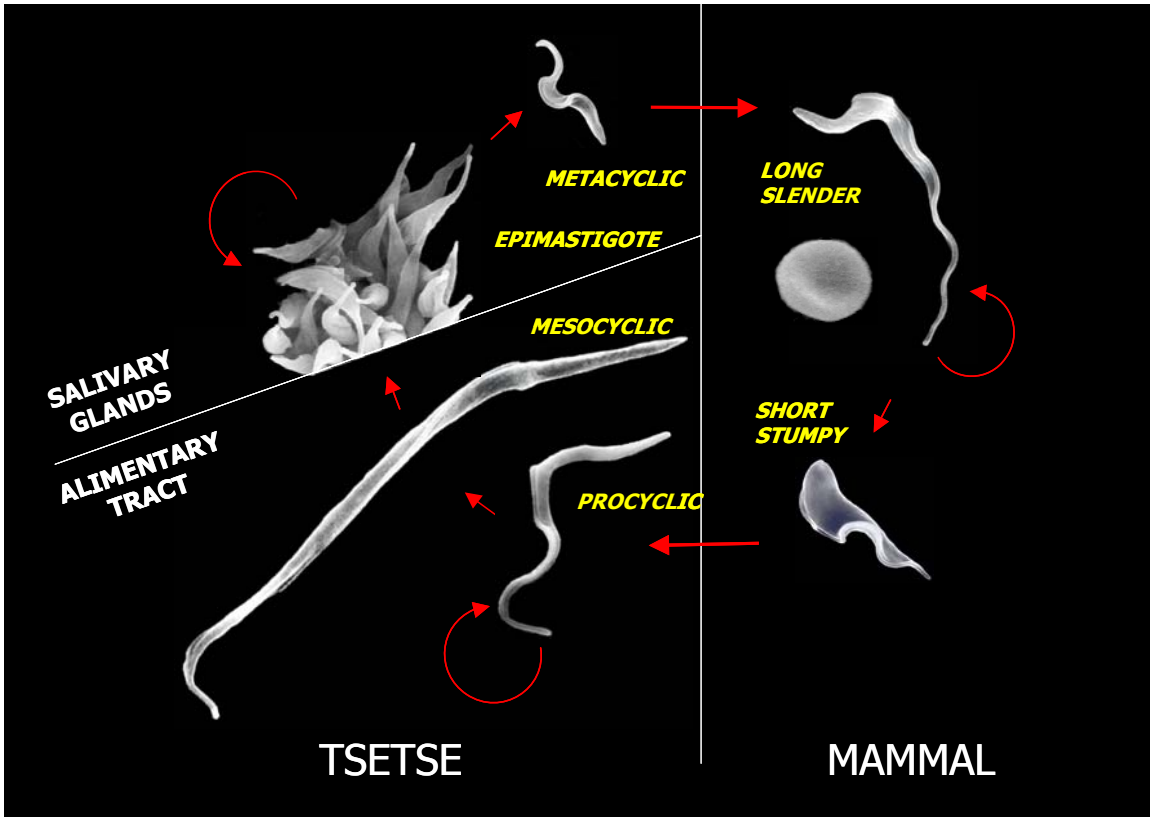


Figure 1.4 *Trypanosoma brucei* lifecycle. Straight red lines indicate the development between different morphological forms, curved lines indicate the multiplication of the proliferative forms. Transient intermediate forms that are exhibited during development between the key stages are not shown. White lines indicate barriers between different environments (tsetse salivary gland, mammalian bloodstream, tsetse alimentary tract). From Barry & McCulloch (2001).

Compared with *T. brucei*, there are some differences in the lifecycle between the species. *T. congolense* does not colonise the salivary glands, rather, the epimastigote stage directly colonises the mouthparts (Peacock *et al.*, 2012). *T. vivax* has a simpler tsetse stage, whereby bloodstream form trypanosomes in the tsetse's meal transform directly into the epimastigote form in the fly's mouthpieces, without any multiplicative stage elsewhere (Aksoy *et al.*, 2003). For those species and strains where tsetse infection is not advantageous—sexually transmitted *T. equiperdum*, most strains of mechanically transmitted *T. evansi* and American *T. vivax* (Uilenberg, 1998), and indeed some laboratory *T. brucei* lines subject to long periods of rapid syringe passage or *in vitro* bloodstream form culture (Turner, 1990)—the bloodstream forms of the parasite

become increasingly 'monomorphic', with differentiation to the non-proliferative tsetse-infective form disfavoured.[7]

1.3.4 Antigenic variation is favoured by the African trypanosome life cycle and is mediated by a glycoprotein surface coat

Transmission by haematophagic arthropod vectors provides powerful selection for mechanisms that prolong parasitaemia, since it is blood-borne parasites that are best placed to infect the vector and spread to new hosts (Barbour & Restrepo, 2000). Many arthropod-transmitted pathogens persist by adopting a strategy of concealment and manipulation: by invading and reprogramming host cells they can escape the brunt of extracellular immune effectors such as antibodies. This is the case for the other mammalian-infective trypanosomatids, *T. cruzi* and *Leishmania* spp.; the Apicomplexans *Plasmodium*, *Babesia* and *Theileria*; and eubacterial *Rickettsia* species, such as *Anaplasma* and *Ehrlichia*.

The African salivarian trypanosomes, on the other hand, spend their mammalian phase extracellularly. Immediate protection against innate immunity is achieved by a uniform glycoprotein 'coat' that densely covers the entire surface of the bloodstream form trypanosome, made from approximately 5.5×10^6 identical dimers of the African trypanosome variable antigen: variant surface glycoprotein (VSG) (Jackson *et al.*, 1985). The VSG surface coat is critical for bloodstream survival: without it, parasites are killed by naive serum (Ferrante & Allison, 1983; Mosser & Roberts, 1982). VSG coats are present on both the long slender and the short stumpy bloodstream forms, and are also pre-emptively expressed by metacyclic stage trypanosomes (Vickerman, 1985). The VSG coat is lost when the short stumpy form differentiates to the procyclic form in the tsetse midgut, to be replaced by another glycoconjugate, procyclin (Roditi *et al.*, 1989).

The VSG coat is the physical form of the trypanosome variable antigen type and its guise whilst in the mammal (Cross, 1975). It is changes to the VSG coat—and serial adaptive immune responses against it—that are the basis of antigenic variation in African trypanosomes. The VSG surface coat is shown in Figure 1.5.

[7] 'Monomorphic' trypanosomes can be contrasted with 'pleomorphic' trypanosomes: those that readily undergo differentiation to the tsetse-infective stumpy form.

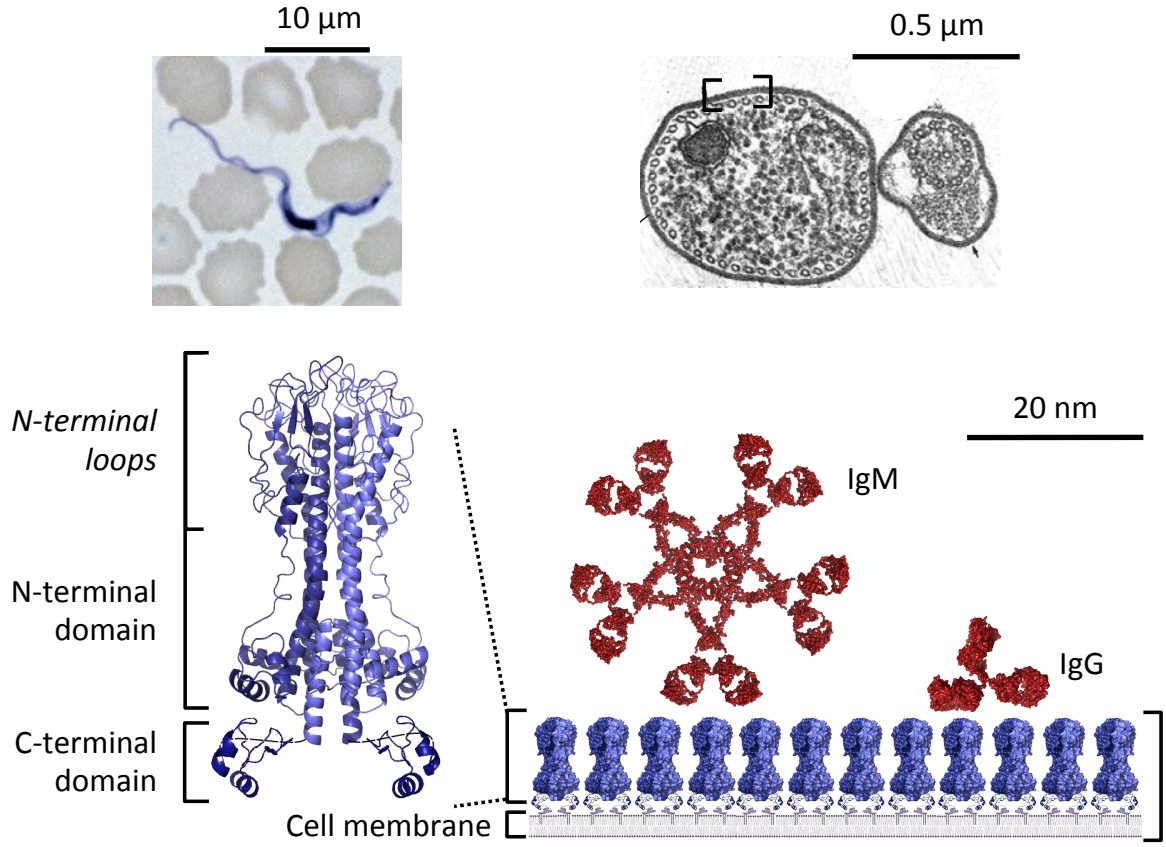


Figure 1.5 The VSG surface coat. (A) The VSG coat is expressed by bloodstream form trypanosomes. Image shows a long slender trypanosome surrounded by red blood cells © Wellcome Library, London. Permission to reproduce this image has been granted by the Wellcome Library. (B) The surface coat can be clearly seen under transmission electron micrography as a dense layer covering the surface of the parasite. Image provided by L. Tetley, University of Glasgow. (C) The assembled surface coat sterically hinders access of host immune effectors to the cell membrane or invariant surface proteins. (D) Each VSG dimer is an extended structure consisting of an N-terminal domain (NTD) and a C-terminal domain consisting of one or two subdomains. The N-terminal loops (indicated) represent the region of the glycoprotein exposed to immune effectors on an assembled surface coat. Images (C) and (D) were assembled using PDB structures 1vsg, 1xu6, 1rcj and 1igt, visualized using Pymol (Schrödinger, LLC), and were inspired by Engstler et al. (2007).

The structure, production and form of the VSG coat will be described first, in section 1.4. I shall then describe the diversity amongst different *VSG*s in the genome in section 1.5, and the variation in *VSG* expression over time will be covered in section 1.6. The focus will be on the *brucei* clade *VSG*s, since they are best characterised.

1.4 The structure of the VSG surface coat

1.4.1 Different VSGs must be antigenically distinct whilst performing the same function

The function of each VSG surface coat is to provide the infecting trypanosome population with a defence against the host immune system (Ferrante & Allison, 1983; Mosser & Roberts, 1982; Ziegelbauer & Overath, 1993). The dual nature of mammalian immunity—non-specific immediate innate immunity, and specific adaptive immunity—places two opposing demands on the population of possible VSG molecules:

a) *Conservation of structure between VSGs.* Different VSG coats require a structure that is able to consistently obstruct immune effectors from accessing the cell membrane or invariant surface antigens.
b) *Variation of structure between VSGs.* Epitope structure must vary sufficiently between different VSG coats to enable antigenic variation.

These demands have been met by a conserved tertiary structure that can be formed by hugely divergent amino acid sequences (Blum *et al.*, 1993; Carrington *et al.*, 1991; Marcello & Barry, 2007a). This tertiary structure gives an extended form to VSG, resulting in a deep, dense coat on the parasite surface (Vickerman, 1969).

1.4.2 VSG structure is defined by long helices adorned with loops

Each VSG unit in the closely packed surface coat is a homodimer, comprising two single-chain polypeptides (Auffret & Turner, 1981). Protease digestion of *T. brucei* VSGs revealed that each VSG monomer consisted of two domains connected by a protease-sensitive 'hinge' region (Johnson & Cross, 1979): a larger N-terminal domain (NTD; in the region of 380 amino acids in length) connected to a C-terminal domain (CTD; itself composed of one or two subdomains each approximately 40 amino acids in length) (Carrington *et al.*, 1991).

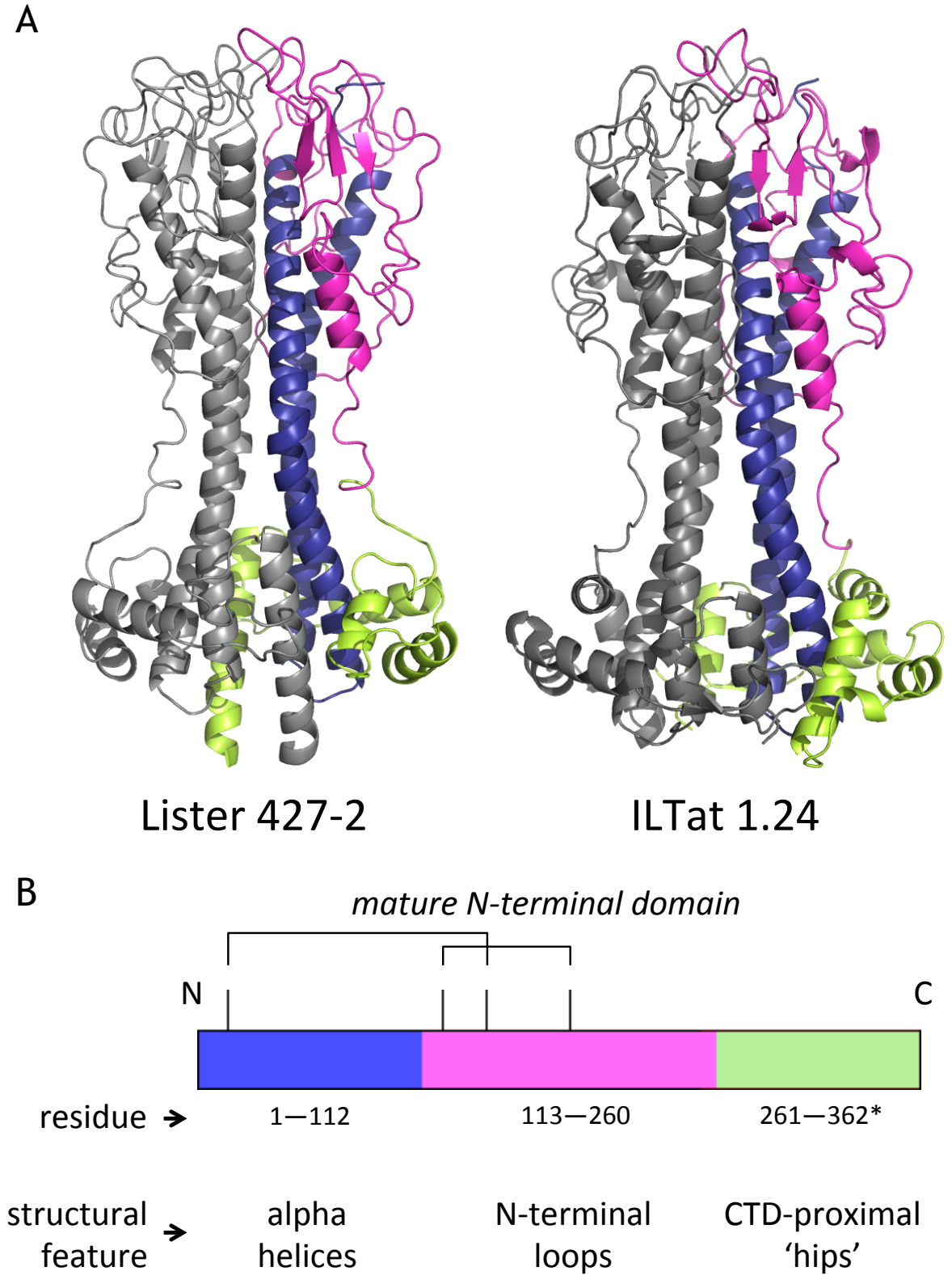


Figure 1.6 The VSG NTD fold. (A) The two structures of VSG N-terminal domains that have been solved—Lister 427-2 (left, PDB identifier 1vsg) and ILTat 1.24 (right, 2vsg)—are shown. For each dimer, one monomer has been coloured grey, the other according to (B). (B) A representation of the mature Lister 427-2 NTD amino acid chain from amino (N) to CTD-proximal (C) end. Bars protruding from the top of the diagram represent conserved cysteine residues, connected by brackets according to the disulphide bridges they form. The chain is divided into three pieces (indicated by 'residue'), and the residues corresponding with each piece are coloured in (A). The predominant structures formed by each piece are indicated ('structural feature'). *ILTat 1.24 NTD is four amino acids shorter than Lister 427-2, and thus only residues 261-358 are coloured green for that structure.

The two VSG NTDs whose structure has been resolved are less than 20% identical at the amino acid level, yet they form a strikingly similar tertiary fold, shown in Figure 1.6 (Blum *et al.*, 1993; Freymann *et al.*, 1990). Essentially, this fold is a framework of two long antiparallel alpha helices per monomer, beginning at the membrane-distal N-terminus and running for approximately 110 amino acids, standing perpendicular to the cell surface with a short loop or kink between them (Blum *et al.*, 1993). The two alpha helices of each monomer interact to mediate VSG dimerization, forming a four-helix coiled-coil bundle on the cell surface (Cohen *et al.*, 1984). This scaffold, coloured blue in Figure 1.6, defines the elongated structure for VSG. Smaller elements of VSG secondary structure hang from this skeleton, tied together by several loops (Blum *et al.*, 1993; Freymann *et al.*, 1990): these begin at the membrane distal end of the NTD, wrapping around to form a 'head' region (coloured pink in Figure 1.6), and end at the bottom of the molecule, forming further short regions of secondary structure sometimes referred to as the 'hips' (J. D. Barry, pers. comm., coloured green in Figure 1.6). It is these elements that decorate the framework alpha helices that are thought to be the primary target of antibodies: they are at the more exposed membrane-distal end of the molecule (Freymann *et al.*, 1990) and this region shows the most variability when comparing otherwise similar VSGs in the field, indicating the effects of positive selection (Hutchinson *et al.*, 2007). The fold is held together by disulphide bridges between conserved cysteine residues (Allen & Gurnett, 1983; Bussler *et al.*, 1998), which may provide some structure to the potentially less ordered N-terminal loops. Biochemical and bioinformatic analyses of other, non-crystallized VSGs give good reason to believe that they too form a similar structure (Bussler *et al.*, 1998; Carrington & Boothroyd, 1996; Cohen *et al.*, 1984). A flexible hinge region connects the NTD to the CTD, which is a compact structure of one or two four-cysteine subdomains whose primary role appears to be to increase membrane-proximal coat density and anchor the glycoprotein in the membrane (Chattopadhyay *et al.*, 2005; Jones *et al.*, 2008).

All VSGs are anchored in the membrane by a glycosylphosphatidylinositol (GPI) group and they are also frequently N-glycosylated on at least one other site in the molecule (Mehlert *et al.*, 2002). N-glycosylation may affect the structure and accessibility of epitopes—either those of VSG itself (influenza

haemagglutinin epitopes are often altered by N-glycosylation (Caton *et al.*, 1982)), or those of other surface molecules by enhancing the coat's barrier function (Mehlert *et al.*, 2002). Additionally, carbohydrate groups have been shown to play a structural role in maintaining VSG dimensions between variants—the absence of a short alpha helix in the membrane proximal region of the ILTat 1.24 NTD is compensated for by an N-linked glycosylation (Blum *et al.*, 1993)—which may be necessary for close packing.

1.4.3 VSG is expressed from a specialised expression site

The large surface area covered by the coat and the dense nature of its packing requires the production by metacyclic and bloodstream form parasites of huge quantities of VSG: in all, 10% of the total soluble protein content in a bloodstream form trypanosome lysate is VSG (Cross, 1990). Achieving this high level of expression requires a large number of *VSG* mRNA transcripts, and this is perhaps one reason why *VSG* is transcribed from a specialised telomeric genomic locus: the active expression site (ES). ESs have the unusual property of transcription by RNA polymerase I (Pol I), African trypanosomes representing the only example in Eukarya of this polymerase transcribing anything other than ribosomal RNA (Günzl *et al.*, 2003).[8] As ribosomal RNA represents the most active transcription in eukaryotic cells (Russell & Zomerdijk, 2005), it is plausible that Pol I transcription has been co-opted to provide the high volume of *VSG* mRNA transcripts necessary to construct and maintain the surface coat.

The ESs used by bloodstream form parasites (bloodstream expression sites, BESs) are telomere-proximal and usually between 40 and 60 kb in size, although some can be shorter (Berriman *et al.*, 2002; Hertz-Fowler *et al.*, 2008). Figure 1.7 shows a diagram of a BES. Besides *VSG*, BESs also contain a number of expression-site associated genes (*ESAG*s) (McCulloch & Horn, 2009). The *VSG* gene is located 5 to 10 kb from the end of the chromosome, with its 3' end several hundred bp upstream of the telomere repeats (Aline & Stuart, 1989; De Lange & Borst, 1982). Upstream of the *VSG* gene, and separating it from the rest of the ES, is a set of imperfect 70-bp AT-rich repeats (Campbell *et al.*, 1984). Transcription of the ES is initiated at a single promoter, and mRNA for

[8] The trypanosome procyclic stage surface glycoprotein, procyclin, is also transcribed by Pol I (Günzl et al., 2003).

each of the ES's genes is produced by trans-splicing of this polycistronic transcript with a 5' 'spliced leader' sequence common to all trypanosome mRNAs (Cully *et al.*, 1985; Parsons *et al.*, 1984; Walder *et al.*, 1986). A trypanosome has numerous BESs, thought to number between five (E. Louis, University of Nottingham, pers. comm.) and 23 (Young *et al.*, 2008b)—the total number is expected to vary between strains—but only one is active at a time, ensuring that only one *VSG* is expressed (see section 1.6.2). Metacyclic form parasites utilise a different set of 'metacyclic' expression sites (MESs), which are much smaller than BES and contain just a *VSG* gene (Barry *et al.*, 1998). MESs likely evolved from BESs as *ESAG*s and pseudogenes are often found in the region upstream of their promoters (Barry *et al.*, 1998).

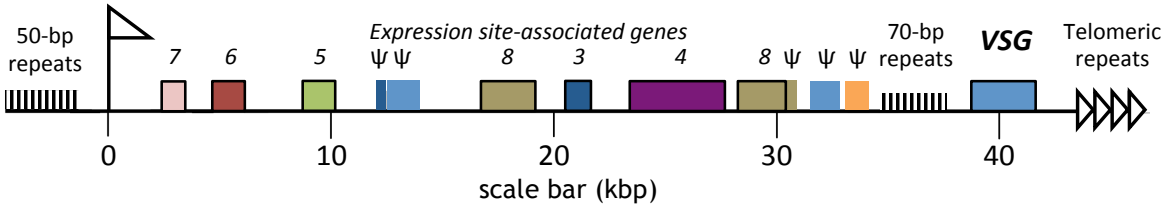


Figure 1.7 A bloodstream expression site (BES). The BES is isolated from the rest of the subtelomere by 50-bp repeats. Downstream of these is the ES promoter (indicated by a flag). The ES contains numerous expression site-associated genes (*ESAG*s), the exact composition of which varies between ESs. Numerous pseudogenes of *ESAG*s and *VSG*s can also be present (indicated by ψ). The *VSG* is separated from the rest of the ES by a series of 70-bp repeats. ES shown in that of GUTat 10.1, from TREU 927 (Berriman *et al.*, 2002).

The newly-translated VSG polypeptide chain contains a hydrophobic N-terminal signal peptide of around 30–40 amino acids (Boothroyd *et al.*, 1981), which is swiftly cleaved (McConnell *et al.*, 1981), directing the protein for secretion. The maturing VSG folds and undergoes post-translational modification: the removal of a C-terminal GPI anchor signal peptide of 17–23 amino acids and the addition of a GPI anchor, and N-linked glycosylation (Ferguson *et al.*, 1986; Mehlert *et al.*, 2002). Mature VSG reaches the surface at the 'flagellar pocket': an invagination of the plasma membrane at the base of the flagellum, where all endocytosis and exocytosis takes place (Overath *et al.*, 1997).

1.4.4 The VSG coat is a fluid, dynamic structure

The complete VSG coat is a bustling monolayer. The means of their anchoring—a relatively small GPI anchor—means that VSGs are able to move freely across the surface of the parasite and also probably spin in place (Buelow *et al.*, 1988). The flexibility of the linkage between the NTD and CTD suggests that they are also capable of bending and stretching (Chattopadhyay *et al.*, 2005). This fluid structure undergoes constant constitutive recycling at the flagellar pocket by endocytosis and exocytosis. It is thought that the entire surface coat is recycled in 12 minutes, and at any one point approximately 9% of total VSG is internal (Engstler *et al.*, 2004). Internalized VSG passes through the endosomal system and is replaced on the surface (Pal *et al.*, 2003; Seyfang *et al.*, 1990). VSG recycling also provides a complementary means of immune evasion: during recycling VSG-antibody complexes are separated and the antibody is proteolytically cleaved (O'Beirne *et al.*, 1998). Antibody clearance is enhanced by the fluidity of the coat: as the parasite swims, antibodies bound to VSGs are preferentially sorted by drag forces to the posterior of the parasite where they are more likely to enter the flagellar pocket and be cleared (Engstler *et al.*, 2007). Whilst hydrodynamic flow-mediated antibody clearance to immune evasion is ineffective at higher antibody concentrations, it nevertheless potentially endows parasites with an additional means of survival in the face of adaptive immunity. The non-switching short stumpy bloodstream form clears surface-bound antibodies more rapidly than the long slender form, perhaps offering them an extra level of resistance amid rising antibody levels (McLintock *et al.*, 1993) and marginally increasing their chances of survival and tsetse uptake. It is also possible that hydrodynamic flow-mediated antibody clearance from VSGs is a side effect, and that it may have a more pertinent functional role in clearing antibodies from partially exposed invariant antigens, or collecting receptor-bound transferrin (Mehlert *et al.*, 2012; Salmon *et al.*, 1997).

1.4.5 Antigenic variation comes about by changing the VSG coat

The above (section 1.4) describes the form and function of an existing VSG surface coat. By changing the expressed *VSG* gene, an individual parasite replaces its VSG surface coat, removing the antigen against which adaptive immunity may be developing, and replacing it, potentially with an antigen to

which the immune system is completely naïve. The source of the variation is the extensive archive of silent *VSG* genes, the nature of which is discussed in the following section.

1.5 *VSG* diversity

1.5.1 African trypanosomes possess hundreds of silent VSGs

The *Trypanosoma brucei brucei* genome is contained on 11 chromosome pairs (known as 'megabase chromosomes' between 1 and 6 Mb in size), a number of intermediate chromosomes (between 300 and 700 kb in size) and approximately 100 minichromosomes (approximately 50 kb in size) (Alsford *et al.*, 2001; Berriman *et al.*, 2005; Melville, 1998). A conspicuous feature of the genome's organisation is its accommodation for *VSG*s. In many eukaryotes, subtelomeres are loci where multigene families associated with variation can expand and diversify (Barry *et al.*, 2003; Linardopoulou *et al.*, 2005). For *T. brucei,* disproportionally large subtelomeres—so variable between homologous chromosomes as to be effectively haploid—contain huge arrays of silent *VSG* genes (Berriman *et al.*, 2005; Van der Ploeg *et al.*, 1982).[9] The basic unit of these subtelomeric *VSG* arrays are '*VSG* cassettes' 3–4 kb long, delimited at the 5' end by imperfect 70-bp repeats, like those found in the ESs, and the 3' end by a region spanning the 3' end of the *VSG* gene, the sequence of which is often well conserved between otherwise divergent *VSG*s (Majumder *et al.*, 1981; Matthyssens *et al.*, 1981; Pays *et al.*, 1981a; Rice-Ficht *et al.*, 1981). *VSG*s are also located immediately proximal to telomeres in ESs and on the ends of the minichromosomes, which seem to exist solely to provide an increased number of telomeres to store silent *VSG*s (Wickstead *et al.*, 2004).

In total, the number of *VSG*s in the *T. b. brucei* subtelomeric arrays is thought to number approximately 1800, although this probably varies slightly between strains.[10] In the lab-adapted monomorphic parasite line (Lister 427) this number

[9] There may also exist one, unusual, chromosome internal *VSG* array (L. Marcello, Ph.D. thesis, 2006, University of Glasgow).

[10] The *T. b. brucei* strain TREU927/4 GUTat 10.1 genome sequencing project sequenced 940 VSGs. The effective haploidy of the telomeres revises this number upwards to between 1500 and 1700, and certainly no more than 2000 (Marcello & Barry, 2007). Assuming each of the 100 minichromosomes contains a *VSG* at each telomere, an approximate figure of 1800 can be reached.

is thought to be even greater due to multiple gene duplications (B. Wickstead, University of Nottingham, pers. comm.), suggesting that selection in the natural life cycle and field population of African trypanosomes may act to restrict the archive's total size. *T. b. gambiense* appears to have a smaller archive, primarily due to its dearth of minichromosomes (Dero *et al.*, 1987; Jackson *et al.*, 2010).

1.5.2 Archive VSG*s can be grouped into different types*

VSGs are, on average, completely divergent at the amino acid level (Marcello & Barry, 2007a). Despite this, silent *VSG*s can be identified and grouped into types by the spacing of codons for conserved cysteine residues[11] as well as features relating to their processing such as their N-terminal signal peptide and GPI signal (Carrington *et al.*, 1991; Marcello & Barry, 2007a). For *T. brucei*, the NTD and CTD are typed separately (Carrington *et al.*, 1991; Jackson *et al.*, 2012).

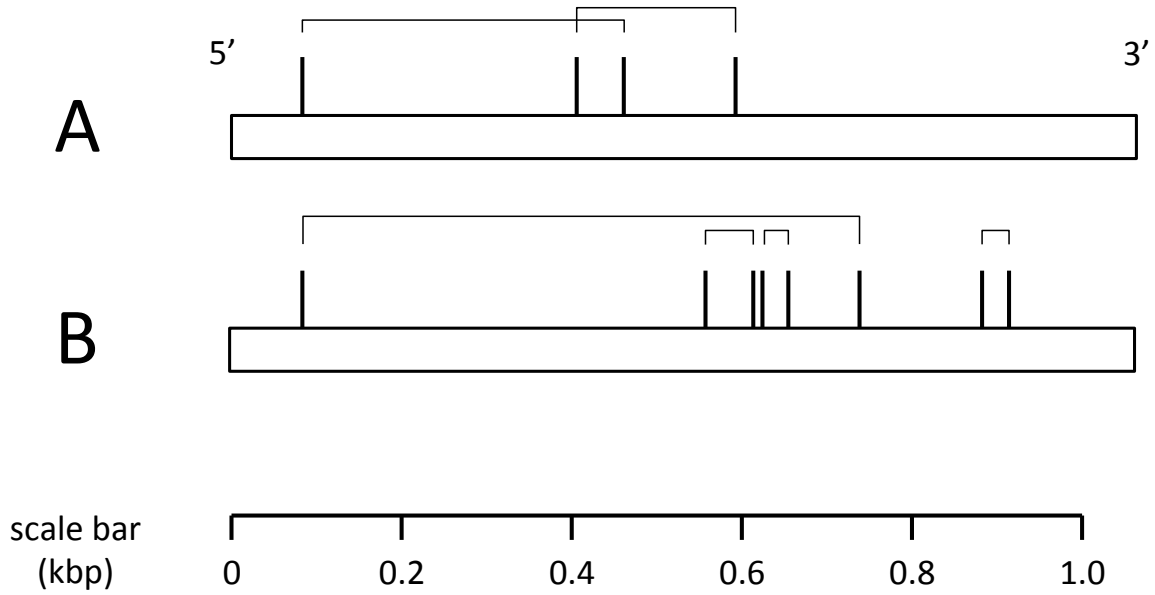


Figure 1.8 N-terminal domain types. Example A and B type NTD-encoding regions are shown. Cysteine codons are indicated by bars projecting above each image. Brackets indicate those cysteines that form disulphide bridges in the mature NTD. Details are from (Marcello & Barry, 2007a), (Carrington *et al.*, 1991) and (Bussler *et al.*, 1998).

[11] The main feature identified when aligning *VSG*s, reflecting their pivotal structural role. Other variable antigens such as influenza haemagglutinin also display conserved cysteine patterns (Air, 1981).

NTDs are typed ‘A’ or ‘B’ by homology, and each has a characteristic pattern of conserved cysteines (Marcello & Barry, 2007a). Examples are shown in Figure 1.8. Within these broad type definitions are numerous groups and subgroups, reflecting the diverse and divergent nature of *VSG*s (the previously described type C can be considered such a group within type A, for example) (Weirather *et al.*, 2012).

For the *T. brucei* CTDs, typing is more difficult (Weirather *et al.*, 2012). Their prevailing distinction is whether they contain one or two four-cysteine subdomains. The spacing of cysteines within these subdomains, and the sequence of the GPI signal they possess at their C-terminus, allow CTDs to be typed 1–6 (Marcello & Barry, 2007a). The relationship between the different *T. brucei* NTD and CTD types is flexible: although there is an apparent bias towards particular combinations, expressed and silent examples of both NTD types (A and B) possess related CTD types (Hutchinson *et al.*, 2003; Marcello & Barry, 2007a). Domain exchange also expresses itself within the CTD: close analysis of the different CTD types reveals that there is promiscuous mixing between CTD types, which is focused towards the N-terminal end of the CTD (Marcello & Barry, 2007a), in contrast to the apparent lack of inter-type mixing seen in the NTD. *VSG* CTD-encoding regions show greater levels of homology than *VSG* NTD-encoding regions (Marcello & Barry, 2007a; Rice-Ficht *et al.*, 1981).

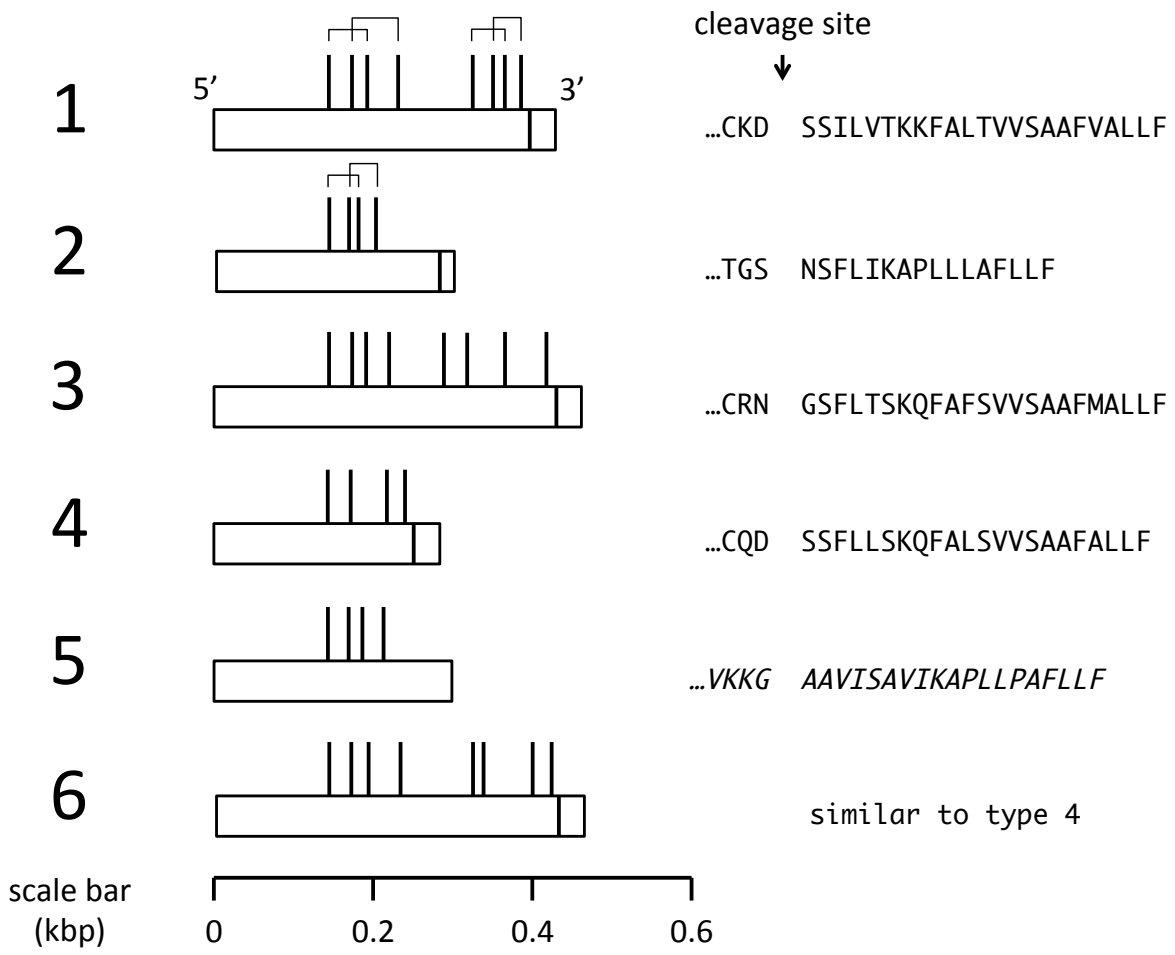


Figure 1.9 C-terminal domain types. The left hand side of the Figure shows example CTDs of each type. Cysteine codons are indicated by bars projecting above each image, and brackets indicate those cysteines that form disulphide bridges in the mature CTD, where known. The division between the CTD- and the GPI anchor signal-encoding regions is indicated by a line spanning the rectangle. The right hand side of the Figure shows example GPI anchor signal sequences for the different types, with a gap indicating the cleavage site between the mature CTD and the signal sequence. Data are from Marcello & Barry, (2007a) and Carrington et al., (Carrington *et al.*, 1991).

1.5.3 Many VSGs *are in subfamilies with one or two other* VSGs

Approximately 40% of archival *T. brucei VSG*s exist in small subfamilies—mostly pairs or triplets—sharing greater than 60% amino acid identity in the NTD (Marcello & Barry, 2007a). *VSG* subfamilies tend to be separated across chromosomes and across the subtelomeric arrays and are not generally located in contiguous loci (Beals & Boothroyd, 1992; Marcello & Barry, 2007a). *T. brucei VSG*s thus have a low-grain level of grouping into different *VSG* types, and a high-grain level of grouping into specific subfamilies. Between these two grades of differentiation there is remarkable divergence in the archive (Jackson *et al.*, 2012; Marcello & Barry, 2007a; Weirather *et al.*, 2012).

1.5.4 The majority of T. brucei VSGs *are pseudogenes*

An unexpected finding from the *T. brucei* genome sequencing was that only 4.5% of the *VSG*s in the subtelomeric arrays are clearly complete and functional (Marcello & Barry, 2007a).[12] As shown in Figure 1.10, the bulk of the subtelomeric archive consists of *VSG* genes damaged by frameshifts, stop codons and fragmentation. Of the pseudogenes, the majority were found to be dysfunctional in their CTDs. Broadly, two explanations exist for the degree of pseudogenicity in the archive. The first is that the archive is degenerate: its size means that the marginal fitness benefit of each individual *VSG* gene retaining its intactness is so low that most have drifted towards pseudogenicity. The second is that damaged *VSG* genes play a functional role, through acting as substrates for segmental gene conversion (see section 1.6.9) or by staggering expression (see section 1.6.12). These two explanations need not be mutually exclusive.

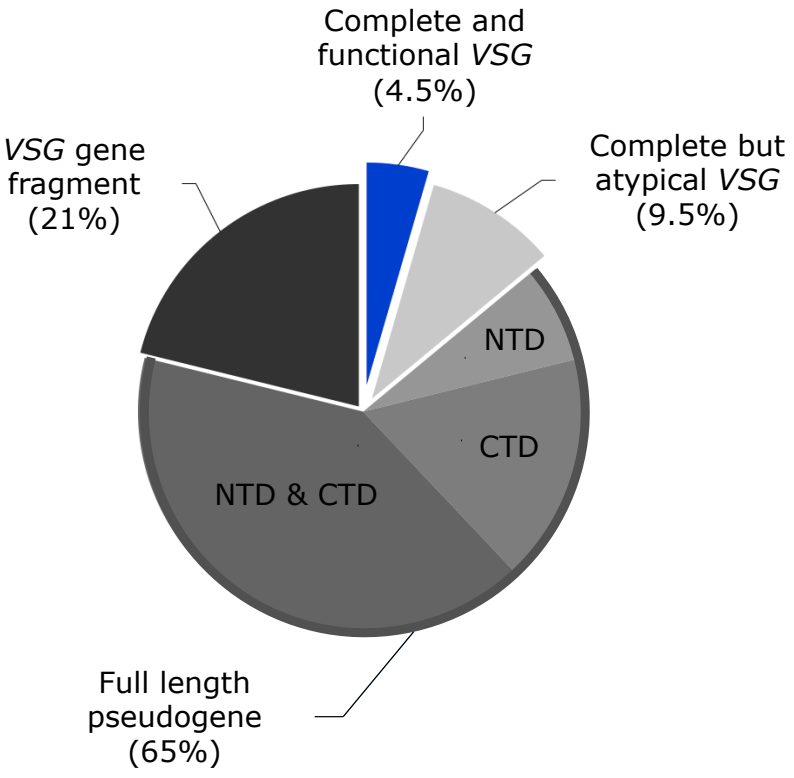


Figure 1.10 *VSG* subtelomeric array composition. Data are from Marcello & Barry, (2007a). For the full-length pseudogenes, the region in which the gene is damaged is indicated: 'NTD' = N-terminal domain, 'CTD' = C-terminal domain.

[12] When atypical *VSG*s – those with unusual GPI signals and cysteine spacing patterns that may compromise their expression – are included, this number is revised upwards to 14%.

1.5.5 VSG *diversity exists across the T. brucei population*

On a broader ecological level, the antigen archives of different trypanosome strains interact as different competing strains are given the opportunity to reinfect or superinfect populations of partially immune hosts (Marcello & Barry, 2007b). Comparisons of field-isolated, expressed *VSG*s from several *T. brucei* strains with *VSG*s in the sequenced genomes found that the majority of *VSG*s isolated have close homologues in different species and strains (Hutchinson *et al.*, 2007). Field populations probably share a common set of *VSG*s that can diverge to become strain specific (Beals & Boothroyd, 1992) and antigenically distinct (Hutchinson *et al.*, 2007). Besides the *VSG* diversity present in a specific trypanosome's archive, there is therefore diversity present across the population of trypanosomes in the field—a field 'metagenomic' archive of *VSG*s. Mating, where it occurs (Macleod *et al.*, 2001), might provide a means by which different strains can reassort their archives and acquire a different complement of *VSG*s.

1.5.6 VSG *diversity is ancestral amongst African trypanosomes*

Like *T. brucei, T. congolense* and *T. vivax* show a wealth of *VSG* diversity, and comparisons between the three species showed no species-specific clustering of *VSG*s, indicating that this diversity was present in their most recent common ancestor (Jackson *et al.*, 2012). Nevertheless, some differences have arisen between the three species as they have evolved. Compared with the more recently diverged *T. brucei* and *T. congolense*, *T. vivax* possesses two further families of *VSG*-like genes besides type A-like and type B-like *VSG*s, indicating that *T. brucei* and *T. congolense* have both passed through bottlenecks, indeed, *T. congolense* has lost A-type *VSG*s altogether (Jackson *et al.*, 2012). The patterns of recombination amongst *VSG*s also vary: neither *T. congolense* nor *T. vivax VSG*s show the CTD mixing patterns of *T. brucei*, but instead each *VSG* NTD type is closely associated with a specific CTD type (Jackson *et al.*, 2012). For *T. congolense*, these CTDs can be short and proline-rich (Rausch *et al.*, 1994; Strickler *et al.*, 1987). When comparing *VSG*s within a genome, another pattern of difference becomes clear: whilst *T. brucei VSG*s are highly diverse with only a high-grain level of grouping into small subfamilies (see section 1.5.3), *T. congolense* and *T. vivax VSG*s often show a gradation of differences between different *VSG*s (Jackson *et al.*, 2012). *T. congolense* and *T. vivax VSG*

repertoires are also less pseudogenic than *T. brucei*, with fewer than 30% of *VSG*s analysed showing pseudogenicity. These patterns hint that the mechanisms that maintain and generate *VSG* diversity may function differently in the three species.

1.6 Expression of the *VSG* archive and antigenic variation

1.6.1 VSG *expression is carefully controlled*

How are different *VSG*s from the archive expressed? There are several broad patterns that will be discussed in this section:

a) Monoallelic expression (section 1.6.2). Of the many available antigenic variants, each parasite exclusively expresses a single *VSG* allele.
b) *VSG* switching. There are two kinds of process by which different *VSG* genes obtain the property of monoallelic expression and become activated: transcriptional switching (section 1.6.4) and recombinatorial switching (section 1.6.5).
c) Segmental gene conversion (section 1.6.6). Recombination can cause segments of different *VSG*s to become combined.
d) Hierarchy (section 1.6.12). Activation of archival *VSG* genes across the population follows a 'semi-ordered' progression.

1.6.2 Only one VSG *is expressed at a time*

Individual trypanosomes express a single *VSG* at a time. This is presumably an advantageous strategy, as simultaneous expression of multiple *VSG*s would result in a heterogeneous coat, potentially exposing all of the expressed VSGs to the immune system. If simultaneous expression were typical, parasites might exhaust their antigen repertoire more quickly, resulting in a shorter infection than had the different antigens been expressed exclusively and consecutively (in *Giardia lamblia*, interference with the mechanism of exclusive expression resulted in a heterogeneous VSP coat and neutralizing responses developed against all exposed antigens (Rivero *et al.*, 2010)).[13]

[13] Dubois *et al.* (2005) argues counter-intuitively that trypanosomes artificially expressing surface coats made from two VSGs induce weaker immune responses and are resistant to immune killing. As discussed in Chapter 5 (section 5.3.1.2), this finding may be an artifact of ectopic

The *VSG* that is exclusively expressed is the *VSG* that is present in the active ES (Vanhamme *et al.*, 2000). Monoallelic expression is therefore achieved by the presence of only one intact *VSG* in an ES, and the activation of only one ES at a time. That there is only one *VSG* in the ES is probably a consequence of the mechanisms of switching rather than a strictly enforced demand, as parasites stably expressing multiple *VSG*s from the same ES at the same time can be artificially created (Muñoz-Jordán *et al.*, 1996).

The exclusive activation of a single ES is more strictly controlled. Experiments using drug selection markers to force simultaneous activation of multiple ESs resulted only in unstable clones rapidly switching between two ESs (Chaves *et al.*, 1999; Ulbert *et al.*, 2002). The rapidity with which the switching can occur, coupled with the lack of any detectable genetic rearrangement associated with the activation of an ES (see section 1.6.4), makes it likely that monoallelic expression is epigenetically maintained and inherited. The most likely mechanism is a single multi-component complex driving transcription, which can contain only one ES at a time. Immunofluorescence analyses show the colocalization of the active ES with a single extranucleolar pol I-containing body, termed the expression site body (ESB) (Navarro & Gull, 2001), whose association with a specific ES is stably inherited. Identifying the causal relationship between the ESB and monoallelic expression remains difficult—a single ESB could be taken as a consequence of monoallelic expression as much as its cause (Horn & McCulloch, 2010)—and there are many questions about its mechanism that remain unanswered. How is a single ESB maintained? What determines its association with one ES and not another? Nevertheless, the case for the importance of the ESB is supported by the close association between the ESB and the mitotic machinery, an association that is critical to the heritability of monoallelic expression (Landeira *et al.*, 2009). Comparisons of active and inactive ESs show that the key distinction of an active ES is not the presence of pol I or the initiation of transcription, but continued mRNA elongation (Vanhamme *et al.*, 2000). Low-level transcription from 'inactive' sites indicates

---

VSG expression. Otherwise, it is surprising that prolonged double expression has not been immunologically selected, as double expressors have been identified naturally (Baltz *et al.*, 1986).

that the role of the ESB in ensuring monoallelic expression is linked more closely to mRNA elongation than to transcriptional initiation *per se*.

If the ESB defines the active ES, complementary processes are at work to silence the inactive ones. The main effect seems to be the telomeric position of those sites: artificially generated telomeres are similarly repressed, although to a lesser extent than inactive ESs (Glover & Horn, 2006). A number of epigenetic mechanisms have been implicated in silencing (Alsford *et al.*, 2012), including:

- chromatin remodelling—the knockdown of various factors, including TbISWI (Hughes *et al.*, 2007) and histone deacetylase DAC3 (Wang *et al.*, 2010) are each accompanied by a de-repression of inactive ESs at the promoter end;
- nuclear location—inactive ESs are clustered at the nuclear periphery where their interaction with nuclear envelope proteins also contributes to their repression (DuBois *et al.*, 2012);
- the *VSG* sequence itself—*VSG* genes possess a conserved 3' untranslated region (UTR), and expression of this sequence from one locus suppresses its transcription from other loci (S. Hutchison pers. comm., P. Batram pers. comm.), perhaps through competition for an mRNA stabilizing factor (Berberof *et al.*, 1995).

The number of candidate factors involved in and affected by monoallelic expression, alongside the intimate and essential arrangement of these systems in a living cell, has complicated identification of 'the key' causative agent. Indeed, it is likely that monoallelic expression and its maintenance is a composite and non-linear network, with many feedback loops and redundancies.[14]

However monoallelic expression is ultimately achieved, it is flexible. The property of exclusive expression is frequently transferred to another *VSG* by switching, and it is to this that we now turn.

[14] In future, mathematical modelling may be able to help assess the relative contributions of the various candidates in such a complex system, but such models would depend on a clear picture of the nature of the interactions between the various components.

1.6.3 Different VSGs *can become exclusively expressed by switching*

VSG switching is the process by which the exclusively expressed *VSG* is changed. Switch rates have been estimated at 10^{-2}–10^{-3} switches/cell/population (Turner & Barry, 1989; Turner, 1997).[15] This rate was measured by assessing the number of parasites surviving after treatment with antiserum, and therefore is an aggregate measure that summarizes the effects of all productive switching events[16] that occur readily in the early stage of infection. A depressed rate of switching is one of the main distinctions of extensively syringe-passaged or laboratory-cultured parasites, which show rates several orders of magnitude lower than those of fly-transmitted pleomorphic lines (Turner, 1997); care must therefore be taken when extrapolating from laboratory strains to field populations, as the infection dynamics and relative contribution of the different *VSG* switching mechanisms may vary (Barry, 1997; Turner, 1990).

There are broadly two ways in which *VSG* switching is achieved (Figure 1.11):

A. Transcriptional switching does not require any genetic rearrangement. As *T. brucei* has multiple BESs, a different *VSG* can be expressed by transferring the property of monoallelic expression from one BES to another (section 1.6.4).
B. Recombinatorial switching involves the copying—or complete migration—of a *VSG* into the active BES (section 1.6.5).
C. Segmental gene conversion is a subset of recombinatorial switching, in which multiple silent genes contribute segments to the active *VSG* (section 1.6.6).

In requiring the convergence of different factors, each of these processes has a different chance of successfully occurring to a given silent *VSG*. The consequent 'activation probability' of each silent *VSG* is the basis of an overall hierarchy of expression across infection (section 1.6.12).

[15] This is the figure for pleomorphic trypanosomes. In lab-adapted monomorphic trypanosomes the rate is much lower, a state that can be reversed following passage through the tsetse (Turner, 1997).

[16] Note that some switching events will be 'non-productive' (that is, events that do not yield a antigenically novel surface coat); these are discussed in section 1.6.12.

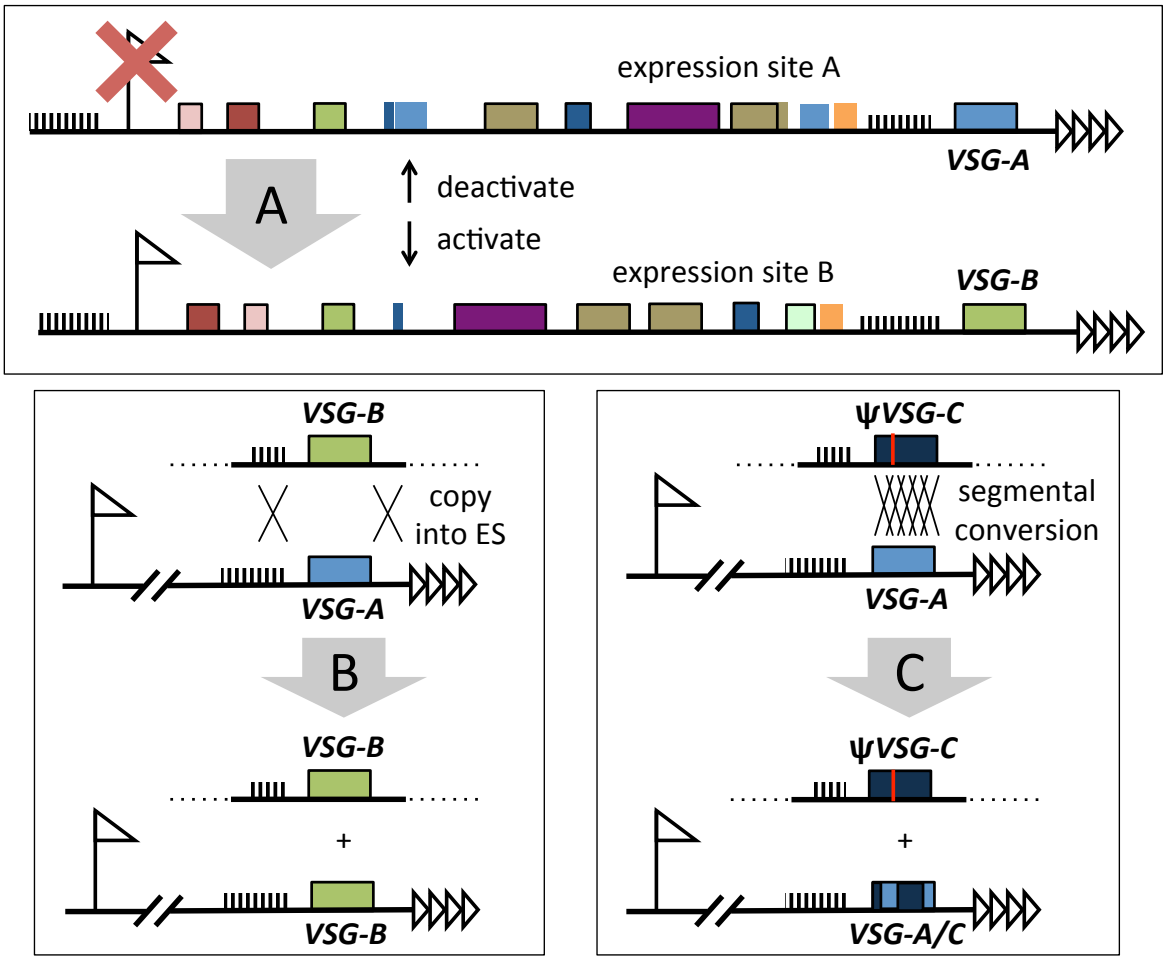


Figure 1.11 VSG switching. *VSG-B* can be activated by: (A) Turning off *VSG-A*'s ES, and turning on an ES containing *VSG-B* (*in situ* switching). (B) Replacing the active-ES-occupying *VSG-A* with a copy of *VSG-B* by recombination. Here a process of gene conversion is shown, leaving the silent copy of *VSG-B* unmodified. (C) It is also possible for segments of a silent *VSG* (ψ*VSG-C*) to be copied into an ES to form a mosaic *VSG* (*VSG-A/C* in the Figure). This process of 'segmental gene conversion' can occur even when the silent *VSG* is damaged by frameshifts or stop codons (indicated by the red bar in (C)).

Whilst changing the expressed *VSG* gene may be rapid, replacement of the glycoprotein coat downstream can take much longer. *VSG* mRNA has a half life of 4.5 hours, and VSG itself has a half life of around 30 hours (Ehlers *et al.*, 1987; Seyfang *et al.*, 1990). There may be mechanisms that operate to hasten the replacement of a surface coat—rapid degradation of old *VSG* mRNA or increase in the rate of coat turnover following a switch, for example—but these have not been defined, and it may be the case that the uniformity of the VSG surface coat is often not absolute (Esser & Schoenbechler, 1985; Seyfang *et al.*, 1990).

1.6.4 VSGs *can be activated* in situ

Switching can occur when the active ES becomes repressed and a silent ES becomes activated, a process that occurs without any necessary concomitant genetic rearrangement (Majiwa *et al.*, 1982; Young *et al.*, 1982; 1983). This 'transcriptional switching'[17] occurs rapidly in the bloodstream form (Ulbert *et al.*, 2002). An elegant model for its mechanism would be the ES body simply moving to a different ES (Navarro & Gull, 2001). However, although that would lead to the expression of a new *VSG*, it appears there are other factors at work. The activation of a new ES and the silencing of the old one are separable processes—the absence of the chromatin remodelling enzyme DOT1B results in the residual expression of the old ES for up to two weeks after switching (Figueiredo *et al.*, 2008). The triggers prompting transcriptional switching are also mysterious, although the close association between the ES body and nuclear architecture hints that mitosis might be one occasion where the ESB-ES complex becomes vulnerable to disruption (Landeira *et al.*, 2009).

Transcriptional switching was found to be a major switching pathway in lab-adapted monomorphic *T. brucei* lines (Liu *et al.*, 1985) but this is may be due to deficiencies in aspects of their recombinatorial switching pathways: in the more field-relevant, rapidly-switching pleomorphic lines it appears to be a minor process (Robinson *et al.*, 1999). Transcriptional switching, by its nature, can only provide access to *VSG*s located in one of the other ESs. What, therefore, is the role for transcriptional switching alongside the more flexible recombinatorial mechanisms? The primary reason seems to lie not with the *VSG*, but with the expression site-associated genes (*ESAG*s): the different *ESAG* collections at different ESs can adapt to the peculiarities of different host species. This might be the case for the transferrin receptor *ESAG6/7* (Bitter *et al.*, 1998), and possibly also the *ESAG4* family of adenylate cyclases (Salmon *et al.*, 2012).[18] The presence of multiple ES may also compensate for the potentially deleterious recombinatorial processes occurring in the active ES (see section 1.6.10).

[17] Also known as *in situ* switching.

[18] However, inconsistent with this hypothesis is the lack of correlation between host range and *ESAG* heterogeneity (Young et al., 2008)

1.6.5 VSGs *can be rearranged to occupy an expression site*

For most *VSG* genes, activation requires genetic rearrangement: the bulk of the archive is located in the subtelomeric arrays. The importance of recombination has experimental support—a study on switching in pleomorphic lines showed almost all switching events in the first two relapses to be accompanied by gene duplication (Robinson *et al.*, 1999) and trypanosomes lacking a key catalyst of recombination, RAD51, are severely deficient in switching even in a monomorphic trypanosome line (McCulloch & Barry, 1999). Recombinatorial switching is therefore the major process in expressing trypanosome antigenic variability during an infection.

Recombinatorial switching occurs through multiple mechanisms. Many early studies found that switching was often accompanied by a duplication-transposition of the gene that is expressed to an ES (Hoeijmakers *et al.*, 1980; Pays *et al.*, 1981b; 1983; Young *et al.*, 1983), a process termed 'gene conversion'.[19] In its best-understood form, gene conversion co-opts the ancient DNA repair mechanism, homologous recombination (HR), to delete the existing ES-occupying *VSG* and replace it with a different *VSG* from elsewhere in the genome (Laurent *et al.*, 1984a; Morrison *et al.*, 2009; Pays *et al.*, 1983). The trigger is thought to be damage to the active ES, for example a double-stranded break in the DNA (Boothroyd *et al.*, 2009): a likely occurrence, given the extremely exposed nature of the active ES's DNA and the physical instability of the AT-rich 70-bp repeats just upstream (Lin *et al.*, 2009b; Stanne & Rudenko, 2010).[20]

In homologous replication, following a double-stranded break, nucleases delete part of the sequence and expose single-stranded DNA. RAD51 (the eukaryotic orthologue of eubacterial RecA), binds to the exposed strand and catalyses its invasion of an intact DNA duplex in a homology-dependent manner (West, 2003) paving the way for replication from the invading strand. There are a number of

[19] This observation gave rise to the distinction between the 'basic copy' (that which is present in both expressor and non-expressor) and 'expression-linked copy' (that whose appearance correlated with expression, Majumder *et al.*, 1981).

[20] The use of non-coding DNA structures to initiate recombination has also been observed in *N. gonorrhoeae* pilus antigenic variation, where a G-quadruplex structure is necessary for recombinatorial antigenic variation to occur (Cahoon & Seifert, 2009).

different ways that such replication might proceed and resolve, depending on the repair apparatus and location of the template (Figure 1.12). In *VSG* gene conversion, the most likely mechanisms are strand-dependent strand annealing (SDSA, Figure 1.12-B) and break-induced repair (BIR, Figure 1.12-C), as double strand break repair (DSBR, Figure 1.12-A) can result in the crossover of flanking sequences and potentially lethal translocations (Morrison *et al.*, 2009). Whatever the process, the result of homologous recombination is the replacement of the original sequence—the *VSG* gene and some of the flanking regions—with that of the template.

HR is underpinned by sequence similarities between DNA sequences, as the invading strand needs to base pair with the template at the 5' boundary of conversion, and in many cases the newly-synthesized copy of the template needs to bind to the sequence that is being repaired at the 3' end. Rates of HR in *T. brucei* are reduced as the length and degree of homology between substrate and template decrease (Barnes & McCulloch, 2007). Sequence examination of ESs which had undergone gene conversion revealed that the 5' boundary of recombination was usually located in the 70-bp repeat regions located upstream of the ES *VSG* (Liu *et al.*, 1983), and the 3' boundary of recombination was usually in the region spanning the 3' end of the *VSG* (Bernards *et al.*, 1981; Liu *et al.*, 1985; Michels *et al.*, 1983; Timmers *et al.*, 1987). Fittingly, these regions correspond with the conserved boundaries of the '*VSG* cassette' (section 1.5.1). A straightforward model of recombinatorial switching therefore sees an archival *VSG* cassette replacing the telomere-proximal *VSG*-containing section of the ES, from the 70-bp repeat region to the 3' UTR of the *VSG*. Divergent sequences (*VSG* genes) are thus exchanged by a mechanism that demands high levels of identity between the exchanged sequences. Alternatively, if a telomeric *VSG* is used as the template, and BIR is the process by which the template is copied, the 3' boundary may not exist as replication could proceed all the way to the telomere end (Morrison *et al.*, 2009).

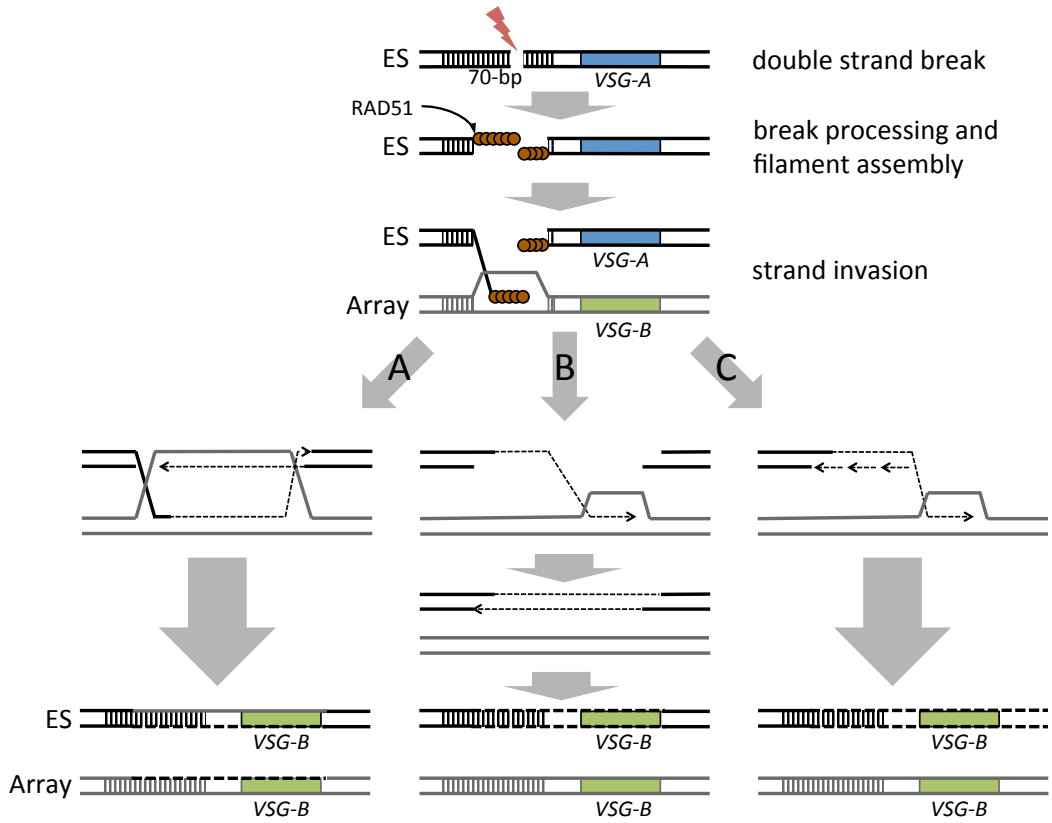


Figure 1.12 Possible recombinatorial switching mechanisms. A double stranded break in the 70-bp repeats is processed by various complexes to reveal single-stranded DNA ends, providing a substrate for a RAD51 filament to form. RAD51 filament assembly is aided by a number of factors (not shown here). RAD51 catalyses homology-dependent invasion of the strand into intact duplex DNA (grey lines) containing a silent *VSG* (green box). Three pathways for the resolution of double strand breaks are known. In each of these images, dashed lines indicate newly-synthesized DNA, and arrowheads indicate the direction of synthesis. (A) Double strand break repair involves the synthesis of DNA from the template, and the strands remain base-paired. This leads to the formation and resolution of Holliday junction intermediates. Here, the reaction is shown with no cross-over of flanking sequences. (B) Synthesis-dependent strand annealing (SDSA) involves the copying of the template sequence onto the end of the invaded strand, which can then re-anneal to the exposed DNA on the other side of the break. Synthesis can then proceed from that strand, using the newly synthesized DNA as a template, and the reaction is resolved by ligation of the DNA backbone. (C) In break-induced replication (BIR) strand invasion promotes formation of a replication fork with lagging and leading strands, which can potentially proceed all the way to the telomere end. Image was copied and modified from (Morrison *et al.*, 2009).

A further pathway is the recombinatorial activation of a *VSG* without any concomitant duplication. Here, classical recombination acts to reciprocally exchange chromosome ends, swapping the active ES *VSG* with the telomere-proximal *VSG* of another chromosome (Pays *et al.*, 1985b). Telomere reciprocal exchange is prominent in monomorphic lines (Rudenko *et al.*, 1996).

RAD51-mediated homologous recombination pathways explain only part of the evidence. Study of residual switch events in the absence of RAD51 revealed a

recombination pathway that required only very short regions of homology—as short as 7–13 bp—and was tolerant of mismatches (Conway *et al.*, 2002; McCulloch & Barry, 1999). Alongside stricter, classical homologous recombination, there are clearly more flexible mechanisms at work.

1.6.6 Parts of different VSGs *can be combined by segmental gene conversion, modifying and repairing them*

The effects of segmental gene conversion can be seen when no single genomic donor can be found for an expressed *VSG*. Rather, two or more donors can be identified, each of which contributes different segments to form the expressed *VSG*.

Two broad patterns of segmental gene conversion can be distinguished in *VSG*: 3' donation (section 1.6.8) and mosaicism (section 1.6.9). These patterns are linked in that they involve recombination boundaries occurring within the coding region, yielding an expressed *VSG* with more than one donor, but it is possible that they are the consequence of different recombinatorial mechanisms. Table 1.3 describes the *VSG*s identified to date that have apparently undergone segmental gene conversion.

Type of variation	VSG name(s)	Reference(s)
Mosaicism	WaTat 1.1 family	Barbet *et al.*, 1982; 1989; Kamper & Barbet, 1992
	VSG20, VSG20bis	Thon *et al.*, 1989; 1990
	VSG78 family	Roth *et al.*, 1986; 1989
	AnTat 1.1B	Pays *et al.*, 1985a
	[7 different mosaics]	Marcello & Barry, 2007a
3' donation (i)	VSG117, VSG118	Michels *et al.*, 1983
	[4 variants]	Marcello & Barry, 2007a
3' donation (ii)	VSG20 family	Thon *et al.*, 1989; 1990
	VSG78 family	Roth *et al.*, 1986; 1989
	AnTat 1.1H, AnTat 1.1C	Pays *et al.*, 1985a
	VSG A	Aline *et al.*, 1994
	VSG-24-09-01	Marcello & Barry, 2007a

Table 1.3 Previously identified segmental *VSG* conversion. (i) and (ii) refer to greater and lesser 3' donation events respectively, see Figure 1.13.

First, segmental gene conversion as a general mechanism in antigenic variation will be considered (section 1.6.7), before describing the patterns of segmental gene conversion seen amongst *VSG* (sections 1.6.8–1.6.10).

1.6.7 Segmental gene conversion is an important process for many antigenically variant pathogens

Segmental gene conversion to generate expressed variable antigens is common to many antigenically variable pathogens. Broadly speaking, segmental gene conversion occurs in two, often complementary, ways:

(i) Variable region cassettes. Variable antigens often possess conserved regions: parts of the molecule that are either functionally constrained or not the target of immune responses. Rather than varying the whole molecule, segmental gene conversion allows just the variable regions of the antigen to be exchanged with archival copies. *Anaplasma marginale* MSP2 and 3 (Barbet *et al.*, 2000; Meeus *et al.*, 2003), *Neisseria gonorrhoeae* pilin (Hamrick *et al.*, 2001), *Treponema pallidum* TrpK (Centurion-Lara *et al.*, 2004), and *Mycoplasma genitalum* MG192 (Iverson-Cabral *et al.*, 2007; Ma *et al.*, 2007) all undergo such segmental gene conversions. Using truncated, pseudogenic, variable region cassettes avoids the duplication of conserved regions (which could potentially suffer from pseudogenizing drift) and makes efficient use of a small genome (Ma *et al.*, 2007).

(ii) Recombinatorial diversity. Successive segmental gene conversion events occurring within the variable region can effect antigenic variation through a process of progressive mosaicism during an infection. *A. marginale* MSP2 (Futse *et al.*, 2005) and *B. burgdorferi* VlsE (Coutte *et al.*, 2009) both show increasingly complex segmental conversion events in their variable regions as infections progress, and combinatorial diversity has been found to mediate immune evasion (Zhuang *et al.*, 2007) and superinfection (Futse *et al.*, 2008) in *Anaplasma*.

Amongst those pathogens that use recombination to access and express an archive of silent genes, segmental gene conversion could be considered typical,

with *Borrelia hermsii* the only well-defined exception to the trend (Dai *et al.*, 2006).[21]

Segmental gene conversion also shapes multi-gene families in their long-term evolution—evidence of its work can be seen in the antigen repertoires of *Plasmodium falciparum* (Bull *et al.*, 2008; Rask *et al.*, 2010) and *Giardia lamblia* (Adam *et al.*, 2010), as well as *T. brucei* (Gjini *et al.*, 2012)—although by copying elements between genes, segmental gene conversion can have an overall homogenizing effect on an archive (Nei & Rooney, 2005), particularly if segments are large (Martinsohn *et al.*, 1999).

1.6.8 VSG *3' donation involves recombination events in the C-terminal domain- and GPI anchor-encoding regions*

VSG 3' donation has occurred when the 3' end of the silent archival *VSG* is different from the 3' end of its expressed form. Examples of 3' donation are shown in Figure 1.13. Here, the term '3' donation' covers a spectrum of exchange, from the very 3' most GPI anchor signal just upstream from the 3' UTR that is the boundary of exchange for 'full-length' *VSG* gene conversion, (marked (i) on the Figure) to the entire CTD (marked (ii) on the Figure).

The observed patterns of 3' donation are usually explained by the 3' boundary of the usual process of gene conversion occurring within the CTD-encoding region of the *VSG* gene (Aline *et al.*, 1994; Michels *et al.*, 1983; Pays *et al.*, 1985a; Thon *et al.*, 1990). This would mean that the newly expressed *VSG* retains part of the 3' end of the previously expressed *VSG*, but since this region encodes the buried, membrane-proximal CTD that is inaccessible to antibodies its retention does not compromise antigenic variation (Schwede & Carrington, 2006). Such retention could, in fact, be advantageous: many silent *T. brucei VSG*s have damage to their GPI signal sequence and towards the 3' end of their CTDs (Marcello & Barry, 2007a). By repairing these genes, 3' donation increases the proportion of the *VSG* archive available for expression. This pattern is somewhat reminiscent of the 'variable region cassette' mechanism outlined above (section 1.6.7). The homology often seen between *T. brucei* VSG CTDs, despite complete

[21] Although segmental gene conversion has been recorded in *B. hermsii*, it is not a prominent pattern (Dai *et al.*, 2006).

NTD divergence, means that the coding region could naturally act as the 3' recombination endpoint instead of the 3' UTR, particularly if just the very 3' most end of the gene is retained (Liu *et al.*, 1985; Majumder *et al.*, 1981). It is also possible that 3' donation occurs to an existing ES *VSG*, as inferred by Marcello & Barry, (2007a).

In *T. congolense* and *T. vivax*, where there are much lower levels of homology between *VSG* CTD-encoding regions, 3' donation may occur less frequently, if at all (Jackson *et al.*, 2012).

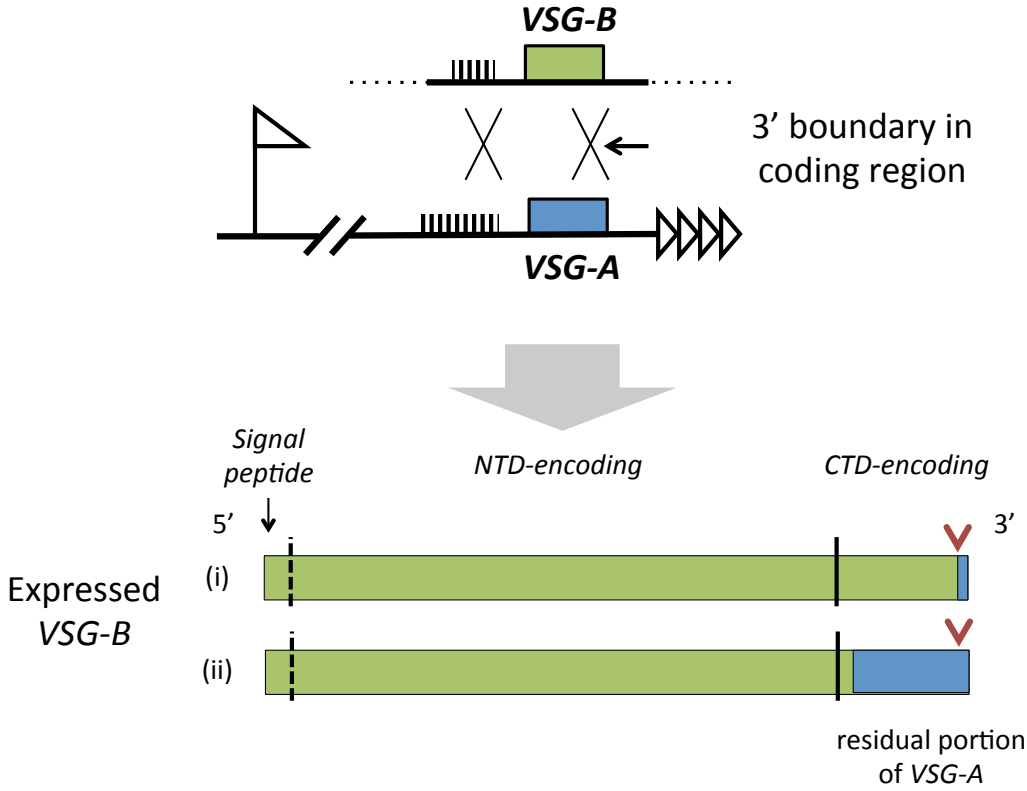


Figure 1.13 3' donation. When the 3' boundary of gene conversion occurs within the open reading frame of the *VSG*, the incoming *VSG* retains the 3' end of the previously resident *VSG*. The bottom half of the image shows two possible outcomes. Each image represents the nucleic acid sequence of a *VSG*, from 5' to 3', with dotted and intact lines indicating respectively the division between signal peptide-encoding region and NTD-encoding region, and NTD- and CTD-encoding regions. A red arrowhead indicates the boundary of the mature CTD: the region 3' of this encodes the GPI anchor signal sequence. The gene contributing to each segment of the *VSG* are coloured according to the top half of the image. In (i), only the GPI-anchor-encoding sequence of *VSG-A* is retained, whereas in (ii) the whole CTD-encoding region is retained.

1.6.9 Mosaicism generates variation across the VSG

Mosaicism is a process of segmental gene conversion that occurs throughout the *VSG*, including in the antigenically important NTD.[22] A set of mosaic *VSG*s identified by Kamper & Barbet (1992) are shown in Figure 1.14 as an example. Mosaic *VSG*s have multiple donors, each of which can contribute multiple segments to form an expressed *VSG* (in the case of WaTat 1.13, 15 segments have been donated by the four different donors).

Compared with straightforward gene activation, mosaicism has two important features that could contribute to antigenic variation:

(i) Accommodation of damaged *VSG*. Like 3' donation, mosaicism could access a greater portion of the silent *VSG* archive, allowing expression of otherwise inaccessible epitopes. As infections can last years, *VSG* repair by segmental gene conversion could become the predominant means of expressing archive *VSG*s (Barbet & Kamper, 1993).

(ii) Recombinatorial variability. By allowing donors to combine in a multitude of different ways, mosaicism can increase their potential for antigenic novelty, increasing the number of antigenic profiles many-fold. Such is the case for *Anaplasma marginale* and *Borrelia burgdorferi* (section 1.6.7).

Could segmental gene conversion amongst a limited set of donors contribute to antigenic variability in African trypanosomes? Analysis of the WaTat 1.1 mosaic family (see Figure 1.14) with monoclonal antibodies showed that the different mosaics shared some epitopes but varied in others (Barbet *et al.*, 1989). As these mosaic variants—with between 0.051 and 0.105 differences/aa between their mature NTDs—were identified by immunological cross-reaction and isolated following a lengthy series of passages, it is possible that diversifying selection imposed by adaptive immunity over the course of an individual infection would favour more antigenically variant mosaics (Barbet & Kamper, 1993).

[22] Importantly, 'mosaicism' here should be distinguished from 'double expression'—the simultaneous appearance of two different VSG isoforms in a surface coat—as double expressors are sometimes referred to 'mosaics' in the literature (Dubois *et al.*, (2005), for example).

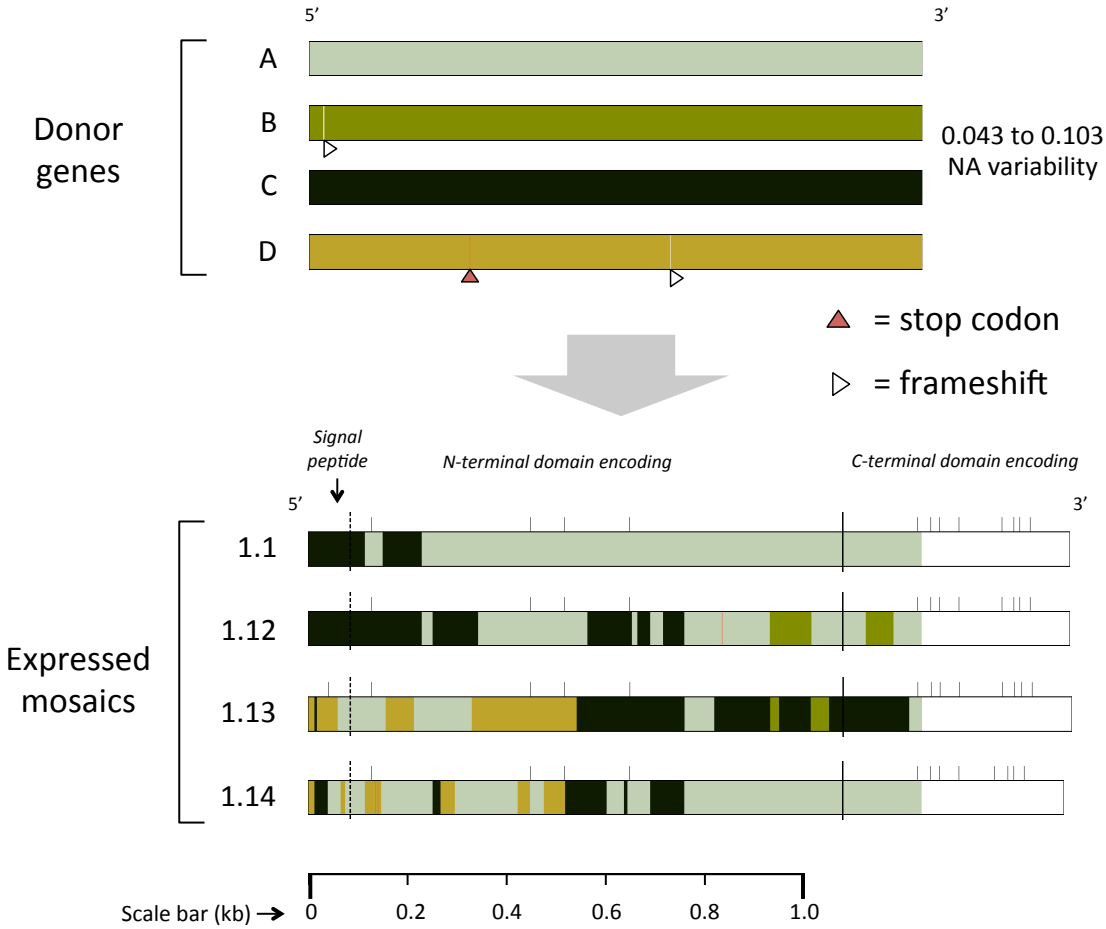


Figure 1.14 Mosaicism in the WaTat 1.1 *VSG* family (redrawn from Kamper & Barbet, (1992)). All images are to scale. The top half of the figure shows the four putative donor sequences. The positions of stop codons and frameshifts, present in donors B and D, are indicated. The nucleic acid (NA) variability between these sequences is given in differences/nt. The bottom half of the image shows four mosaic *VSG* sequences that were generated from these donors via a rabbit infection and a series of lengthy passages. Each image represents the nucleic acid sequence of a *VSG* cDNA from 5' to 3', with dotted and intact lines indicating respectively the division between signal peptide-encoding region and NTD-encoding region, and NTD- and CTD-encoding regions. Different regions of each mosaic *VSG* are coloured according to the most probable donor sequence for that region. Segment boundaries were arbitrarily considered to be the midpoint of the region of identity between the two adjacent segments. Sequence data were not available for the 3' ends of the donor sequences, and thus the expressed mosaics are coloured white in this region. Cysteine codons are represented as bars protruding from the top of the diagrams and red bars spanning the diagram represent positions that match none of the donors identified. For this figure, GenBank sequences M83694–M83702 were used.

1.6.10 Segmental gene conversion depends on homologies, and complex mosaics may be constructed by sequential events

The mechanisms underlying segmental gene conversion are unclear, but sequence analyses suggest the process is promoted by homologies between donor sequences. For mosaic *VSG*s, donors show sequence similarities—they are members of the same *VSG* subfamily (section 1.5.3)—but the boundaries of

segmental recombination lack the long regions of perfect identity that are associated with classical homologous recombination (Barnes & McCulloch, 2007) and the size of the contributed segments can also be very small (14 bp, in the case of WaTat 1.14 (Kamper & Barbet, 1992)). Thus construction of mosaic *VSG*s may occur by a pathway or pathways different from that of full-length gene conversion: one that is held to be inefficient, due to the late appearance of mosaic *VSG*s in infection (Morrison *et al.*, 2009), but better-able to cope with imperfect homologies. The less-stringent Rad51-independent pathway is the most promising candidate to date (Conway *et al.*, 2002), but direct experimentation has been hampered by the late appearance of mosaic *VSG*s in infection and the difficulties associated with transfecting pleomorphic bloodstream form trypanosomes.

By examining the sequences of mosaic *VSG*s, the process of their construction can be inferred. The best model to date holds that mosaics are 'built up' in an ES, as various donors recruited by homology contribute segments in succession to form a 'string' of mosaics (Barry *et al.*, 2005). Here, this model will be referred to as 'progressive mosaicism'. It draws on the work of Pays et al., (1985a), who observed, over a number of passages and clonings, repeated interactions with the same set of donor genes with an expressed copy to generate a number of related mosaics. The data of Marcello & Barry, (2007a) are also consistent with this model, as from a single infection two mosaics with some conserved, and some independent, segmental conversions could be identified and from these a putative 'lineage' deduced. This model is analogous to the patterns defined in *A. marginale* MSP2 (Futse *et al.*, 2005) and *B. burgdorferi* VlsE (Coutte *et al.*, 2009), where donors serially contribute segments to construct increasingly complex antigen genes during infection.

One of the appeals of progressive mosaicism is that it allows mosaics to be selected for function—and potentially antigenic novelty—at each step in their development. The expressors of functional, antigenically distinct mosaics can go on to proliferate, creating a larger pool of mosaic-expressors and increasing the chances for further segmental gene conversion events, using the same donors, to occur. Therefore, the initial event in generating a mosaic—for example

repairing a pseudogenic donor—is a high barrier to entry that, once overcome, may pave the way for further mosaicism to occur readily.

Assuming that one of the donors initiates the mosaic lineage, progressive mosaicism requires a dysfunctional *VSG* to readily occupy an ES (in the case of (Roth *et al.*, 1989)). Similarly, it is possible that segmental gene conversion events introduce errors (for example if they are out of frame or if they involve the damaged part of the donor). How would a cell cope with a damaged *VSG* in its ES? Studies where RNAi was used to deplete parasites of *VSG* mRNA induced global translational repression, indicating that the lack of *VSG* transcripts has a profound effect on cell biology (Smith *et al.*, 2009). Although this example is distinctly artificial, it hints at mechanisms that may feed back to the ES to perhaps increase recombination or trigger transcriptional switching. The pathways that trypanosomes possess for sensing and degrading nonsense mRNA (Delhi *et al.*, 2011) and the large number of proteins involved in facilitating VSG folding and quality control (Field *et al.*, 2010) are potential candidates for the mechanisms by which problems with VSG production might be detected. The error-prone nature of segmental gene conversion could explain the presence of multiple BES: providing a source for easy-to-access backup *VSG*s to protect against dangerous segmental experimentation at the active ES. It is also possible that silent BES are 'sandboxes'—loci where multiple segmental recombination events might occur safely without compromising transcription of the active *VSG* gene. These silently-assembled mosaics could then be activated *in situ* (Barry & McCulloch, 2001). Silent gene conversion (that is, not associated with activation) has been frequently recorded in *T. brucei* (Aline *et al.*, 1989; Aline & Stuart, 1985; Myler *et al.*, 1988).

1.6.11 Point mutations are probably a minor source of antigen variability

Could point mutation also contribute to VSG antigenic variation? Studies have reported single nucleotide variation in the expressed *VSG* when comparing it with its archival copy (Lu *et al.*, 1994; Rice-Ficht *et al.*, 1982), and in chronic infection studies these mutations appear to accumulate preferentially in regions of the NTD that may form epitopes (Marcello & Barry, 2007a). Point mutation variation is unlikely to be routinely generated by the process of gene conversion

itself, since there are many examples of perfect *VSG* copying (Graham & Barry, 1996; Pays *et al.*, 1985a) but the exposed chromatin structure around the active ES may make it particularly vulnerable (Stanne & Rudenko, 2010). Whether such nucleic acid variation translates to meaningful changes in epitope structure is another question. Variable monoclonal antibody binding as a consequence of point mutation has been reported (Baltz *et al.*, 1991; Lu *et al.*, 1994) but these variants, obtained by artificial immunological selection, may bear no resemblance to the consequences of immunological selection *in vivo*. Whilst in field populations over longer time periods point mutation could be a useful way of generating variation (Hutchinson *et al.*, 2007) it seems unlikely that the build up of point mutations in the expressed *VSG* occurs rapidly enough to play an important role during an infection (Graham & Barry, 1996).

1.6.12 The appearance of different variable antigen types during an infection follows a hierarchy

There is a semi-predictable ordering to the *VSG*s that are expressed over time in an infection, which can be reset when those trypanosomes are used to infect a naive host (Gray, 1965). This is termed the 'hierarchy' of *VSG* expression. The presence of a antigen expression hierarchy has been explained with arguments similar to those made for monoallelic expression: activation of archival *VSG* genes *across the population* must follow some order; otherwise the trypanosomes would rapidly expose their entire archive to the immune system and run out of antigens to express (Morrison *et al.*, 2005) (this is likely the case for *Plasmodium*, which possesses a much smaller archive (Recker *et al.*, 2011)). Another theory is that uncontrolled antigen variability would result in overwhelming parasite density and host exhaustion as the immune system is left unable to cope (Turner *et al.*, 1995): extended periods of parasitaemia can exert a crippling energy burden on the host (Seed *et al.*, 2003). Neither of these theories is completely satisfying in light of differentiation-controlled infection dynamics (see section 1.7.5), as a multitude of variants could theoretically persist beneath the threshold of immune induction, avoiding host exhaustion by timely stumpy differentiation. Perhaps expression hierarchy has been subject to more complex selection, involving a balance between competition amongst trypanosome strains and the demands for flexibility in different host types

(Turner, 1999) (see Chapter 6 section 6.2.4–6.2.6 for further discussion on this matter).

A number of theories have been advanced to explain the mechanics of expression hierarchy (ranging from immune cross reaction to differential growth rates (Morrison *et al.*, 2005)), but the best-supported and most elegant explanation is simple: that different *VSG*s have different activation probabilities. As the infection progresses, immune responses against the readily-expressed *VSG*s—those with high activation probability—constantly select *VSG*s that have lower activation probability (Morrison *et al.*, 2005). One of the consequences of this model is that individual *VSG* switches can be classified as 'productive' (activating a *VSG* against which antibodies have not arisen), or 'non-productive (activating a *VSG* to which the host has developed immunity) (Turner, 1999).

What is the genetic basis for a *VSG*'s activation probability? As stated above (section 1.6.1), for a *VSG* to be expressed it must reside in the active ES. Therefore, the activation probability of a specific *VSG* reflects the number of steps required to get to that locus, and the likelihood that they will occur. As the majority of *VSG* activation events are rooted in homology-based gene conversion, the homologies between the flank regions of the active ES and the *VSG* under consideration are likely to be an important determinant of activation probability (Morrison *et al.*, 2005). For those *VSG*s resident at telomeres, telomere interactions promote recombination (Barry *et al.*, 2003) and repeats downstream of the *VSG* represent regions of 3' homology with the active ES that the *VSG*s in the arrays do not have (Aline & Stuart, 1989).[23] Accordingly, telomere-resident *VSG*s appear to be preferentially activated in the early stages of infection (Liu *et al.*, 1985; Robinson *et al.*, 1999). Intact array *VSG*s rely on shorter, more variable homologies between their 3' region and that of an ES *VSG* (Marcello & Barry, 2007a), but can be copied into an ES in a single step (as a 'cassette'). Pseudogenic *VSG*s, requiring inefficient segmental conversion for their activation, become prominent later in infection (Pays, 1989; Roth *et al.*, 1989).

[23] In fact, as described above, if BIR is used to copy telomeric *VSG*s into the ES, replication could proceed all the way to the telomere end, making 3' homology unnecessary. 3' homology is also unnecessary for reciprocal telomeric exchange.

An extra layer of complexity to the model comes from the fact that every change to the ES (for example the copying in of a new *VSG* cassette) can change the degree of homology that any archival sequence has with the ES, and thus their potential to interact (Pays *et al.*, 1985a). Although there is no evidence of homology-dependent switching taking place early in infection—the activation probability of a silent *VSG* seems to have little bearing on the currently-expressed *VSG* (Morrison *et al.*, 2005)—later on, during the progressive generation of mosaics, coding sequence homologies probably become more important (Thon *et al.*, 1990). The principle of activation-probability-driven hierarchy is shown in Figure 1.15.

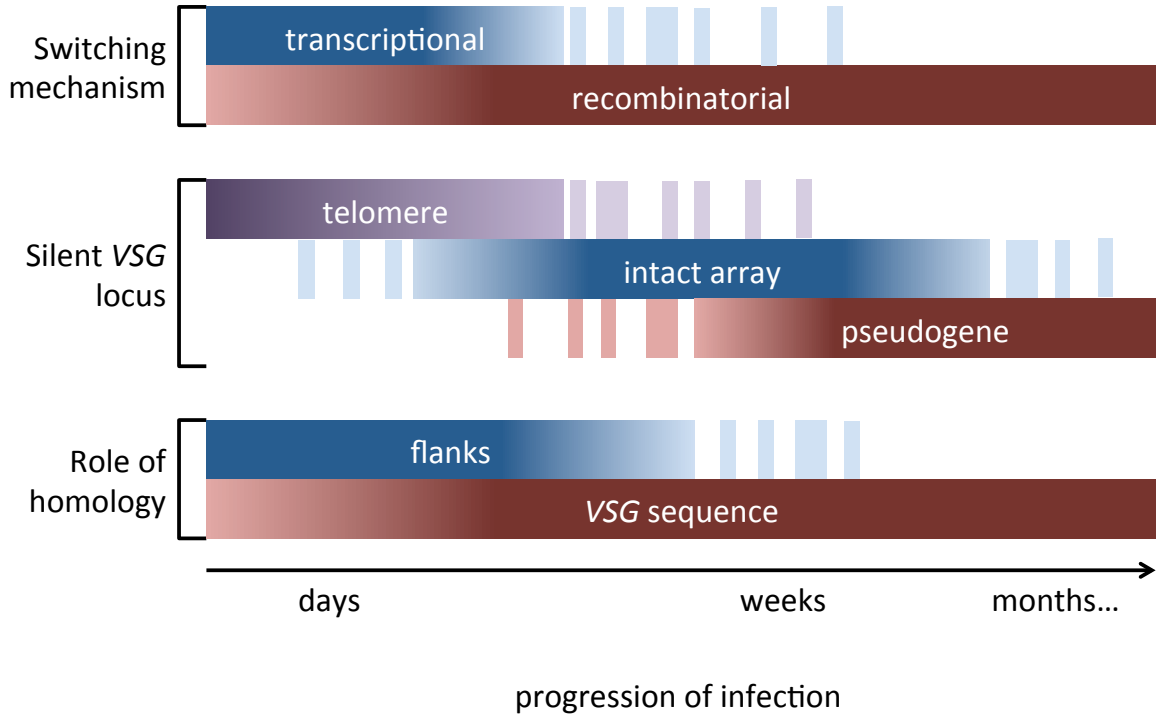


Figure 1.15 *VSG* switching hierarchy. The intensity of shading indicates the effective switching rate, that is, the relative importance of each process in expressing novel *VSG*s. Transcriptional switching is likely to occur in the early stages of infection (although this has not been experimentally verified, and transcriptional switching may provide a supporting role throughout, see section 1.6.10), whereas recombinatorial switching is used throughout. Telomeric *VSG*s are activated more readily than intact genes in the subtelomeric arrays, which in turn are activated more readily than pseudogenes in the arrays. In the early stages of infection, homologies in the flanks, such as the 70-bp repeats, mediate recombination, whereas as the infection progresses, segmental gene conversion requires homologies in the *VSG* coding sequence. The scale bar at the bottom is approximate: timing likely varies between hosts. Image copied and adapted from Morrison *et al.* (2009).

Hierarchy is also flexible, with previously 'late' *VSG*s able to occupy more easily-activated genomic locations and thus taking an earlier position in the hierarchy (Laurent *et al.*, 1984b). Such reorganisation may be important in field cases of

reinfection or superinfection, where hosts might already have antibodies against the common early-expressed *VSG*s.

These patterns described in section 1.6 are the parasite-intrinsic processes behind antigenic variation. The *raison d'etre* of antigenic variation is its interaction with the host. It is to this extrinsic process that we now turn.

1.7 Host-parasite interactions in African trypanosome infections

1.7.1 The VSG surface coat is the point of contact between parasite and host

To the immune system, a living bloodstream form trypanosome is essentially invisible but for its surface coat, a fluid, repetitive array of identical VSG subunits. How does the immune system 'see' VSG, and respond to it? The structure and orientation of VSG on the surface and its close packing makes the N-terminal loops of the glycoprotein the most likely interface with the immune system (Mehlert *et al.*, 2002). It is likely that the epitopes they form are conformational, complicating efforts to identify them by mutagenesis or tryptic digestion (Pinder *et al.*, 1987). Studies using monoclonal antibody competition to map the number and structure of the epitopes on an intact surface coat estimated that each VSG has between one and four distinct epitopes (Clarke *et al.*, 1987; Hall & Esser, 1984; Masterson *et al.*, 1988; Miller *et al.*, 1984a, b; Pinder *et al.*, 1987; Theodos *et al.*, 1990) and analyses using mosaic *VSG*s identified an epitope corresponding with a loop on the side of the VSG (Hsia *et al.*, 1996).

Besides the form of VSG on a live parasite, VSG can be released from the parasite surface (Black *et al.*, 1982; Seyfang *et al.*, 1990), and damaged or dying parasites are also likely to be present during an infection. The role that these forms of VSG play in infection outcome are unclear: soluble VSG appears not to bind anti-trypanosome antibodies strongly (Black *et al.*, 1982), but the activation of B-cells may be assisted by damaged trypanosomes, perhaps because the parasite's motility is compromised (Black *et al.*, 2010; Sendashonga & Black, 1982).

1.7.2 Antibodies against VSG drive the host response

Antibodies against VSG clearly play the central role in the host adaptive response to trypanosome infection (Campbell & Phillips, 1976; Guirnalda *et al.*, 2007; Magez *et al.*, 2008; Morrison *et al.*, 1982). The huge range of antibodies available to the mammalian immune system is thanks to a mechanism of recombination and mutagenesis across the developing B-cell population, selecting for immense variation within structural constraints (Tonegawa, 1983). Generation of antibody diversity is thus mirrored by the mechanisms that antigenically variable pathogens have evolved to evade it.

Antibody levels rise rapidly, related to parasite burden, and remain high over the course of infection (Barry & McCulloch, 2001). Decavalent IgM isotype antibodies are thought to be the initial response to a VSG, although bivalent IgG is also seen (Radwanska *et al.*, 2000; Taylor, 1998). Whilst the kinetics of the response against a specific VSG are probably variant-independent, there exists a lower bound of sensitivity, whereby low levels of a specific antigen are not able to trigger a corresponding antibody response (Morrison *et al.*, 1982): a phenomenon that may have consequences for the antigenic variation as minor variants may be able to persist undetected.

VSG-bound antibodies kill parasites by activating complement, which can either lyse parasites directly (Van Meirvenne *et al.*, 1995) or activate parasite-killing macrophages (Guirnalda *et al.*, 2007; Holmes *et al.*, 1979; Pan *et al.*, 2006). Antibody responses have been associated with spontaneous self-cure in livestock infections (Nantulya *et al.*, 1984; 1986; Penchenier *et al.*, 2005). Variant-independent responses, mediated by trypanotoxic cytokines (Magez *et al.*, 1999) and nitric oxide (Vincendeau *et al.*, 1992), have also been proposed; however, their *in vivo* effects are unclear and their overall contribution to the dynamics of infection is probably minor (Hertz & Mansfield, 1999).

1.7.3 Host factors influence infection outcome

Host-intrinsic factors have a powerful impact on the outcome of infection. Different species and breeds of animal host are differently affected by infection.

Some hosts are completely refractory, baboons, for example (Wheeler, 2010), whilst amongst others there is a broad spectrum of outcomes, ranging from exceptional susceptibility (horses and European cattle breeds, for example) to varying degrees of 'trypanotolerance', where symptoms are mild and the parasite population is maintained at very low, often undetectable numbers (Guirnalda *et al.*, 2007; Mulla & Rickman, 1988; Taylor, 1998)

Innate immunity clearly plays the pivotal role in refractory hosts, with a trypanolytic serum factor implicated (apolipoprotein L-1 (Vanhamme *et al.*, 2003)).[24] The spectrum of susceptibility and tolerance is harder to dissect, but using different species, breeds and genetic knockouts some broad patterns have been identified. Control of pathogenesis (where parasitaemia can be sustained with few ill effects) and control of parasites (where the parasite load is repressed) are likely linked through the inflammatory cytokines TNF-α and IFN-γ (Magez & Caljon, 2011). Higher levels of IFN-γ and TNF-α stimulate immunity and are required for controlling the initial parasite load (Hertz *et al.*, 1998; Magez *et al.*, 1999) but as the infection progresses the suppression of their activity through IL-10 is important to minimise immunopathogenesis (Guilliams *et al.*, 2007). Indeed, IL-10-negative mice are exceedingly susceptible to trypanosomiasis, probably due to immunopathogenic responses (Baetselier *et al.*, 2001; Namangala *et al.*, 2001). Although much of our knowledge of the immunology of trypanosomiasis comes from model systems, particularly mice, there are indications that these findings do have veterinary and medical relevance (Kennedy, 2007; Magez & Caljon, 2011).

1.7.4 African trypanosomes can induce immunosuppression in their hosts

Trypanosome infections are known to induce immunosuppression in their hosts, linked to the activity of suppressor macrophages whose release of immunomodulatory cytokines may skew the immune response away from productive antibody responses and towards immunopathogenesis (Askonas *et al.*,

[24] It is interesting to note that the resistance of *T. b. rhodesisense* to human trypanolytic factor is mediated by a VSG-like protein, encoded by the 'serum resistance associated' (*SRA*) gene located in the ES (DeGreef & Hamers, 1994; Campillo & Carrington, 2003; Xong *et al.*, 1998). The ES may therefore act as a general contingency locus (Pays et al., 2001) and the structure of the VSG fold itself may be one that is easily co-opted to perform other roles (Carrington & Boothroyd, 1996).

1979; Mansfield & Paulnock, 2005). In *T. brucei* infections of mice, considerable B-cell dysfunction causes serious disruption to antibody responses. As infections progress, trypanosomes induce apoptosis in marginal zone B-cells, rendering hosts susceptible to re-challenge with previously encountered antigens, including VSG coats (Radwanska *et al.*, 2008). It is likely that this phenomenon is unusually exaggerated in hosts that sustain an exceptionally high parasitaemia for their body mass, such as mice (La Greca & Magez, 2011): in other hosts B-cell dysfunction and its effects may be less extreme.[25]

1.7.5 Density-dependent differentiation controls parasite numbers and provides transmission-ready parasites

Alongside the extrinsic force of the immune system, trypanosomes have an intrinsic mechanism of population control. In contrast to monomorphic trypanosomes, pleomorphic trypanosome infections of immunocompromised animals do not rapidly overwhelm or kill their host (Magez *et al.*, 2008; Seed & Sechelski, 1988). Instead, parasitaemia is maintained at constant, high levels for many weeks, thanks to the irreversible differentiation to non-proliferative tsetse-infective 'short stumpy' form. Stumpy differentiation is critical for tsetse transmission—and hence trypanosome persistence in its natural habitat—therefore mechanisms that maximise and prolong the presence of stumpy forms are probably subject to strong selection. Infections routinely show an abundance of short stumpy trypanosomes (MacGregor *et al.*, 2011; Robertson, 1912). The soluble, parasite-produced signal for differentiation, 'stumpy induction factor' (SIF) is unknown, but it has a density-dependent effect on the parasite population (Reuner *et al.*, 1997). Short stumpy differentiation is therefore a negative feedback loop, which stabilises total parasitaemia, preventing early death of the host (Seed *et al.*, 2003) and providing an plentiful source of short stumpy form cells to maximise tsetse-infectivity.

Stumpy differentiation has important consequences for antigenic variation. Because they cannot proliferate, stumpy form trypanosomes do not change their VSG coat, indeed, *VSG* transcription is repressed within the limit of detectability

[25] If, on the other hand, absolute B-cell dysregulation is more widespread amongst hosts, the selection pressures favouring the evolution of the elaborate system of trypanosome antigenic variation would be mysterious, since parasites would only need to possess a handful of distinct antigens if they were able to abrogate immunological memory altogether.

(Amiguet-Vercher *et al.*, 2004). The total population of parasites undergoing *VSG* switching is therefore reduced to a smaller set of 'stem cell'-like proliferative slender form cells, potentially restricting antigenic diversity across the trypanosome population (MacGregor *et al.*, 2011). In addition, if the effects of SIF are variant-independent, stumpy differentiation could affect rarer *VSG* expressors as well as the predominant *VSG*. This non-specificity might repress the emergence of some variable antigen types, possibly allowing their persistence 'below the radar' of immune sensitivity (Gjini *et al.*, 2010).[26] The dynamics of differentiation are host-dependent, with the parasite population stabilising at different levels in different hosts (Seed & Sechelski, 1988): host processes may influence parasitaemia indirectly, perhaps by producing, sequestering, or removing SIF (or SIF analogues).

1.8 The dynamics of antigenic variation

1.8.1 Mathematical models help to describe relationships between processes in antigenic variation

A recent mathematical model by Gjini et al. (2010) has expanded on previous modelling attempts (Frank, 1999; Lythgoe *et al.*, 2007) to shed some light on the relationships between antigenic variation, *VSG* switching hierarchies, and (variant-independent) population control through density-dependent differentiation. Briefly, the model showed that as the number of easily-activated variants increases (relative to the sensitivity of the immune response) the greater the role of differentiation in controlling parasitaemia, since the number of different variants at low concentration cannot all induce sufficiently powerful specific immune responses for their rapid elimination. Similarly, the larger the size of the trypanosome population, the greater the likelihood that harder-to-activate variants will appear in an infection, since switching is parasite-intrinsic. As the balance of infection tilts towards differentiation-based control, opportunities for tsetse transmission increase, but so does the risk of premature host death as a consequence of persistent high parasitaemia (Seed *et al.*, 2003). These key interactions are summarized in Figure 1.16. By partitioning the archive into 'blocks' of distinct activation probabilities—an

[26] Similar effects might also result from variant-independent host immune responses, although without concomitant production of the tsetse-infective form.

expression hierarchy—a hybrid strategy is adopted: peaks of parasitaemia dominated by differentiation, punctuated by periods of remission. This model is able to broadly recreate experimentally observed patterns of infection (Barry, 1986; Morrison *et al.*, 2005).

Another interesting prediction of the model is that partial cross-reaction between variants can actually increase the period that the later variant is present by preventing those parasites from reaching the numbers required to trigger a late-variant-specific immune response. This could have special relevance in progressive mosaicism, where antibodies raised against an early variant may be inefficient at (but still capable of) killing a later, related mosaic (Kamper & Barbet, 1992).

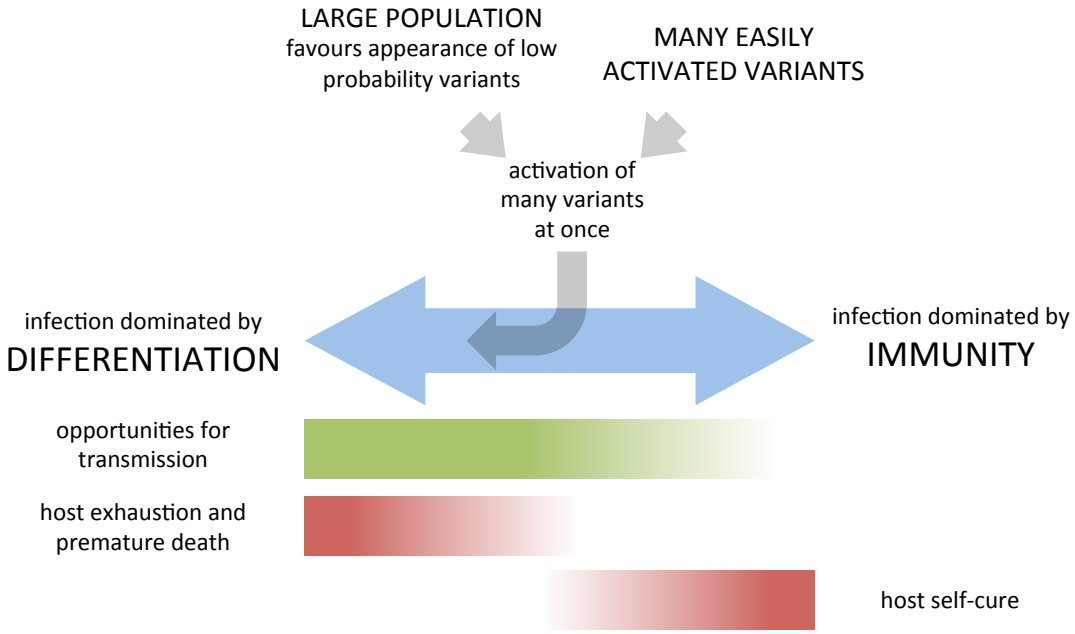


Figure 1.16 Dynamics of antigenic variation. As described in the text, the activation of many variants means that the dynamics of a trypanosome population are controlled primarily by differentiation. Processes that favour and disfavour trypanosome persistence in an ecosystem are indicated by green and red bars respectively. The opportunities for transmission, and probabilities of host exhaustion or self-cure, are indicated by the intensity of shading.

1.8.2 Antigenic variation exists in a complex ecosystem

From a parasite point of view, host responses, and host effects on differentiation, are unknown quantities that can vary considerably between hosts. These forces set the boundaries within which antigenic variation and differentiation interact (Barry, 1986). A broad ecological picture of African

trypanosomes reveals an environment where competing parasite strains straddle geographically discontinuous populations of mammals and tsetse flies (Krafsur, 2008; Munang'andu *et al.*, 2012). Trypanosomes' obligatory parasitism demands flexibility in their survival strategies. Their system of antigenic variation has evolved to reflect these demands, and should be considered in this context (Marcello & Barry, 2007b).

1.9 Outstanding questions and aims of the project

1.9.1 Research question

Segmental gene conversion remains a somewhat elusive process in African trypanosomes. It is known to occur later in infection, but only a few mosaic, expressed *VSG*s have been identified, often following immunological selection, and almost all of these from separate infections. The facts that pseudogenicity dominates the *VSG* archive (Marcello & Barry, 2007a), and that segmental gene conversion is capable of generating sufficient antigenic novelty to sustain chronic infections of *A. marginale* and *B. burgdorferi* (Coutte *et al.*, 2009; Zhuang *et al.*, 2007) indicates that segmental gene conversion may play a prominent role during African trypanosome infections, particularly when the host has mounted responses to easily-activated *VSG*s.

The question at the centre of this project is: what is the contribution of segmental gene conversion to *T. brucei* antigenic variation? Does segmental gene conversion readily access damaged archival *VSG*s and introduce antigenic variability into expressed *VSG*s? Answering these questions requires a broad knowledge of the *VSG*s expressed during infection, and how they relate to the silent archive. To investigate whether segmental gene conversion contributes directly to antigenic variability during an infection, the changes in epitope structure caused by segmental gene conversion also need to be understood.

1.9.2 Experimental approach

To do this, I monitored *VSG* expression in individual infections over time. Longitudinal samples of trypanosomes were taken from several mouse infections, and the sequences of *VSG*s they were expressing were retrieved. These samples

focused on the later stages of infection, where segmental gene conversion was previously identified as a prominent process in antigenic variation (Marcello & Barry, 2007a). Expressed sequence data were compared with the genome sequence and with one another to get a general understanding of *VSG* variability during an infection and how it changes over time (Chapter 3). Closer analyses of the *VSG*s sequenced allowed inference of segmental gene conversion events and possibly the identification of progressive mosaicism occurring within an infection (Chapter 4). Different-but-related mosaics retrieved from a chronic infection were subjected to serological analysis to investigate whether mosaicism could contribute directly to antigenic variation (Chapter 5). The results will be discussed in the context of other host and parasite processes, as well as the broader evolution of antigenic variation, in Chapter 6.

Chapter 2

Materials and Methods

2 Materials and Methods

2.1 Procedures related to running chronic mouse infections and growing trypanosomes *in vivo*

2.1.1 Trypanosome strains used

Two different parasite strains were handled: bloodstream form TREU 927/4 GUTat 10.1 (Turner *et al.*, 1990) *in vivo*, and bloodstream form Lister 427 13-90 (Wirtz *et al.*, 1999) *in vitro* and *in vivo*. The strains were kindly given by L. Marcello and C. Harkins respectively.

2.1.2 Host immunosuppression

Immunosuppression of Balb/c or ICR mice to assist trypanosome growth (when initiating infections from blood straw stabilate, for example) was performed by intraperitoneal administration of cyclophosphamide (250 $mg.kg^{-1}$ body weight) 1-24 h prior to trypanosome injection. Cyclophosphamide solution was made (Sigma-Aldrich, Gillingham, UK, cat. no. C7397-1G) in phosphate buffered saline (PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4{\cdot}2H_2O$, 1.76 mM KH_2PO_4, pH 7.2) and passed through a 0.2 µm pore size syringe filter. Single-use ampoules were prepared and stored at -20ºC for up to three months.

2.1.3 Trypanosome growth and collection

All infections were initiated by intraperitoneal injection of parasites in a volume no greater than 0.5 ml. Parasites in blood straw stabilates were first thawed in approximately 100 µl incomplete HMI-9 medium (see section 2.3.1) before injection. Parasites from cell culture were collected by centrifugation at 320 *g* for 10 mins and resuspended in 150 µl complete or incomplete HMI-9 for injection. Where necessary, parasites from blood were diluted using incomplete HMI-9.

To prepare parasites for clonal infections, trypanosomes were optically cloned. All materials and samples were pre-cooled to 4ºC to prevent evaporation, and the cloning procedure was carried out in a 4ºC cold room. Blood containing trypanosomes was diluted iteratively with trypanosome dilution buffer (TDB; 5

mM KCl, 80 mM NaCl, 1mM $MgSO_4$, 20mM Na_2HPO_4, 2mM NaH_2PO_4, 10 mM glucose, pH 7.4 (Cross, 1975)) to achieve a working concentration. A Terasaki plate (Greiner, Stonehouse, UK) was prepared by lining its edges with tissue wet with deionised H_2O, and a paperclip end was used to place one small drop of the trypanosome suspension into each well. The drops were examined by phase-contrast microscopy at 400x magnification. Those wells containing drops holding single trypanosome were verified by a second party, flooded with 10 µl complete HMI-9, and the contents transferred to a microfuge tube containing 150 µl complete HMI-9 for immediate injection.

2.1.4 Collection of blood samples from infections

To monitor parasitaemia, small (~1 µl) blood samples were collected by tail venepuncture and a thin film prepared by mounting the sample on a slide beneath a glass coverslip. By observation at 400 x magnification, parasitaemia could be estimated according to the 'rapid matching' method of Herbert & Lumsden (1976). Where more accurate counts were required (for example when initiating infections) blood was diluted 1:20 with 0.85% w/v ammonium chloride and incubated at room temperature for 10 mins to lyse erythrocytes. The number of parasites in the sample could then be counted using an Improved Neubauer Haemocytometer (Hawksley, Lancing, UK).

For larger samples (for extraction of mRNA, gDNA or plasma), blood was collected from ongoing infections by tail venesection, or from terminal infections by exsanguination by cardiac puncture under terminal anaesthesia. Blood was collected in Carter's Balanced Salt Solution (CBSS; 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, pH 7.4) supplemented with 5% w/v tri-sodium citrate as an anticoagulant. For on-going infections, no more than 10% of blood volume was collected at any one time, and no more than 15% of blood volume was collected over a 28 day period. Blood volume was calculated as ~5.5 ml per 100 g of body weight (Eppig *et al.*, 2012).

2.1.5 Immunization of animals for the purpose of raising antibodies

The standard procedure for infection to raise antibodies was as follows: 1 x 10^6 cells were injected intraperitoneally into a Balb/c mouse. When the parasitaemia exceeded antilog 7.2 ml^{-1} blood (scored by rapid matching), mice were cured by a dose of 20 $mg.kg^{-1}$ cymelarsen. Cymelarsen solution was prepared prior to injection by dissolving the powdered drug (Rhône Mérieux, now Merial, Duluth GA, U.S.A.) in PBS and passing through a 0.2 µm pore size syringe filter. Plasma and splenocytes were retrieved and processed 5–7 days after drug cure.

2.1.6 Stabilates from blood samples

Stabilates were prepared by mixing blood 2:1 with 22.5% v/v dimethyl sulphoxide (DMSO) in CBSS as a cryopreservant. In seven cases, triladyl (Minitüb, Tiefenbach, Germany, cat. no. 13500/0250) was used as a cryopreservant, by mixing blood 1:1 with 25% v/v triladyl in CBSS. In either case, the blood-cryopreservant mixture was injected into fine-bore polyethylene tubing, cut to size to form 'straws', and placed into a cryotube (Alpha Labs, Eastleigh, UK, cat. no. LW3532) perforated to allow contact between the liquid nitrogen and the straws. Stabilate tubes were insulated by wrapping in cotton wool and transferred to a -80ºC freezer overnight. Frozen stabilates were then transferred to a liquid nitrogen tank for long-term storage.

2.2 Preparation of blood samples

2.2.1 Preparing plasma from blood samples

To purify plasma from total blood, blood samples were centrifuged at 14,000 *g* for 10 mins. The supernatant was transferred into a clean microfuge tube and again centrifuged at 14,000 *g* for 10 mins. The supernatant was transferred into a further clean microfuge tube and sodium azide was added as a preservative, to a final concentration of 0.05% w/v.

2.2.2 Erythrocyte lysis

For the extraction of gDNA or RNA from blood-borne parasites, the erythrocytes present in the blood sample were first osmotically lysed. Five volumes of Erythrocyte Lysis Buffer (ELB; Qiagen, Crawley, UK, cat. no. 79217) were added to one volume of blood, and the sample incubated on ice for 15 mins. Every five mins the sample was mechanically agitated by inversion. The sample was centrifuged at 320 *g* for 10 mins and the supernatant discarded. The pellet was resuspended in 400 µl ELB and collected by centrifugation again at 320 *g* for 10 mins. The supernatant was again discarded, at which point the pellet was ready for gDNA or RNA extraction.

Retrieving both parasites and plasma from the same blood sample was possible through an initial centrifugation at 320 *g* for 10 mins. The supernatant fraction, containing the plasma, could be purified according to section 2.2.1 above; the pellet could be resuspended in five volumes of the original blood sample volume of ELB and processed according to the protocol in the previous paragraph.

Parasites prepared in this manner were subjected to nucleic acid extraction as described in sections 2.4.1 and 2.4.11 below.

2.3 *In vitro* culture and manipulation of trypanosomes

2.3.1 Culturing trypanosomes

In vitro culture of bloodstream form Lister 427 13-90 trypanosomes was achieved by using 'complete' HMI-9 medium (Hirumi & Hirumi, 1989). Pre-formulated HMI-9 powder was provided by Gibco (Paisley, UK, cat. no. 074-90915N) and supplemented with 3% w/v $NaHCO_3$ and 200 µM beta-mercaptoethanol to produce 'incomplete' HMI-9; medium could be completed by the replacement of 20% v/v with foetal bovine serum (FBS; Invitrogen, Paisley, UK, cat. no. 10108-157). Medium was sterilised by filtration through a 0.2 µm bottle-top filter and incomplete medium was stored at 4ºC for up to six months; complete medium for no longer than a week. When working with parasites collected from *in vivo* experiments, medium was supplemented with penicillin-streptomycin (Sigma-Aldrich, Gillingham, UK, cat. no. P4333) at a 1:1000 dilution. Where

appropriate, the growth medium was supplemented with drugs for selection: hygromycin (Source Bioscience, Nottingham, UK, cat. no. ant-hg-1) at 5 $\mu g.ml^{-1}$, neomycin (Sigma, Gillingham, UK, cat. no. G8168) at 2.5 $\mu g.ml^{-1}$, puromycin (Merck, Darmstadt, Germany, cat. no. 540222-100MG) at 1 $\mu g.ml^{-1}$, blasticidin (Merck, cat. no. 203351-10ML) at 5 $\mu g.ml^{-1}$. Cultures were kept at 37ºC in a humidified 5% CO_2 incubator, and subcultured regularly to avoid over-growth (the cell density was not permitted to exceed 2.5 x 10^6 ml^{-1}).

2.3.2 Transfecting trypanosomes

Transfections were carried out using the AMAXA Nucleofactor apparatus (Lonza, Basel, Switzerland), following a protocol provided online by G. Cross (http://tryps.rockefeller.edu/Protocols/transfection_amaxa.pdf). 90 ml of complete HMI-9 was supplemented with the drugs appropriate for parental cell growth and divided amongst three tubes, so that the first tube (A) contained 30 ml of medium and the two other tubes (B & C) contained 27 ml of medium. The medium in the tubes was prewarmed to 37ºC. Approximately 30 x 10^6 cells were harvested by centrifugation at 320 *g* for 10 mins at room temperature, and resuspended in 100 µl Amaxa Human T-cell Nucleofactor solution (Lonza, cat. no. VPA-1002) at 4ºC. The cells were added to a cuvette alongside 10 µl of linearized plasmid DNA in distilled water (0.5–1 $\mu g.\mu l^{-1}$ concentration) and transfected according to Program X-001. The contents of the cuvette were transferred to tube A of prewarmed medium and mixed well. From tube A, 3 ml of the cell suspension were transferred to tube B and mixed well (giving a 1:10 dilution). From tube B, 3 ml of the cell suspension were transferred to tube C and mixed well (giving a 1:100 dilution). The contents of each tube were spread across the wells of a 24 well tissue culture plate, giving three 24 well plates with 1 ml per well. Between six and 18 h later, 75 ml complete HMI-9 was supplemented with the drugs appropriate for parental cell growth, and the selective drug at double the normal concentration, and pre-warmed to 37ºC. 1 ml of the selective medium was added to each well of the three plates. The wells were inspected for growth six days after transfection, and the clonality of the transfectants could be estimated by counting the number of positive wells per plate with reference to the Poisson distribution (fewer than 30% of wells showing growth indicates that >80% of the positive wells were seeded with only

one drug-resistant transfectant). Where necessary parasites could be recloned as described below.

2.3.3 Cloning trypanosomes

Trypanosomes were cloned by diluting parasites to ~0.5 ml^{-1} in fresh complete HMI-9 supplemented with the appropriate drugs, and spreading across the wells of a 96 well plate, 200 µl per well. Wells were inspected for growth 6 days after cloning, and the clonality of lines could be estimated as described in section 2.3.2 above.

2.3.4 Preparation of cultured trypanosomes for molecular biology procedures

Parasites were collected from a culture in log phase growth and pelleted at 320 *g* for 10 mins. The supernatant was discarded and the pellet was washed in 1 ml PBS or TDB, and pelleted again at 320 *g* for 10 mins. On discarding the supernatant, the pellet was ready for use.

2.3.5 Stabilates of in vitro cultures

In vitro stabilates were prepared from cultures in the log phase of growth, with a density between 1–2 x 10^6 parasites ml^{-1}. Glycerol, previously sterilized by autoclaving, was added as a cryopreservant to a final concentration of 10% v/v and the suspension was transferred to a cryotube (Alpha Labs, Eastleigh, UK, cat. no. LW3532) and frozen as described in section 2.1.6.

2.4 Basic laboratory procedures

2.4.1 RNA isolation

The cell pellets obtained as described above (sections 2.2.2 and 2.3.4) were resuspended in 350 µl Buffer RLT (provided with the kit) supplemented with 1% v/v beta-mercaptoethanol and processed using the RNeasy kit (Qiagen, Crawley, UK, cat. no. 74104) according to the manufacturer's instructions. Approximately 0.5–1 x 10^7 cultured parasites were used for RNA extraction. DNA digestion was carried out using the on-column DNase kit (Qiagen, cat. no. 79254). RNA yield

and quality were assessed photospectometrically using a Nanodrop 1000 (Thermo Scientific).

2.4.2 cDNA synthesis

RNA (maximum 1 µg) was used as the template for cDNA synthesis. cDNA synthesis was carried out using the SuperScript III First Strand Synthesis System kit (Invitrogen, Paisley, UK, cat. no. 18080-051) according to the manufacturer's instructions. For each sample, a negative control to which no reverse transcriptase was added was also performed: this would exclude the possibility that downstream PCR products were amplified from residual contaminating genomic DNA. For all cDNA synthesis reactions the primer used was oligo$[dT]_{20}$.

Those cDNA samples from which full-length *VSG* sequences would be amplified (as described in 2.4.3 below) were purified using a column purification kit (PCR Purification Kit, Qiagen, Crawley, UK, cat. no. 28104).

2.4.3 Polymerase Chain Reaction (PCR)

To amplify full-length *VSG* sequences for cloning and sequencing, the proofreading polymerase Herculase II Fusion (Agilent, Wokingham, UK, cat. no. 600675) was used to maximise sequence accuracy. For this reaction, primers corresponding to the spliced leader and the conserved 3' 16-mer were used (Appendix 7.1). Reactions comprising of 2–4 µl of cDNA, 2 µl of each primer at a working concentration of 10 µM, 1 µl of a dNTP mix (each dNTP at 10 mM), 10 µl 5x Herculase Reaction Buffer (provided with enzyme kit), 5 µl MasterAmp 10x PCR Enhancer (from Cambio, Cambridge, UK, cat. no. ME81201), 1 µl Herculase II Fusion enzyme, and made up to 50 µl with deionized H_2O. Deionized H_2O (dH_2O) was provided by a MilliQ purification machine (Millipore, Billerica MA, USA). Reactions were cycled using a Robocycler PCR machine (Stratagene); reaction conditions were 5 mins at 95ºC, followed by 30 cycles of 1 min at 95ºC, 2 mins at 40ºC and 2 mins at 68ºC, and a 5 min final extension at 68ºC. Products were analysed by agarose gel electrophoresis and purified by gel extraction.

For other PCR reactions, *Taq* polymerase provided by New England Biolabs was used (NEB, Hitchin, UK). Each 25 µl reaction contained template, 1 µl of each

primer from a working concentration of 10 mM, 0.5 µl of a dNTP mix (each of dATP, dCTP, dGTP and dTTP at 10 mM), 2.5 µl of 10x Thermopol reaction buffer, 0.25 µl *Taq* enzyme at 5 $U.\mu l^{-1}$ and made up to 25 µl with dH_2O. Reaction conditions were generally as follows: 5 mins at 95ºC, followed by 30 cycles of 50 seconds at 95ºC, 50 seconds at an annealing temperature between 50 and 63ºC, 1 min at 68ºC, and a 5 min final extension at 68ºC. The annealing temperature used would vary according to each primer pair; for most reactions it was 50ºC.

All primers for PCR were provided by Eurofins MWG Operon (Ebersburg, Germany) and their sequences can be found in Appendix 7.1.

2.4.4 Restriction digestion

Restriction digestions were performed using enzymes supplied by New England Biolabs (Hitchin, UK) in the recommended buffers provided. Usually, 10 units of enzyme were added per µg of DNA, and reactions were incubated at 37ºC for 1–2 h. Where necessary, restriction digests were purified by gel extraction, column purification, or phenol-chloroform extraction.

2.4.5 Annealing primers

To generate the *Sbf*I sites required in the pVSG plasmid (Chapter 5), complementary oligonucleotides containing the *Sbf*I site flanked by *Hin*dIII and *Eco*RI sites were ordered from Eurofins MWG Operon (see Appendix 7.1 for details). 4 µl of each oligonucleotide (100 µM concentration) were added to 58.67 µl dH_2O and 133.33 µl 1.5x annealing buffer (100 mM Tris HCl pH 7.5, 1 M NaCl, 10 mM EDTA pH 8.0). The reaction was incubated as follows: 85ºC for 2 mins, 65ºC for 15 mins, 37ºC for 15 mins, 4ºC for 15 mins. The products were electrophoretically separated on a 4% agarose gel (Lonza, Basel, Switzerland, cat. no. BMA 50101), and purified by gel extraction.

2.4.6 Agarose gel electrophoresis and gel extraction

Separation of DNA by electrophoresis was achieved by loading on 1% agarose gels (Invitrogen, Paisley, UK, cat. no. 15510-027) run at 30–120 V in 1 x TAE buffer (40 mM Tris, 19 mM acetic acid, 1mM EDTA). For routine analysis of DNA, 1%

w/v agarose in 1 x TAE buffer gels were used. SYBR-Safe DNA stain (Invitrogen, cat. no. S33102) was used at 1:10,000 dilution to allow visualization of DNA under ultraviolet light. Gels were visualised using a Bio-Rad GelDoc (Bio-Rad). Images were collected using Quantity One (Bio-Rad) contrast adjusted using the GNU Image Manipulation Program (GIMP Development Team, gimp.org).

Where separated DNA was to be extracted from the gel, a sterile scalpel blade was used to carefully excise the desired band and place it into a sterile microfuge tube. It was processed using the Qiagen Gel Extraction Kit (Qiagen, Crawley, UK, cat. no. 28706) according to the manufacturer's instructions.

2.4.7 Southern blotting

To blot DNA from an agarose gel, the gel was prepared by incubating in 125 mM HCl for 10–15 mins, rinsing in dH_2O, and soaking in denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 30 mins. The gel was then transferred to neutralising solution (1 M Tris base, 1.5 M NaCl, 186 mM HCl) and incubated for 30 mins. The gel was then soaked in 20x SSC solution (3 M NaCl + 0.3 M sodium citrate). All incubations were carried out at room temperature. DNA was blotted from an agarose gel onto a nylon membrane (Hybond-XL, GE Healthcare, Chalfont St Giles, UK, cat. no. RPN303S) by capillary action overnight. DNA was crosslinked to the membrane by ultraviolet radiation (UV Stratalinker 2400, Stratagene, two rounds of the ‘auto-crosslink’ program).

For the preliminary dot blots, the membrane was prepared by soaking in 6x SSC solution (diluted 20x SSC in dH_2O) for 10 mins. Each plasmid sample was adjusted to give a final concentration of 6x SSC, 0.4 M NaOH, 10 mM EDTA, boiled at 100ºC for 10 mins and placed on ice. Each sample was spotted onto the wetted membrane, 2 µl at a time. The membrane was dried, rinsed briefly in 2x SSC, and air dried before crosslinking as with gel blots.

Non-radioactive probes were made from PCR amplified DNA using the Amersham CDP-Star kit (GE Healthcare, cat. no. RPN3680), following the manufacturer's instructions. This kit utilises the activity of a probe-bound thermostable alkaline phosphatase which catalyses the breakdown of a dioxetane substrate, releasing light which is captured on an X-ray film. The kit instructions were followed for

hybridization of the probes to the blots, washing the blot, and application of the detection reagent. Blots were wrapped in Saran wrap (Saran, Racine WI, U.S.A.) and exposed to X-ray film (Kodak, Geneva, Switzerland, cat. no. 8143059) in light-sealed cassettes. Films were developed using a Konica SRX101A auto-developer machine.

2.4.8 DNA plasmid cloning

The full-length *VSG* sequences were cloned using a TOPO TA cloning kit (Invitrogen, Paisley, UK, cat. no. K4500-01). Because Herculase II Fusion has proofreading capability, their products required overhanging deoxyadenosine (A) residues to be added by treatment with *Taq* polymerase before cloning. 8.7 µl of gel extracted PCR product was added to 1 µl 10x Thermopol buffer, 0.2 µl dNTP mix and 0.1 µl *Taq* polymerase (reagents as used for PCR, section 2.4.3). The reaction was incubated at 70ºC for 15 mins. These products were TOPO cloned without any additional purification step. TOPO TA cloning was carried out according to the manufacturer's instructions, with a 15 min room temperature ligation reaction for maximal diversity of cloned PCR products. These ligation reactions were used to transform chemically competent TOP10 *Escherichia coli* cells (provided with the kit) as described in section 2.4.9.

For the ligation reactions carried out in the generation of the pVSG plasmids, plasmid backbones were first treated with calf intestinal phosphatase (CIP, acquired from New England Biolabs, Hitchin, UK, cat. no. M0290S) to remove reactive phosphate groups that could contribute to plasmid recircularization. 0.5 units of CIP were added to each µg of DNA in 1x NEBuffer 3 (supplied with the enzyme) and the reaction was incubated at 37ºC for 1 hour. For ligation, purified digested DNA of both insert and vector were combined at a molar ratio of 3:1 insert:vector in 8 µl dH_2O. 10 x ligase buffer (provided with the enzyme) and 1 µl 400 $U.\mu l^{-1}$ T4 DNA ligase (New England Biolabs, cat. no. M0202S) were added to the reaction. Ligation reactions were usually incubated for one hour at room temperature before being used to transform chemically competent *E. coli* cells as described in section 2.4.9.

2.4.9 Escherichia coli transformation

For the products of TOPO cloning reactions, TOP10 *Escherichia coli* cells were used (provided with the cloning kit). For general transformations, either DH5-α or EL-2925 *Dam- E. coli* cells were used. For both cell types the transformation reaction was the same. 2–5 µl of ligation reaction, or ~25 pg of purified plasmid was added to a vial containing approximately 150 µl competent cells and incubated on ice for 25 mins. The vial was then heat shocked at 42ºC for 40 seconds. The cells were left to recover on ice for approximately five mins, before 250 µl SOC recovery medium (5 g yeast extract, 20 g tryptone, 0.5 g NaCl, 10 ml 1M $MgCl_2$, 10 ml 2M glucose, 10 ml 1M $MgSO_4$ in 1 L), was added and the cells incubated at 37ºC in a shaking incubator for one hour to allow antibiotic resistance gene expression. Transformed cells were spread on LB agar plates containing 100 $\mu g.ml^{-1}$ ampicillin (Sigma, Gillingham, UK, cat. no. A9518). LB agar was made using L-Broth (see section 2.4.10) to which 20 $g.L^{-1}$ agar had been added. The volume of transformed cells spread was adjusted to ensure convenient spacing of colonies. Where blue-white screening was required (to test for insertion in the TOPO clones, for example), the plates were prepared beforehand by spreading 40 µl of a 40 $mg.ml^{-1}$ X-gal (Melford, Ipswich, UK, cat. no. 7240-90-6) in dimethylformaldehyde (DMF) solution and incubating at 37ºC for one hour. Plates spread with transformants were incubated at 37ºC overnight and stored inverted at 4ºC.

2.4.10 Plasmid isolation

Colonies containing the desired plasmid were used to inoculate 6 ml overnight cultures of lysogeny broth (L-Broth: 5 g yeast extract, 10 g tryptone, 10 g NaCl in 1 L H_2O, autoclaved) supplemented with 100 $\mu g.ml^{-1}$ ampicillin. Transformants containing desired plasmids were generally identified by colony PCR. A number of colonies could be picked, and each resuspended in 30 µl dH_2O. 10 µl of this resuspension was used as a template for PCR using appropriate primers, the remainder, if the PCR gave a positive result, was used to inoculate a 6 ml overnight culture. Where larger plasmid yields were required, multiple 6 ml overnight cultures inoculated with the same colony were set up. Approximately 18 h later, the cells in the overnight culture were collected by centrifugation at 2000 *g* for 10 mins and processed using a small-scale plasmid purification kit

(Miniprep kit, Qiagen, Crawley, UK, cat. no. 27106, carried out according to the manufacturer's instructions). Plasmid yield and quality was routinely assessed photospectrometrically using a Nanodrop 1000 (Thermo Scientific).

2.4.11 Genomic DNA extraction

Cell pellets obtained as described above (section 2.1.3) were resuspended in 500 µl trypanosome lysis buffer (50 mM Tris pH 8, 1 mM EDTA, 100 mM NaCl, 1% SDS) containing 100 µg.ml^{-1} proteinase K (Invitrogen, Paisley, UK, cat. no. 25530015) The reaction was incubated at 37ºC for 24–48 h, before proceeding with phenol-chloroform extraction and ethanol precipitation. Genomic DNA (gDNA) yield and quality was assessed photospectrometrically using a Nanodrop 1000 (Thermo Scientific). gDNA was stored in Tris-EDTA buffer (TE Buffer: 10 mM Tris HCl, 1 mM EDTA, pH 8.0) at 4ºC or -20ºC.

2.4.12 DNA purification by phenol-chloroform extraction and ethanol precipitation

To purify genomic DNA or digested plasmid DNA for transfection, phenol-chloroform extraction was carried out. One volume of phenol:chloroform:isoamyl alcohol 25:24:1 (Sigma-Aldrich, Gillingham, UK, cat. no. P3803) was added to each volume of DNA and mixed well by vortexing. The mixture was centrifuged at 14,000 *g* in a benchtop centrifuge for 10 mins and the aqueous phase carefully retrieved by Gilson pipette and transferred into a clean sterile microfuge tube. One volume of phenol:chloroform:isoamyl alcohol was added to the retrieved aqueous layer, the mixture vortexed and again centrifuged at 14,000 *g* in a bench- top centrifuge for 10 mins, and the aqueous (top) phase again taken into a fresh microfuge tube. Phenol-chloroform extraction was followed by ethanol precipitation.

Ethanol precipitation was used to purify DNA from salt and solvent contaminants, and to concentrate DNA (for transfection, for example). Two volumes of ethanol precipitation mix (made from 40 ml 100% ethanol and 2 ml 3 M sodium acetate and supplemented with 20 µl 20 mg.ml^{-1} glycogen (Fermentas, St. Leon-Rot, Germany, cat. no. R0561)) was added to the sample and the mixture incubated at -80ºC for at least one hour, promoting precipitation of the

DNA. The precipitated DNA was pelleted by centrifugation at 14,000 *g* for 30 mins in a 4ºC chilled centrifuge, and the supernatant discarded. The DNA was washed in ice-cold 70% ethanol in deionized H_2O and again pelleted by centrifugation. The supernatant was removed completely and the pellet left to dry at room temperature. It was then dissolved in either dH_2O or TE buffer.

2.4.13 SDS-PAGE

For the analysis of crude cell lysates in Chapter 5, SDS-PAGE was used to separate the proteins and determine the relative molecular weight of the variant band. SDS-PAGE was carried out using precast NuPage 10% Bis-Tris acrylamide gels in 1x MOPS SDS running buffer (Invitrogen, Paisley, UK, cat. nos. NP0301BOX and NP0001 respectively). Crude cell lysates were made by washing a cell pellet twice in TDB, and resuspending it in protein loading buffer (PLB, 2% SDS w/v, 0.05M Tris, 10% glycerol v/v, 0.1% bromophenol blue w/v for 1x), to a final concentration of 1x, mixing well, and boiling at 100ºC for 10 mins. Aggregates in the samples were collected by centrifugation at 14,000 *g* immediately before loading the samples to prevent the bands from running unevenly. Approximately 5 x 10^6 cell-equivalents were loaded in each lane. Gels were photographed as described in section 2.4.6 above.

For those bands subjected to tryptic digest mass spectrometry analysis, the indicated bands were excised from the gel using a clean, sterile scalpel and placed in a sterile microfuge tube. Mass spectrometry was carried out by Richard Burchmore at the University of Glasgow facility, and peptide masses were used as queries in the MASCOT database (www.matrixscience.com).

2.4.14 Sequencing

All sequencing was performed by the Sequencing Service at the University of Dundee (Dundee, UK) and Eurofins MWG Operon (Ebersburg, Germany).

2.5 Generation of hybridoma lines for the production of monoclonal antibodies

2.5.1 Media for the generation and maintenance of hybridomas

RPMI (Roswell Park Memorial Institute) incomplete medium was used to wash cells and as the basis of cell culture medium. Incomplete RPMI consisted of RPMI 1640 (Gibco, Paisley, UK, cat. no. 31870-025) supplemented with 25 mM HEPES at a pH of 7.2. Batches of this medium were prepared and filter sterilised, and maintained at 4ºC for up to six months.

RPMI complete medium was used as the growth medium for myeloma and hybridoma cells. To produce 500 ml of this medium, 445 ml of incomplete RPMI were supplemented with 250 µl 0.1 M beta-mercaptoethanol, 5 ml 200 mM L-glutamine (Sigma-Aldrich, Gillingham, UK, cat. no. G7513-100ML) and 50 ml fetal bovine serum (Sigma-Aldrich, cat. no. F9665). Complete medium was stored at 4ºC for no more than 7 days.

RPMI-HFCS complete medium was used to promote the growth of hybridoma cells at low concentration (for example after fusion or cloning). To produce 100 ml of this medium, 2 ml hybridoma fusion and cloning supplement (HFCS, Roche, Mannheim, Germany, cat. no. 11 363 735 001) were added to 97 ml incomplete RPMI supplemented with 1 ml 200 mM L-glutamine and 50 µl 0.1 M beta-mercaptoethanol. Where required, antibiotics (0.2 ml 10 $mg.ml^{-1}$ gentamycin sulphate in dH_2O (Sigma-Aldrich, cat. no. G1264-250MG)) and antifungals (0.8 ml 250 $\mu g.ml^{-1}$ amphotericin B (Sigma-Aldrich, cat. no. A2942-100ML)) could replace an equivalent volume of incomplete RPMI.

HAT (hypoxanthine-aminopterin-thymidine) medium was used as the selection medium for hybridoma cells. Samples of HAT 50x concentrate (MP Bio, Illkirch, France, cat. no. 091680849) were stored at -20ºC. To produce 1x HAT, 2 ml of 50x concentrate were added to 98 ml RPMI-HFCS complete medium.

Polyethylene glycol (PEG) solutions were used to induce cell fusion. PEG of molecular weight 1500 was obtained from MP Bio (cat. no. 151915). Two solutions, A and B were made up as follows. Solution A, 50 grams of PEG 1500

were dissolved in 55 ml RPMI incomplete medium. 15 ml dimethyl sulphoxide (DMSO) was added, and the solution sterilised by passing through a 0.2 µm pore size syringe filter. This gave 41.6% PEG/15% DMSO. Solution B, 50 grams of PEG 1500 were dissolved in 150 ml RPMI incomplete medium and the solution sterilised as with solution A. This gave 25% PEG. PEG solutions were stored at 4ºC for no more than three months.

2.5.2 Myeloma cell lines

All hybridoma lines generated were produced using the p3-x63-ag8 clone 653 myeloma cell line (source stabilate prepared by S. Terry, University of Glasgow). Stabilates of this cell line were kept frozen in liquid nitrogen at a concentration of 3×10^6 ml^{-1} in 90% fetal bovine serum/10% DMSO until ready for use. One week before fusion, an ampoule of cells was removed from liquid nitrogen and heated to 37ºC. The contents were added to 8 ml RPMI incomplete medium and centrifuged at 200 *g* for 7 mins. The cell pellet was resuspended in 5 ml RPMI complete medium, transferred to a tissue culture flask and incubated at 37ºC, 5% CO_2 in a humidified atmosphere. Cells were examined daily and regularly subcultured to keep them in log phase growth at approximately $1–5 \times 10^5$ $cells.ml^{-1}$. Viable cell counts were performed by trypan blue exclusion (1:1 cell suspension:trypan blue; trypan blue acquired from Fluka, Gillingham, UK, cat. no. 93595-50ML). For fusion, 10^7 myeloma cells were required, which were obtained from up to 100 ml of cells in log phase growth.

2.5.3 Fusion of splenocytes and myeloma cells

The spleens of mice immunized as described above (section 2.1.5) were collected under sterile conditions into 5 ml incomplete RPMI. The spleen was placed in a petri dish containing 5 ml incomplete RPMI at room temperature and placed between two sterile pieces of nitex gauze (a kind gift from C. Hansell, University of Glasgow). Using the plunger of a syringe, the spleen was gently strained through the gauze, and the gauze was removed along with any remaining clumps of fat and connective tissue. The cell suspension remaining in the petri dish was transferred to 50 ml of RPMI incomplete medium and pelleted at 200 *g* for 7 mins at room temperature. The supernatant was discarded and the pellet washed again in 50 ml RPMI incomplete medium. The cell pellet was

resuspended in 10 ml RPMI incomplete medium and viable cells counted by trypan blue exclusion as described above; viability was always greater than 90%. For fusion, all solutions were pre-warmed to 37ºC. 10^8 spleen cells and 10^7 myeloma cells were mixed in 50 ml of RPMI incomplete medium and pelleted by centrifugation at 200 *g* for 7 mins. The supernatant was removed completely. To the pellet, 0.5 ml PEG solution A was added dropwise over a 30 second period. During this process, the tip of the pipette was used to disrupt the pellet. The tube was gently rocked for another 30 seconds. 0.5 ml PEG solution B was then added, and the pellet mechanically agitated using the tip of the pipette for 2–3 mins. 40 ml HAT selection medium supplemented with gentamycin and amphotericin B was added drop-wise over a 10 min period with constant rocking of the tube, after which time 60 ml HAT selection medium supplemented with gentamycin and amphotericin B was added. The cell suspension was added to the wells of five flat-bottomed 96 well microtitre plates, 200 µl per well, and incubated at 37ºC in 5% CO_2 in a humid atmosphere for 10 days without feeding.

On day 10, the plates were observed macroscopically from below. Colonies of hybridoma cells were visible in approximately 40–60% of the wells, and wells showing growth were recorded. 15 µl of supernatant from the wells showing growth were removed for antibody screening. Cells in wells secreting desired mAbs (as determined by indirect immunofluorescence described in section 2.6.2 below) were removed by gently tritiating the well and transferring the resulting cell suspension into 1 ml of prewarmed RPMI-HFCS complete medium.

2.5.4 Hybridoma cloning

Hybridoma lines were cloned by limiting dilution as soon as possible after fusion. Viable cell counts were carried out by trypan blue exclusion. Cells were diluted with RPMI-HFCS and placed in a 96 well plate so that 16 wells received approximately 1 hybridoma cell each, eight wells received approximately 0.1 hybridoma cell each and eight wells received approximately 0.01 hybridoma cell each. Wells showing growth at the greatest dilution were selected for screening, carried out as described in 2.6.2. The clonality of the wells could be estimated by reference to the Poisson distribution as described in section 2.3.2.

Hybridoma lines secreting the desired mAbs were cloned twice to ensure clonality.

2.5.5 Maintenance of hybridoma cell lines and production of mAbs

Cultures of hybridoma cells were examined daily and regularly subcultured to keep them in log phase growth at approximately 1–5 x 10^5 cells.ml^{-1}. Viable cell counts were performed by trypan blue exclusion. To produce a stock of desired mAbs, 50 ml cultures of hybridoma cells were set up, and allowed to overgrow. Cell death could be estimated by microscopic examination of the cultures: when cells began to detach from the bottom of the plate (after approximately 1 week) the supernatant was collected by centrifugation at 3200 *g* for 10 mins to remove cells and debris. Sodium azide was added to the supernatants as a preservative at a final concentration of 0.05%.

2.5.6 Cryopreservation of myeloma and hybridoma cells

To prepare cells for cryopreservation, 3 x 10^6 cells per stabilate were collected by centrifugation and resuspended in 1 ml of 90% fetal bovine serum/10% DMSO. The cell suspension was transferred to a cryotube and frozen as described in section 2.1.6. Removal of hybridoma cells from stabilate was performed as described for myeloma cells, above.

2.6 Analysis of antibody responses

2.6.1 Preparation of acetone-fixed slides

Acetone-fixed slides of in vitro parasite cultures were prepared by resuspending cultured parasites in complete HMI-9 at a concentration of 2 x 10^7 parasites.ml^{-1}. Thin films of this suspension were spotted on the surface of a clean glass slide using a Gilson pipette. The thinness of the film was critical to ensure rapid drying of the slide. As many as 12 spots could fit on a slide. When the slide had dried, it was submerged in a Coplin jar containing acetone at room temperature for 10 mins. The slide was then air dried. For longer-term storage, fixed slides could be wrapped in tissue and placed in a sealed plastic bag with silica gel (4–7

mesh, Fisher Scientific, Loughborough, UK, cat. no. S/0760/53) as a desiccant, and frozen at -20ºC.

2.6.2 Immunofluorescence assay on acetone-fixed slides

Immunofluorescence screening using acetone-fixed trypanosomes was the assay used to screen hybridoma lines. Undiluted supernatant from hybridoma cultures was used as the primary antibody in this assay. A paint pen ('Mark-Tex', Menke, Santa Fe CA, U.S.A.) was used to draw hydrophobic circles around the spots on the slide, forming wells to use as reference and preventing cross-contamination of antibodies during the preparation of the slide. Undiluted hybridoma culture supernatant was added to the slide, and the slide incubated in a humid chamber for 30 mins at room temperature. The slide was washed twice with PBS. Secondary antibodies, at a 1:2000 dilution in PBS supplemented with 1% BSA (Sigma-Aldrich, Gillingham, UK, cat. no. A4503) were applied to the slide. The secondary antibodies used in all cases were Alexa 488-conjugated goat anti-mouse IgG H+L (Invitrogen, Paisley, UK, cat. no. A-11001) which reacts against IgG heavy chains and both classes of immunoglobulin light chain from mouse, and, where isotype differentiation was required, Alexa 594-conjugated goat anti-mouse IgM (μ chain) (Invitrogen, cat. no. A-21044) which is specific for the IgM isotype. The slide was again incubated in a humid chamber for 30 mins at room temperature, before being washed twice with PBS and mounted using Vectashield + DAPI (Vector Laboratories, Peterborough, UK, cat. no. H-1200). Slides were examined using a Zeiss Axioscop 2 mot + microscope at 1000x magnification. Each slide would routinely include parasites expected not to react, as a negative control for non-specific antibody binding. Images were collected using Openlab 5.0 software (Improvision, Waltham MA, U.S.A.) and contrast adjusted using GNU Image Manipulation Program (GIMP Development Team, gimp.org).

2.6.3 Live cell immunofluorescence

To test whether antibodies are able to recognize the intact VSG surface coat, the primary antibody was applied to live cells and fixed using formaldehyde. For this assay, 1 x 10^6 cells were required per reaction. All centrifugation steps were carried out at 1200 *g* for 7 mins at 4ºC. Preparations of TDB or PBS

supplemented with 1% BSA (Sigma-Aldrich, Gillingham, UK, cat. no. A4503) were filter sterilised using a 0.2 µm pore size syringe filter before use. 8% formaldehyde solution was prepared in a fume hood in advance. 16 g paraformaldehyde powder (Honeywell Riedel de Haën, Seelze, Germany, cat. no. 16005) was dissolved in 50 ml water at 60ºC (using 10 µl 10M NaOH to aid solubility), 10 ml 10x PBS was added and the volume made up to 100 ml. The pH of the solution was measured using pH paper and dilute HCl added to reach a final pH of 7.2. The prepared formaldehyde solution was distributed into single-use tubes and frozen at -20ºC and used within 6 months.

Cells from culture were resuspended in TDB and collected by centrifugation. Cells were resuspended in 100 µl primary antibody solution. For polyclonal antibodies from plasma, this solution was 1:25 dilution in TDB + 1% BSA, and for monoclonal antibodies the primary antibody solution was undiluted supernatant. Cells were incubated on ice for 10 mins. 100 µl 8% formaldehyde solution was added and mixed well, and the reaction incubated on ice for 10 mins. The cells were collected by centrifugation and washed twice with 200 µl PBS. The cells were then resuspended in 100 µl secondary antibody solution. For this assay, secondary antibodies (as described above) were at 1:500 dilution in PBS + 1% BSA. The cells were incubated on ice for 15 mins, washed twice in PBS, mounted on a glass slide and examined as described above. This method is a modified version of that kindly provided by M. Engstler (Universität Würzburg, pers. comm.).

2.6.4 Complement-mediated lysis and agglutination assay

To test whether antibodies had functional activity, complement-mediated lysis was carried out. Complement-competent plasma was prepared from freshly collected guinea pig blood by centrifugation at 14,000 *g* for 10 mins, retrieving the supernatant into a fresh tube, removing any remaining blood cells by an additional 10 min 14,000 *g* centrifugation, and again retrieving the supernatant into a fresh tube. This purified plasma (guinea pig plasma, GPP) was divided into small samples, frozen, and stored at -80ºC until ready for use. On the day of the experiment, GPP was thawed and centrifuged at 14,000 *g* for 10 mins to separate any aggregates that may have formed. Representative polyclonal antibodies directed against the various antigens under consideration were

serially diluted in the guinea pig plasma, and spread across the wells of a Terasaki plate (Greiner, Stonehouse, UK), 5 µl per well. 5 µl of trypanosomes resuspended in GPP were added to each antibody dilution. For this assay, 50,000 cells were required per well (that is, 1 x 10^7 parasites.ml^{-1}). The reaction was incubated at room temperature for one hour before observation under phase contrast microscopy at 400x magnification and scoring the wells for cell death. For each parasite line, cells were incubated at the same concentration in guinea pig plasma to control for non-specific killing.

To test whether antibodies were able to agglutinate parasites, agglutination assays were set up. These were broadly similar to the complement-mediated lysis assay, above, except that antibodies were diluted, and parasites were resuspended, in TDB. For this assay, 100,000 cells were required per well (that is, 2 x 10^7 parasites.ml^{-1}). The reaction was incubated at room temperature for 30 mins before observation under phase contrast microscopy and scoring the wells for agglutination. For each parasite line, cells were incubated at the same concentration in TDB to control for non-specific aggregation.

2.7 Bioinformatics and *in silico* sequence manipulation

2.7.1 Manual sequence manipulation

Sequences were assembled, visualised, compared and analysed using CLC Genomics Workbench (CLC Bio, Aarhus, Denmark), eBioX (developed by Erik Lagercrantz, available at www.ebioinformatics.org/ebiox), custom Ruby and BioRuby scripts I wrote (Goto *et al.*, 2010) and Aquamacs (developed by David Reitter, available at aquamacs.org).

For the *VSG* clone sequences, details of manual sequence manipulations were recorded in an Excel spreadsheet (Appendix 7.2.3). Full-length assembled sequences were saved as FASTA files corresponding with their unique clone name (XX-YYcZZ, where XX was the infection number, YY the day of collection and ZZ a numeric identifier) and placed together in a directory.

2.7.2 Analysis of sequences

2.7.2.1 Grouping of sequences into 'sets'

VSG clone sequences were placed into 'sets' according to their BLAST matches to a database of genomic *VSG* sequences (Altschul *et al.*, 1990) and their alignments with one another. The genomic *VSG* database is described in Appendix 7.2.4. Clone sequences were aligned automatically using Clustal W (Larkin *et al.*, 2007). Details of each set—its constituent clones and best-matching genomic copies, alongside notes and other comments—were recorded in separate flat text files. The format of these text files was standardized so they could be parsed easily by Ruby script. For each set, alignments of the full-length nucleic acid sequences, and mature N-terminal domain (identified as described in section 2.7.2.2) amino acid sequences, were stored in separate directories. In some cases, alignments were altered by hand to improve matching to putative donors.

2.7.2.2 Identifying features of VSG sequences

For each VSG, there were considered to be 50 amino acids in the C-terminal domain before the first conserved cysteine (Marcello & Barry, 2007a). This assumption was used to assign the junction between N- and C-terminal domains (the first position of the first codon encoding the C-terminal domain) for each clone; this datum was recorded for each clone in a script-parsable text file. The SignalP 4.0 server (Petersen *et al.*, 2011) was used to predict the length of the N-terminal signal peptide for each clone, this datum (the third position of the last codon of the signal peptide) was stored alongside the information on domain junctions. The position of the N- and C-terminal domain junction, and the position of the N-terminal signal peptide cleavage, were usually the same for all clones of a set. Where they were not, the positions were standardized for each set (according to the mode), to simplify the analyses. Where the position of GPI anchor signal cleavage was required, the length of the GPI anchor was determined by reference to known VSG GPI anchor signal sequences (Marcello *et al.*, 2007), or using the big-PI (Eisenhaber *et al.*, 1999) or Fraganchor (Poisson *et al.*, 2007) servers.

2.7.2.3 Identification of 'read' donors

Where one or more *VSG* clones had no match in the genomic *VSG* database for part or all of their sequence, putative 'read' and 'mc' donors could be identified and assembled as follows. The part(s) of the clone for which no donor could be found were used as queries in BLAST searches of several 'reads' databases (described in Appendix 7.2.5). Matching reads were retrieved from these databases, and compared with the clone sequence using CLC Genomics Workbench. Reads that were similar to the clone and different from all identified genomic copies were assembled to form a contig. Occasional mismatches between the reads were treated as follows. If the mismatch occurred towards the ends of a read, it was considered to be the result of poor quality sequencing, and was corrected in favour of the other read(s). If the mismatch occurred in the middle of the read, but more than one other read did not contain the mismatch, it was corrected in favour of the other reads. If the mismatch occurred in the middle of the read, and there was only one other read with which to compare it, the mismatch was corrected in favour of the corresponding clone sequence. Putative read assemblies were saved as FASTA files and were placed in a separate directory. They are described in Appendix 7.2.6.

2.7.2.4 The semi-automated analysis of sequences using scripts

With the text files describing each set's constituent clones and their putative donor sequence(s), the multiple sequence alignment for that set, and the position of the N-terminal signal sequence and N- and C-terminal domain boundary for each clone of that set in place, an automated BioRuby script was written to describe the features of each clone and its relationship with the donor(s). This script is 'vsg_analysis_from_aln.rb' and is freely available on request. Essentially, the script took an alignment in FASTA format and identified clone sequences and donor sequences according to their names. Each clone sequence was considered in turn. The clone sequence was examined, and each position was compared with each donor sequence at that position. For each position, a 'piece' object was created, which contained a list of the matching donors at that position. These 'pieces' were accumulated to form 'blocks'. The list of blocks simply starts as the list of pieces, but it is

progressively condensed in an iterative process: if adjacent pieces match the same donor(s), those pieces are accumulated in a block, corresponding with their shared donor. Blocks expand to accumulate all possible pieces. This process continues across the sequence, until blocks can accumulate no more pieces. At the end of the process, ambiguous blocks (blocks that still match more than one donor) are weighted towards the overall best-matching donor for that clone. Following the process of accumulation, blocks are corrected to give the junctions. Pieces at the edges of the blocks that are ambiguous between that block's donor and the donor of an adjacent block constitute the junction. These pieces are equally distributed between the two contesting blocks. The script also removes gaps from the alignment, records the position of cysteine codons and positions where the clone sequence matches none of the putative donors (and the non-synonymity of that mismatch), calculates the best matching donor, and extracts the data on the N- and C-terminal domain boundary and N-terminal signal peptide from the data file. All this information is transcribed into a script-parsable text file, which was placed in a different directory. Variations on 'vsg_analysis_from_aln.rb' were used to 'strip out' data of interest for various bioinformatic studies. Semi-automated analyses were complemented by at least one manual analysis to control for bugs in the scripts.

2.7.2.5 Graphical representation

The script-parsable data files produced in section 2.7.2.4 were described graphically using the Ruby script 'vsg_diagram.rb' (freely available on request). This program simply read in the information in one or more of the script-parsable data files and transcribed this information as a Scalable Vector Graphics (.svg) file that could be visualised using many internet browsers and image manipulation programs. These diagrams were used as tools to indicate what sequence alignments should be investigated manually for segmental gene conversion. Variations on 'vsg_diagram.rb' were also used to represent other data where required. The analysis of sequences was an iterative process: the graphical representation of the data indicating where reanalysis was required. These graphical representations formed the basis for the *VSG* diagrams provided in this document; where necessary they were manipulated further using the GNU Image Manipulation Program (GIMP Development Team, gimp.org).

Chapter 3

VSG expression in chronic mouse infections

3 *VSG* expression in chronic mouse infections

3.1 Introduction

3.1.1 What VSG*s are expressed during infection?*

Previous findings have hinted at the importance of segmental gene conversion in generating expressed *VSG*s as *Trypanosoma brucei* infections progress (Barbet & Kamper, 1993; Marcello & Barry, 2007a). Segmental gene conversion could offer several key benefits to the trypanosome population. Firstly, it allows expression of *VSG*s damaged by pseudogenicity, which comprise the bulk of the archive (Marcello & Barry, 2007a). Damaged *VSG* activation could become important as responses accumulate against easier-to-activate undamaged *VSG*s. Secondly, it introduces variability in expressed variants, which could contribute directly to antigenic variation. Other antigenically variant pathogens use the combinatorial power of segmental gene conversion to generate huge antigenic diversity (Brayton *et al.*, 2001; Coutte *et al.*, 2009; Palmer *et al.*, 2009). Segmental gene conversion could also affect the rate at which silent *VSG*s are activated, helping to bring about a hierarchy of expression (Pays, 1989).[27]

To understand the role of segmental gene conversion in *Trypanosoma brucei* antigenic variation, a detailed picture of the *VSG*s expressed over the course of infection was required. A pilot study investigating infections at a single timepoint suggested a prominent role for mosaic *VSG*s in late stage infection (Marcello & Barry, 2007a). The approach of Marcello & Barry (2007) was expanded, with longitudinal sampling allowing individual infections to be followed over time, enabling changes in the antigenic profiles of the resident trypanosome population to be identified. By comparing expressed *VSG*s with one another and with those present in the genome, the extent of segmental gene conversion—in particular expressed mosaic *VSG* formation—could be estimated. These data also provided material for further analyses in subsequent chapters.

[27] These possibilities are discussed further in Chapter 6 (section 6.4), in light of the results presented here.

3.1.2 Key hypotheses to be tested

To investigate the distinctive features of *VSG* expression during infection, repeated samples were taken longitudinally from a number of infections to address the following hypotheses:

(i) Early appearance of intact *VSG*s. Each infection is hypothesised to display unique characteristics, but at earlier timepoints and in numerous infections, a subset of frequently-activated *VSG*s will be detected. Intact, full-length, and telomere proximal genes have a greater activation probability as they require fewer genetic rearrangements to become a functional, active-expression-site-occupying *VSG* and hence are more likely to be expressed earlier on in numerous infections (Aline *et al.*, 1985; Laurent *et al.*, 1984a; Morrison *et al.*, 2005). Telomere-proximal *VSG*s are also more likely to physically interact with the telomere-proximal BES (Barry *et al.*, 2003).

(ii) Late appearance of damaged *VSG*s, through segmental gene conversion. As an infection progresses the host develops responses to the subset of easily accessible genomic copies. Antibodies might select for rarer genetic rearrangements—segmental gene conversions that have repaired damaged silent *VSG*s—leading to an increased detection of mosaic *VSG*s at later timepoints (Marcello & Barry, 2007a; Pays, 1989; Roth *et al.*, 1989).

(iii) Progressive mosaicism. As infections continue through the late stage, repeated incidences of segmental gene conversion will ultimately result in the expression of increasingly complex mosaic antigens (as seen in Anaplasma marginale by Palmer *et al.*, 2007). When such segmental conversions change the epitope structure of the expressed VSG, a series of related yet antigenically distinct mosaic VSGs may persist over time in an infection: a ‘string’ of mosaics (Barry *et al.*, 2005).

3.2 Experimental approach

3.2.1 VSG *sequences were amplified, cloned and sequenced from longitudinal samples*

To address these questions, individual infections of mice were sampled at different points in time, and the expressed *VSG* sequences were retrieved. Mice provide an accessible model of mammalian infection: they reliably maintain a high parasitaemia with *Trypanosoma brucei* (depending on strain), they seldom self-cure, and they are easy to work with and maintain. In addition, the range of immunological knockouts available means that further studies could easily investigate the impact of different aspects of the host immune system on any phenomena identified here.

The rapidity with which parasites switch antigen expression (estimated to be as high as 0.01 per cell per division (Turner & Barry, 1989; Turner, 1997)) precludes the isolation of different antigens by cell cloning, as cloned parasites would likely switch before they reach required numbers. Therefore, antigen mRNA sequences were isolated and amplified for sequencing. Currently, the short read length approach employed by 'next-generation' Illumina sequencing would make assembly of mosaic *VSG*s not unambiguous, so *VSG* sequences selectively amplified from reverse-transcribed RNA were cloned using a commercial PCR subcloning kit and sequenced by the Sanger chain-termination method as described in Chapter 2 (section 2.4.14). Sequences could then be analysed bioinformatically to reveal patterns of *VSG* expression across infection.[28]

3.2.2 Assumptions of experimental approach

As a way of investigating the antigenic profile of the parasite population, this mRNA cloning and sequencing approach makes several assumptions about *VSG* expression and trypanosome population structure, which although justified should be made explicit.

[28] The small blood samples necessitated by ethical and legal considerations meant that only mRNA could be reliably retrieved from longitudinal samples (although see Chapter 2 section 2.2).

3.2.2.1 *VSG* mRNA is a valid proxy for the VSG surface coat.

Despite strong evidence that all detectable full-length *VSG* mRNA is produced from the active ES (reviewed in Chapter 1, section 1.6.2) it is possible that mRNA can be transcribed but not culminate in being expressed as a surface coat. A plausible example would be a cell in which a *VSG* gene with a premature stop codon or frameshift has entered the active expression site: transiently, these cells would transcribe a *VSG* that cannot be translated to protein. This state will be resolved either by the cell switching expression from the damaged gene to an intact one, by the cell's inability to produce a complete surface coat (resulting in elimination by the innate immune system (Ferrante & Allison, 1983)), or possible cell death through global translational repression (Smith *et al.*, 2009). Due to the preferential degradation of nonsense mRNA transcripts (Delhi *et al.*, 2011) and the need for the cell to replace the constantly turning-over coat (Seyfang *et al.*, 1990) it was assumed that the effects of mistranscription would be minor. Stumpy form trypanosomes downregulate transcription from their BES (Amiguet-Vercher *et al.*, 2004) and so it is possible that their surface coats will be underrepresented in the mRNA if their population is not representative of the proliferative *VSG*-transcribing long slender forms.

3.2.2.2 All *VSG* mRNA contains primer binding sites for the *VSG* amplifying primers.

The forward primer binds to the spliced leader sequence trans-spliced onto the 5' end of all trypanosome mRNAs (Parsons *et al.*, 1984; Walder *et al.*, 1986). The reverse primer binds to a conserved 16-mer found in the 3' untranslated region of all *VSG* mRNAs (Matthyssens *et al.*, 1981; Merritt *et al.*, 1983; Michels *et al.*, 1983). There is also experimental evidence that this 16-mer has a pivotal role in silencing non-active *VSG*s (see Chapter 1 section 1.6.2), further supporting the assumption that the primers are sufficient and exclusive for the expressed *VSG* (S. Hutchison pers. comm., P. Batram pers. comm. Berberof *et al.*, 1995). Prior to these experiments being carried out (March 2009), the exact 16-mer was identified in all but two of the sufficiently long *VSG* mRNA sequences retrieved from the NCBI nucleotide database (www.ncbi.nlm.nih.gov, those that differed, differed by only a single nucleotide).

3.2.2.3 Blood sampled by tail venesection is representative of the whole infection.

Bottlenecks restricting the movement of either parasites or immune effectors between tissues would likely result in the course of antigenic variation to run semi-independently either side of the bottleneck. The most pertinent example of this would be the immunoprivileged central nervous system (Jennings *et al.*, 1979). However, experiments addressing this question have found similar patterns of VATs occurring in different tissues, indicating that the potential effects of differential tissue location would probably be small (Turner *et al.*, 1986). As peripheral blood[29] is the easiest and most humane tissue to collect regularly, it forms the most immediately accessible experimental model.

3.2.2.4 Amplification and cloning of *VSG* sequences is not grossly biased or mutagenic.

There is abundant evidence suggesting that the amplification of certain sequences proceeds more quickly or effectively than others (Arezi *et al.*, 2003). It is therefore possible that amplification and cloning inherently favours certain *VSG*s over others. To minimise this possibility, an enzyme with proven efficacy at amplifying diverse sequences, Herculase II Fusion polymerase, was used for amplification for cloning (Agilent, Wokingham, UK). In addition, further PCR reactions were carried out on cDNA to examine the potential for cloning and sequencing bias (section 3.5.2). It is also possible that the process of amplification and cloning introduces mutations, discussed in more detail in Chapter 4 (section 4.2).

3.3 Retrieval of *VSG* sequences

3.3.1 Samples were taken to focus on the later stage of infection

To follow changes in *VSG* expression over time, multiple samples from individual infections were required. Humane restrictions on the total volume of blood sampled over the course of infection (maximum 15% of total blood volume over a 28 day period) and a desire to strike a balance between breadth and depth of

[29] The terminal samples in all cases were collected by exsanguination by cardiac puncture under terminal anaesthesia, to maximise the size of the sample.

sampling meant that each mouse could be sampled only a limited number of times. The later, chronic stage of infection was of particular interest, as natural infections of mammals often proceed for many months before resolution (Guirnalda *et al.*, 2007; Nantulya *et al.*, 1984; Penchenier *et al.*, 2005; Songa *et al.*, 1991). Previous studies have broadly described the patterns of the first relapse peak of infection (occurring around day 10–14 (Miller & Turner, 1981; Myler *et al.*, 1988; Robinson *et al.*, 1999)). Furthermore, a previous pilot study on which the present analyses have been based has indicated the predominance of less-well-characterised mosaic *VSG*s from approximately day 21 onwards (Marcello & Barry, 2007a). Therefore, samples were collected from eight mice to focus on this later stage of infection, between days 20 and 34. Three further infections also cover earlier timepoints. Details of the infections are shown in Table 3.1.

Mouse #	Inoculum source	Inoculum size	Infection length (days)
01	Mouse FF	40,000	32
02	Mouse FF	40,000	32
03	Mouse FF	40,000	32
04	Mouse FF	40,000	31
05	Mouse FF	40,000	31
06	Mouse FF	40,000	29
08	Mouse F2	1	34
09	Mouse F2	1	34
10	Mouse F3	10,000	30
11	Mouse F3	10,000	30
12	Mouse F3	10,000	30
LM_XX	Mouse F	1,000,000	3–28

Table 3.1 Chronic mouse infections. The relationship between the inocula is shown in Appendix 7.2.1. Inoculum size was measured by haemocytometer for all cases except for mouse 08 and mouse 09 where the inoculum was optically cloned. 10,000–40,000 parasites was considered sufficient to reliably establish infection by intraperitoneal inoculation. LM_XX refers to 11 infections analysed previously by Marcello & Barry (2007a).

Some of these infections were carried out to address specific experimental questions (for example the primary clonal infections in mice 08 and 09) but were subsequently used as a source of material for the *VSG* expression analysis described here; in these cases the sequences retrieved will be discussed both in

the context of the other infections, and separately in the context of the specific experimental questions being addressed. As 37 *VSG* sequences described in previous work (Marcello & Barry, 2007a) will be frequently referred to in this and the following chapter, the infections from which they were obtained are also described in the Table. All chronic mouse infections were carried out using the 'genome strain' of *T. brucei*—specifically a recently recloned TREU 927/4 GUTat 10.1 stabilate kindly provided by L. Marcello. The pedigree of the parasites used in these infections is shown in Appendix 7.2.1.

3.3.2 Parasitaemias followed a broadly similar pattern in different infection

The parasitaemias of the 11 infections were monitored by rapid matching (Herbert & Lumsden, 1976). Graphs of parasitaemia over time are shown in Figure 3.1, and for clarity the exemplary parasitaemia of mouse 05 is shown separately in Figure 3.2 (the individual parasitaemias for the other mice are shown in Appendix 7.2.2).

The parasitaemias followed a broadly similar pattern. In all infections except one of the clonal infections (mouse 08), a 'first peak' of parasitaemia was seen between days four and six, marked A in Figure 3.2, reaching as high as antilog 8.1 parasites.ml^{-1}. The parasitaemia then fell and remained reasonably stable between antilog 7.2 parasites.ml^{-1} and antilog 7.8 parasites.ml^{-1} for the next 15 to 20 days of infection (marked B on Figure 3.2). Interestingly, all but two of the infections studied showed a dramatic (ten-fold) drop in parasitaemia approximately 3 weeks into the infection (occurring approximately between days 21 and 26, marked C on Figure 3.2).[30] This state lasted for one or two days, after which the parasitaemia recovered. A similar transient drop in parasite numbers at approximately week 3 of infection was also observed in the infections carried out by (2007a). Possible explanations for this phenomenon are discussed below, in sections 3.5.4 and 3.5.5.

[30] The exceptions are mouse 02, which sees a dip in parasitaemia later (day 29), and mouse 10, which sees a dip in parasitaemia earlier (around day 16).

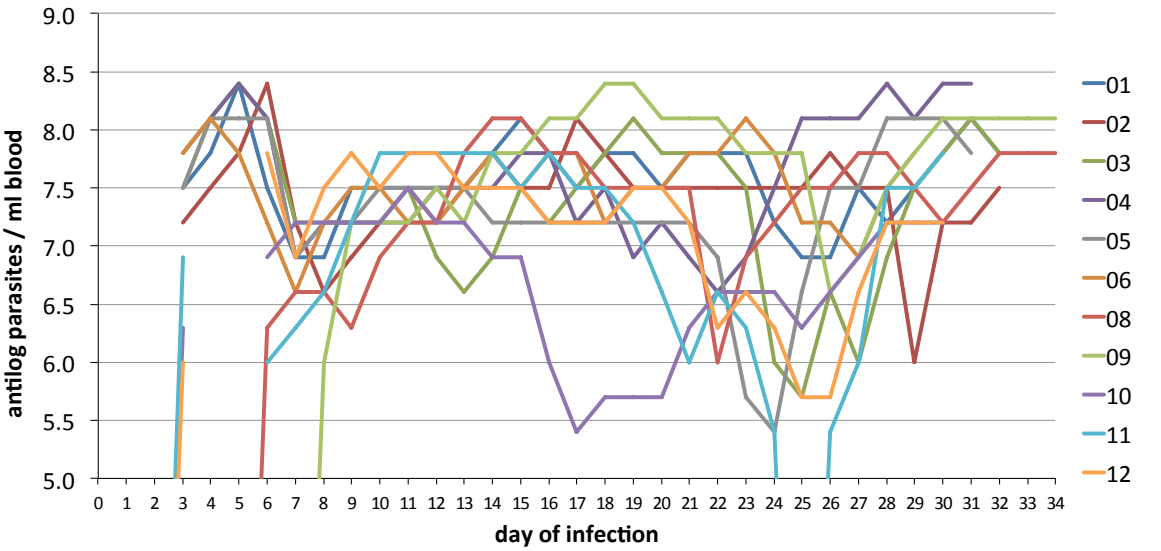


Figure 3.1 Parasitaemias. Mice were examined daily from day 4 of infection onwards and parasitaemia estimated by rapid matching (Herbert & Lumsden, 1976).

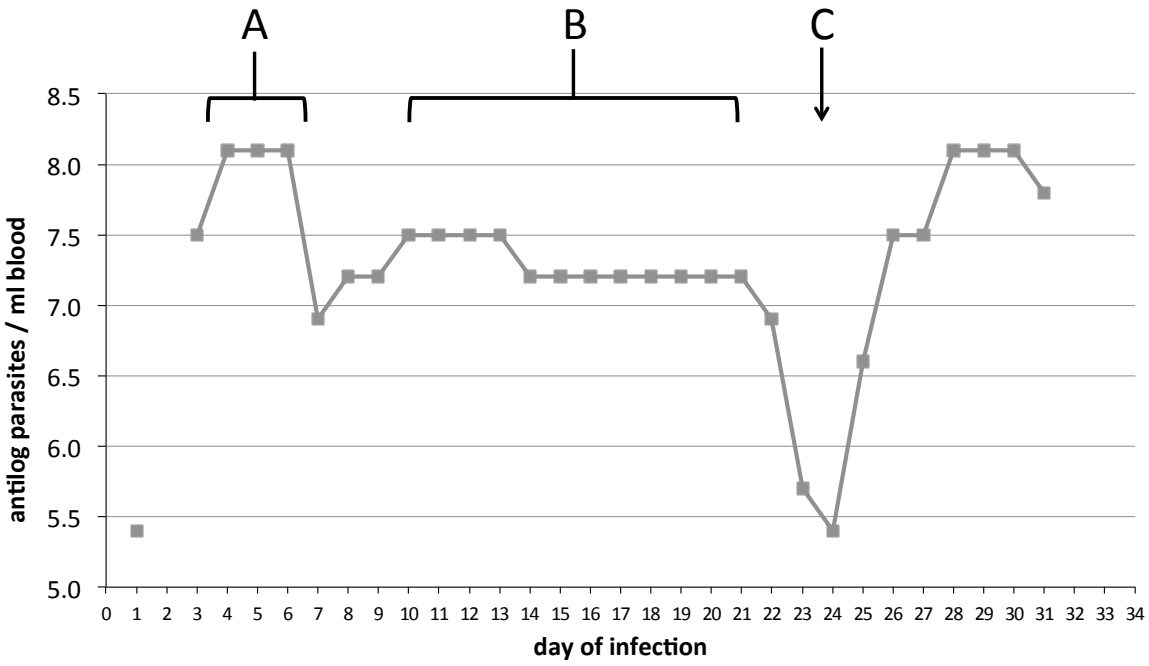


Figure 3.2 Mouse 05 parasitaemia. Points A, B and C are discussed in the text.

Between days 29 and 34 parasitaemia could reach very high levels. In one case, observed ill effects led to the euthanasia of the animal at day 29 (mouse 06). The median life expectancy for a Balb/c mouse infected with a pleomorphic trypanosome line[31] is 35 days (Magez *et al.*, 2008), and all experiments were terminated before then (the median length of infection was 31 days).

The lack of distinct peaks and generally high parasitaemia indicates that density-dependent differentiation is likely to be the dominant factor controlling the parasite population in these infections (Gjini *et al.*, 2010; MacGregor *et al.*, 2011). Short stumpy form trypanosomes were often observed when assessing parasitaemia microscopically, although their abundance was not formally measured.

3.3.3 VSG *sequences were readily amplified from trypanosome mRNA*

From individual infections, blood samples were taken by repeated tail venesection[32] into CBSS buffer containing 5% w/v sodium citrate anticoagulant. Erythrocytes were osmotically lysed, total RNA was extracted and reverse transcribed using an oligo[dT] template, and the synthesized cDNA was column purified. Details of the protocol are in Chapter 2 (sections 2.2 and 2.4).

cDNA corresponding to full-length *VSG* mRNA was selectively amplified by PCR. To control for the possibility of genomic DNA (gDNA) contamination, each reaction was also carried out on the products of a 'cDNA synthesis' reaction to which no reverse transcriptase had been added. No products could be seen in any *VSG* PCR gDNA control. Indeed, because the primers used target the trans-spliced leader sequence, no products of the correct size are seen even if the PCR is carried out on purified gDNA (data not shown). A portion of each PCR reaction was separated on a 1% agarose gel, shown for mouse 01–06 and FF in Figure 3.3. The remainder of each reaction, run on a separate gel, was purified by gel

[31] In the case of Magez et al. (2008) the trypanosome line in question was AnTat 1.1E, but a similar trajectory is seen in infections with TREU 927/4 (C. M. R. Turner, pers. comm.)

[32] The exception was the terminal sample, which (as described above) was collected by exsanguination by cardiac puncture)

extraction, ensuring that all of the gel corresponding to the 1300–1700 bp size markers was collected. Samples from mice 10–12 were prepared by H. Wang.

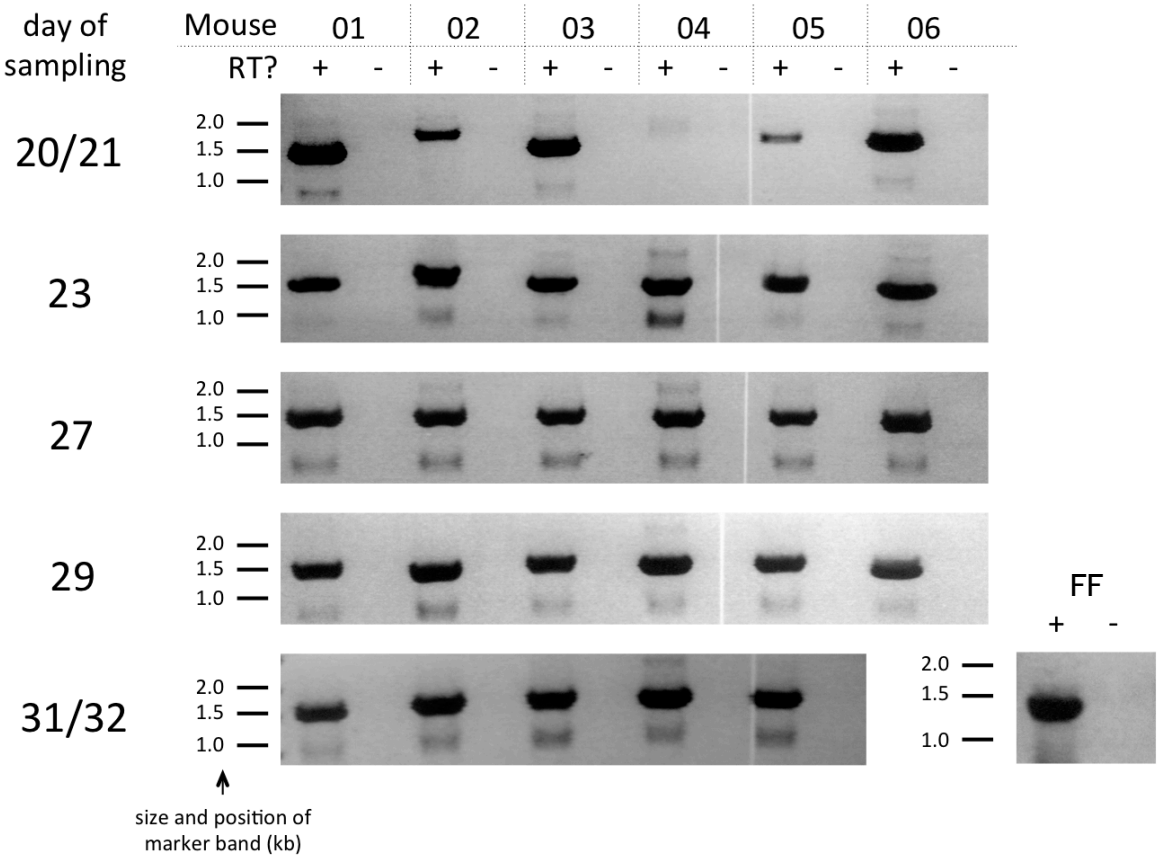


Figure 3.3 Amplification of *VSG* cDNA. The products of *VSG* amplification reactions from cDNA and controls for gDNA contamination. FF is the inoculum source (day 2 of infection). The smaller band seen in each case is likely due to the 16-mer primer binding to an alternative, unknown template, as the primer is short and reaction annealing temperature is low (40ºC), smaller products were previously seen in similar reactions by Merritt *et al*., (1983).

To compare the sizes of the *VSG* amplicons over the course an infection, a sample of gel-purified PCR product from each timepoint of mouse 01–06 was run in adjacent lanes on a 1% agarose gel. The result is shown in Figure 3.4. The band can clearly be seen to change apparent size as each infection progresses, and at some timepoints numerous bands, or a tight smear, can be seen. This indicates the presence of different-sized amplicons at different timepoints, or within a timepoint, as would be expected from a product that is known to vary considerably during infection.

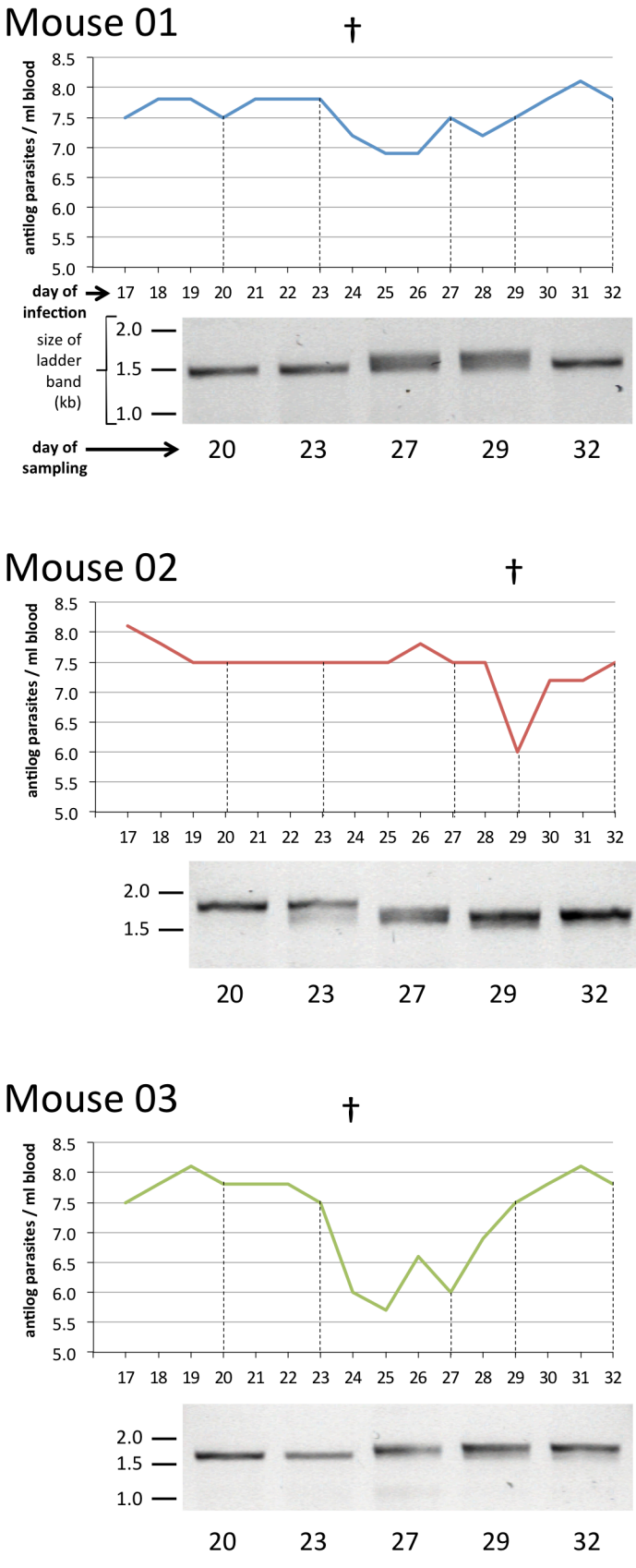
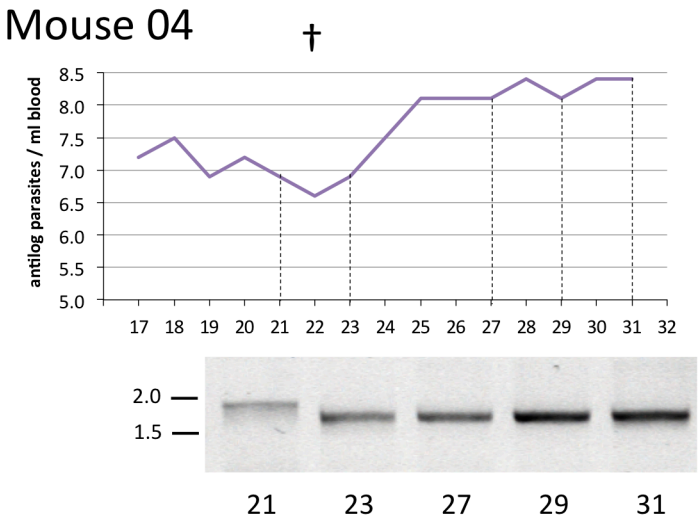
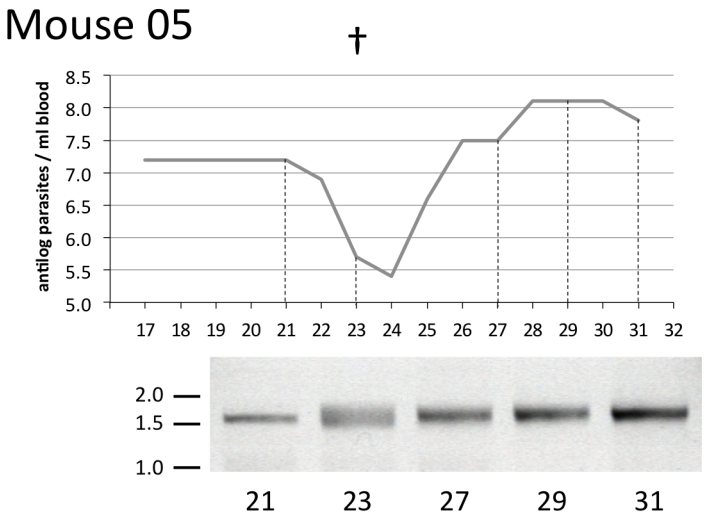


Figure 3.4 *VSG* amplicon sizes change over course of infection. Purified products from the *VSG* amplification reactions were analysed for each infection (01–06) by separating in adjacent lanes of a 1% agarose gel. The associated parasitaemias for each infection are shown. Dotted lines represent sampling times. A dagger above the parasitaemia indicates the point where the parasitaemia undergoes a transient dip (point 'C' in Figure 3.2), which varies slightly in time and extent between infections.





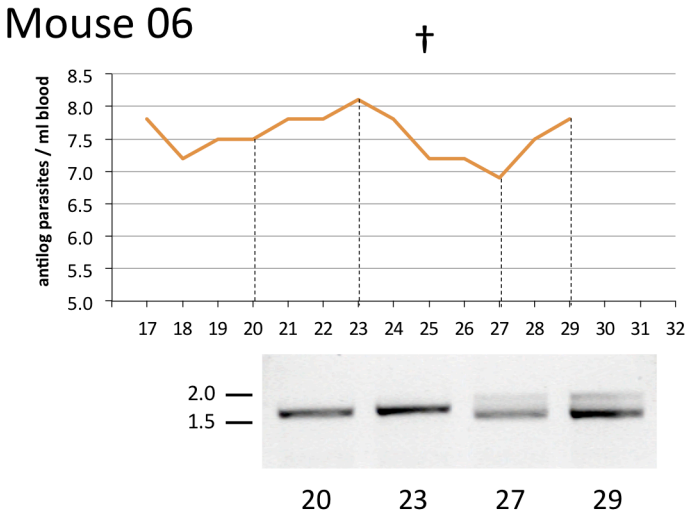


Figure 3.4 (continued)

3.3.4 664 full-length VSG *clones were sequenced*

To retrieve the nucleic acid sequences of the *VSG*s expressed during infection, the PCR products were subcloned, screened and sequenced as described in Chapter 2 (sections 2.4.8, 2.4.9 and 2.4.14). Briefly, the products were A-tailed and ligated into the vector, and used to transform competent *E. coli* cells. White colonies were screened for insert-containing plasmids by PCR using M13 forward and M13 reverse primers. Plasmids from those colonies yielding a product of estimated size 1200–1700 bp were purified using a small-scale plasmid DNA purification 'miniprep' kit (Qiagen, Crawley, UK) and was tested by *Eco*RI digest. Samples of each isolated plasmid had their insert sequenced by Eurofins MWG Operon (Ebersberg, Germany) using the M13 forward and M13 reverse primers that bind on either side of the multiple cloning site. Clones from mice 10–12 were prepared by H. Wang.

Cloning efficiency varied between attempts but was generally good. For example, in the initial round of cloning, of the 372 colonies screened, 312 were found to contain an insert (an average of 10 positive per 12 colonies tested). This is an improvement on that recorded previously (approximately one positive in 10 colonies tested, L. Marcello, Ph.D. thesis, 2006, University of Glasgow), likely due to the additional cDNA purification step here improving PCR efficacy.

Table 3.2 shows the total numbers of *VSG* plasmid clones prepared and sequenced for each mouse-timepoint. Eleven plasmid clones were prepared from mouse FF, the mouse used to amplify parasites to initiate infections in mice 01–06. In an attempt to capture the total antigenic variability present, a greater number of clones were sequenced from two infections (mice 04 and 05). For most clones, one forward and one reverse read gave a sufficiently long (~200 bp) region of exact overlap allowing their sequences to be confidently assembled to give the sequence of the full-length clone. In 56 cases, internal sequencing reactions were performed to confirm the correct sequence of the assembly. Ambiguities, when present, were always corrected in favour of minimising the number of differences between similar clones and between the clone and its putative genomic copy (see section 3.4.2). For 19 clones, a failure in one of the sequencing reactions resulted in a partial sequence. These partial sequences

could nevertheless be identified by reference to other complete clones retrieved from that timepoint or infection.

Infection	Day of sampling																			Total
	2	3	7	9	10	14	17	20	21	22	23	24	27	28	29	30	31	32	34	
FF	11																			11
1								4			11		9		9			11		44
2								6			10		11		9			11		47
3								11			10		11		14			12		58
4									35		33		30		20		35			153
5									24		25		32		31		31			143
6								12			12		12		11					47
8								11											19	30
9								11											16	27
10					12											8				20
11			5		7	8	12		6				6			22				66
12					11		11					9				6				37
LM_01		2																		2
LM_02		2																		2
LM_03				4																4
LM_04				3																3
LM_05						2														2
LM_06						1														1
LM_07										4										4
LM_08									2											2
LM_09												1								1
LM_10														5						5
LM_11														11						11
Total	11	4	5	7	30	11	23	55	67	4	101	10	111	16	94	36	66	34	35	**720***

Table 3.2 Sequenced clones retrieved from various timepoints of chronic mouse infections. 19 of these are incompletely assembled. *Includes 37 data from Marcello & Barry (2007a) (LM_01–11).

Each assembled sequence was verified as being that of a *VSG* by BLAST identity to a *VSG* present in the 'genomic *VSG* database' (see section 3.4.2 and Appendix 7.2.4 for details) or by comparison of the spacing of conserved cysteine residue codons with that of known *VSG*s (Carrington & Boothroyd, 1996; Marcello & Barry, 2007a). Of the 684 clones analysed, just one was not identifiably a *VSG* and so was excluded from further analysis. Primers mismatching early in the PCR reaction may have produced this non-*VSG* amplicon.[33]

[33] Preliminary studies carried out without the cDNA purification step showed a much greater number of non-VSG sequences being cloned (data not shown). Non-specific amplification was likely due to residual oligo[dT] from first-strand synthesis acting as an alternative primer to the 16-mer, allowing amplification of all cDNAs.

For each of the 47 mouse-timepoints, between four and 35 *VSG* clones were identified, and in total, 664 full-length sequences were retrieved.[34] The median number of clones retrieved per infection was 45.5, and the median number of clones retrieved per timepoint was 11. Each clone was given a three-part name: XX-YYcZZ, where XX is the mouse number, YY is the day of infection from which that clone was amplified, and ZZ is a numerical identifier.

The clone sequences were manually manipulated prior to further analysis. The sequences were cropped to the putative open reading frame (ORF), assumed to begin with the first ATG codon 3' of the spliced leader at the 5' end of the sequence, proceed with no in-frame stop codons, and end with an in-frame stop codon, usually TAA, approximately 1500 bases downstream. Occasionally the ORF could not be found. In 43 cases failure to identify the ORF was due to an insertion or deletion mutation that disrupts the frame of the gene. Such 'errors' were recorded and corrected with reference to other, undamaged sequences from a similar timepoint or infection. In 12 other cases failure to identify the ORF was due to a substitution mutation creating a stop codon in the middle of the gene, which was usually left uncorrected since it seldom disrupted downstream analyses. All manual edits to clone sequences are described in Appendix 7.2.3, and the names of the clones themselves suffixed with '-add', '-del' or '-edit' according to their modification. The sources of the errors are described and discussed in more detail in Chapter 4 (section 4.2.3). In one case an assembly of the clone could not be completed, and the draft assembly sequence is suffixed '-draft'. The 19 clones for which full-length sequences were not determined could be partially analysed, and these clone sequences were suffixed '_x'. The open reading frames of the cloned *VSG*s ranged in length between 1300 and 1710 bp, with most being between 1460 and 1600 bp long.

The regions of each *VSG* clone encoding the N- and C-terminal domains was predicted by reference to similar annotated genomic copies in VSGdb (Marcello *et al.*, 2007). For those clones with no clear matches in VSGdb, the region encoding the C-terminal domain was defined using the positions of conserved

[34] 684 clones analysed, minus the one that was not a *VSG*, minus the 19 for which only one sequencing reaction was achieved.

cysteine residues as described in Chapter 2 (section 2.7.2.2). The length of the N-terminal signal peptide, which is cleaved from the mature VSG, was predicted using the SignalP 4.0 algorithm (Petersen *et al.*, 2011).[35] Both the FragAnchor algorithm (Poisson *et al.*, 2007) and the big-PI predictor (Eisenhaber *et al.*, 1999) were used to predict the functionality of GPI anchor signal sequences when required.

3.4 Identification and classification of expressed *VSG*s

3.4.1 Expressed VSGs *can be grouped into sets*

At a given point in time during an infection, we might expect to find many related parasites expressing identical or closely related *VSG*s, having descended from an ancestor that switched to express this *VSG*, evaded the immune response and proliferated. We might also expect to find similar *VSG*s expressed in different infections, representing different populations activating the same silent *VSG*. To examine the relationships between the sequenced expressed *VSG*s, the clones were first compared with one another. These comparisons revealed two broad patterns of variation between the clones. Clones group in 'sets': members of one set of clones are much more similar to one another than they are to any member of another set. To illustrate this grouping, Figure 3.5 shows a midpoint-rooted tree drawn from a Clustal alignment of the 137 different *VSG* clones retrieved between days 20 and 32 from mouse 05, which constitute 30 sets. The variation between clones of one set and clones of another is referred to as 'between-set' variation. The other pattern of variation can be seen when comparing different clones within a set—in the Figure, Set_14 shows the most widespread 'within-set' variation, whereas Set_21 shows little within-set variation.

[35] In the rare cases where they did not exactly coincide, the N-terminal signal peptide cleavage site and the boundary between N-and CTDs was standardized for similar *VSG*s (that is, *VSG*s of the same set) to simplify analyses.

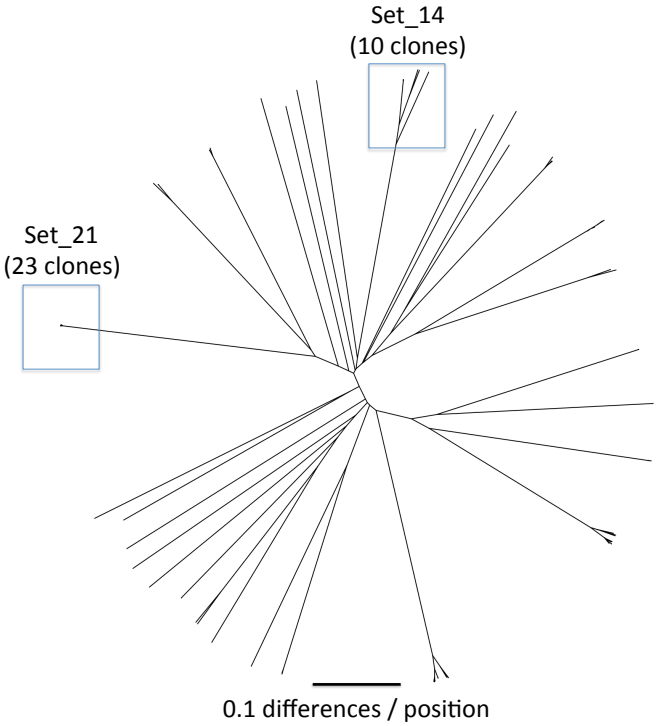


Figure 3.5 Relation between 137 *VSG* sequences retrieved from mouse 05. Nucleic acid sequences were aligned using Clustal and the corresponding phylogenetic tree visualised with FigTree (A. Rambaut, University of Edinburgh). Set_14 and Set_21, showing greater and lesser degrees of within-set variation respectively, are highlighted.

To analyse these complementary kinds of variation, all 683 expressed *VSG* sequences from the 11 experimental infections were grouped into sets based on sequence similarity. The maximum amount of variation seen when comparing the full-length nucleic acid sequences of two set members (maximum within-set variation) was 0.25 differences/nt, whereas the minimum amount of variation seen when comparing two sequences of different sets (minimum between-set variation) was 0.41 differences/nt.[36] Assigning sequences to sets was therefore unambiguous. In total, the 683 expressed *VSG* sequences formed 81 distinct sets. Some sets were very highly represented amongst the clones, whereas others had just one clone; the number of clones comprising each set varies between one (42 sets had just one member) and 123 (Set_08).

The 37 sequences retrieved by Marcello & Barry (2007a) from 11 experimental infections were reanalysed in a similar way. These formed 20 sets,[37] 18 of which

[36] The mean between-set variation was 0.59 differences/nt, not substantially less than what is seen comparing randomized sequences (~0.75 differences/nt).

[37] Marcello & Barry (2007) used the term 'VSG' to group the expressed *VSG* sequences for analysis. This was disfavoured here, to emphasise that (i) for many sets there is not necessarily a one-to-one relationship with a single genomic copy, and that (ii) there can be biologically-relevant variation amongst members of the set (Chapter 4).

overlapped with the 81 sets described above. The two sets that did not overlap comprised *VSG*s detected at timepoints underrepresented in the data presented here (day 9 of infection). Bringing together all the expressed *VSG* sequence data collected from TREU 927/4 GUTat 10.1 infections to date, 83 different sets have therefore been identified. Each set was given an arbitrary number (Set_01–Set_84).[38] All the expressed *VSG* clones isolated here and those described by Marcello & Barry (2007a), and the sets they form, are described in Appendix 7.2.3.

First, the patterns of, and processes contributing to, between-set variation will be considered. Between-set variation constitutes the most dramatic variation between *VSG* sequences, and therefore presumably contributes most to expressed antigen diversity over the course of an infection. Then, in Chapter 4, the patterns and processes of within-set variation will be investigated.

3.4.2 Sets correspond with genomic copies of VSGs

In its simplest form, we might consider different members of a particular set as being the expressed forms of one or more specific silent archival *VSG* genes (the 'donors'), copied into a BES and then being activated. *VSG* sequences belonging to the same set were identified in separate infections, consistent with the hypothesis that these expressed sequences represent independent activation of the same silent *VSG*. Between-set variation thus represents the activation of different silent *VSG*s.

To investigate the relationship between expressed *VSG* sequences and sequenced genomic *VSG*s, each expressed *VSG* clone was compared to the 'genomic *VSG* database' (the source of which is described in Appendix 7.2.4) using BLAST (1990). All genomic *VSG*s that gave a BLAST E-value of less than 1e-100 were investigated. Particular focus was put on similarities in the putative NTD-encoding part of the gene, due to greater homologies between CTDs (Rice-Ficht *et al.*, 1981), the occurrence of 3' donation (covered in Chapter 4, section 4.4), and the fact that the CTD is not exposed to the immune system (Schwede *et al.*,

[38] To be consistent with the previous study, sets 01—21 were given numbers corresponding to the 'VSG' nomenclature of Marcello & Barry (2007a). As VSG 14 and VSG 15 of that study qualify as being members of the same set, they were subsumed into Set_14, so there is no 'Set_15'.

2011). In addition to the genomic *VSG* database, a database of unassembled reads was consulted (the 'reads database', described in Chapter 2 section 2.7.2.3), representing minichromosomal sequencing, intermediate chromosome sequencing, and unassembled reads from the megabase chromosome sequencing project). Candidate contiguous reads found in this way were assembled to form putative set-matching donors; those reads contributing to each assembly are shown in Appendix 7.2.6.

As expected, all clones within a set identified the same genomic *VSG*s. For some sets, a number of potential donors were found. In some cases these represent families of related genomic *VSG*s, whereby a sequence with identity to one member of the family will likely have similarly high levels of identity to other family members (40% of *VSG*s are in such subfamilies (Marcello & Barry, 2007a)). In other cases, it may be that more than one donor genomic *VSG* is contributing to the expressed *VSG* to form a mosaic *VSG*. All matching potential genomic copies were recorded for closer analysis.

There remained eight sets for which no matching genomic sequence could be found at all. These sets probably match genomic *VSG*s which are not present in any publicly-available sequencing attempt to date, and are most likely from minichromosomal telomeric loci which are currently under-represented in genome sequence data.

3.4.3 Expressed VSGs *correspond imprecisely with their associated genomic copies*

To examine in more detail the relationships between the expressed forms of the *VSG* gene and their putative silent genomic form(s), multiple sequence alignments of each retrieved clone sequence and all its putative donors were carried out. For each position in the alignment, identities between the clone and each genomic copy were recorded. The results were plotted diagrammatically for each clone, and diagrams and alignments studied by eye. When comparing each clone sequence with its matching genomic copies, four different kinds of variation were seen. Examples of each are shown in Figure 3.6:

(i) Single base mismatches. Single base mismatches were seen both when comparing an expressed *VSG* sequence with the genome, and when comparing expressed *VSG* sequences within the same set. It may be that single base mismatches are artefacts generated by the process of amplification and cloning; this possibility is considered in depth in Chapter 4 (section 4.2).

(ii) 3' donation. In many cases, an expressed gene matched its putative genomic copy very well in the region encoding the N-terminal domain, but very poorly towards the 3' of the gene, in the region encoding the C-terminal domain. This pattern could also be seen when comparing similar expressed clones within a set: they are identical in the ~1000 bp at the 5' end of the gene, but show high levels of variation towards their 3' end. This kind of variation is likely to come about by 3' donation, a process of segmental gene conversion in which the N-terminal-domain-encoding region of one *VSG* is joined to a different C-terminal-domain-encoding region (see section 1.6.8).

(iii) Mosaicism. Often, a gene would show high levels of identity with a putative genomic copy in parts of its sequence, but match that copy poorly in other parts. These areas of mismatch were often complemented by one or more different genomic copies. One explanation for these complementary patterns of matching and mismatching is that segments of one gene are replacing segments of the other to generate the expressed *VSG*, as described previously (see section 1.6.9). This process of segmental gene conversion forms a 'mosaic' *VSG*, with multiple contributing donors.

(iv) Missing donor sequence. Some genes showed one or more promising genomic copies following BLAST analysis, but alignments showed considerable variation between the expressed *VSG* and all putative genomic copies. Comparing members of this set found in different infections or at different timepoints often revealed identical patterns of mismatch to the putative genomic copies. The most likely explanation for this was that the true genomic copy contributing to this set has not been found; instead the BLAST search identified a related member of the *VSG* family, against which the comparisons were done.

In many cases, combinations of these different types of variation were seen, for example a clone having undergone 3' donation and possessing single base mismatches in the 5' NTD-encoding region, or a clone having apparently been constructed by segmental gene conversion, but the database missing one of the donor sequences. Because of the possibility that patterns of variation between an expressed *VSG* set and its genomic copies could be more parsimoniously explained by the presence of an as-yet-unknown additional silent *VSG*, within-set variation between expressed clones from different infections was examined; this approach was the favoured means of identifying patterns of within-set variation.[39] Identical patterns of apparent 'point mutation', '3' donation' or 'N-terminal mosaicism' in these clones would hint at such an unidentified donor. Detailed analysis of all of these patterns is covered in Chapter 4.

Variation between the expressed *VSG* and its best-matching putative genomic copy was prevalent and often considerable. Indeed, only 5 of the 664 clone sequences matched a genomic copy perfectly.

[39] Southern blot hybridizations would be the most rigorous method of establishing the number of related silent copies in the genome. This was carried out for one set (section 4.5.9).

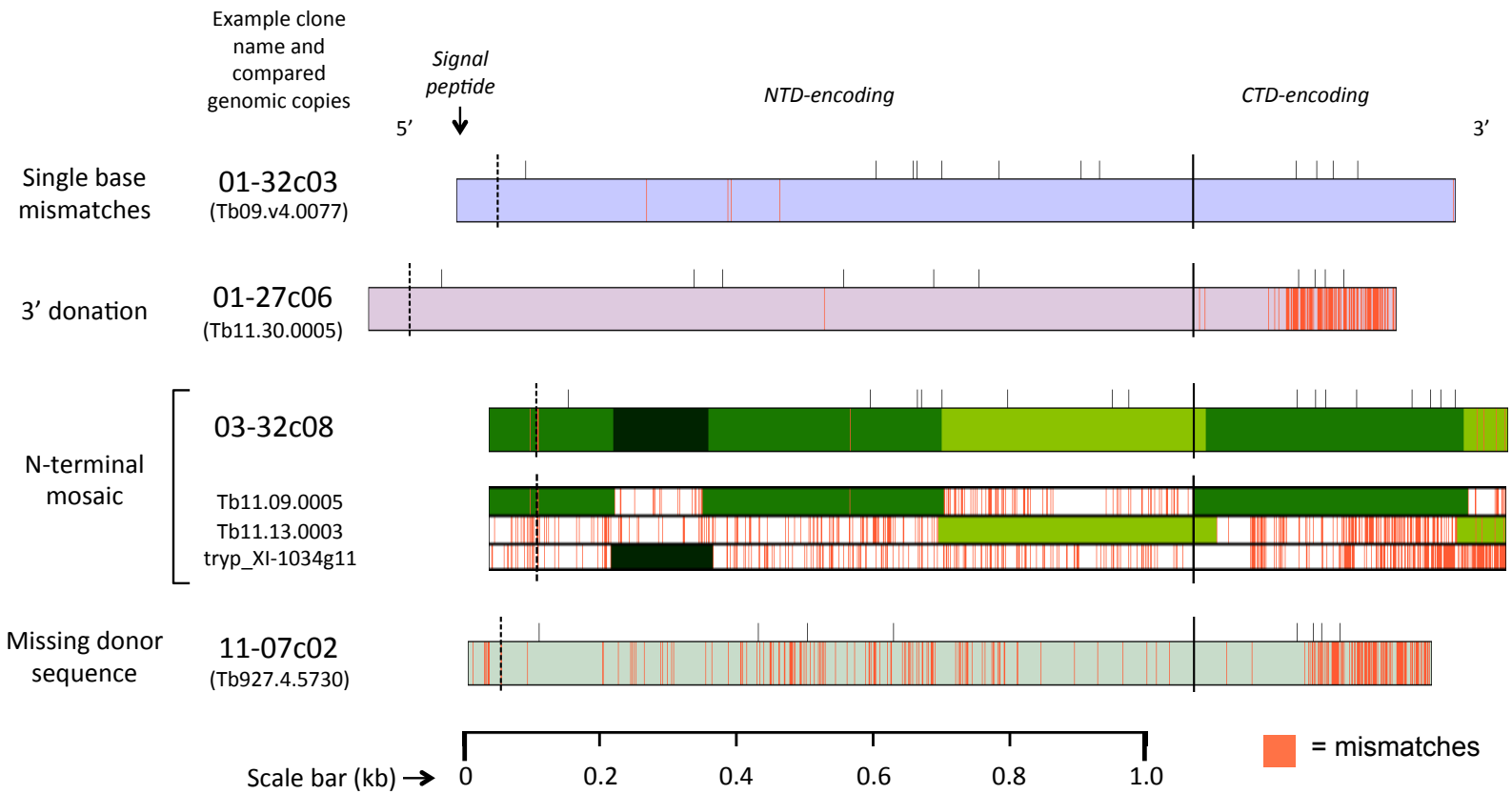


Figure 3.6 Examples of different patterns of divergence from putative genomic copies. Each diagram shows a different cDNA clone, with predicted signal peptide and N- and C-terminal domain encoding regions marked. Cysteine codons are shown as bars projecting above the diagram. Each clone was aligned with putative genomic copies (described under the clone name) and mismatches recorded. Mismatches are shown as red bars spanning the diagram. For the N-terminal mosaic more than one donor is shown. Mismatches with each donor were recorded (lower image) and segments matching each donor are coloured by donor in the upper image. Junctions between segments are drawn at the midpoint of the region of identity between donors. For more detail see Chapter 4 section 4.5.1.

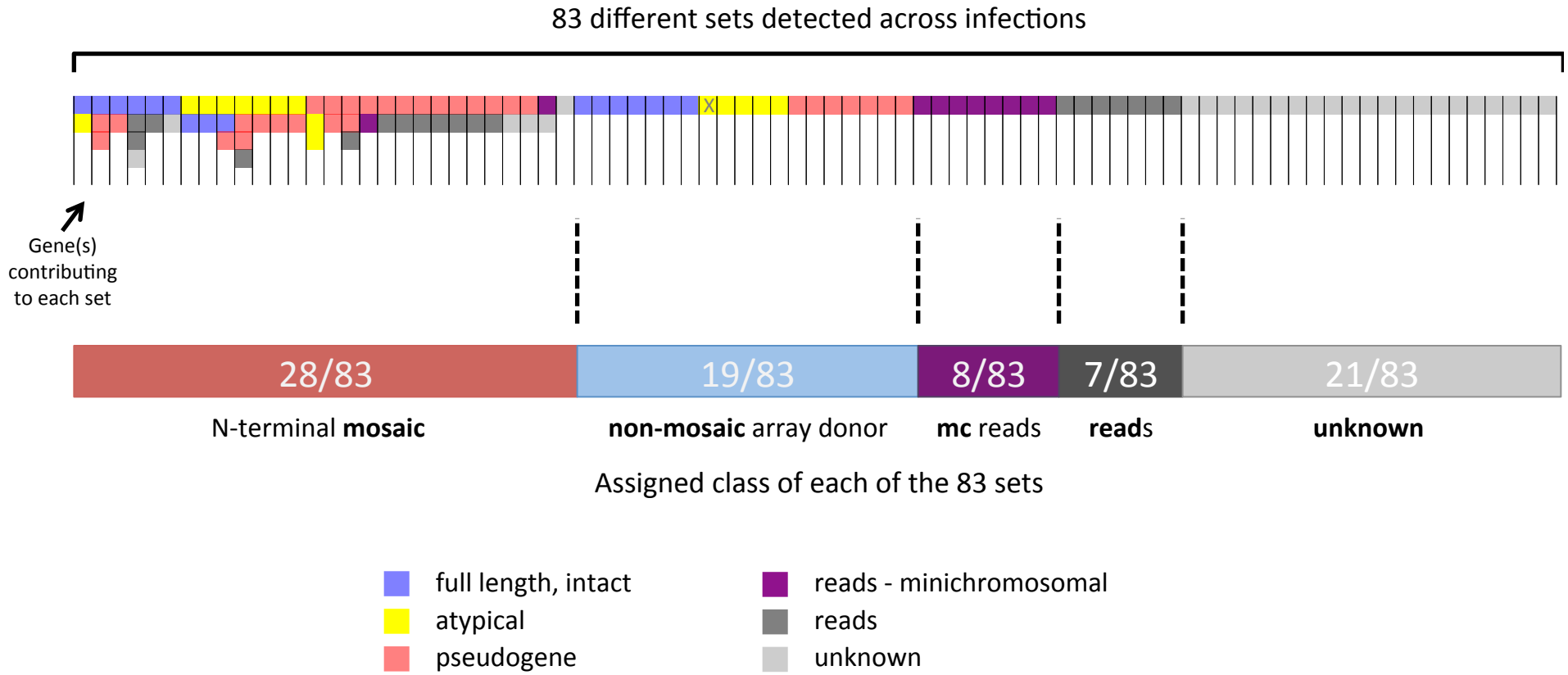


Figure 3.7 The expression of genomic *VSG*s during infection. The upper diagram shows, for each of the 83 sets detected across all 22 infections, the properties of the putative genomic copies. Based on the number and properties of the associated genomic copies, the expressed *VSG* sets could be classified, as shown in the bar chart underneath. The fraction of sets classified in each way is overlayed on the bar chart. There was one set for which patterns of mosaicism could be inferred but no associated genomic copy could be found (see Chapter 4 section 4.5.1). The 'X' indicates Set_01, which corresponds with Tb10.v4.0001, also known as GUTat 10.1, BES-occupying at the beginning of infection and also possessing a silent telomeric copy (J. D. Barry, pers. comm.)

Together, these results allowed the likely genomic copies for each set to be assigned, and their properties to be investigated. Genomic copies were classified as 'full-length intact', 'atypical', or 'pseudogene', using the annotations present in VSGdb. Where the donor had been assembled using reads from the 'reads database', the donor was classified 'read', unless the read was annotated as minichromosomal, in which case the donor was classified 'minichromosomal'. Where a donor could not be found it was classified 'unknown'.

Each set of expressed *VSG*s was itself classified based on the classification and number of contributing donors. Those sets whose members often show patterns of mosaicism were classified as 'N-terminal mosaic'. Those sets showing the contribution of only one *VSG* to the N-terminal domain, and a putative genomic copy of that donor could be found in the subtelomeric arrays, were classified as 'non-mosaic array donor'.[40] Those sets showing the contribution of only minichromosomal reads to the N-terminal domain were classified as 'minichromosomal'. Those sets for which only a (non-minichromosomal) read or set of reads could be identified were classified 'read', and those sets for which no satisfactory donor(s) could be found were classified 'unknown'.[41] The involvement of 3' donation was detected to varying degrees over all set classes, including those classified 'non-mosaic array donor'. Figure 3.7 shows the classification of the genomic copies contributing to each expressed *VSG* set, and the set's classification.

3.4.4 Activation of damaged VSGs *is not disfavoured*

Eleven out of 83 of the total sets detected (~13.5%) represent the non-mosaic[42] activation of an intact or atypical *VSG* from the *VSG* arrays. Considering the fact that other array *VSG*s must undergo varying degrees of repair for their expression, this is comparable with the proportion of full-length, intact/atypical silent *VSG*s in the archival arrays (14%, (Marcello & Barry, 2007a)). On the other

[40] It is possible that there are additional, possibly telomeric, copies of the 'non-mosaic array' donors.

[41] Partial donors could be found for some 'unknown' sets, but they were not sufficient for the set to be reclassified.

[42] However, some of these show 3' donation that may have come about by mosaic-like events accumulating in their 3' ends (discussed further in Chapter 4).

hand, those full-length, intact *VSG*s whose expression is reported here are just a fraction of the total full-length, intact silent *VSG*s in the arrays (known to number at least 43, (Marcello & Barry, 2007a)), many of which were not assigned to any expressed sets at all.

There are certainly other full-length, intact genomic copies, currently unassembled, that directly correspond to some of the other sets. For example, those minichromosomal, read and unknown sets detected early on in infection probably correspond with intact telomere-proximal *VSG*s (as described in section 3.5.5). It is also possible that other full-length, intact silent *VSG*s were expressed during the infections but were not detected—perhaps the time of their appearance was underrepresented in the sampling timepoints. Here, times of sampling focussed on the later stage of infection, with the majority of clones (629/720) coming from days 20–34 when mosaic *VSG*s were known to predominate (Marcello & Barry, 2007a). However, considering the 4–5 week infections as a whole, full-length, intact genomic copies do not appear to be 'exhausted' before the expression of damaged *VSG*s becomes prevalent, counter to the hypothesis that antibodies raised during infection are required to select for 'rare' segmental gene conversions.

Seven sets detected in these infections directly correspond with specific damaged[43] genomic copies (pseudogenic towards their CTD-encoding region), and the involvement of up to 43 damaged genomic copies could be detected.[44] Damaged by frameshifts, stop codons and fragmentation, these genes cannot be straightforwardly expressed. Examining these sets in closer detail, it is apparent that processes of mosaicism and 3' donation have enabled (partial) expression of these silent *VSG*s. Twenty-eight out of the 83 sets show patterns of N-terminal domain mosaicism, and 3' donation is detectable in 47 of the sets. The integrity of the silent *VSG* is apparently not the main determinant of its expression probability.

[43] That is, non-mosaic array donors annotated 'pseudogene'.

[44] This figure includes 31 array pseudogenes, three minichromosomal read assemblies and nine other read assemblies. For more details see Appendix 7.2.9.

Besides enabling expression, the processes of mosaicism and 3' hybridism have generated variation in the expressed clones. This variation is clear not only when comparing the expressed clone with its assigned donors, but also when comparing similar expressed clones with one another (within-set variation), suggesting that activation of some genomic copies goes hand-in-hand with the generation of diversity in the expressed antigen population. Could this diversity also contribute directly to antigenic variation by generating novel epitopes? This question will be considered in detail in Chapters 4 and 5.

Given the extreme variation between expressed *VSG* clones from different sets at the amino acid level, it is reasonable to believe that different sets are antigenically distinct from one another. Compared with within-set variation, variation between clones of different sets is consistently greater. The activation of different sets by the parasite population—by any means—is therefore the most obvious source of antigenic variability over the 4–5 week infections presented here. It is the patterns of appearance and disappearance of distinct sets, corresponding with the activation of completely different silent *VSG*s, that represent the classical process of African trypanosome antigenic variation, and it is to these patterns that I now turn.

3.5 Patterns of antigenic variation

3.5.1 Different VSG*s were detected at different times*

The classical model of African trypanosome antigenic variation predicts that, as an infection progresses, the successive immune responses mounted by the host against parasite surface coats constantly select those parasites which have switched to express an 'unused' *VSG*. Within an infection, we might therefore expect to find different sets being expressed at different timepoints, as the host eliminates those parasites that have not switched.

To investigate the patterns of between-set antigenic variation over the course of infection, bubble charts were drawn for each infection and are shown in Figure 3.8. Each chart shows time on the x-axis and the different expressed *VSG* sets on the y-axis. For each timepoint, a circle indicates the presence of an expressed *VSG* clone from that set. The area of the circle is proportional to the

fractional abundance of clones from that timepoint belonging to that particular set. The figures inside the bubbles indicate the total number of clones belonging to that set at that timepoint.[45] These charts show three broad patterns in set expression:

(i) Variation in the different sets detected over time within an infection. For example, in mouse 05, clones belonging to Set_08 were found at day 20, but were not seen in any subsequent timepoints, whereas clones belonging to Set_21 were not found at the earlier timepoints, and were seen only from day 27 onwards. These patterns might be expected, as antibodies clear earlier variant sets and later variant sets expand to fill their place.

(ii) Persistence of specific *VSG* sets over time within an infection. For example Set_14 in mouse 04 and mouse 05 is found in all samples between days 21 and 31 (mouse 04), and 23 and 31 (mouse 05). The processes leading to apparent persistence are considered in the broader context of antigenic variation in section 3.6.3.

(iii) Occasional reappearance of a *VSG* set in an infection. For example, Set_08 is present at days 20, 23 and 32 in mouse 01, but is not detected at the intermediate timepoints. Reappearance of a *VSG* set in the later stages of infection may be due to that variant's high activation probability: in periods of high parasitaemia, many parasites will undergo futile switching to an easily-activated *VSG* and will be rapidly killed by the previously-developed immune response (Turner, 1999). Other processes that may contribute to reappearance are similar to those of persistence, and are discussed in section 3.6.3.

[45] Note that the relationship between the number of clones of a particular set retrieved at a timepoint and the abundance of that particular set in the original mRNA is not known: increased numbers of clones of a particular set may simply represent sequences that are preferentially amplified and/or cloned. However, no difference was found in the GC content (a factor commonly contributing to inefficient PCR) between sets for which 10 or more clones were retrieved and those for which only one clone was retrieved (data not shown).

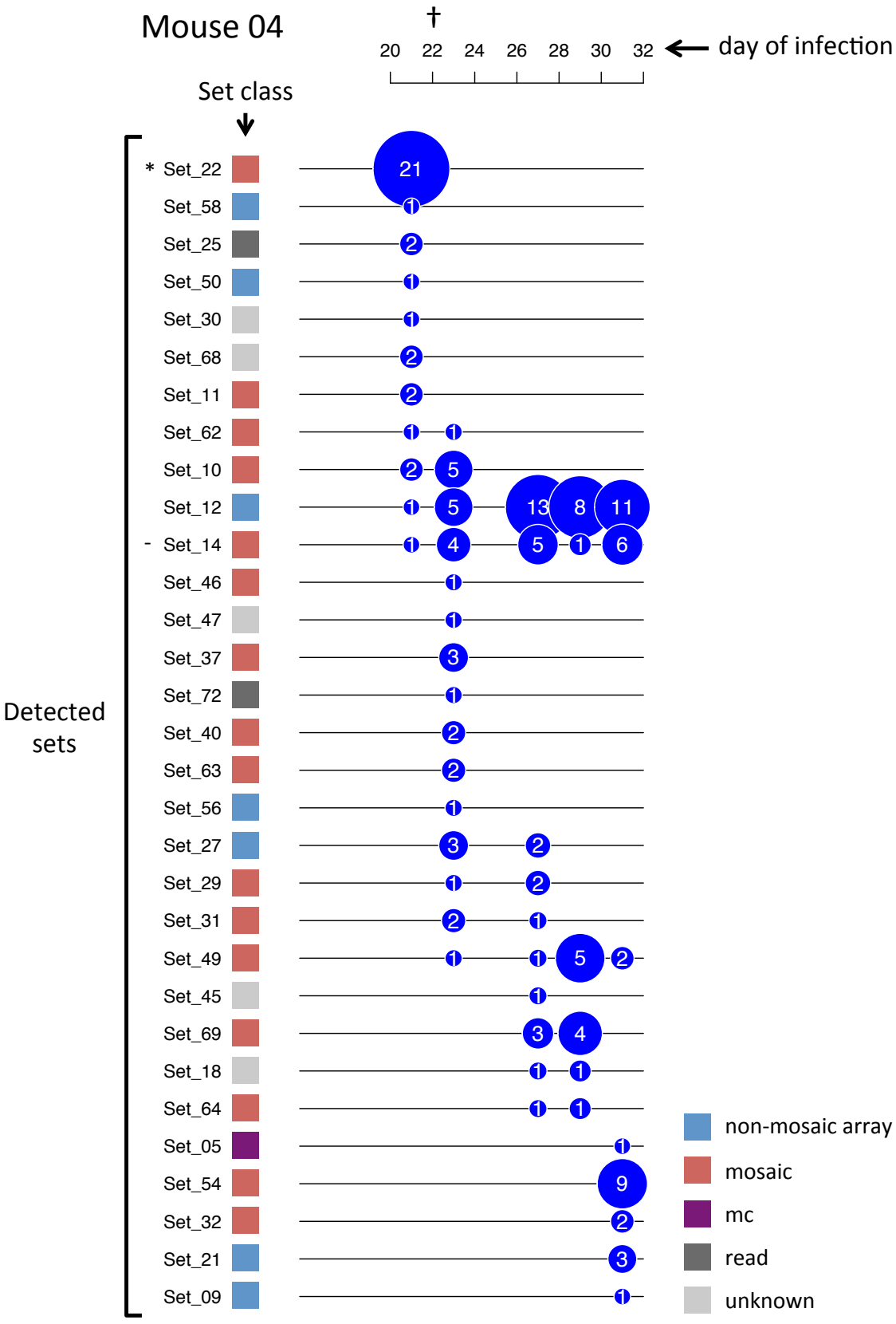


Figure 3.8 Bubble charts of *VSG* sequence detection. Different charts have been drawn for each infection. *VSG* sets and their classification are on the y-axis, and the time of sampling is on the x-axis. The area of each bubble is proportional to the fractional abundance of clones of that set at that timepoint. The actual number of clones is overlaid on the bubble. The dagger indicates the transient dip in parasitaemia marked 'C' on Figure 3.2 (see Appendix 7.2.2 for parasitaemias). Set_08 and Set_22 are indicated by ' * ' and Set_14 is indicated by '–'.

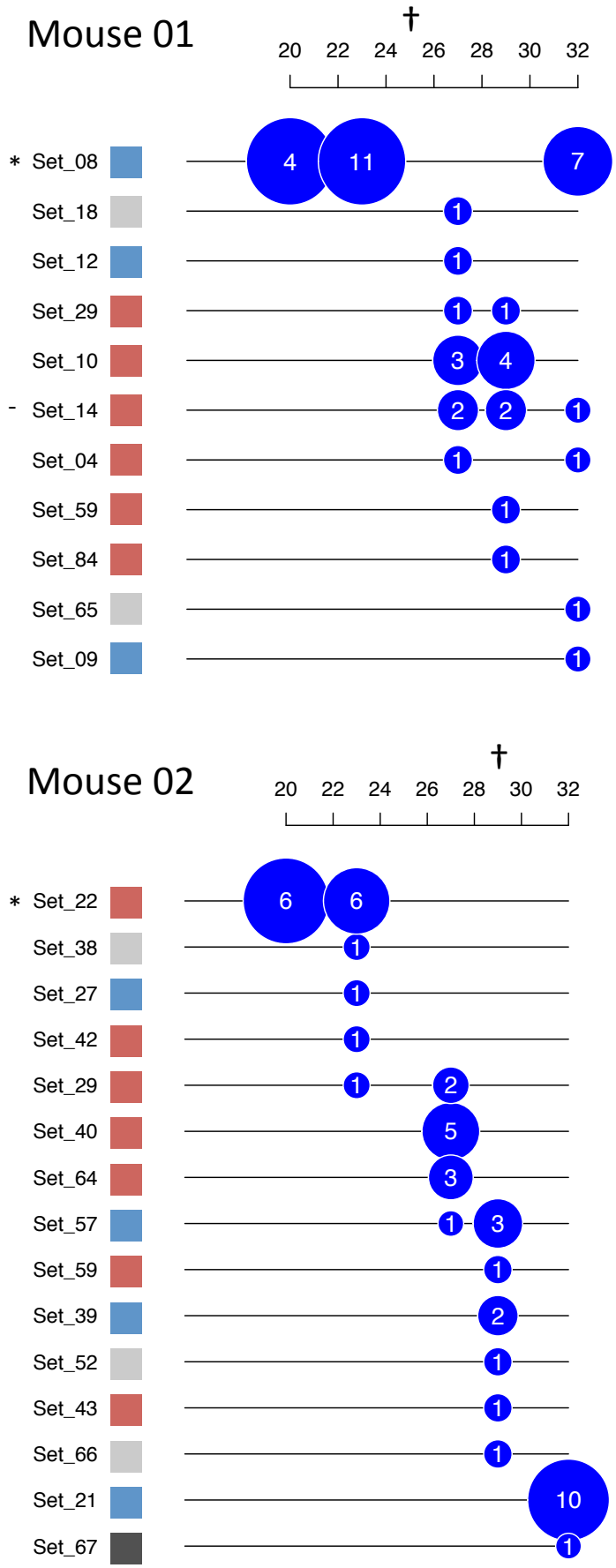
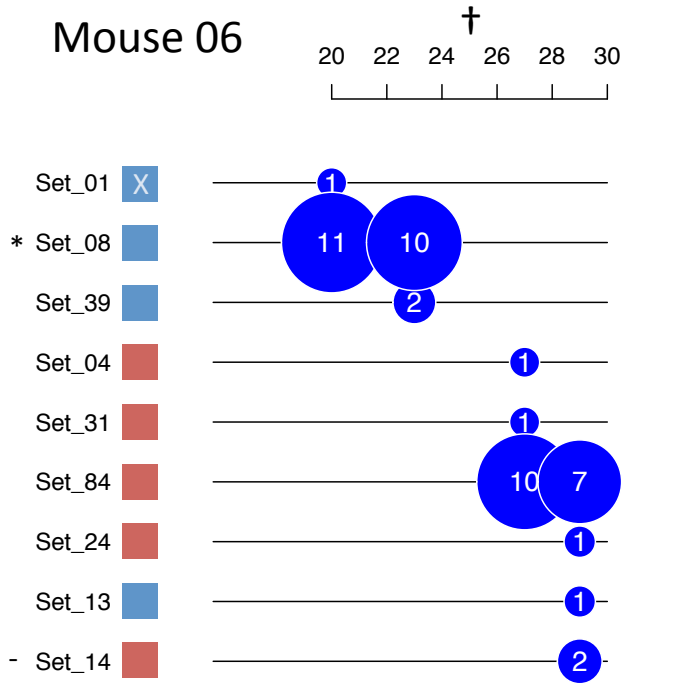


Figure 3.8 (continued)



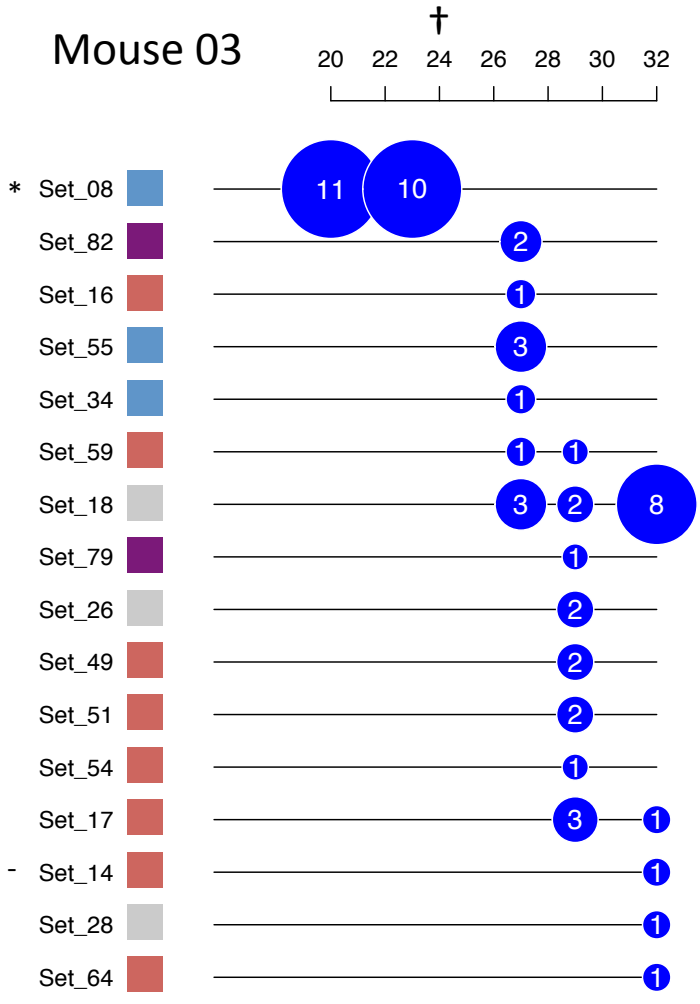


Figure 3.8 (continued)

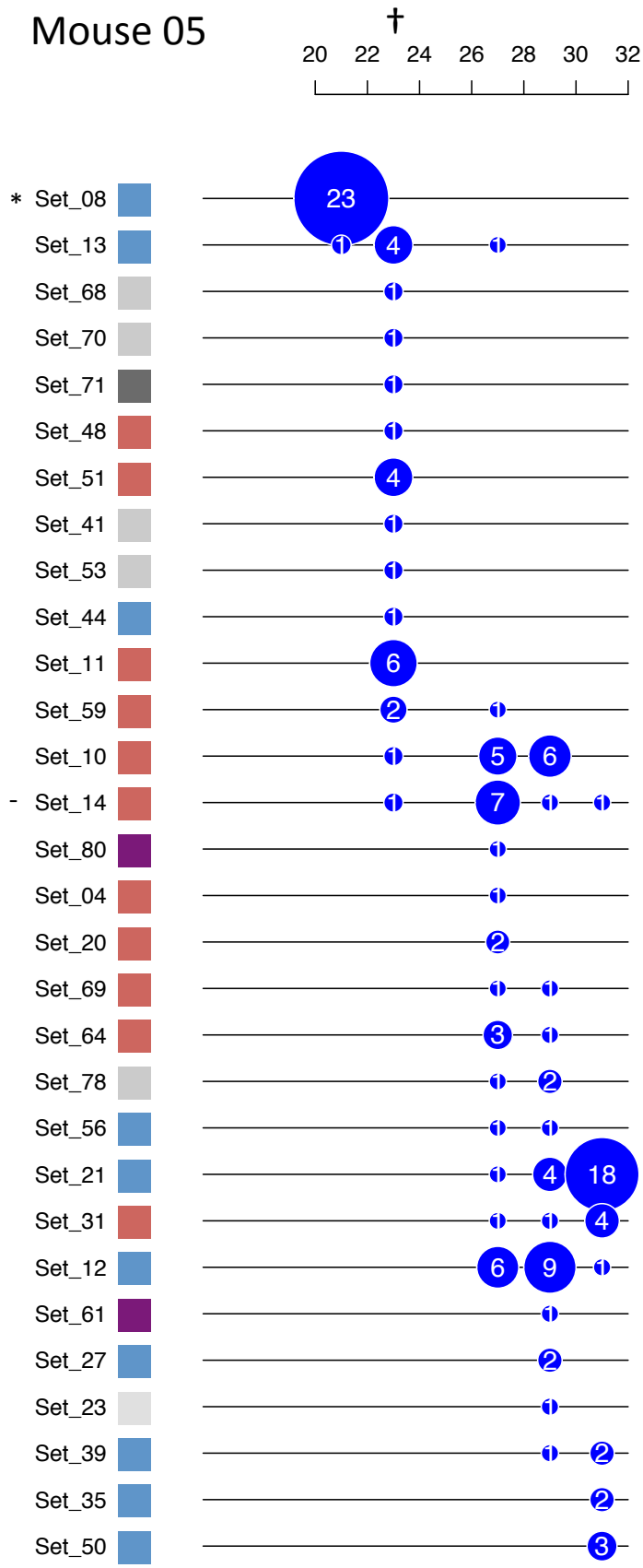


Figure 3.8 (continued)

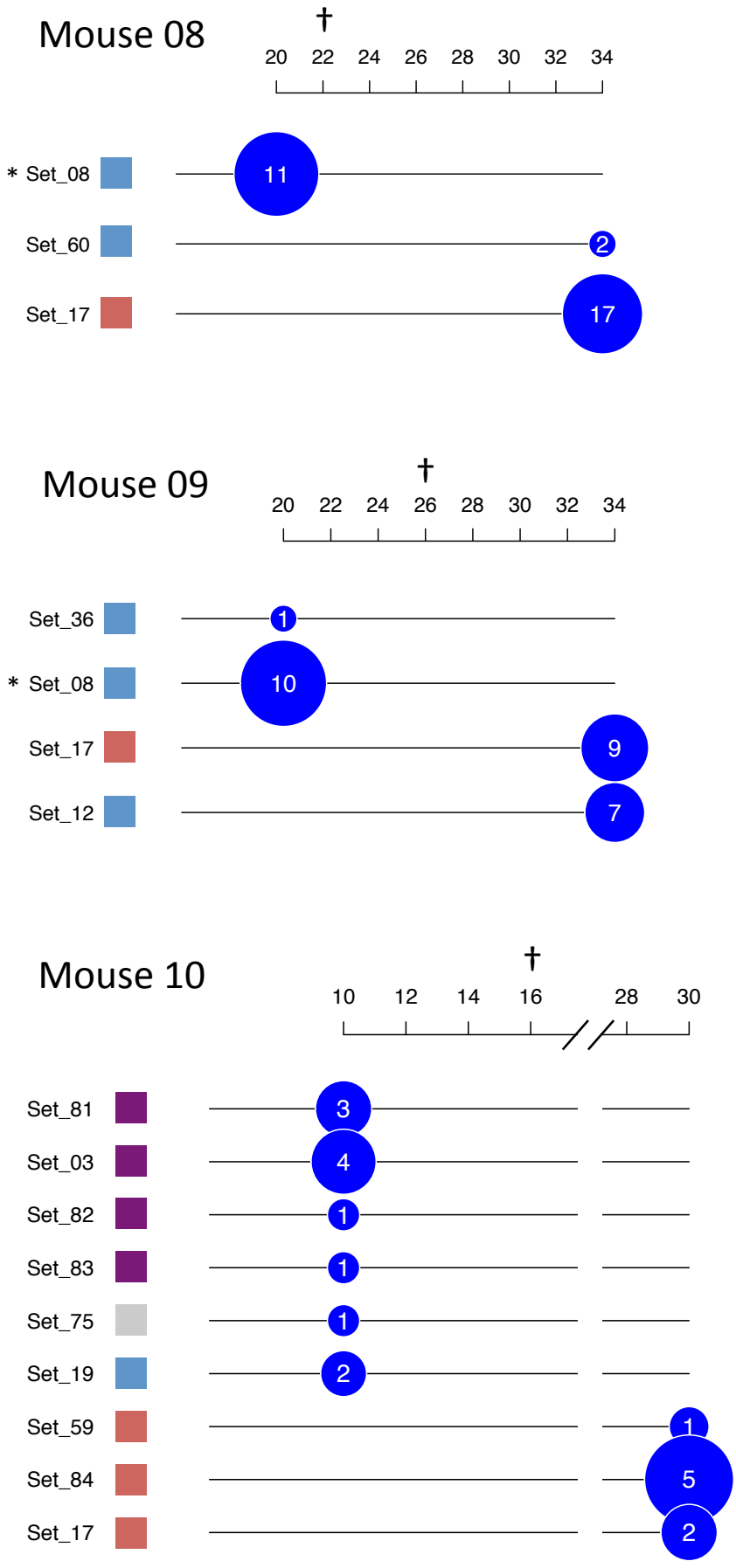
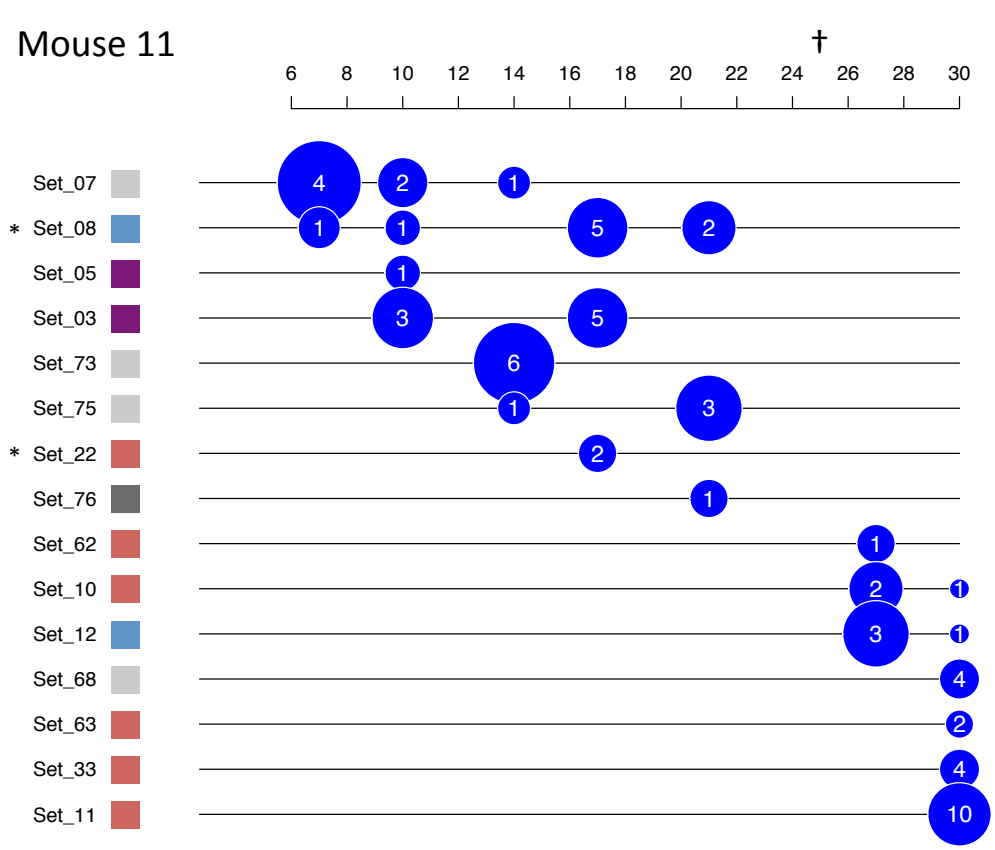


Figure 3.8 (continued)



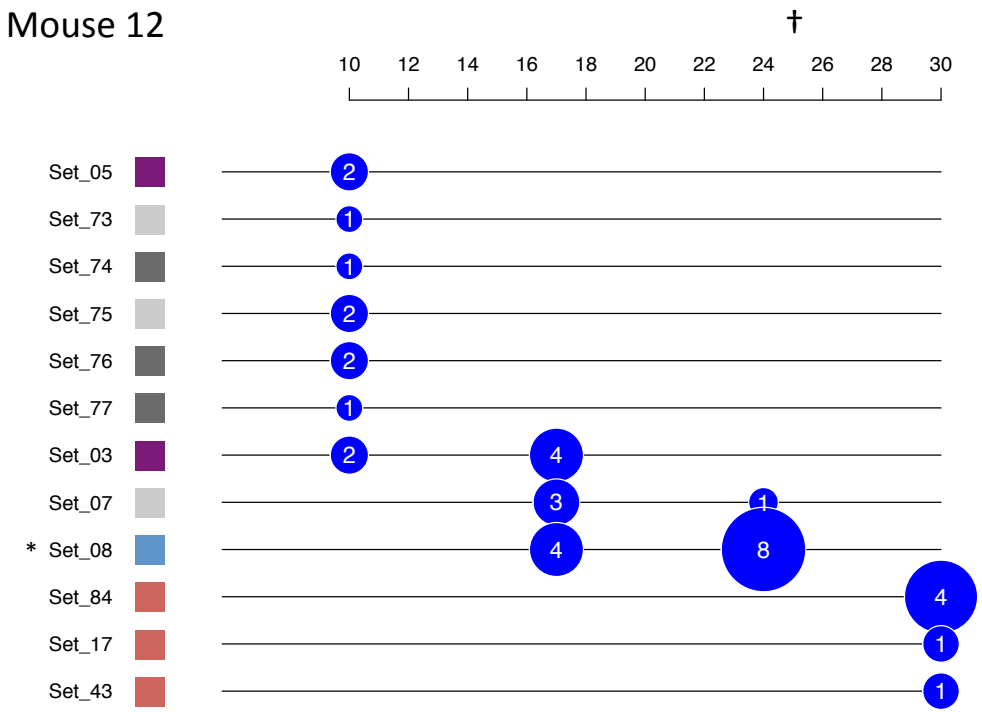


Figure 3.8 (continued)

3.5.2 PCR analysis was consistent with clone sequencing

Although the detection of a member of a specific set at a timepoint confirms its presence, its non-detection does not necessarily indicate its absence. As clones were selected for screening arbitrarily, it may have simply been the case that at those timepoints where the set was not detected, the set was present but clones were not selected. This was a particular concern for those timepoints where relatively few clones were sequenced.

To control for this possibility, PCRs were carried out directly on the cDNA product using primer pairs specific for different *VSG* sets. Primers were designed using all available sequence information, taking into account known within-set variation to allow annealing to any member of the set. Each primer pair amplified a region of the gene ~500–1000 bp long. Primer pairs were designed for six representative sets, as shown in Table 3.3, and tested for specificity on genomic DNA and plasmid clone DNA (data not shown). PCRs were performed on cDNA and plasmid DNA controls from the five timepoints of three separate infections (mouse 01, 04 and 05), as well as from the infecting population (feeder mouse FF). The reaction products were run on a 1% agarose gel. For each timepoint, primers directed against tubulin were used to control for equal sample loading. For these experiments *Taq* polymerase was used. The results of the PCRs are shown alongside the corresponding bubble charts to allow easy comparison are shown in Figure 3.9.

The PCRs broadly corresponded with the cloning shown in the bubble charts, confirming the process of between-set variation over the course of infections. For example, the sets detected by cloning and sequencing only at day 21 in mouse 04 and mouse 05 (Set_22 and Set_08) did not give a strong PCR product at any other timepoint in either of those infections. Similarly, Set_21, which was identified in the sequences only at the later timepoints, was only amplified from the later cDNA samples.

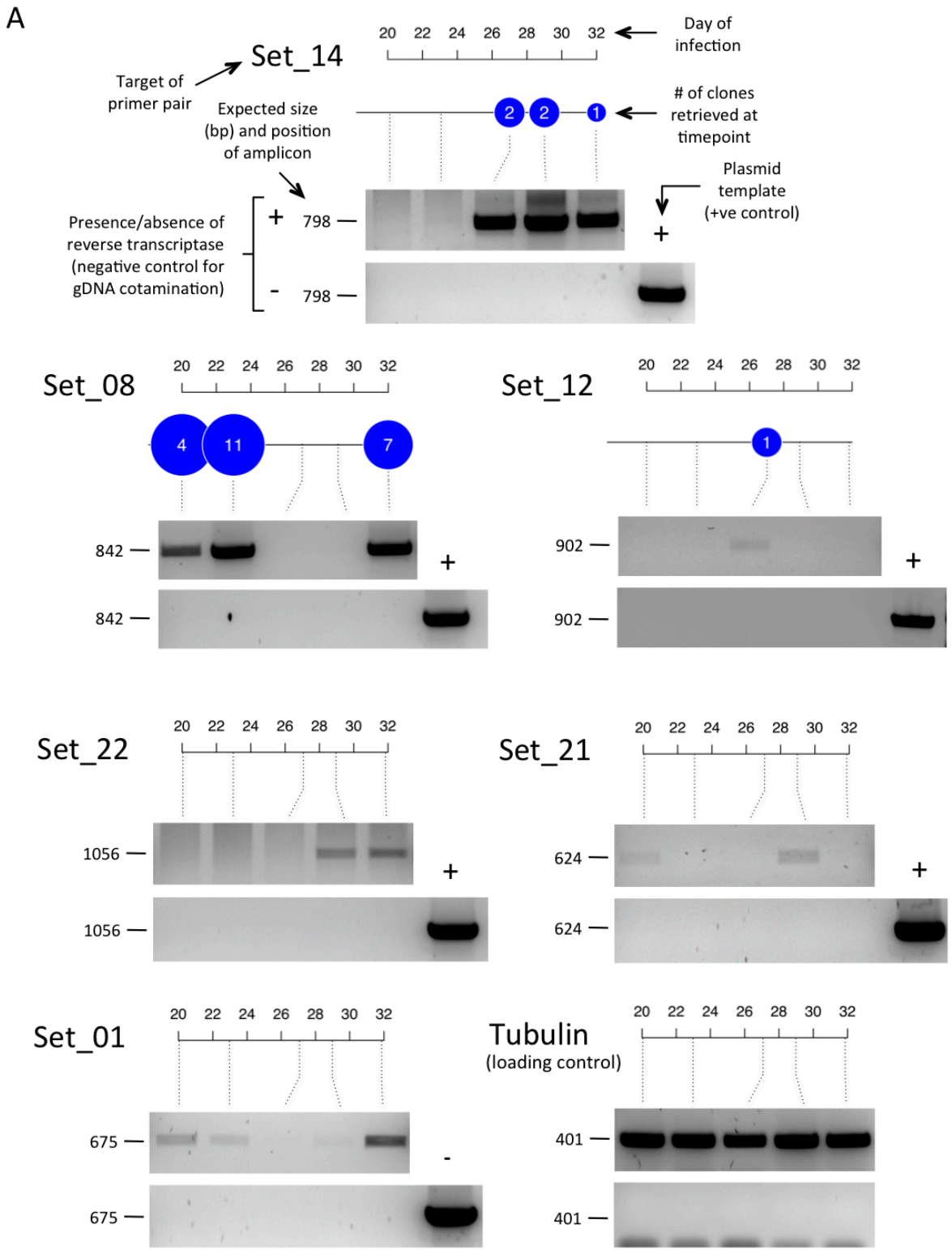


Figure 3.9 Antigenic variation PCRs. Seven primer pairs were used to test expression of six antigen sets and tubulin (to control for equal template loading). Samples from four infections were tested: mouse 01 (A), mouse 04 (B), mouse 05 (C) and mouse FF (D). Reactions were carried out on total cDNA, as well as the product of a reaction from which reverse transcriptase was absent, as a control for gDNA contamination. For each reaction, the corresponding line from the bubble chart (Figure 3.8) is shown, so the results can be easily compared.

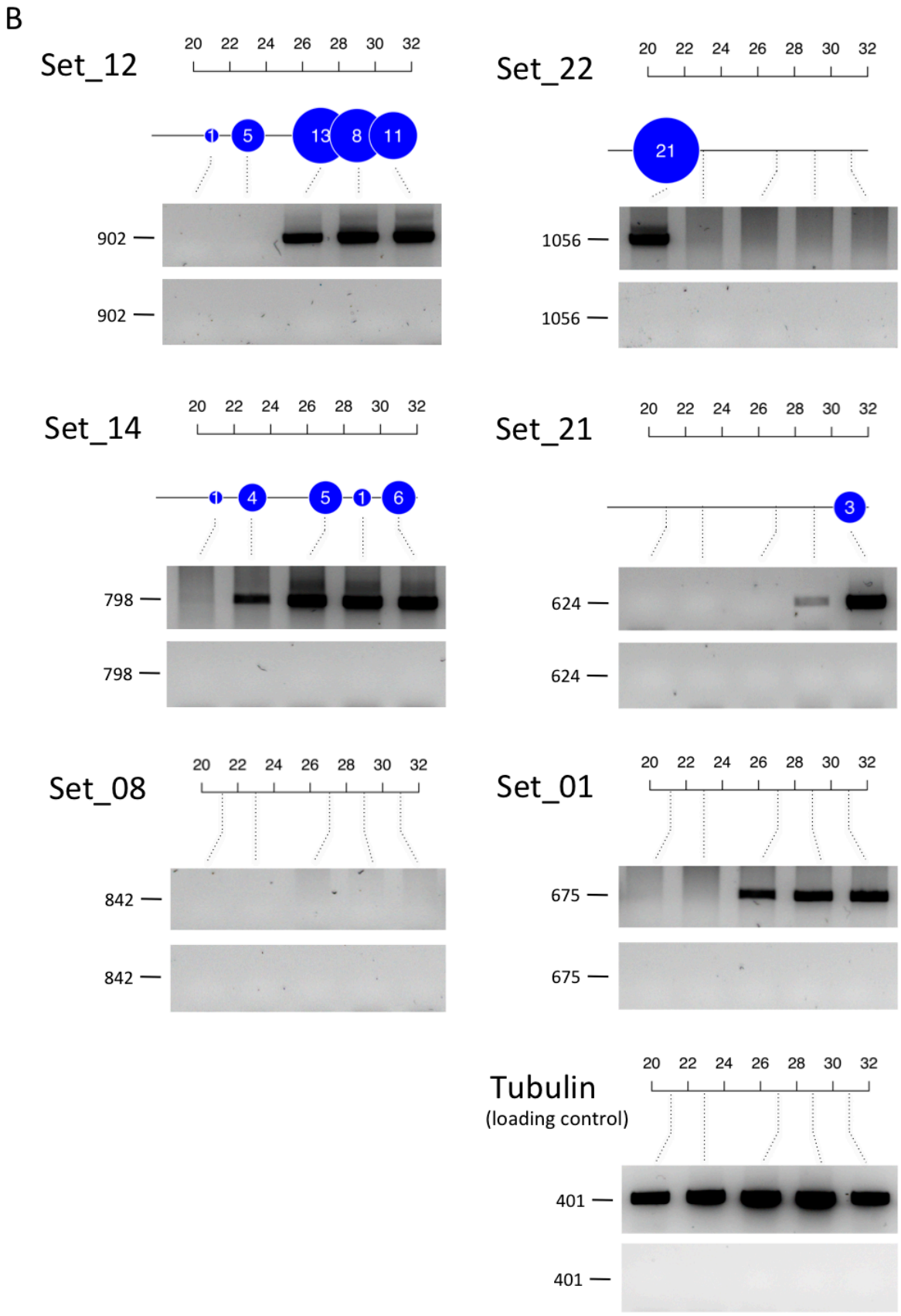


Figure 3.9 (continued)

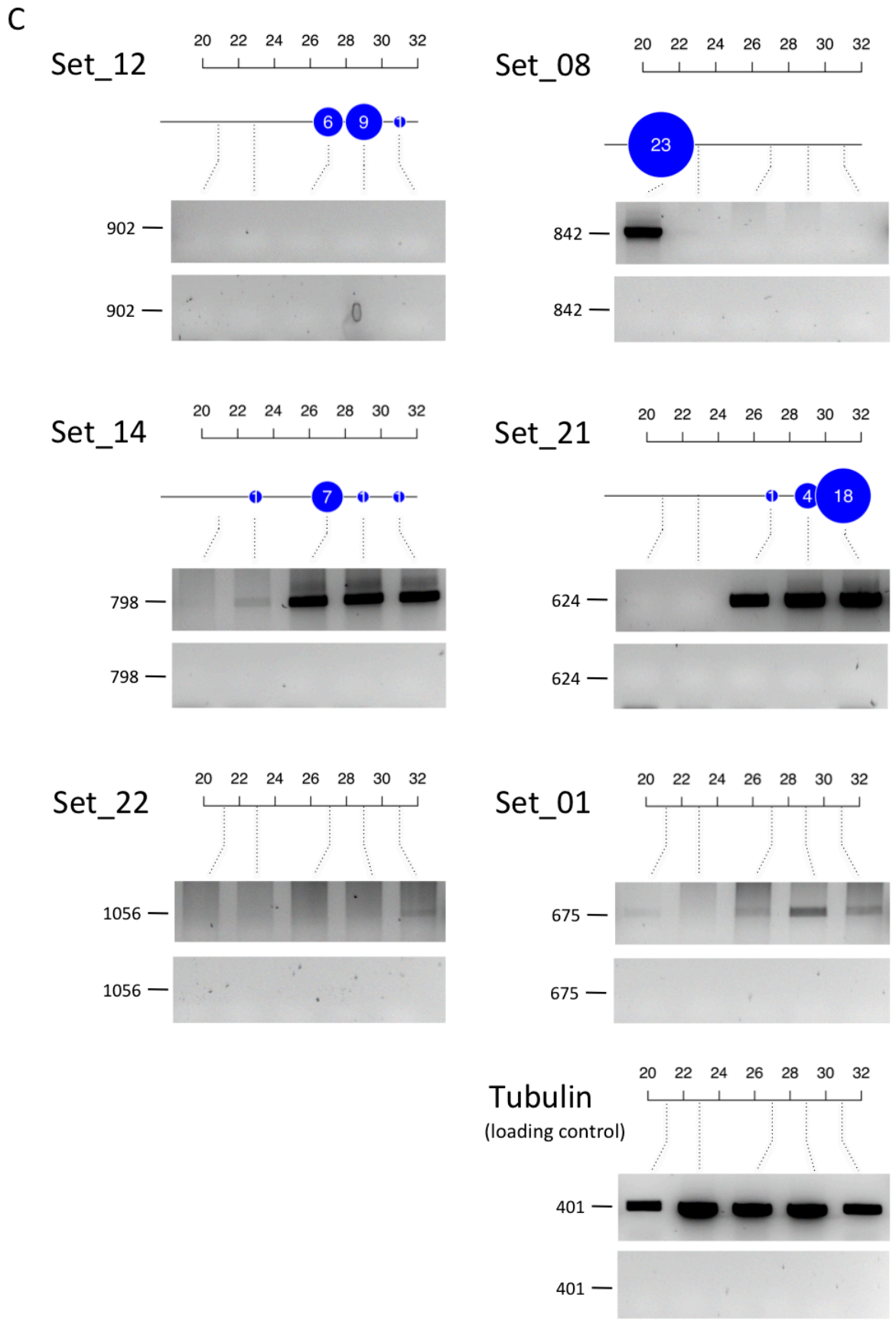


Figure 3.9 (continued)

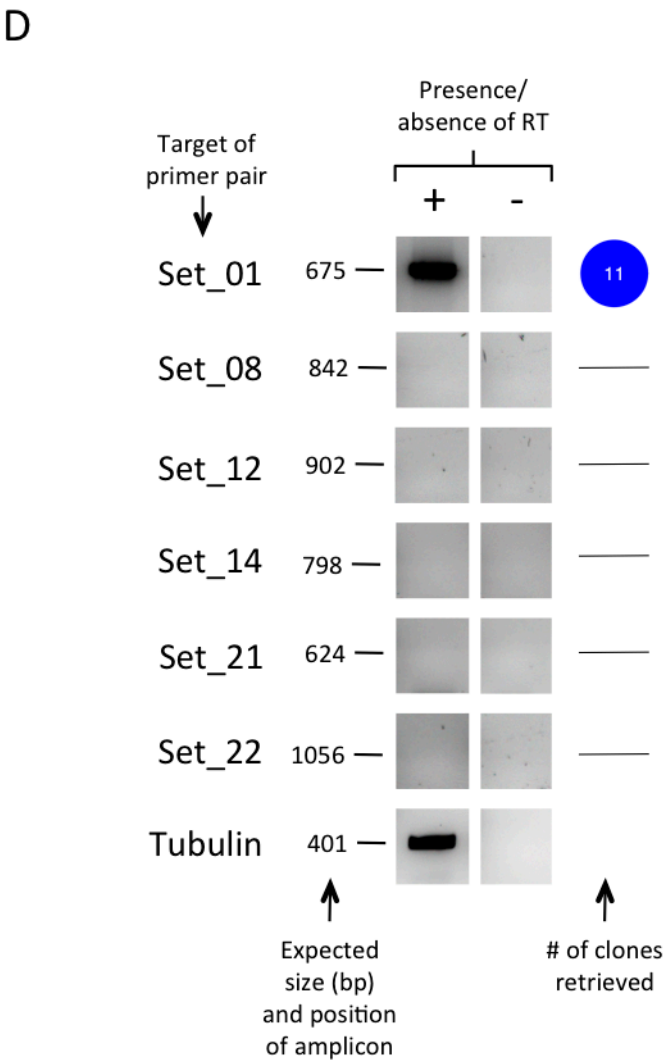


Figure 3.9 (continued)

Set	Class	Time of appearance	Amplicon size
Set_01	Non-mosaic	Initial (expressed by inoculum)	675 bp
Set_08	Non-mosaic array*	Early-mid	842 bp
Set_22	Mosaic	Mid	1056 bp
Set_12	Non-mosaic array	Mid-late	902 bp
Set_14	Mosaic	Mid-late	798 bp
Set_21	Non-mosaic array	Late	624 bp
Tubulin	N/A	Constant (loading control)	401 bp

Table 3.3 *VSG* sets tested by PCR. *Set_08 may have an additional, non-array copy at the beginning of infection (see Chapter 4 section 4.4.3).

In some cases, there was a discrepancy between the PCR analysis and the cloning. PCR analysis revealed the presence of *VSG* sets not detected by cloning (for example, Set_21 was not detected by cloning in mouse 01). This was most

frequently seen towards the later timepoints. PCR analysis also identified the reactivation of *VSG* sets not detectable at intermediate timepoints, the most striking of which is Set_01, the only set identifiably expressed by the parasites used to initiate the infections[46] (material collected from the feeder mouse FF). The processes leading to reactivation of early *VSG*s are discussed in section 3.6.3. On the other hand, some sets represented by many clones were identified only very weakly by PCR if at all, for example Set_12 in mouse 05. Poor amplification might have been due to differences in primer binding or the relative processivity of the different polymerases used for the PCR reactions (Arezi *et al.*, 2003).

3.5.3 Infections showed unique patterns of VSG *expression*

The *VSG* sets detected varied uniquely between each infection. Figure 3.10 shows these patterns as a network, in which a connection between nodes represents detection of different *VSG* sets within an infection. Some sets were uniquely connected to infections, whereas others were shared amongst infections. This result is consistent with the hypothesis that *VSG* expression is '*semi*'-ordered, rather than strictly orchestrated, perhaps providing flexibility in a partially-immune host.

It might be the case that deeper sampling and sequencing would reveal further *VSG* sets that are shared between infections, but are being kept below the level of detection here by short-stumpy differentiation. Mouse 04 and 05, from which more clones were retrieved than the other infections, showed a greater number of uniquely-detected sets.

[46] Also known as GUTat 10.1, and Tb10.v4.0001.

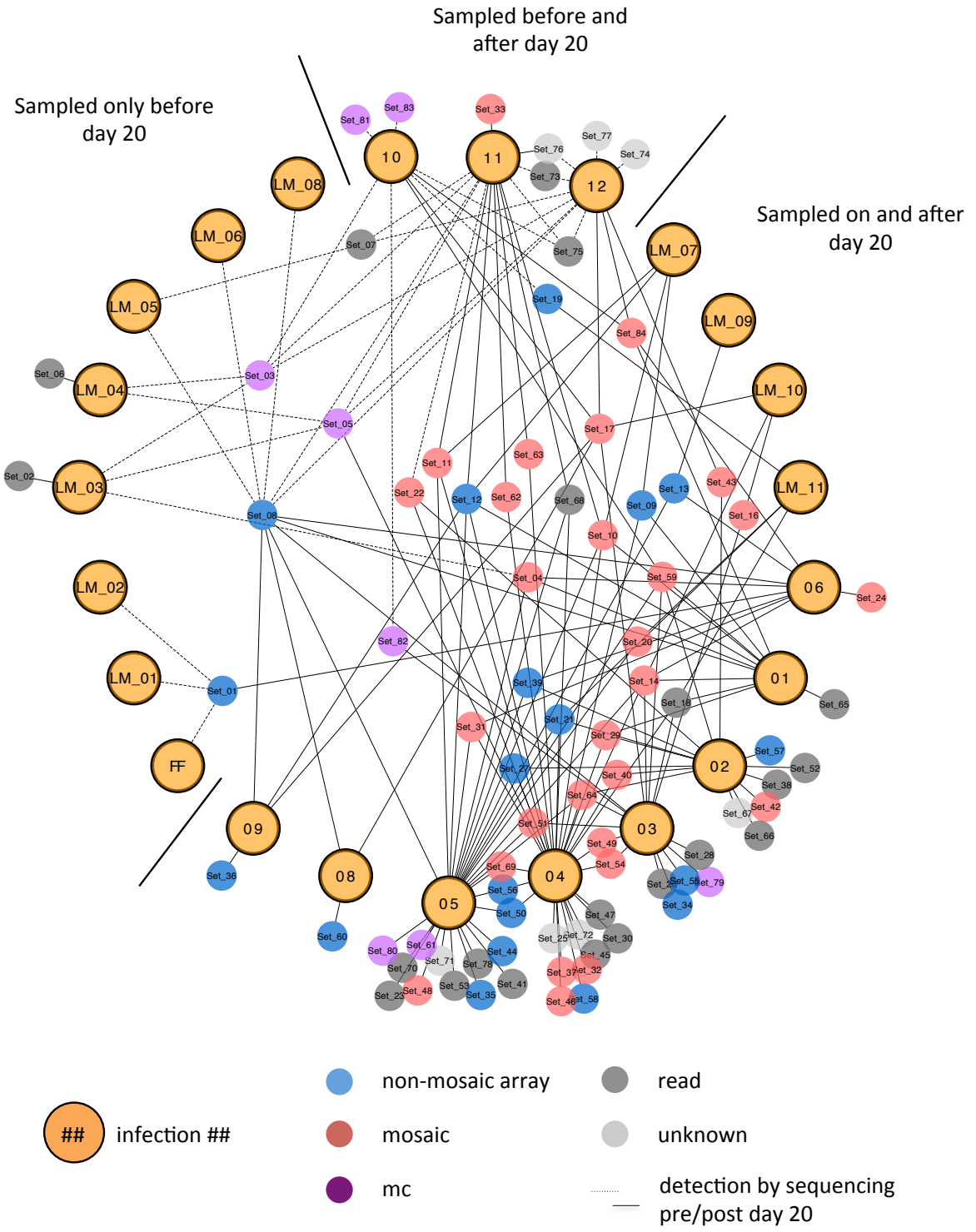


Figure 3.10 Expression of *VSG* sets in different infections. The larger, yellow nodes represent the different infections. Smaller nodes represent *VSG* sets. Edges indicate that the set was detected in that infection by sequencing, with dotted and intact lines representing identification before and after day 20 respectively. Infection nodes have been grouped according to their time of sampling, as earlier timepoints have not been sampled in so great a depth. Sets detected in more than one infection are shown inside the ring of infection nodes, whereas sets unique to an infection are outside.

3.5.4 Infections showed great expressed VSG *diversity*

The bubble charts presented in section 3.5 show that a wide range of different *VSG* sets could be detected in a small blood sample collected at a single point of the infection. The number of different variable antigen types present in an infection is referred to as the 'variant richness' of the population. Variant richness can be considered alongside 'variant evenness'—the relative abundance of different variable antigen types—as measures of expressed antigen diversity (Hill, 1973).

A crude estimation of variant richness could be achieved by counting the number of sets present at different timepoints. Considering richness or unevenness as concerned simply with the number and relative abundance of different sets ignores the presence of extensive within-set variation that also contributes to diversity (discussed more fully in Chapter 4). However, this simplification lends confidence to the assertion that the estimated variant diversity is functional in antigenic variation, as many cases of within-set variation (in particular, point mutations and 3' donation) will probably not affect the epitope structure of the antigen.

To compensate for a lack of saturation in sequencing—at most timepoints, many sets were represented by just one clone, leaving it likely that further sequencing would identify additional sets at that timepoint—the Chao 1 richness estimator (Chao, 1987) was calculated using the statistical package EstimateS (R. Colwell, http://purl.oclc.org/estimates). This approach estimates total variant richness using the number of 'singletons' and 'doubletons' in the population (the sets represented by one or two members respectively) as a way of gauging the likelihood of identifying additional sets with further sequencing. For this analysis, the diversity of each sample of mouse 04 and mouse 05 was estimated, as similar numbers of clones were obtained from each sample (between 20 and 30) and the increased number of clones sequenced from each sample gave a better idea of total richness (diversity estimates depend greatly on the size of a sample (Hill, 1973)).

Infection #	day	# of clones	Observed # of sets	Singletons	Doubletons	Estimated richness	95% lower bound	95% upper bound
04	21	35	11	6	4	14.0	11.5	29.0
	23	33	15	7	3	20.3	16.0	42.6
	27	30	10	5	2	13.3	10.5	32.1
	29	20	6	3	0	9.0	6.4	30.6
	31	35	8	2	2	8.3	8.0	14.0
05	21	24	2	1	0	2.0	2.0	2.0
	23	25	13	9	1	31.0	17.0	94.3
	27	32	14	9	1	32.0	18.0	95.3
	29	31	13	8	2	22.3	14.9	58.0
	31	31	7	2	2	7.3	7.0	13.0

Table 3.4 EstimateS calculations. The richness estimate was Chao 1, calculated using correction for bias, and relates to total predicted antigen richness (number of variants present). The large margin of error (particularly amongst mouse 05 timepoints) is reflective of the proportion of singleton and doubleton sets.

Evidently, the antigen profile of the parasite population at a single point in an infection can be remarkably heterogeneous. Some samples showed little variant richness, particularly those taken at day 20/21. At these timepoints, samples were dominated by one of two sets: Set_22 in mice 02 and 04, and Set_08 in mice 01, 03, 05, 06, 08 and 09 (this can be seen in Figure 3.8, where Set_08 and Set_22 are indicated by a ‘ * ’). It is possible in these cases that unevenness in the population is crowding out minor variants that are nonetheless present: deeper sequencing of the sample taken at day 21 of mice 04 and 05 showed smaller variant subpopulations existing alongside a dominant set. Crowding out might be expected in cases where differentiation is the predominant factor controlling infection dynamics—minor variants are suppressed below the level required for immune stimulation by stumpy induction factor released by the dominant variant (Gjini *et al.*, 2010).

Following day 20/21, many infections showed an apparent increase in diversity. This is seen most clearly in mice 01, 02, 03, 05 and 06, which show Set_08 or Set_22 sets at day 20/21, and their absence at day 23 or day 27 with a number of other sets becoming detectable in their place (witness the large number of sets becoming detectable at day 27 in the bubble charts, and in Figure 3.11). For many infections, this change in expressed diversity corresponded with the transient but substantial dip in parasitaemia. Perhaps the crash in parasite numbers was due to the bulk of the parasite population—expressing either Set_08 or Set_22 VSG surface coats—being eliminated by an immune response,

exposing minor variant subpopulations that were detected as they expanded over the subsequent days.

3.5.5 VSG *expression followed a hierarchy loosely based on genomic locus*

The classical model of African trypanosome antigenic variation predicts that different genomic *VSG*s have different activation probabilities, which are believed to be a consequence of their genomic location (*VSG*s located on telomeres are activated more readily), their identity to existing expression-linked copies (greater homology increases chance of activation) and their functionality (damaged *VSG*s are harder to activate) (see Chapter 1 section 1.6.12). Previous studies have demonstrated a 'hierarchy' in *VSG* expression, whereby some *VSG*s were detected earlier in infection than others, linked to their genomic locus (Morrison *et al.*, 2005).

To investigate the expression hierarchy of the *VSG* sets expressed by this *T. brucei* line, the times at which each of the 83 sets were detected were plotted. The results are shown in Figure 3.11.

First, the 'early' timepoints, before day 20, will be considered. These data came from fewer infections and thus represent a smaller set of samples (just 91 clones) than the later timepoints. Nevertheless, these infections were broadly consistent with one another, with many of the sets detected in more than one infection (see Figure 3.10). Eighteen sets were identified prior to day 20. The majority (11/18) were classified as minichromosomal, or unknown. It is likely that these correspond to telomeric donors, as minichromosomal *VSG*s are telomere-proximal and the telomeres of all chromosomes are underrepresented in the sequencing data. Three early-appearing sets matched *VSG*s present in the subtelomeric arrays. Set_01 corresponds with Tb10.v4.0001/GUTat 10.1, the dominant VAT of the parasites initiating the infection, a *VSG* that is present in the active ES of those parasites but also has a copy in the subtelomeric *VSG* array of chromosome 10 (Tb10.v4.0001).[47] It is possible that Tb09.v4.0077 and Tb927.5.5080, which correspond with Set_08 and Set_19 respectively, also had

[47] There is also a silent telomere-proximal copy of Tb10.v4.0001 (J. D. Barry, pers. comm.)

extra telomere-proximal copies present at the start of infection, facilitating their early activation.[48] Two sets classified as 'mosaic' appear before day 20: Set_04 and Set_22. Not all donors for Set_04 have been identified, and the Set_04 clone appearing before day 20 (identified by Marcello & Barry, (2007a) at day 9) is potentially a mosaic predecessor rather than the product of segmental gene conversions itself (this is investigated further in Chapter 4 section 4.5.8). Set_22 was not detected until day 17, and then only in one infection (mouse 11), although its predominance in mice 02 and 04 at their earliest-sampled point (days 20 and 21 respectively) suggests that this is a mosaic set that has a tendency to appear earlier than other mosaics (and indeed many array genes). Bearing these exceptions in mind, the early stage of infection was associated primarily with the expression of *VSG*s whose genomic locus is either unknown or telomere-proximal (minichromosomal).

The later stages of infection were sampled in much greater depth (days 20–34, 629 clones). Longitudinal samples from the three infections for which early-stage data was available (mice 10, 11 and 12) showed that the sets present in the later stages of those infections are broadly similar to the sets detected in infections for which only later-stage data was available (see Figure 3.10) consistent with the premise that all infections are following similar patterns. Despite the greater number of sequences retrieved from each timepoint of these infections, samples taken on days 20/21 and 23 remained dominated by sets which were detected earlier in other infections. Set_08 was detected in mouse 11 and 12, and several infections by Marcello & Barry (2007a), with its earliest appearance at day 7 (mouse 11). Set_22 was detected in mouse 11 at day 17. It therefore appears to be the case that at this point in time, around day 20, the trypanosome population of each infection was predominantly expressing an early-type *VSG* set.

[48] Patterns of within-set variation for Set_08 suggest this may be is the case, see Chapter 4 section 4.4.3.

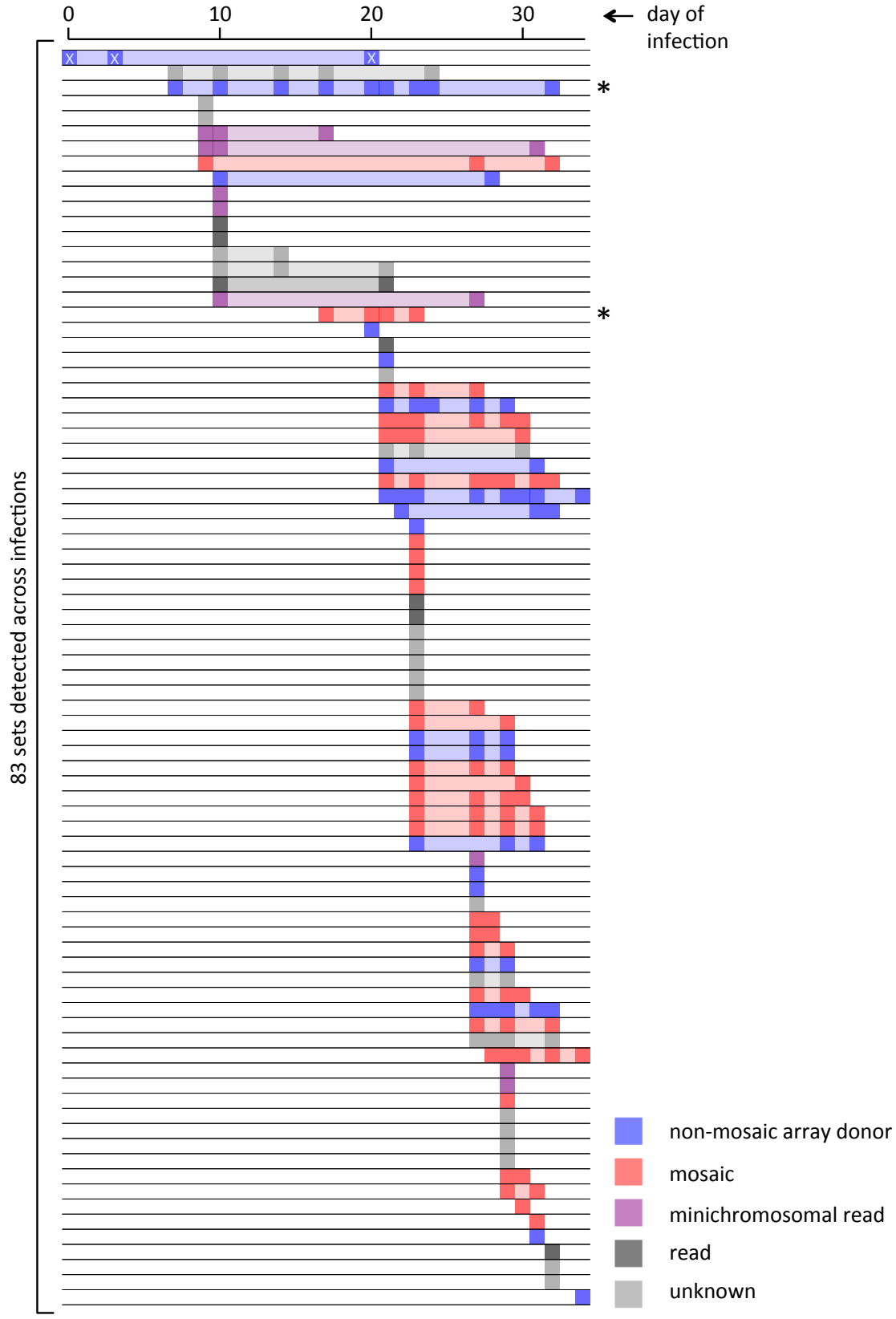


Figure 3.11 Different sets appear at different points in infection. The 83 sets detected across 22 chronic mouse infections are on the y-axis (ordered by time of appearance), and time is on the x-axis. Sets are coloured by their classification. A square indicates detection of that set at that timepoint in one or more infections. The range of detection of a set is highlighted in a lighter shade. This figure also includes the 37 sequence data from the 11 infections carried out by (Marcello & Barry, 2007a), which contribute 19 unique data points. 'X' indicates Set_01, which is known to be telomere-proximal, and ' * ' indicates Set_08 (blue) and Set_22 (red).

As discussed in sections 3.3.2 and 3.5.4, the parasite population often drops at around day 23, which coincides with the detection of a number of *VSG* sets, the majority of which are not identified at any earlier timepoints. Accurate identification of activation hierarchy is made difficult here as the detection of a set corresponds with its expansion in the population rather than directly to its activation. Where the dynamics of infection are dominated by differentiation, the activation of a set may not immediately be followed by its expansion, as variants expressing an antigenically novel *VSG* can still be suppressed by differentiation. Nevertheless, variant sets corresponding with *VSG*s that are easier to activate are best placed to replace a dominant variant when it is cleared by immunity. Polyclonal activation of easier-to-activate variants (Lee & Van der Ploeg, 1987; Timmers *et al.*, 1987) would make theirs a larger subpopulation, better able to outcompete other variants, provided they have not yet induced a specific immune response themselves (see Chapter 6 section 6.2.3 for further discussion).

Sixty-five variant sets first appear after the early stage of infection (that is, on or after day 20 in any infection). Of these, 26 sets showed patterns of N-terminal mosaicism.[49] Of the 39 that did not show N-terminal mosaicism in any instance, 16 matched *VSG*s present in the subtelomeric arrays, seven matched read assemblies, and for 16 sets no convincing genomic copy could be found. In the data here, there was no obvious difference in time of appearance between mosaic *VSG*s and the non-mosaic activation of array donors, hinting that mosaic *VSG*s—many of which are formed from pseudogenic donors—are no more difficult to generate than straightforward expression-linked copies of many subtelomeric *VSG*s. The dominance of mosaics is consistent with the findings of section 3.4.4 that, in the later stage of infection, damage to genomic *VSG*s is not an exceptional impediment to their expression.

[49] The 28 sets described in Chapter 4 section 4.5.4, minus Set_04 and Set_22. As discussed in Chapter 4 (section 4.5.8) there was no clear pattern of progression for many mosaic sets—apparent mosaic predecessors arising from single donors were identified rarely (Set_04 being an exception, above)—so the appearance of a 'mosaic' set was generally coincident with the appearance of its mosaicism.

3.6 Discussion

3.6.1 Antigen expression can be followed in vivo

The results presented in this chapter show that retrieving samples longitudinally from chronic infections is a valid method of analysing the patterns of antigen expression in *Trypanosoma brucei* infections. In this section, the main conclusions will be summarized and discussed in relation to the dynamics of *T. brucei* antigenic variation.

Two approaches were used to identify *VSG* expression in infections: cloning and sequencing of expressed *VSG*s, and PCR on total cDNA using variant-specific primers. Using these approaches, changes in the expression patterns of different *VSG* sets were identified. These patterns were consistent with the premise that over the course of *T. brucei* infection, the trypanosome population undergoes antigenic variation: different silent *VSG*s were activated and antibodies were raised against them. There was also evidence to suggest that differentiation plays an important role in the dynamics of antigenic variation in infections with this combination of trypanosome strain and host species.

3.6.2 Key findings of this chapter

3.6.2.1 Segmental gene conversion is involved in expression of many VSGs

Most expressed *VSG*s did not match genomic copies exactly, a discrepancy arising in many cases from apparent segmental gene conversion: mosaicism and 3' donation. Amongst all the *VSG* sets detected over 4–5 weeks of infection, one third showed patterns of mosaicism in the N-terminal domain. Many more sets showed clones that had undergone some 3' donation in their expression. Divergence from the genomic copies was often necessary for expression, as most of the genomic *VSG*s associated with expression were damaged. If these genomic copies were indeed the donors that became expressed (for reasons of time, direct association by Southern blotting, cloning and sequencing was not tested) it appears that damaged *VSG*s were readily expressed during infection, thanks to segmental gene conversion, and were not clearly disfavoured when compared with intact array *VSG*s.

If silent *VSG* expression is frequently accompanied by the generation of diversity, not all expressed antigenic variability need be genomically encoded at the start of infection. Mosaic *VSG* formation could introduce antigenic variability through recombination of a small set of donors.

3.6.2.2 There is considerable diversity in the variant antigen population during infection

In a single infection, mouse 04, the trypanosome population showed a total of 31 different sets, many of which showed substantial variation amongst their members. Therefore, despite the isolation of 153 *VSG* clones, it is highly unlikely that the full antigenic diversity of this infection has been captured. Similar patterns were seen in all infections studied, indicating that *T. brucei* populations can show exceptional antigenic variability over the course of infection. MacGregor *et al.* (2011) predicted that the low frequency of proliferative slender form trypanosomes present in a differentiation-controlled infection would result in a low rate of productive antigenic variation, inconsistent with the data presented here. The most likely reason for their underestimation of antigenic variability is that the linear, sequential switching pattern employed by their model precludes the presence of minor variant subpopulations in an infection (MacGregor *et al.*, 2011). In fact, the dominance of differentiation-based control may enhance the antigenic diversity present in an infection by suppressing minor variant subpopulations below the threshold for antibody induction (Gjini *et al.*, 2010). Formal comparison with an experimental model where immunity-based control is dominant is now required to test this hypothesis.

3.6.2.3 The later stage of infection shows expression of subtelomeric array *VSG*s and frequent segmental gene conversion

It was only in the later stages of infection, after approximately day 20, that the sets that appeared were clearly linked with *VSG*s from the subtelomeric arrays, and that patterns of mosaicism became frequent. Before this point, the early stages of infection were associated with *VSG*s that likely correspond with telomere-resident donors, which are known to be activated more readily (Morrison *et al.*, 2005). The transition between these two stages, around day 23, often coincided with a transient dip in parasitaemia, which could represent

the clearance of parasites expressing the most easily activated 'early' variants, allowing 'late' variants to expand in the following days.

3.6.3 Why do some VSG *sets persist over time?*

When stimulated under ideal conditions, anti-VSG antibodies develop rapidly, reaching a peak approximately six days post-stimulation (Reinitz & Mansfield, 1988). In this study, some variant sets were found to persist in the population for longer periods. How might this persistence have come about?

(i) Insufficient numbers to stimulate immune response. The dominance of differentiation-based control could have worked to suppress the expansion of a variant subpopulation, keeping it below the threshold required to induce an antibody response (Morrison *et al.*, 1982) and allowing it to persist in the population without being cleared.

(ii) The set was easily activated. *VSG*s with a high activation probability that appear early in infection will induce a specific antibody response that makes subsequent activation of that *VSG* futile. However, when sampling enough parasites there will be a transient window in time when mRNA for the *VSG* is present but there is not yet enough mature VSG in the surface coat for the parasite to be killed. If the parasitaemia is particularly high (as it was towards the end of infection, section 3.3.2), the total number of individual trypanosomes futilely activating an easily-activated *VSG* would be proportionally high, which could explain the continued presence of *VSG* sets, and would also explain the (occasionally observed) reappearance of *VSG* sets. These patterns were predicted by the model of Lythgoe *et al*. (2007) (J. D. Barry, Pers. Comm.).

(iii) Immunosuppression. As African trypanosome infections progress, mice increasingly suffer from immunosuppression that could relieve the extrinsic selection force of the immune system. A weak, absent or ineffective immune response could have allowed variant expressors to persist without clearance. The apparent clearance of variants suggests that, in general, immunosuppression is probably not absolute. Immunity may, however, become less sensitive or slower to respond as infections go on. Lymphadenopathy and splenomegaly, which may disrupt the efficient development of an immune response, were observed during these infections (data not shown). Immune responses against the

numerous different variants present must also compete for resources. In conjunction with differentiation-based control or easy variant activation, partial immunosuppression could help explain the patterns of persistence.

(iv) Within-set variation was providing sufficient variation in the antigenic profile of the *VSG* set to enable immune evasion. The NTD amino acid sequences of two VSGs within a set can vary by as much as 0.219 differences per amino acid. This variability may, in some cases, result in changes to the epitope structure of the antigen, enabling evasion of adaptive immune effectors through progressive mosaicism. Persistence through progressive mosaicism may be enhanced if an ineffective immune response raised against an earlier mosaic variant can inhibit the development of an effective immune response against its successors (Gjini *et al.*, 2010). However, many of the sets persisting here did not show exceptional within-set variation in the NTD. The ability of within-set variation to contribute to antigenic variation is considered further in Chapters 4 and 5.

These possibilities are not mutually exclusive: it is possible that all four can help to explain the patterns of persistence of a set. Further experiments are required to assess their relative contributions to infection dynamics in this host.

3.6.4 How might these findings have differed in other experimental models?

African trypanosome antigenic variation is a strategy that has evolved flexibility to adapt to different hosts. The number of different processes, host and parasite, that interact to produce the dynamics of antigenic variation can produce a phenotype that can vary considerably between different models of infection. How might the key findings summarized in section 3.6.2 have varied under different conditions?

The expression of *VSG* is a parasite-intrinsic process: expression of different variants occurs in culture (Doyle *et al.*, 1980). The appearance of a variant in an infection depends on the host insofar as different hosts can harbour different-sized parasite populations (carrying capacity), and their immune systems

respond to those populations at different rates (Gjini *et al.*, 2010). Hosts with a smaller carrying capacity (rabbits, for example) and a more responsive immune system might be expected to show less variant diversity during an infection and a clearer pattern of parasitaemic peaks than the experimental mice here (Barry, 1986). Patterns of activation hierarchy might be easier to distinguish, as immunity-based control becomes predominant.

The relationship between expressed *VSG*s and the archive is almost certainly host-independent, and thus patterns of within-set variation would likely appear in other experimental models (Barbet *et al*. (1989), Thon *et al*. (1989) and Roth *et al*. (1989) all identified mosaicism amongst trypanosomes retrieved from infections of rabbits). Segmental gene conversion is frequent in the later stage of infection as array *VSG*s become expressed. Not only does segmental gene conversion allow expression of damaged array *VSG*s, but it also introduces another layer of diversity in the expressed *VSG* population.

3.6.5 Further experiments

Now that the key *VSG* sets expressed during TREU 927/4 GUTat 10.1 infection have been identified, further experiments could accurately quantify the proportion of parasites expressing different *VSG*s over the course of infection, for example using quantitative RT-PCR. To assay the host responses, protein microarrays could be employed: preliminary collaborations with the Felgner Lab at the University of California, Irvine to produce a protein microarray of key *T. brucei* antigens is underway. If the dynamics of density-dependent differentiation were monitored alongside (MacGregor *et al.*, 2011), the predicted interactions between differentiation, immunity and antigenic variation (Gjini *et al.*, 2010) could be tested directly.

Much of the inference of the status and location of genomic copies is based on an incompletely sequenced genome, missing some of the *VSG* arrays of homologous chromosomes and *VSG*s present on the mini- and intermediate chromosomes (Berriman *et al.*, 2005). A more comprehensively sequenced genome, and in-depth analyses of genomic DNA by Southern blotting and cloning would clarify the number, location and identity of different associated genomic copies, and confirm the relationship between the expressed *VSG*s and their

genomic copies. It is also possible (but unlikely, see Chapter 4) that a more comprehensive database may reveal further, as-yet-unknown full-length intact genomic copies corresponding to the mosaic sets. Deeper and more frequent sampling could help build up a more comprehensive picture of *VSG* expression, particularly at points where *VSG* expression shows extensive richness and evenness. Although this study favoured cloning individual *VSG* sequences to get unambiguous high-resolution information about individual sequences and to simplify assembly (particularly of mosaic *VSG* clones), it is possible that a high-throughput next-generation RNA sequencing approach could be developed to capture the full picture of antigenic diversity within a sample.

3.6.6 Concluding remarks

Considering just between-set antigenic variation, it seems clear that these infections are far from exhausting their intact *VSG* repertoire—35 intact, full length genomic copies were not identified by sequencing from any of the infections analysed. There were also some 81 atypical *VSG*s, whose unusual GPI anchor signal sequences could be easily overcome by a short 3' donation, that were also not identified.[50] Yet these infections the expression of many silent *VSG*s was apparently accompanied by the introduction of variation consistent with mosaicism.

Why might a parasite population with an apparently strong hand start playing an increasingly risky game, combining and recombining its antigen genes? The primary explanation, that segmental gene conversion is an occasional event that occurs solely to allow expression of damaged *VSG*s, and hence is selected as the host develops antibodies against the products of intact *VSG*s, is not supported by either of two independent lines of evidence. Firstly, across all infections analysed to date, it does not appear that all, or even most, of the most-accessible *VSG*s are expressed (and therefore neutralised) before segmentally-converted *VSG*s appear. Secondly, within-set variation is frequently seen *within* a timepoint. Were it the case that segmental gene conversion was a very unusual event endowing its bearer access to a valuable antigen, we would

[50] Both of these figures were obtained by collecting the names of all 'functional' or 'atypical' *VSG*s from VSGdb, and subtracting from the list those *VSG*s whose expression was detected (in any form) in the infections presented here.

expect all segmentally-converted *VSG*s within a timepoint to be identical: clones descended from the single parasite which had made the daring switch. But this is not the case. These patterns of within-set variation are considered in the following chapter.

Chapter 4

Variation amongst expressed *VSG*s

4 Variation amongst expressed *VSG*s

4.1 Introduction

4.1.1 Variation is present between similar expressed VSGs

As described in Chapter 3, the *VSG*s expressed in the experimental infections could be grouped into 'sets'. The members of each set were more similar to one another than to any members of any other set, and each set is thought to represent the expression of one or more specific genomic *VSG*s. Variation between sets—switching expression between discrete genomic *VSG*s—corresponds with the classical process of African trypanosome antigenic variation. However, it was also found that, alongside 'between-set' variation, expressed *VSG*s belonging to a set frequently varied from their putative genomic copy and from one another. Broadly, three patterns of 'within-set' variation could be observed in the sequence data: point mutation, 3' donation, and mosaicism. Each of these patterns has been identified in expressed *VSG*s by previous studies (see Chapter 1 sections 1.6.6–1.6.11 for a detailed discussion). The latter two represent processes of segmental gene conversion, as they are gene conversion events where the boundaries of recombination occur within the *VSG* coding region to produce a composite *VSG*.

What is the role of within-set variation in trypanosome antigenic variation? As described in the previous chapter, differences between expressed *VSG*s and their related genomic copies produced by segmental gene conversion allows the expression of pseudogenic genomic *VSG*s. Yet variation amongst related *VSG*s can also be identified within an infection. By introducing a further layer of variability amongst expressed *VSG*, within-set variation could potentially play a role in immune evasion, by producing antigenically distinct but sequence-related variants.

4.1.2 Key hypotheses to be tested

To assess the contributions of mosaicism, 3' donation and point mutation to *VSG* expression and antigenic variation, the patterns of within-set variation in the

720 *VSG* sequences described in Chapter 3 were analysed bioinformatically and experimentally with regards to the following hypotheses:

(i) Within-set variation between expressed *VSG*s—and between an expressed *VSG* and its putative genomic donor(s)—is routinely generated over the course of infection. That is, within-set variation does not simply represent the straightforward activation of related *VSG*s already present in the genome of the inoculum, or experimental artefact.

(ii) Variation between the expressed *VSG* and its putative genomic copy allows the expression of damaged silent *VSG*s.

(iii) Patterns of within-set variation are consistent with a role in changing epitope structure. By modifying the exposed NTD sufficiently, within-set variation could directly contribute to immune evasion.

(iv) Within-set variation accumulates over the course of infection. If within-set variation is playing a role in antigenic variation in the infections here, trypanosomes that have accumulated events in their expressed *VSG* by segmental gene conversion and/or point mutation are more likely to survive and become detected at later timepoints.

4.1.3 Investigating within-set variation

To investigate within-set variation, sequence-related expressed *VSG*s were aligned with one another and with their putative donors, and the alignments were studied both by eye and automatically using custom BioRuby scripts (Chapter 2, section 2.7.2.4). As described in Chapter 3 (section 3.4.3), patterns of within-set variation were identified in two complementary ways. First, by comparing expressed *VSG* clones with their putative donor(s), regions where the expressed sequence did not match the donor sequence were identified (variation 'from donor'). Second, expressed *VSG* clones of a set were compared with one another. Regions where the expressed clones varied from one another could be identified (variation 'amongst clones'). Variation amongst clones reveals changes to the expressed form of *VSG*, and comparing variation amongst clones retrieved from a single infection indicates the possible contribution of such variation to antigenic variation. In addition, variation amongst clones was useful for inferring the involvement of unannotated genomic copies, as identical patterns appearing in different infections indicates the involvement of an

unidentified template and/or divergence between the clone used for genome sequencing and the clone used for these infections.

The different patterns of variation—point mutation, 3' donation and mosaicism—were identified as follows:

(i) Where mismatches (either from the donor or amongst clones) were isolated, and no alternative template containing a region with these differences could be found, they were investigated as point mutations.

(ii) Where mismatches were clustered towards the 3' end of the gene, they were investigated as cases of 3' donation. In some cases, donors for the 3' end could be found.

(iii) Where clusters of mismatch appeared within the gene they were investigated further as potential mosaics. Additional donors containing regions corresponding exactly with the cluster of mismatching bases could often be found. Occasionally, additional donors corresponding with the cluster of mismatching bases could not be found. In this case, the set was not classified as a mosaic (instead it was classified as 'unknown' and was not investigated further), except where mosaicism could be inferred by comparison amongst clones (described in section 4.5.1).

4.1.4 Each pattern of within-set variation will be considered in turn

This chapter will consider each of the different patterns of within-set variation in turn, ending each section with a short summary. First, patterns of point mutation will be described: whether they are generated over the course of infection, and what their contribution to antigenic variation might be (section 4.2). Then, segmental gene conversion will be considered. The question of whether the observed patterns of segmental gene conversion are likely to be genuine will be considered separately (section 4.3) before moving on to discuss 3' donation (section 4.4) and mosaicism (section 4.5). Finally (section 4.6), the results of the previous sections will be brought together to address the key question of this chapter: whether and how within-set variation might contribute to antigenic variation.

4.2 Point mutations

4.2.1 Point mutations were present in VSG *sequences*

Many viruses depend on the accumulation of point mutations in their antigens for immune evasion within a host, or across host populations (Boni, 2008; Sodroski *et al.*, 1998). Under selective pressure, variation can arise in an individual base that changes VSG epitope structure sufficiently to allow escape from a monoclonal antibody (Baltz *et al.*, 1991). Were this process rapid and extensive enough, evasion of a polyclonal antibody response might be possible. Is there any evidence in the collected sequence data that immune evasion by the accumulation of point mutations occurs over the course of the trypanosome infections?

Many expressed sets showed the presence of 'point mutations'[51]: isolated single differences found both amongst expressed clone sequences, and between an expressed clone sequence and its putative genomic copy. Point mutations were the only source of within-set variation identifiable in the *VSG* N-terminal domain (NTD)-encoding region for 30 of the 83 sets.

4.2.2 Some point mutations were probably generated over the course of infection

To investigate whether observed point mutations are experimental artefacts, the estimated rate at which single-base mismatches might be artefactually introduced by amplification and cloning was calculated. Published fidelity measurements for the two polymerases employed in sequence amplification (SuperScript III (Invitrogen, Paisley, UK); Herculase II Fusion, (Agilent, Stockport, UK)) were combined to calculate a predicted error rate per sequence.

$$P(error) = l.E(rt) + l.n.E(h)$$

[51] 'Point mutation' is generally taken to mean variation at a specific position introduced by an enzyme as it copies a template strand. However, the possibility that the apparent point mutations observed here are in fact introduced by short gene conversion events with unidentified non-homologous donors cannot be ruled out.

where $E(rt)$ is the error rate of SuperScript III per base (1 / 37037), $E(h)$ is the error rate of Herculase II per base (1 / 770000), l is the average length of a *VSG* nucleic acid sequence (1500 bp) and n is the average number of times a sequence has been copied in an ideal, 30 cycle PCR reaction (29).

To predict the proportion of sequences containing one or more mismatches, the Poisson distribution was used.

$$P(n) = \frac{m^n . e^{-m}}{n!}$$

where m is probability of event ($P(error)$) and n is number of events (0, that is, sequences with no errors). The probability of one or more errors was calculated:

$$\begin{aligned} P(>1) &= 1 - P(0) \\ &= 1 - e^{-m} \\ &= 0.09243 \end{aligned}$$

or, 9.2% of sequences would be predicted to have one or more errors, were these two polymerases the only source of variation.

Assuming that the published fidelity rates for these two polymerases are dependent on numerous, unaccounted for reaction variables (such as salt concentration, template composition, and so on), the same calculation was carried out using error rates ten times greater than published. Under these conditions, 62.1% of sequences would be predicted to have one or more artefactual errors and 1.1% of sequences two or more artefactual errors. This value is still markedly less than the observed frequency of point mutations (see section 4.2.4) suggesting that experimental artefact alone cannot account for the variation observed.

Another line of evidence that point mutations genuinely existed in expressed *VSG*s came from the identification of the same point mutations in independent amplification reactions made from samples taken from the same infection. Examples can be found in mouse 05 Set_13 and Set_21, shown in Appendix 7.3.1.

The different patterns of point mutations seen in samples taken from different infections (including the primary clonal infections) hints that point mutation variation was generated independently in different infections.

Together, these analyses suggest that point mutations did occur over the course of infection, although further experiments are required to formally verify their appearance, such as cloning from genomic DNA at the beginning and end of infection.

4.2.3 Some point mutations resulted in non-functional VSG *sequences*

It was noted during assembly of the *VSG* sequences that 61 expressed *VSG* clones contained apparent mutations that would prevent their functional translation or vary the structure of the GPI anchor signal sequence: insertions and deletions causing frameshifts, and substitutions generating internal stop codons. Most of these errors were corrected during sequence assembly by reference to similar, undamaged clones from the same sample, as described in Appendix 7.2.3.[52] Close examination of these sequences show that in 35 cases these errors have been introduced by the expansion or contraction of a poly-A/T tract: regions that are particularly prone to artefactual damage during amplification as the polymerase can 'slip' as it passes through them. The CTD-encoding portion of many of the *VSG*s under consideration here contain poly-A/T tracts. Previously, Rice-Ficht *et al*. (1982) found the artefactual contraction of a poly-A tract in one of their sequencing reactions from genomic DNA. That none of the non-functional *VSG* sequences was found in independent transcription-amplification reactions is consistent with the occurrence of some artefactual point mutations in the sequence data. It is therefore likely that there are both genuine and artefactual point mutations in the sequence data. As a consequence the patterns of point mutation should be interpreted with caution.

[52] In addition, three clones showed the loss of a fragment of their sequence, a phenomenon that could be ascribed to residual recombination occurring in the *E. coli* cells used to amplify the sequence-containing plasmid.

4.2.4 Does point mutation contribute to antigenic variation?

The variation introduced into the expressed NTD by point mutation is shown in Table 4.1. The maximum amino acid distance between related mature NTDs was calculated for all sets where a clear genomic copy could be identified and point mutation was the only pattern of NTD variability. Point mutation introduced variability amongst related VSG NTDs of up to 0.057 differences/aa. Within an infection, variability was lower, with most sets showing fewer than 0.010 differences/aa.

Is there any evidence that point mutation contributes to antigenic variation? Nine sets were chosen for closer analysis. These sets were chosen as they matched their putative genomic copy well in the NTD-encoding region (there was no evidence of identical patterns of mutation in different infections, and no evidence of N-terminal mosaicism), they had multiple clones which could be compared with one another, and they were representative of the totality of sets in that some appeared early, some late, and some across infection. The chosen sets also showed generally higher levels of variation than other sets so were considered a strong test of the null hypothesis that patterns of point mutation do not contribute substantially to antigenic variability. For these analyses, just the NTD-encoding region was considered, to isolate point mutation variation from variation caused by 3' donation (which was frequently observed, section 4.4). First, the distribution of point mutations appearing independently in different infections was tested, to see whether point mutations generally cluster in particular regions of the NTD. For each set, the position of all differences amongst expressed *VSG* sequences was recorded. The maximum number of differences occurring within a region of defined size ('the sliding window') was counted, and their position was recorded. This is the observed maximum. To assess whether these differences were non-randomly distributed, 10,000 random distributions of the same number of differences were created for each set, and the maximum number of differences occurring within the sliding window was counted for each of these. The rank of the observed maximum amongst the simulated maxima corresponds with a p-value. The results of these analyses are shown in Table 4.2.

	Set	Classification	# of infxn	# of clones	Max set	Max infection	Max from don
*	Set_21	non-mosaic	4	36	0.057	0.023	0.032
*	Set_09	non-mosaic	3	3	0.051	0	0.039
*	Set_19	non-mosaic	2	8	0.044	0	0.027
*	Set_13	non-mosaic	3	8	0.043	0.04	0.031
	Set_55	non-mosaic	1	3	0.032	0.032	0.028
	Set_68	unknown	3	7	0.03	0.027	-
	Set_18	unknown	4	18	0.029	0.029	-
*	Set_27	non-mosaic	3	6	0.029	0.023	0.015
*	Set_08	non-mosaic	11	127	0.027	0.024	0.015
*	Set_56	non-mosaic	2	3	0.027	0	0.03
*	Set_82	mc	2	3	0.026	0	0.026
*	Set_39	non-mosaic	3	6	0.024	0.021	0.015
	Set_07	unknown	3	12	0.018	0.015	-
	Set_03	mc	5	20	0.016	0.016	0.02
	Set_75	unknown	3	6	0.013	0.01	-
	Set_25	read	1	2	0.009	0.009	-
	Set_57	non-mosaic	1	4	0.006	0.006	0.015
	Set_05	mc	5	6	0.006	0.003	(-)
	Set_81	mc	1	3	0.003	0.003	0.003
	Set_01	non-mosaic	4	14	0.003	0	0.009
	Set_73	unknown	2	7	0	0	-
	Set_78	unknown	1	2	0	0	-
	Set_76	read	2	2	0	0	0
	Set_44	non-mosaic	1	1			0
	Set_34	non-mosaic	1	1	-	-	0.012
	Set_36	non-mosaic	1	1	-	-	0.003
	Set_72	read	1	1	-	-	0.006
	Set_74	read	1	1	-	-	0.014
	Set_77	read	1	1	-	-	0.003
	Set_80	mc	1	1	-	-	0.009
	Median				0.024	0.009	0.015

Table 4.1 Expressed N-terminal domain variability introduced by point mutation. Mature NTD amino acid sequences from sets where point mutation was the only identifiable source of NTD variability were compared. '# of infxn' = number of infections in which that set was detected; '# of clones' = total number of clones obtained. The maximum difference between related sequences ('Max set'), related sequences expressed within an infection ('Max infection'), and sequences and their putative donor ('Max from don') were calculated using ClustalW. The measure of the differences is differences/aa in the mature NTD. The top ten sets are highlighted in red. Sets referred to in Table 4.2 are indicated by a '∗'.

Set	NTD type	Window size (nt)	Mutations		Rank	p
			Max	Position		
Set_09	A	30	7	516..546	26	**0.0026**
Set_21	A	30	9	651..681	55	**0.0055**
Set_39	A	30	4	201..231*	1420	0.142
Set_19	A	30	5	224..254	1801	0.1801
Set_08	B	30	15	270..300	4	**0.0004**
Set_13	B	30	8	682..712	12	**0.0012**
Set_56	B	30	5	963..993	45	**0.0045**
Set_82	B	30	4	672..702	308	**0.0308**
Set_27	B	30	4	464..494*	3446	0.3446

Table 4.2 Distribution of N-terminal domain point mutations. The positions of non-synonymous point mutations were analysed by a sliding window of indicated size. The region where most point mutations were located was recorded ('Max' = the number of mutations in this region, 'Position' = the position in the nucleic acid alignment). 'Max' was compared with the maximums obtained from 10,000 simulations in which point mutations were randomly distributed, and 'Rank' describes its rank amongst these simulations. Significance at the 0.05 level is indicated in bold. Similar results were obtained when considering total point mutations, except in those sets indicated by a '✱' (p-values were not improved), and/or by increasing window size to 100.

Set	Donor	Prop. donor sites ns	# clones analysed	Clone point mutations		Total dn/ds
				Max (ns)	Prop. ns	
Set_82	927mc_VI-6c05	0.76	3	21 (12)	0.57	0.75
Set_39	Tb09.v4.0177	0.75	6	12 (5)	0.65	0.86
Set_56	Tb11.v4.0035	0.77	3	16 (11)	0.69	0.90
Set_27	Tb09.142.0470	0.75	6	8 (6)	0.71	0.94
Set_08	Tb09.v4.0077	0.77	127	7 (5)	0.74	0.97
Set_19	Tb927.5.5080	0.76	8	12 (10)	0.74	0.98
Set_21	Tb10.v4.0088	0.76	36	13 (11)	0.80	**1.05**
Set_09	Tb09.244.1580	0.75	3	19 (14)	0.80	**1.06**
Set_13	Tb09.v4.0102	0.76	8	13 (10)	0.84	**1.10**

Table 4.3 (Non-)synonymity of N-terminal domain point mutations. For each set under consideration, the proportion of non-synonymous sites in its corresponding donor's NTD-encoding region was calculated ('Prop. donor sites ns'). Each clone of that set was considered in turn, and the number of synonymous and non-synonymous point mutations in the NTD-encoding region recorded. For each set, the maximum number of total point mutations seen in a clone, and the number of these that were non-synonymous in brackets, is given in the column 'Max (ns)'. The proportion of non-synonymous point mutations (Prop. ns) was calculated for each set by taking all the point mutations seen in that set together, and dividing the number of non-synonymous point mutations by the total. This value was divided by the proportion of non-synonymous donor sites to give a total dn/ds ratio. dn/ds ratios > 1 are shown in bold.

Non-random distribution of point mutation variation amongst clones was found, but was set-specific. There was no region of the domain in which point mutations uniformly accumulated, either when considering all domain types together, or separating into A and B NTD types.

If point mutations causing a change in VSG epitope structure were favoured, the dn/ds ratio (the ratio of non-synonymous mutations per non-synonymous site to synonymous mutations per synonymous site) amongst members of a set might be expected to be significantly greater than 1: point mutations that change protein structure being more successful, and hence more frequently detected, than those which have no effect. The number of non-synonymous and synonymous sites for each set's putative donor was counted and compared with the total number of non-synonymous and synonymous mutations observed for each clone of that set, and the dn/ds ratio calculated. The results are shown in Table 4.3. As can be seen, there is not a general trend towards non-synonymous over synonymous mutations. It is possible that by considering the entire sequence of the NTD-encoding region as a whole, the selection for non-synonymous mutations specific codons is being masked by no, or negative selection, at other codons. To test this, a fixed effects likelihood model (Kosakovsky Pond & Frost, 2005) was used, which tests for the presence of positive or negative selection at each individual codon. In none of the sets examined[53] was there significant positive selection detectable at any codons (data not shown).

Together, there is no evidence that point mutations played an important role in varying the antigenic profile of the parasite population during these infections. Their possible contribution to the biology of African trypanosome antigenic variation is discussed in more detail in Chapter 6 (section 6.3.2).

[53] There were insufficient unique sequences to analyse Set_82.

4.3 Segmental gene conversion

4.3.1 Segmental gene conversion probably occurred during infection

Before discussing the patterns of segmental gene conversion observed here, the question of whether they are generated over the course of infection must be addressed. Both mosaicism and 3' donation have been analysed in depth by PCR, Southern blotting and genomic cloning numerous times (see Chapter 1 for a review) demonstrating that these segmental gene conversion events readily occur during *VSG* expression. For reasons of time, similar analyses could not be performed on any but a few of the sequences reported here. Nevertheless, bioinformatic analyses of the data could go some way to addressing the question of whether the segmental gene conversion variation observed generally occurs during infection.

First, the possibility that apparent segmental gene conversions are an experimental artefact will be considered. Second, the possibility that apparent segmental gene conversions represent variation pre-existing in the genome of the inoculum will be considered.

4.3.1.1 Was segmental gene conversion an experimental artefact?

Studies on reverse transcriptase and thermostable DNA polymerases have shown that 'copy choice' can cause recombination between sequences during polymerisation, as the DNA polymerase switches between templates (Odelberg *et al.*, 1995; Zaphiropoulos, 2002). Template switching as the main cause of apparent segmental gene conversion was considered, but discounted here for three reasons.

(i) The Herculase II Fusion polymerase used for PCR amplification is based on *Pfu* DNA polymerase, which does not show any evidence of template switching (Shafikhani, 2002).

(ii) Were template switching the main cause of apparent segmental gene conversion, it would be unusual to find identical patterns of segmental gene conversion from different transcription-amplification reactions. Here, *VSG*s having undergone 3' donation or mosaicism could often be

found from independent transcription-amplification reactions. Examples can be found in Appendix 7.3.4.

(iii) Were it the case that template switching was occurring at an appreciable level, we would expect to see, from the same reverse transcription-amplification reaction, identical regions in clones from different sets, representing the templates between which the polymerase had switched. The observed patterns of segmental gene conversion were inconsistent with this model.[54]

These observations make template switching an unconvincing explanation for the patterns and prevalence of segmental gene conversion observed here, however, further experiments would be necessary for each sequence to rule out template switching as a contributing factor altogether.

4.3.1.2 Were identified mosaic and 3' donation *VSG*s already present in the inoculum genome?

The parasites used to initiate the infections came from the same clone used for the genome sequencing (Berriman *et al.*, 2005). Nevertheless, it is likely that there were further unsequenced or unannotated *VSG* genes in the inoculum that were not present in the genomic *VSG* database (Marcello & Barry, 2007a). The patterns of segmental gene conversion were studied to assess whether the apparent segmental gene conversion events observed here represented the straightforward expression of as-yet-unknown full-length *VSG*s present in the genome of the inoculum. Identical patterns of segmental gene conversion in different infections indicate either that identical gene conversion events occurred independently in different populations, or, more likely, that these variants were present in the genome of the parasites used to initiate the infections. Close analysis of sequences showed some identical patterns of segmental gene conversion occurring in different infections, hinting at the contribution of unannotated genomic variation. However, the number of *VSG*s apparently formed by segmental gene conversion made unannotated genomic variation an implausible explanation for most of the segmental gene conversion

[54] In some cases (see section 4.4.4), clones from different sets had similar 3' regions, but in almost all cases the similar 3' region-bearing *VSG*s of different sets were retrieved from different reverse transcription-amplification reactions.

events described here, as it would require the parasites to possess huge families of closely related *VSG*s, and often activate them all at a similar time. Genome analyses of *VSG* subfamilies suggests that this is not the case (rather, 40% of *VSG*s are in small subfamilies, most of which contain two members (Marcello & Barry, 2007a)).

The parasites used for most of the infections had a common source, cloned six passages previously (~72 generations, assuming a doubling time of six hours and approximating time in each mouse at three days, see Appendix 7.2.1). It is possible that over this period the population became heterogeneous, with some individual parasites undergoing segmental gene conversion events (possibly in a silent locus) that become expressed in the experimental infections. In this case, although the observed variation might have arisen rapidly from an individual parasite, technically it would not have developed over the course of the specific experimental infection under consideration. To assess whether a potentially polyclonal inoculum population could have influenced the patterns of segmental gene conversion observed, primary clonal experimental infections (initiated by a single parasite) were analysed. The variety of different mosaic *VSG*s derived from a single infecting parasite (described in section 4.5.3) suggests that segmental gene conversion occurs readily over the course of a 4–5 week infection, lending support to the premise that the segmental gene conversion events observed here generally occurred over the course of an infection.

4.3.2 Segmental gene conversion was analysed as 3' donation and mosaicism

Patterns of presumed segmental gene conversion were separated here into 3' donation and mosaicism. Broadly, 3' donation introduces specific variation in the CTD-encoding region of the expressed *VSG*, whereas mosaicism introduces variation within the NTD-encoding part of the gene. These patterns of variation are likely a consequence of homologies between donor sequences. The donors combining in 3' donation (where they have been found) shared little overall nucleotide identity (with short regions of homology located in the CTD-encoding region, or the region encoding the hinge between N- and CTDs), whereas donors combining to produce mosaics showed homology across their entire length. Roth et al. (1986) made the distinction between 'hybrid' *VSG*s, constructed from *VSG*

genes that do not have extensive sequence homology, and 'mosaic' *VSG*s produced from members of a highly related gene family. A distinction based solely on donor homologies was not possible here, due to the difficulty in identifying full-length donors for many 3' donations (see section 4.4.4). Therefore, '3' donation' was made a more inclusive term than 'hybridism', covering all putative recombination events that modify the 3' end of the expressed *VSG*. Where identified, specific 3' donors indeed had little homology with the N-terminal donor of the expressed *VSG*. On the other hand, 'mosaicism' included all putative segmental conversion events occurring in the NTD-encoding region, which incidentally always involved related donors (section 4.5.5). The involvement of mosaic donors in the CTD-encoding region of a *VSG* was also considered mosaicism.

Mosaicism and 3' donation may be united in a common homology-driven recombination mechanism but their distinction was found to be a useful way of categorizing the patterns of within-set variation observed here, allowing variation that affects the epitope-containing NTD to be considered separately from that affecting just the buried CTD.

4.4 3' donation

4.4.1 3' donation was frequently observed

3' (CTD) donation in *VSG*s comes about when the region of the *VSG* gene encoding the membrane-distal NTD of the protein is joined to a different sequence in the CTD-encoding region (see Chapter 1 section 1.6.8). 3' donation was identified by comparing expressed clones with their putative genomic copy donors and finding abundant mismatches towards the 3' end of the gene, and by comparing the members of a set with one another and finding expressed *VSG* clones with high levels of identity in their NTD-encoding region with numerous differences from one another in their CTD-encoding region.

The most likely means by which a *VSG* comes to acquire a different 3' end is during the formation of its expression-linked copy, when the 3' boundary of recombination occurs within the coding region of an expression-site-occupying *VSG*. The 3' end of the previous *VSG* is therefore retained and joined to the

NTD-encoding region of the incoming *VSG*. However, it is also likely that events can accumulate in the CTD-encoding region of a *VSG in situ* (Marcello & Barry, 2007a).

Inclusive criteria for 3' donation (specific variation from a donor *or* variation amongst clones) found 3' donation occurring in 47/83 sets. A more conservative definition (specific variation from a donor *and* variation amongst clones) found 3' donation in 23/83 sets (see Appendix 7.3.2 for details). Thus the expressed form of a *VSG* often acquired a 3' end that was different from its silent form.

4.4.2 3' donation allowed the expression of damaged VSGs

In replacing the 3' end of a silent *VSG*, 3' donation can repair damage to that *VSG*'s CTD or GPI anchor signal sequence. Of the silent *VSG*s corresponding with the 47 sets for which 3' donation was observed, 16 relied on 3' donation to repair pseudogenic CTD-encoding regions,[55] and 10 silent *VSG*s whose CTD function was uncertain ('atypical' *VSG*s) underwent 3' donation. Several of these pseudogenic or atypical *VSG*s were the primary contributing donors to mosaic sets. By undergoing 3' donation, these mosaic donors acquired 3' ends that had little or no overall homology with any of the other mosaic donors. The different degrees of 3' donation, and the nature of the silent copies undergoing 3' donation, are summarized in Figure 4.1.

In one case an expressed *VSG* (04-21c25, the sole member of Set_58) acquired an apparently dysfunctional 3' end, lacking a GPI anchor signal sequence due to a premature in-frame stop codon. This clone may represent the stochasticity of segmental gene conversion and *VSG* activation—occasionally damaged *VSG*s may occupy active expression sites, although (presumably) only transiently (see Chapter 3 section 3.2.2.1).

[55] Four of these were read assemblies, two of which were a minichromosomal read assemblies.

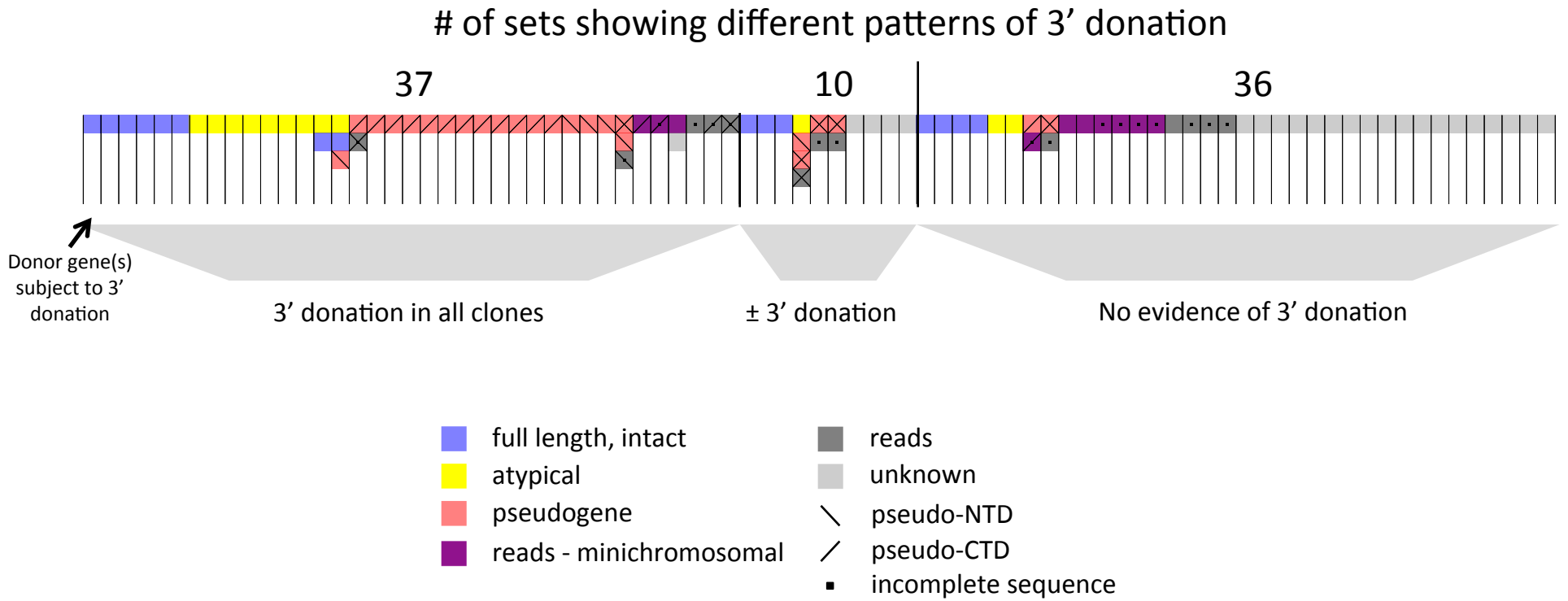


Figure 4.1 Summary of 3’ donation events. The diagram shows, for the 83 sets detected across all infections, the properties of the primary genomic copy. Compared with the diagram in Chapter 3 (Figure 3.7), donor genes that apparently contributed only to the NTD-encoding regions of mosaic sets are not shown, and this diagram has been reordered according to presence of 3’ donation (‘± 3’ donation’ indicates 3’ donation identified in *some* clones). The pseudogenicity of each of the donors is indicated by a diagonal line across the gene marker. A small black square for read assemblies indicates that this assembly didn’t cover the whole expressed VSG, and therefore it is possible that it is damaged further. Note that one set shows a damaged donor CTD but no 3’ donation; in this case damage towards the 5’ end of the CTD had been repaired by mosaicism and thus was not counted. This diagram was drawn using the more inclusive definition of 3’ donation (variation from donor *or* variation amongst clones).

4.4.3 The extent of 3' donation varied

The extent of 3' donation varied. Some expressed *VSG*s showed substantial donation, in which their whole CTD-encoding region was replaced; others showed variation solely towards the 3' end of the CTD-encoding region, and others had changes to just their most 3' end, in the GPI anchor-encoding region. Examples of different degrees of 3' donation are shown in Figure 4.2.

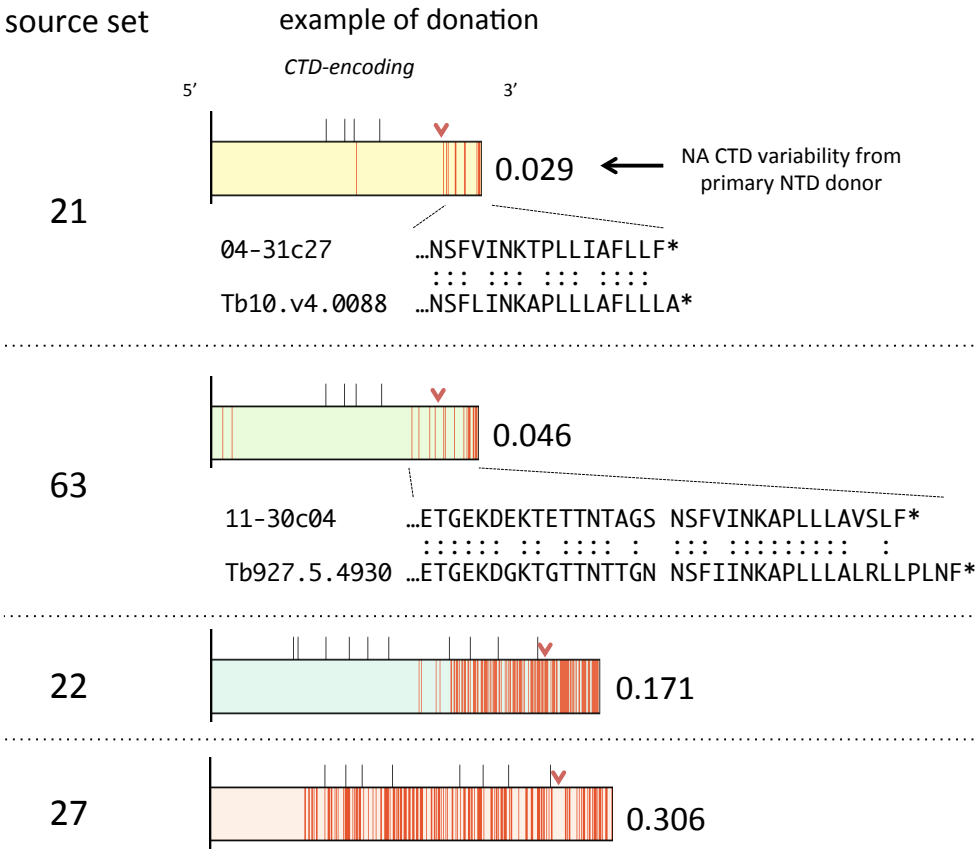


Figure 4.2 Degrees of 3' donation. Diagrams showing the CTD-encoding regions of four expressed VSG clones demonstrate different degrees of 3' donation events. Each diagram shows a summary of the alignment (5'–3') between the CTD-encoding regions of an expressed VSG clone and its primary donor. Cysteine residues are shown as bars extending above the diagram, and a red arrowhead indicates the beginning of the cleaved GPI anchor signal sequence. Mismatches between the clone and the donor are shown as red bands extending across the diagram. The CTD-encoding nucleic acid variability (differences/nt between expressed VSG and primary donor) is given for each example. For minor donation events, a portion of the translated alignment is shown.

Although the period before day 20 is covered by fewer data (91 clones, compared with 629 clones on and after day 20) making direct comparison with the later period qualified,[56] it is interesting to note that only 3/18 sets from this period of infection, Set_08, Set_74 and Set_22, showed patterns of 3' donation. Of these, Set_08 showed identical patterns of variation in all eleven infections in which it was detected, including both of the clonal infections, hinting that this particular 3' variant was pre-existing in the inoculum.[57] The later stage of infection showed 3' donation in 46/75 sets. This observation is consistent with a model in which 3' donation becomes more prominent in the activation and expression of *VSG*s as an infection progresses.[58]

4.4.4 The donors of 3' ends often corresponded with likely *telomere-proximal* VSGs

Preliminary attempts to identify the 3' donor were only successful in cases where a substantial part of the CTD-encoding region had been replaced, as *VSG*s show generally high levels of homology towards their most 3' end, particularly in the GPI anchor-encoding region (Majumder *et al.*, 1981; Matthyssens *et al.*, 1981; Pays *et al.*, 1981a; Rice-Ficht *et al.*, 1981). Donors were therefore only investigated for cases where there was at least 75 bp of 3' donation. The involvement of a mosaic *VSG*'s donors in the CTD-encoding region (C-terminal mosaicism) of that *VSG* was also excluded from this analysis, as this pattern is considered alongside N-terminal mosaicism in section 4.5.[59] BLAST searches of the genomic databases enabled identification of putative 3' donor sequences for 13 of the different 3' donation variants. Matches in the 3' region were also found by comparing *VSG*s from different sets, allowing the identification of a further six 3' variants. Identified 3' variants are shown in Table 4.4. As is clear

[56] As 3' donation was identified in part by variation between clones, the greater the number of sequences retrieved, the greater the probability of identifying 3' donation, thus this analysis is somewhat biased against finding 3' donation in the earlier timepoints.

[57] Note added in proof: a Southern hybridization carried out following submission of this thesis gave results inconsistent with this hypothesis (data not shown).

[58] For this analysis one pre-day 20 set, Set_04, was discounted from the pre-day 20 count as 3' variation in this set was detected only at day 27. 3' donation in sets that had appeared pre-day 20 were counted again if they showed 3' donation on or after day 20 (that is, Set_08 and Set_22).

[59] However, instances where mosaic VSGs have acquired 3' donors in a similar manner to non-mosaic *VSG*s, (that is, involving donor sequences that have overall low homology with the primary donor), were investigated.

in the table, the same 3' end could be a donor in a number of different sets, as also shown in Figure 4.3.

Donor		Source set	Notes	TTAGGG-like?	Recipient sets	Pattern
a	trypA22d6.q1t-rev	Set_70	d23		08, 13, 17, 32, 56, 59, 63	
b	927mc_VII-5d10.q1k-rev	-	mc	*	22, 31, 56	
c	unknown	-			(23), 64	
d	Tb927.5.5080 / unknown	Set_19	d10		64	
e	tryp_XI-452h04.p1k	-			64	
f	unknown	Set_04	d9		55, 14	
g	unknown	-			14, 60	
h	927mc_VII-7g12.p1k	-	mc	*	27, 34	
i	927mc_VIII-11d07.p1k-rev	-	mc		27	
j	unknown	-			27, 79	
k	927mc_VI-7e09.q1k-rev	-	mc	*	22	
l	927MC_VII.0.0152-rev	-	mc	*	22	
m	927mc_III-7c06.pk 927MC_III_0.438	-	mc		50	
n	927mc_VI-4f07.p1k 927MC_VI.0.382-rev	-	mc	*	31	
o	927mc_V-6d05.p1k-rev	Set_81	mc / d10		37	
p	unknown	-			37, 57	
q	tryp_XI-430a10.q1k	-			13	
r	read AZ217061	-			14	
s	tryp_X-135d07.q1c-rev	-			58	

Table 4.4 Identified 3' variants. Each identified 3' donation variant is given an identifier (a–s, first column). 3' ends matching *VSG*s from another set are indicated under 'Source set'. The time of first detection of that set is indicated under 'Notes', as are those reads found in the minichromosomal read database. For Set_23, only one clone was found, and no donor could be found for it, but it matched Set_64 in the 3' end. In the case of variant d, matches were not for the entire 3' donation. Those reads showing 'TTAGGG'-like sequence 3' of the VSG are indicated with a '*'. Example patterns of 3' donation variants used in figures elsewhere in this thesis are provided as a reference.

The majority of identifiable 3' donors correspondeded with minichromosomal reads and *VSG* sets expressed early on in infection, both of which are likely to represent telomere-proximal *VSG*s. Indeed, TTAGGG-like repeats were identified downstream of the coding region in many of the reads shown in Table 4.4: these may represent the telomere repeats themselves, or more likely a

region of homology often found upstream (Aline & Stuart, 1989). This finding is consistent with a model in which the 3' donation was acquired during the *VSG*'s duplication into a telomere-proximal locus (Michels *et al.*, 1983; Pays *et al.*, 1985a) although it is also possible that sequence exchange could have occurred by telomere interaction following gene conversion.

The failure to identify 3' donors for all sequences was probably due to the fact that telomere-proximal *VSG*s are underrepresented in the genome sequence data. Furthermore, efforts to identify donors may have been complicated by successive 3' donation events at a telomere each leaving small portions of residual sequence, producing a composite 3' end of the expressed *VSG* whose complex construction could not be dissected.

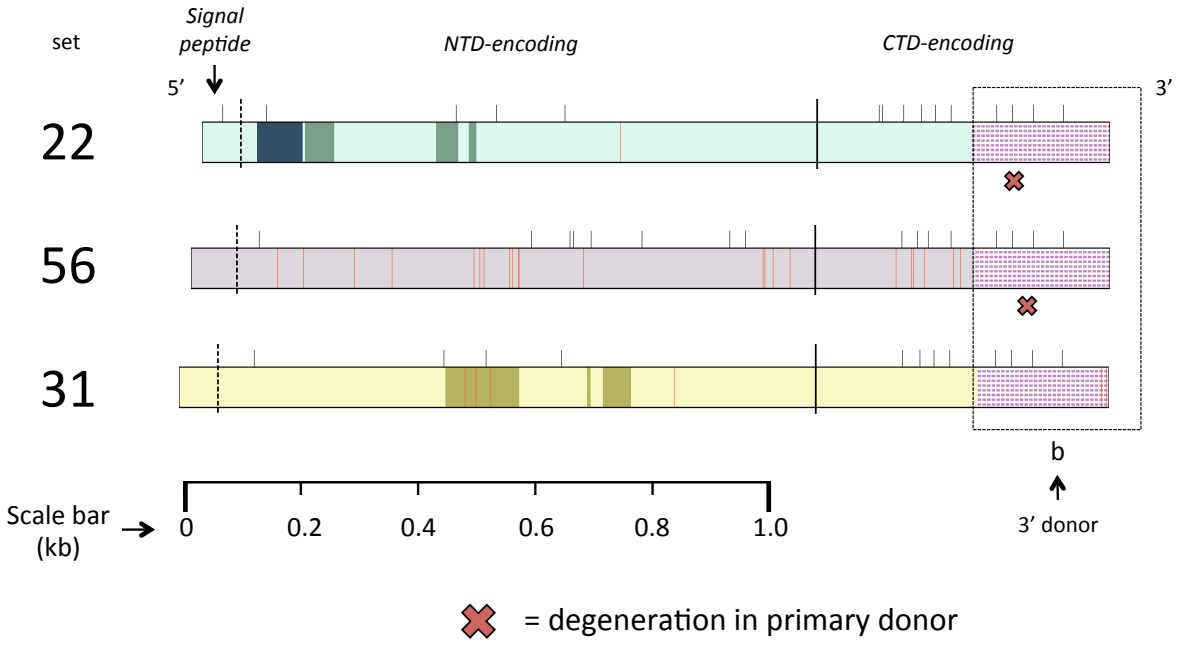


Figure 4.3 Many sets, same 3' end. Three diagrams representing cDNA clones belonging to three different sets (Set_22, Set_56 and Set_31) are shown. The division between N- and CTD-encoding regions is indicated by a black bar. A thick dotted line indicates the division between the N-terminal signal peptide-encoding region and the mature NTD-encoding region. Cysteine codons are represented by a bar projecting above the diagram. For the mosaic *VSG*s from Set_22 and Set_31 segments matching different donors are shown in different colours. Mismatches with all donors are shown as red bars spanning each diagram (see Figure 4.8 for details on how diagrams of mosaics were drawn). The primary donors of Set_22 and Set_55 are pseudogenic in the CTD: the location in the alignment where degeneration begins is indicated by a red cross. 3' variant b (see Table 4.4) contributed the 3' end for each of these *VSG*s: the boxed region (233 nt) contains only two nt differences between the three sequences (<0.01 differences/nt), whereas the unboxed region has variability between 0.492 and 0.539 differences/nt. Each of these sequences was identified in independent reverse transcription-amplification reactions from that infection. A Set_22 3' donation event identical to 02-20c06 was also identified in mouse 04, indicating that this particular 3' variant may have been present in the genome of the inoculum (see Figure 4.13).

4.4.5 The same VSG *could acquire many different 3' ends*

While the previous section (4.4.4) showed that that different silent *VSG*s could be activated in similar ways, it was also observed that the same silent *VSG* could be activated in different ways. Different 3' donations to produce different members of Set_27 are shown in Figure 4.4. In some cases, the acquisition of different 3' ends could be identified within an infection. The most likely explanation for this pattern is that the same *VSG* became polyclonally activated, acquiring a different 3' end each time it was copied into an expression site (Michels *et al.*, 1983; Timmers *et al.*, 1987).

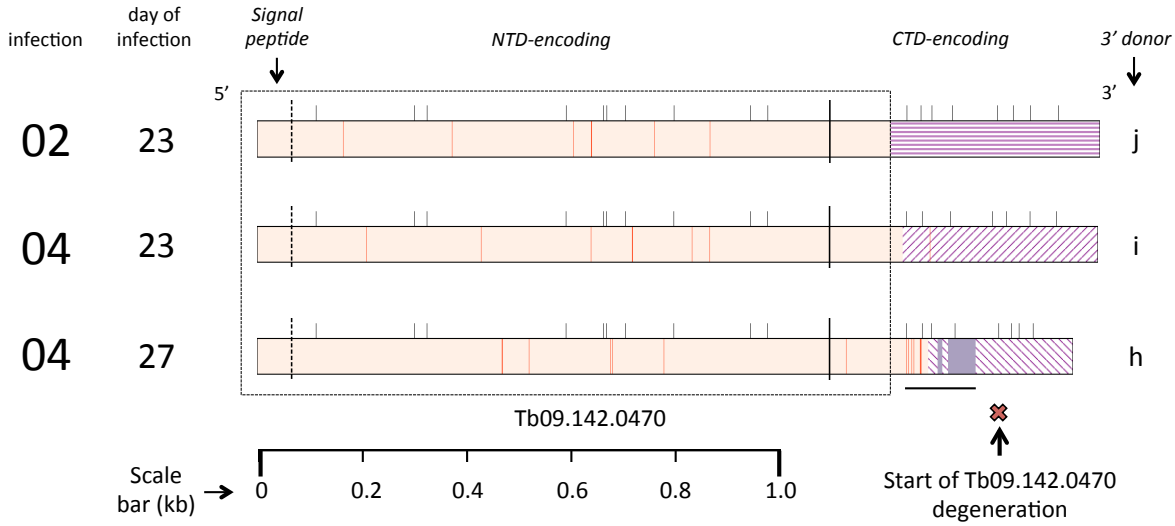


Figure 4.4 Same set, many 3' ends. Three diagrams representing cDNA clones belonging to the same set, Set_27, are shown. The diagrams are annotated as in Figure 4.3. Tb09.142.0470, the donor of Set_27, is pseudogenic in the CTD, and a red cross indicates the position in the alignment where the degeneration begins. For each clone, the damaged donor has been repaired with a different 3' donor variant. The identifier of the 3' variant type contributing to each clone is shown to the right of each sequence. In the boxed region, the clones show variability between 0.008 and 0.011 differences/nt, whereas they show 0.357 to 0.458 differences/nt in the 3' end. In the case of the day 27 clone from mouse 04, a black bar under the sequence indicates a region of low identity with the Set_27 donor and the 3' donor sequence, with may correspond with a residual sequence from multiple 3' donation events at this site. A Set_27 3' donation event identical to day 23 mouse 04 variant was also identified in mouse 05, indicating that this particular 3' donation variant may have been present in the inoculum genome.

Set_12 *VSG*s from mouse 04 showed great variety in their CTD-encoding regions (more than six distinct variants). Some examples are shown in Figure 4.5. Due to the lack of saturation in the sequencing, it is likely that even the observed degree of variation was an underestimation of the total population diversity in this set. Donors could not be found for the extreme 3' end, but Tb11.v4.0021

(12-B) contributed segments to the 3' part of the gene (see also footnote 66). It is possible that a lineage within the mouse 04 infection silently acquired a telomere-proximal copy of the Set_12 donor (Tb11.30.0005, 12-A), which was easily and frequently activated in a polyclonal manner to produce the variety of 3' donation events observed. The identification of Set_12 at many timepoints in this infection (Set_12 was detected in all mouse 04 samples between days 21 and 31) is consistent with this theory, as the mRNA of an easily activated *VSG* may continue to be detectable even in the presence of antibodies. However, it is also possible that variation accumulated in the 3' end of an expression-linked Set_12 *VSG*, perhaps through mosaic-like processes: 12-B and 12-A show greater identity in their 3' ends than they do throughout the rest of the sequence (see footnote 66). For other sets, the number of different variants seen within an infection was usually three or fewer.

4.4.6 3' donation can change the C-terminal domain type of a VSG

When comparing expressed *VSG* clones with their putative genomic donors, 3' donation was apparently able to exchange different CTD types. CTDs were typed by eye, using Carrington et al. (1991) and Marcello & Barry (2007a) as guides (see Chapter 1, section 1.5.2), as current alignment-based approaches have been unable to reliably classify the different domain types (Weirather *et al.*, 2012).

This process of CTD exchange is consistent with the well-mixed nature of the archive: although there is some bias in the archive between N- and CTD type combinations, all combinations of different N- and CTD types are present (Marcello & Barry, 2007a). The different domain type combinations seen in the relatively short periods presented here allows us to speculate that the expression site is a locus where mixing of domains can take place, although it is currently unclear how recombined expression-linked *VSG*s become archival copies (see Chapter 6 section 6.4.7). There is no functional difference ascribed to different CTD types, but it is possible that combining different N- and CTDs has a role in determining activation probabilities of expressed *VSG*s (since homology in the CTD-encoding region can act as a substrate for recombination), or in the generation and evolution of diversity in the *VSG* archive.

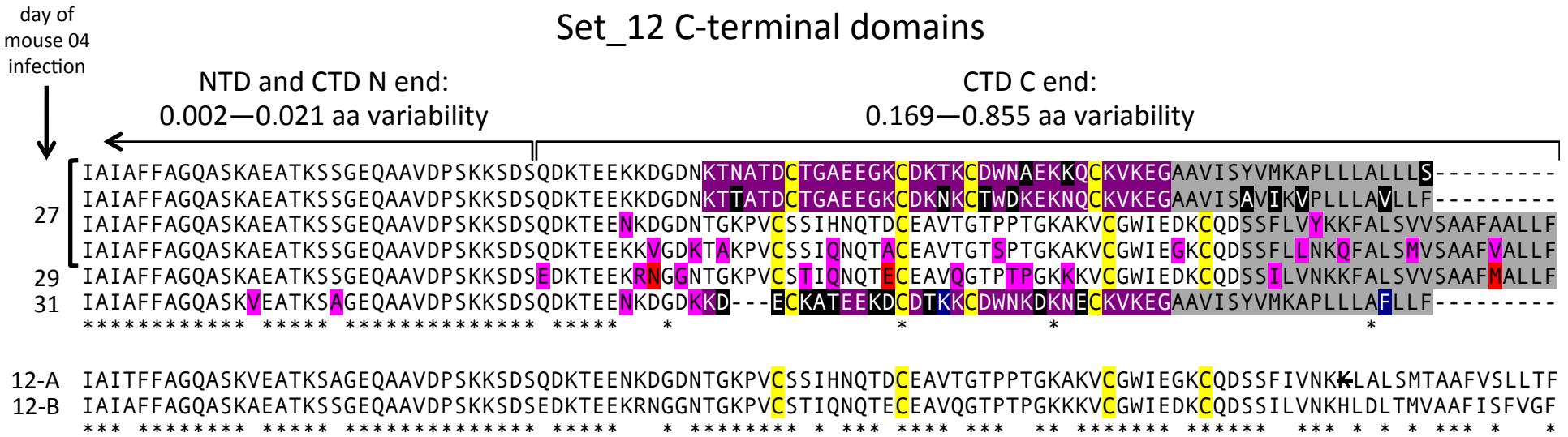


Figure 4.5 Variety amongst Set_12 C-terminal domains. Amino acid sequences of the CTDs of representative Set_12 variants from mouse 04 are shown. Cysteine residues are highlighted in yellow, and the predicted GPI anchor signal sequences are coloured in dark grey. Amongst these clones, there are two broad patterns: a type 5 CTD (indicated in dark purple) and a type 4 CTD. Sequences were aligned and arbitrarily coloured light purple, red, blue or black to emphasise differences amongst expressed sequences. Stars under the alignment indicate identity amongst all expressed sequences. The variability amongst these clones was calculated for the NTD and the N-terminal end of the CTD, and for the C-terminal end of the CTD and are given above the clone sequences (measured in amino acid differences/position). Below the clone sequences, sequences of the main donor (12-A; Tb11.30.0005) and a C-terminal donor that contributes to many variants (12-B; Tb11.v4.0021 contributes most substantially to the day 29 variant here) are given. Stars under the donor sequences indicate identity between the donors. 12-A has a frameshift towards its most 3' end, which was corrected to give the sequence here; the position of the frameshift is indicated by a double-strikethrough.

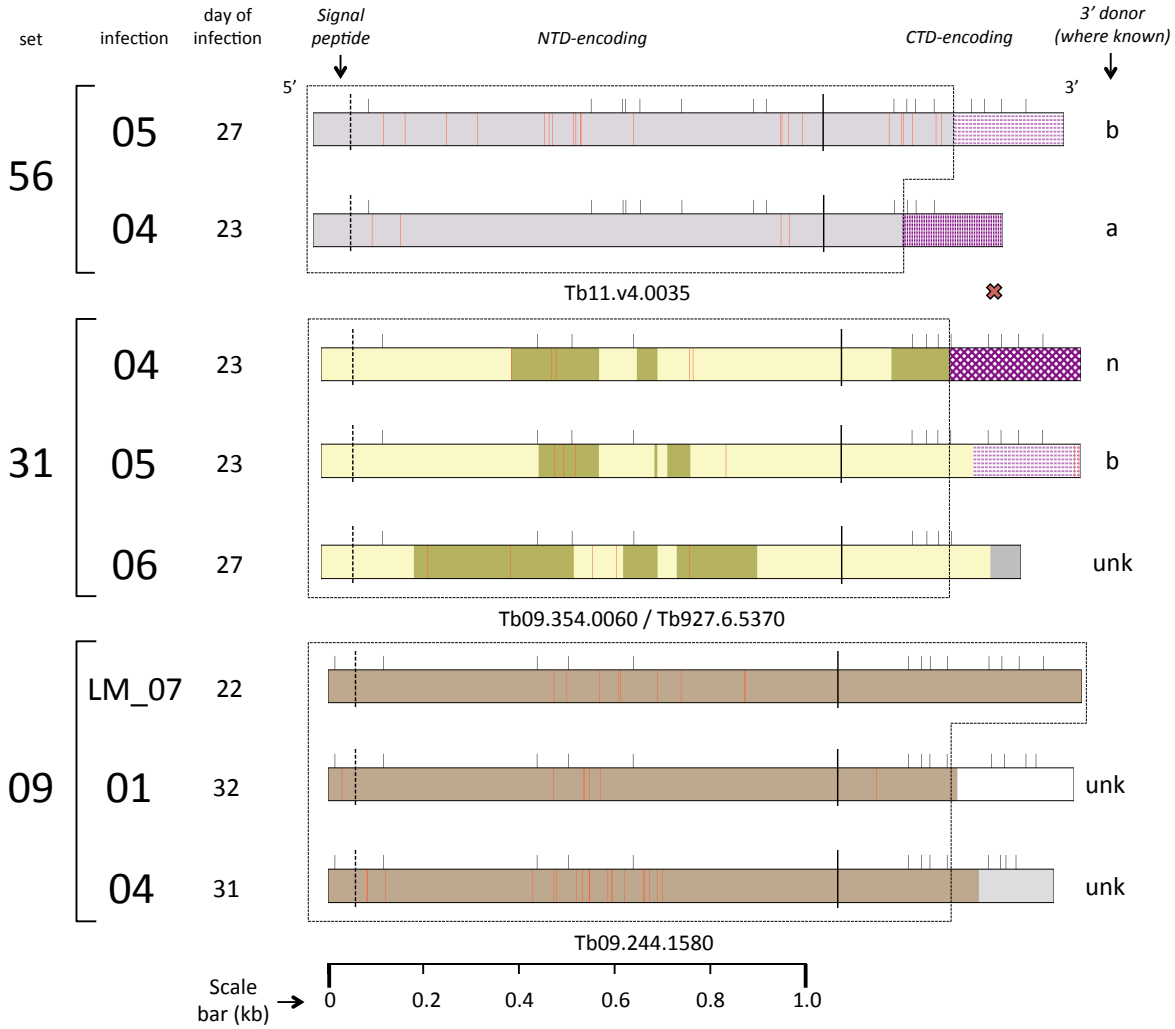


Figure 4.6 C-terminal domain exchange. Diagrams representing cDNA clones are annotated as in Figure 4.3. The position in the alignment where the Set_56 donor becomes degenerate is indicated by a red cross. 'unk' = unknown 3' donor.

4.4.7 Boundaries of recombination

The boundaries of 3' donation were not evenly distributed throughout the CTD-encoding region. For each set, the locations of 3' donation boundaries were recorded. The 3' donation boundary was considered to be the region where an expressed *VSG* sequence begins to diverge from the primary donor. Boundaries were grouped into regions according to structural features of the CTD in which they occurred. The different regions were: 5' of the tetra-cysteine subdomains, amongst the cysteine codons of the first subdomain, between subdomains, amongst the cysteine codons of the second subdomain, between the subdomain and the GPI signal-encoding region, within the GPI signal-encoding region. The locations of the 3' donation boundaries are shown in Figure 4.7.

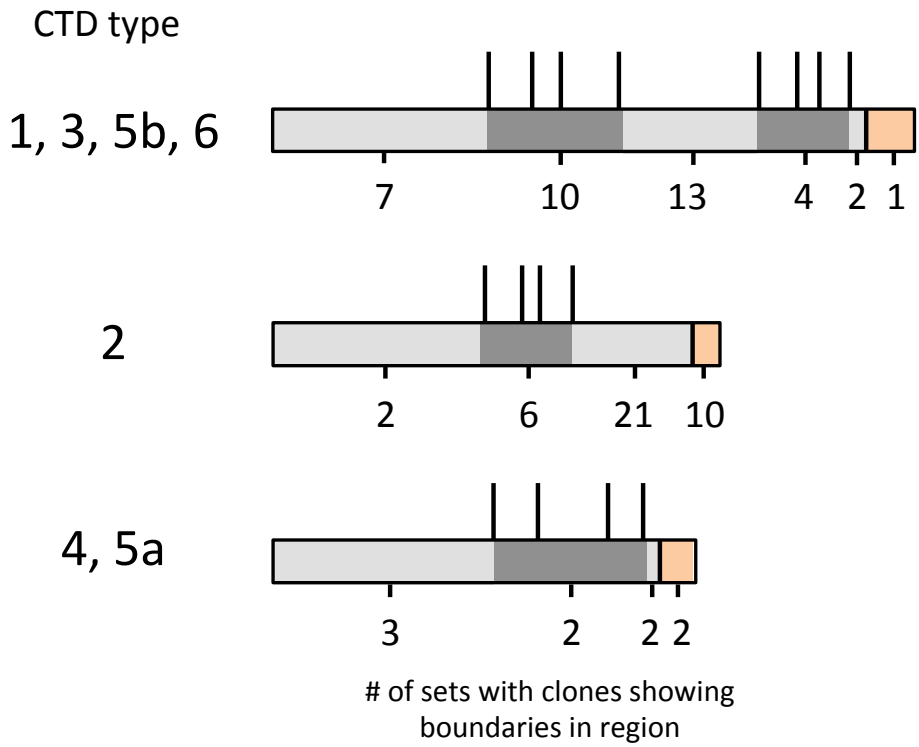


Figure 4.7 Location of 3' donation boundaries. For each set, the location of 3' donation boundaries was recorded. This diagram shows the number of sets with one or more clones possessing 3' donation boundaries in the region of the CTD identified. Some sets showed differing degrees of 3' donation, thus the figures do not total 47. Conserved cysteine residues are shown as bars projecting above the diagrams, and the putative GPI signal-encoding region is shown in orange. Type 2 CTDs possess a longer region between the last cysteine and the GPI signal sequence, so they were considered separately from types 4 and 5a.

Chi-squared tests indicated that it is unlikely that 3' donation boundaries were randomly distributed across the CTD-encoding region.[60] For the two subdomain CTDs, 3' donation boundaries primarily occurred in the first subdomain and in the region between the subdomains, whereas for type 2, boundaries occurred most frequently in the region between the subdomain and the GPI anchor. These patterns probably reflect archive homologies: previous studies of archive *VSG*s show considerable mixing between the first subdomain of types 1, 3 and 6 (L. Marcello, Ph.D. thesis, 2006, University of Glasgow)

The features of the boundary of 3' donation were examined in 16 cases. These cases were chosen as both the primary and the 3' donor could be identified, and

[60] For this test, types 1, 3, 5b and 6 were considered together, to give sufficient data. There was insufficient data to test types 4 and 5a. 'Expected' values were obtained by calculating the proportions of the CTD occupied by the different regions. For this, the CTDs of ILTat 1.24 (for types 1, 3, 5b and 6) and MITat 1.2 (for type 2) were used as approximate standards. For type 1, 3, 5b and 6, $p < 0.05$; for type 2, $p < 0.001$.

they belonged to several different sets. In each case, the boundary of 3' donation occurred in a region of identity 7–27 bp long (median = 12) amidst a region of <0.1 differences per nt of 9–107 bp long (median = 35.5). The donors had 0.307–0.487 differences per nt across their entire lengths (median = 0.444). Details are in Appendix 7.3.3.

4.4.8 How could 3' donation contribute to antigenic variation?

For most sets, 3' donation was the greatest source of within-set variation. Considering that variation in the CTD cannot directly contribute to antigenic variation, this finding initially seems counterintuitive. However, the substantial variation between sequence-related, expressed *VSG*s towards their 3' ends likely corresponds with the different *VSG*s that they displaced as their telomere-proximal copies were formed. Variation in expressed *VSG*s at their 3' ends is therefore consistent with a model of *VSG* activation in which the 3' boundary of gene conversion is often within the coding sequence. The role for 3' donation in antigenic variation is therefore as a route for the expression of different silent *VSG* NTDs—including those with dysfunctional CTDs—rather than itself being a means of generating antigenic novelty. This is analogous to the use of hypervariable cassette regions by other antigenically variant pathogens (see section 1.6.7).

4.5 Mosaicism

4.5.1 Mosaics were constructed from multiple donor sequences

Mosaicism in expressed *VSG*s is the result of recombination events—segmental gene conversions—with boundaries occurring throughout the *VSG* gene, including the NTD-encoding region. Mosaic *VSG*s identified to date involve two or more related *VSG* donors (Roth *et al.*, 1986). Here, for brevity and familiarity, each donor of a mosaic set was named XX-Y, where XX is the number of the set to which the donors contribute and Y is a unique capital letter identifier (A–D). A key is provided in Table 4.5.

Putative mosaics identified as described in section 4.1.3 were investigated in depth by multiple sequence alignment. The sequences of putative mosaics and

their donors were aligned, and for each position in the alignment, identity between a clone and each donor was recorded. These patterns of matching and mismatching were examined. In many cases, blocks of matching and mismatching with the donors could be seen: in some parts of the sequence, donor A might match the clone identically and donor B match the clone poorly, in other parts donor B might match much better than donor A, as shown in Figure 4.8 (A). The diagrams were simplified further, by describing the mosaic as being composed of the blocks from each donor. In these diagrams, the boundary between the blocks is arbitrarily located at the midpoint of the region of identity between the expressed clone and the two combining donors (the 'mosaic junction'). Blocks that were required because of a single base difference between a clone sequence and the donor sequence contributing the surrounding blocks were generally not included.

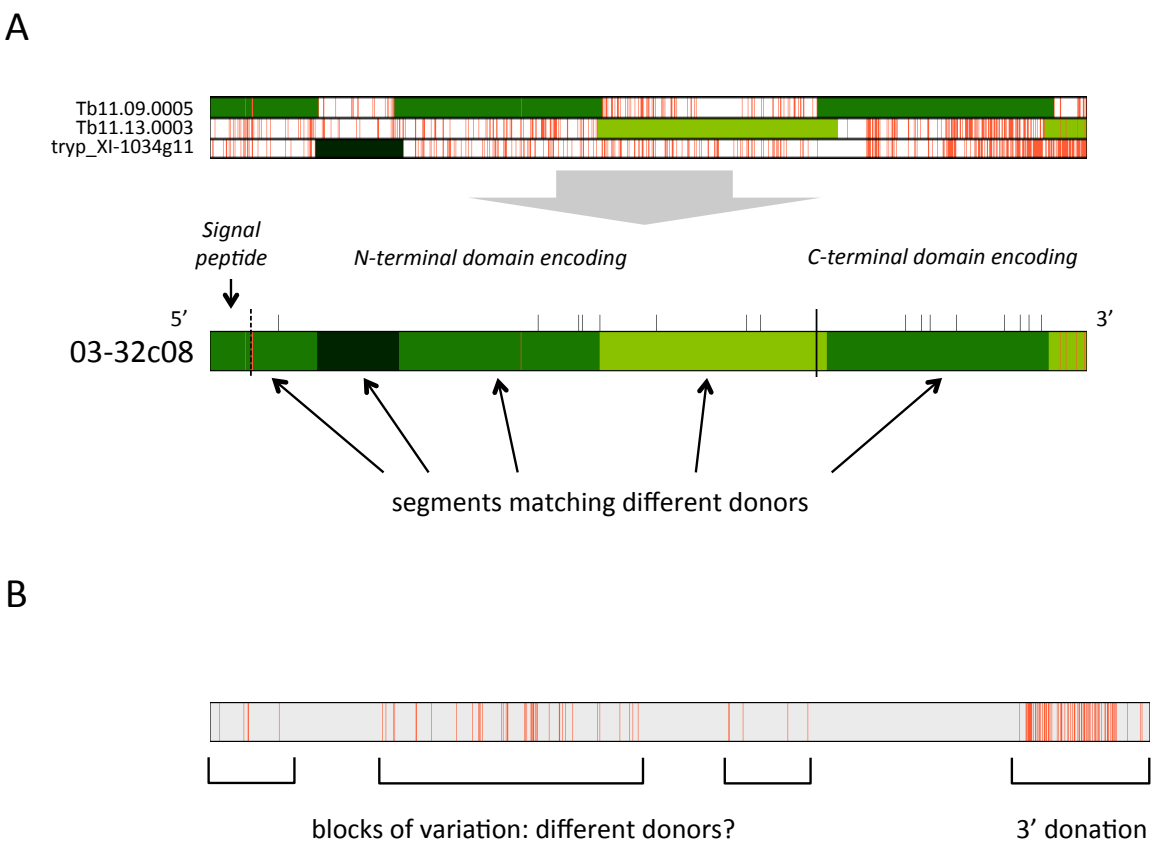


Figure 4.8 Identifying mosaicism. (A) A multiple sequence alignment with Tb11.09.0005, Tb11.13.0003 and tryp_XI-1034g11 showed blocks of matching and mismatching. By assembling these blocks, a diagram of the full length sequence can be drawn, where different colours represent the contributing blocks. (B) Comparison of two clones from Set_69 shows variation similar in pattern and greater in extent to that seen when comparing two clones belonging to a verified mosaic set, although no donors could be found for Set_69.

Comparing putative mosaic sequences with one another aided identification of mosaics by indicating the involvement of unknown genomic copies. For example, it remains possible that putative mosaics are, in fact, the straightforward expression of an unknown genomic *VSG*. Were many identical apparent 'mosaic' *VSG*s identified in different infections, we might doubt that these clones were indeed true mosaics. In more complicated cases, comparing clones with one another revealed possible further unidentified donors. For example, if expressed putative mosaic clones (particularly from different infections) had a region of high identity with one another that had low identity to all of the known donors, a further donor may have contributed to the mosaic at this position. Second, comparing clones with one another revealed mosaicism in sets for which only one—or even no—genomic copy could be found. For example, Set_69 clones showed patterns of identity and mismatch with one another reminiscent of mosaic sets, even though none of their associated genomic copies could be found. This example is shown in Figure 4.8 (B). Mosaicism could therefore be defined conservatively, considering only those sets for which donors could be found *and* variation amongst clones was observed, or more inclusively, extending the definition to those sets showing variation consistent with mosaicism but lacking identification of all donors, and those sets for which comparison with putative donors but no comparison amongst clones could be performed. For the most part, the more inclusive definition of mosaicism has been used in this thesis. The means by which each mosaic set was identified is included in Table 4.5.

4.5.2 Mosaics were verified by PCR

To test directly whether mosaic *VSG*s are indeed formed *de novo* during an infection and are not a product of in vitro template switching, PCR analyses were carried out on Set_14.[61] Primers were designed to bind to both putative donors either side of a junction observed in expressed clones, as shown in Figure 4.9. PCR was carried out on genomic DNA from pre-infection parasites and on genomic DNA from parasites collected at the end of the infection[62] (days 29–32)

[61] Further investigations into segmental gene conversions were carried out following submission of this thesis. See 'Note Added in Proof', Appendix 7.4.2 for further details.

[62] Mosaics possessing the Set_14 junction were observed in the cDNA made from terminal samples of infections of mouse 01, 03, 04, 05 and 06.

using different combinations of the four primers, and the products were separated by agarose gel electrophoresis. The results are shown in Figure 4.10. The reaction designed to detect the mosaic junction was able to amplify a product only from post-infection gDNA, and even then, only from infections where that mosaic had been identified by cloning and sequencing. The mosaic junction was not detectable in pre-infection gDNA.

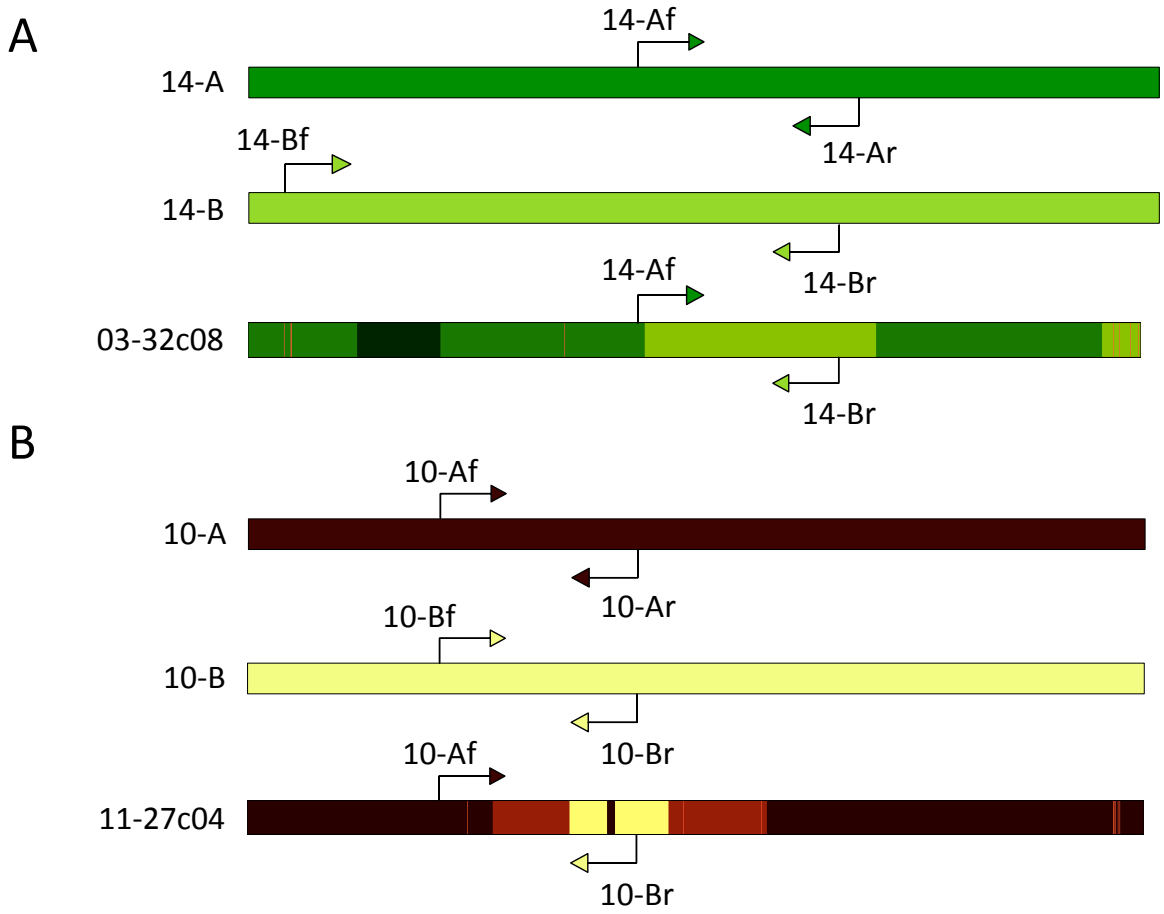


Figure 4.9 Design of PCR. Primers designed to bind to specific donors corresponding to Set_14 (A) and Set_10 (B) were used to investigate mosaicism. For each primer set, the two donors and a mosaic VSG sequence are shown. Arrows indicate the primer binding sites on each sequence.

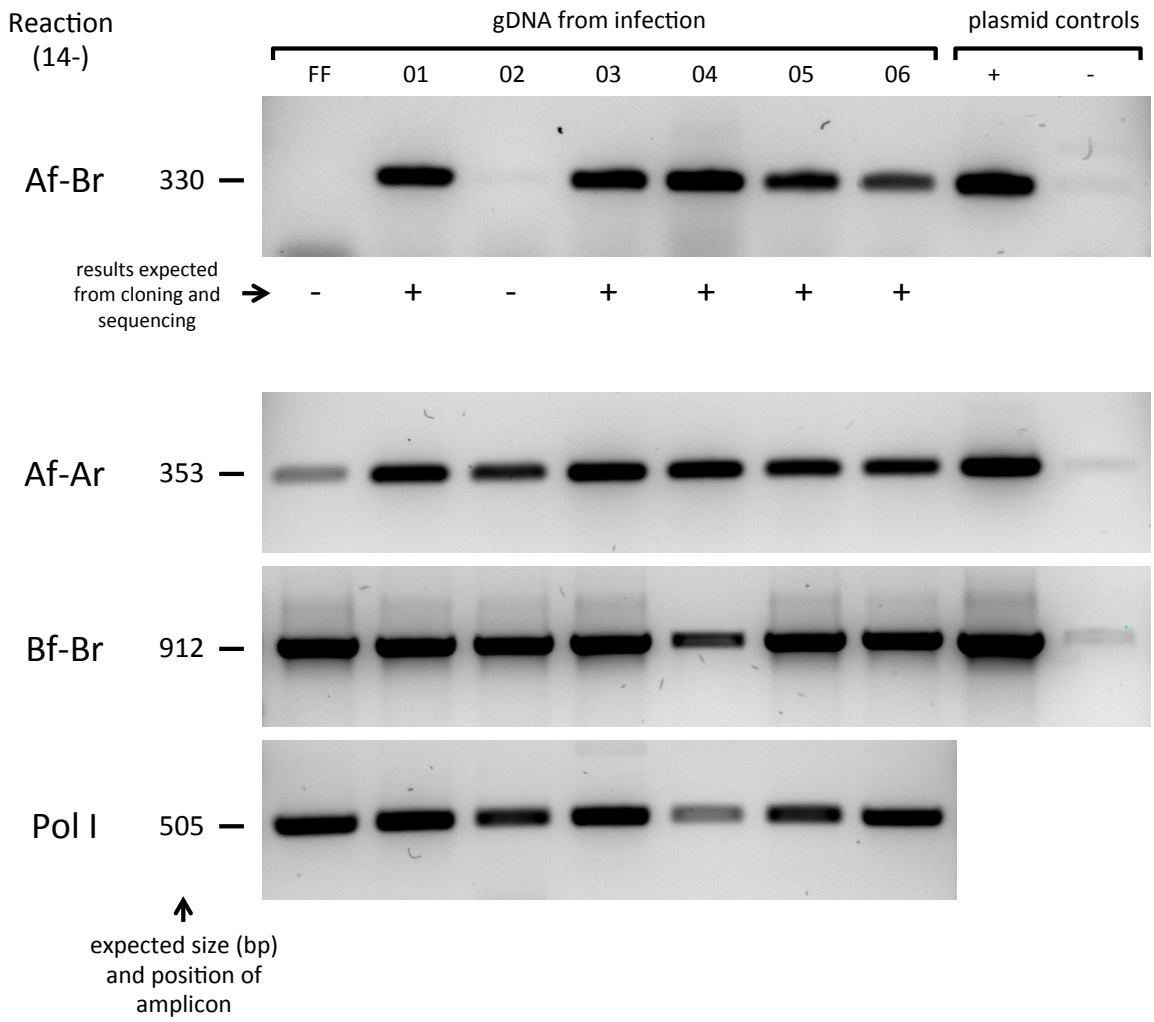


Figure 4.10 Results of PCR (1). PCR was carried out on ~30 ng gDNA template using different combinations of the primers shown in Figure 4.9 (A). In addition, primers directed against RNA polymerase I were used as a control for reaction input. Plasmids containing related mosaic *VSG*s that either did (+), or did not (-), contain that particular junction were used as controls. Non-specific primer binding due to high identity between primer and template sequence and high concentration of template means that a small amount of product can be seen in the negative controls. 'FF' indicates gDNA made from the source of the inoculum (see Appendix 7.2.1 for the pedigree). The terminal gDNA samples from infections 01–06 were retrieved between days 29 and 32.

Similar experiments were designed for a Set_10 mosaic junction, and preliminary results from an experiment performed by H. Wang show that this mosaic junction was also not detectable in the pre-infection genomic DNA (Figure 4.11).

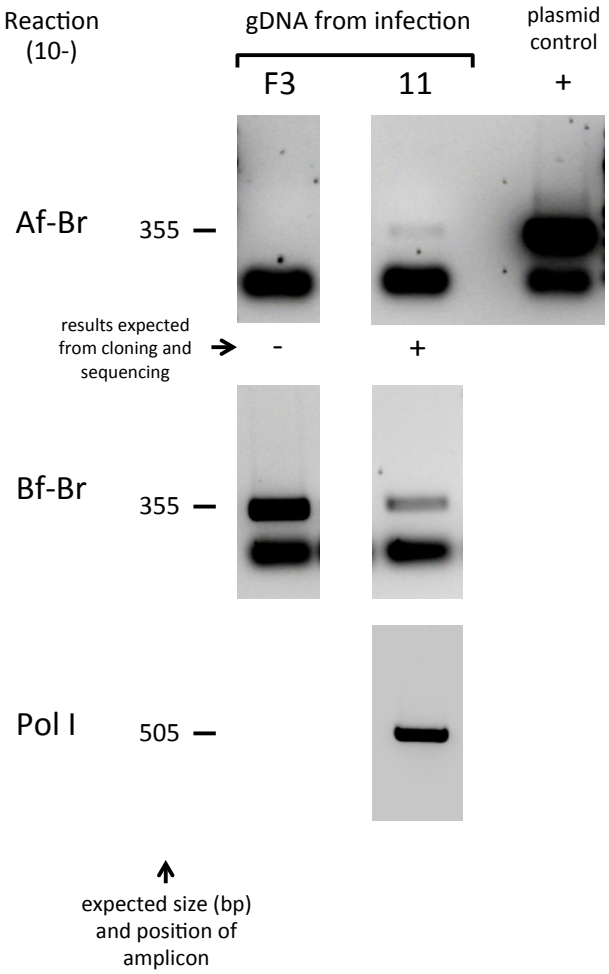


Figure 4.11 Results of PCR (2). PCR was carried out on ~30 ng gDNA template using different combinations of the primers shown in Figure 4.9 (B), and using primers directed against RNA polymerase I as a control for equal loading. A smaller non-specific product can be seen in the reactions, probably a consequence of primer dimerization. For reasons of time the reaction was performed only on two samples, and controls were not available for all reactions. 'F3' indicates gDNA made from the source of the inoculum (see Appendix 7.2.1 for the pedigree). This experiment was performed by H. Wang. See also Figure 7.36 (Appendix 7.4.2).

4.5.3 Related mosaics appear in primary clonal infections

Although the parasites used to initiate all infections had recently been cloned, it is possible that the line had become heterogeneous since cloning, potentially accumulating numerous silent segmental gene conversions amongst its members. To investigate the ability of *T. brucei* to produce mosaic diversity during an infection, primary clonal infections (initiated by a single parasite) of mouse 08 and 09 were set up.

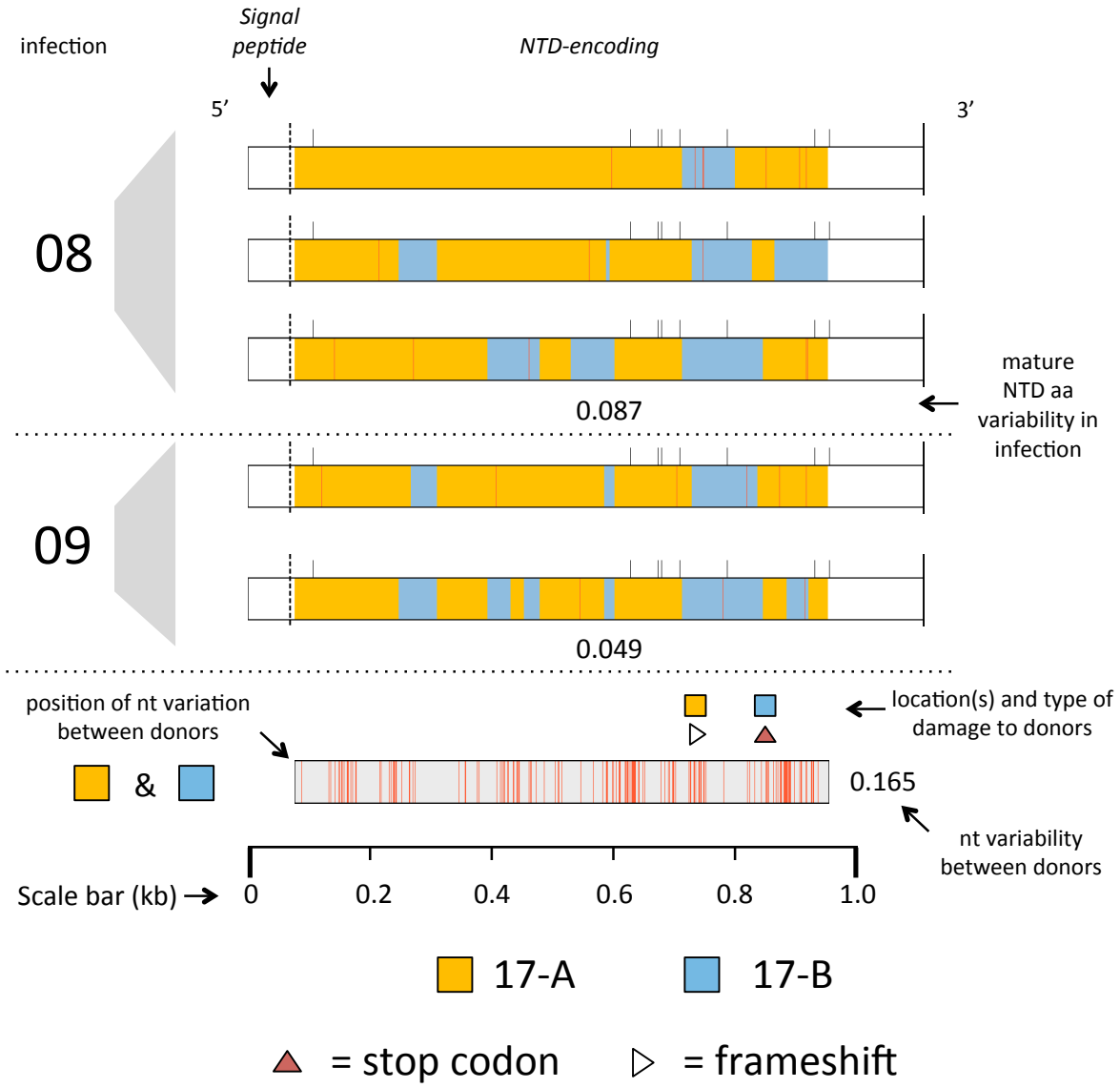


Figure 4.12 Mosaics in clonal infections. Diagrams of the NTDs of mosaic *VSG*s retrieved from clonal infections 08 and 09 are shown. Diagrams of each mosaic were drawn according to Figure 4.8 (A). Only a portion of the mosaic is shown, as the full sequence for 17-A was not available. The maximum variability amongst mature NTDs within each infection, measured in differences/aa, is shown ('mature NTD AA variability'), to give an idea of the antigenic variability introduced by mosaicism. Below the mosaic diagrams is a diagram showing the locations of nucleic acid differences between the two putative donors, and the number of differences/nt ('NA variability between donors') to give an idea of the sequence identity between donors. All the diagrams are aligned, allowing the amount of variation introduced by each segmental conversion to be assessed. Both of the donors are pseudogenic: the type of damage and its location in the alignment is shown in between the mosaics and donors diagrams. 17-A is tryp_XI-1058d01_assembly, 17-B is Tb11.57.0032.

Set_17 mosaic *VSG*s were identified at day 34 in both primary clonal infections. Each clonal infection showed the presence of more than one distinct mosaic variant, and each variant was unique to the infection in which it was detected. Different Set_17 *VSG*s were also detected in mouse 03, mouse 10 and mouse 12, and in the pilot study (Marcello & Barry, 2007a); these showed patterns different

from those seen in the clonal infections. The genomic copy corresponding to the tryp_XI-1058d1 read assembly (17-A) appeared to be the primary donor of the mosaic set, with Tb11.57.0032 (17-B) contributing smaller segments and Tb09.244.0090 (17-C) contributing less frequently. Unfortunately, the read assembly was only 881 bp long, so only a part of the NTD could be analysed with reference to the genome. Figure 4.12 shows diagrams showing patterns of NTD mosaicism for example mosaics from this set. For the most part, each infection showed a different pattern of mosaicism despite using the same donors, consistent with the hypothesis that mosaic sets progress independently in each infection.

The fact that mosaic junctions could not be detected in pre-infection gDNA (section 4.5.2) together with the variety of related mosaics that could be identified from primary clonal infections lends support to the hypothesis that mosaicism is readily generated over the course of infection, and confidence to the apparent mosaic sequences retrieved. With this in mind, the patterns of mosaicism can now be considered.

4.5.4 Mosaicism was frequently observed, and various mosaic sets were identified

Mosaicism could be identified in 28/83 sets, with 16 of these fulfilling the more conservative definition of mosaicism (variation from putative donor sequences *and* variation amongst clones[63]). First, selected mosaic sets will be described. Then, the general patterns of mosaic *VSG*s will be summarized.

The following subsections will explore selected mosaic sets. These sets were selected as a number of clones for each set were available and donors could be identified. Full diagrams for all mosaic variants can be found in Appendix 7.3.4. Diagrams of mosaics were coloured arbitrarily, although an effort was made to maximise contrast. Mosaic sets previously identified (Marcello & Barry, 2007a) were coloured consistently with that work.

[63] Of these, four sets had donors for which a full-length sequence was unavailable, thus under an even stricter definition of mosaicism (full-length donor sequences identified) the count is revised downwards to 12.

Figure 4.13 Set_22 mosaics. Diagrams of each mosaic were arranged and annotated similarly to Figure 4.12. The corresponding 3' donation variants are indicated to the right of each mosaic diagram. For this set, three donors were implicated so each diagrams showing the locations of nucleic acid differences between each pair of donors are given. 22-A is Tb05.5K5.330, 22-B is Tb11.21.0004 and 22-C is Tb927.5.4840.

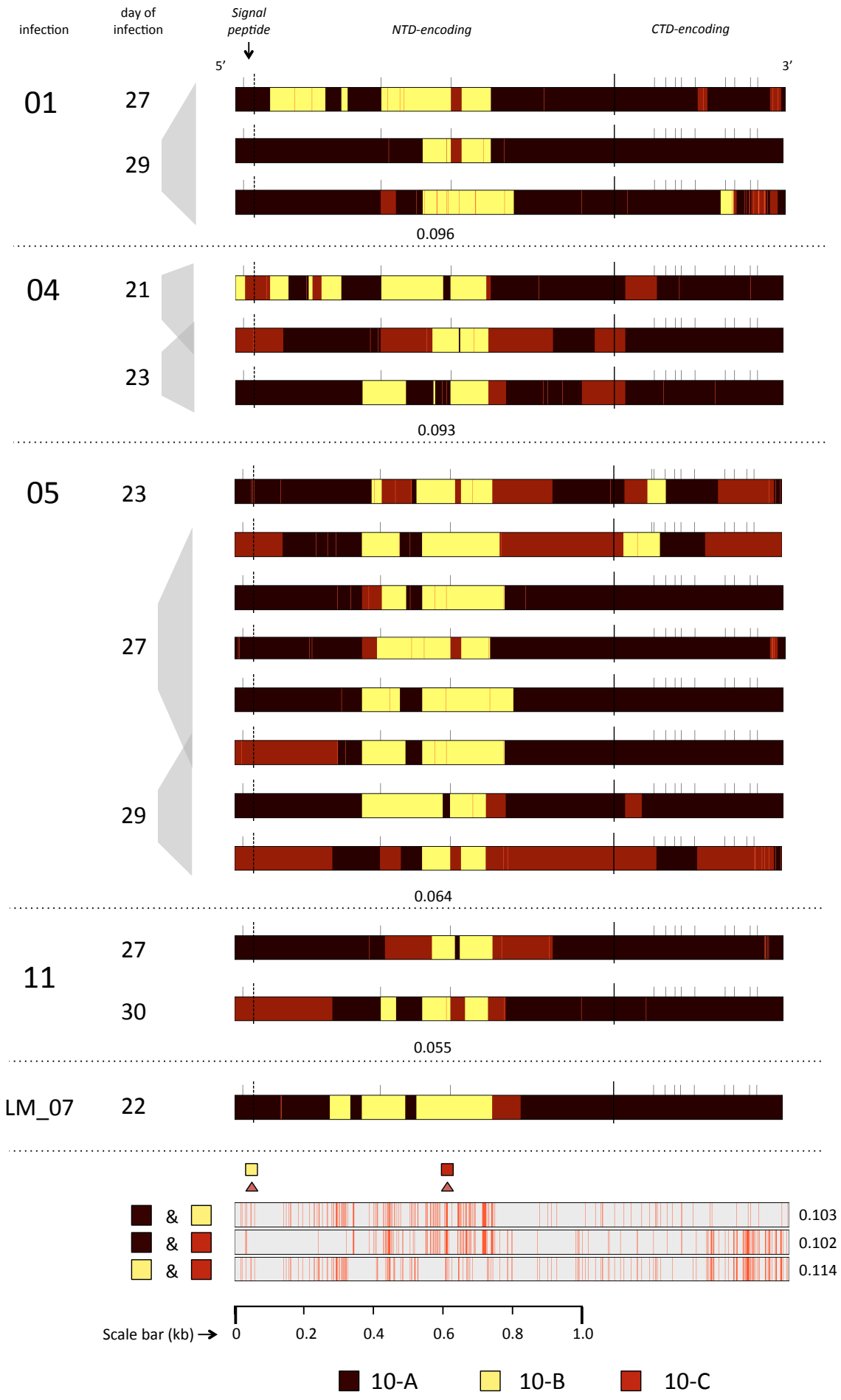


Figure 4.14 Set_10 mosaics. Diagram is arranged and annotated similarly to Figure 4.13. 10-A is Tb11.31.0001, 10-B is Tb10.v4.0161 and 10-C is Tb11.14.0001.

4.5.4.1 Mosaic Set_22

Set_22 (Figure 4.13) showed the earliest-appearing mosaicism,[64] identified between days 17 and 23 in three infections. The patterns of segmental conversion in Set_22 can best be described as the activation of the C-terminally pseudogenic 22-A thanks to 3' donation, followed by varying patterns of segmental involvement of 22-B and 22-C towards the 5' end of the gene. There was some consistency in the 3' donation events between mosaics appearing in different infections, hinting that a copy of the donor may have acquired a functional 3' end before the infections began and may help to explain this mosaic set's relatively early appearance (see section 3.5.5). The patterns of variation in the NTD-encoding region were distinct to infections.

4.5.4.2 Mosaic Set_10

Set_10 (Figure 4.14) is a large mosaic set, showing diverse variants both within an infection and across different infections. Its main donor[65], 10-A, was an intact *VSG*, which appeared to have acquired segments of the NTD-pseudogenic 10-B and 10-C in its NTD over the course of infection. 10-B appeared particularly prone to contributing a segment around the second cysteine of the mature NTD. In some cases, 10-C also contributed segments to the CTD-encoding region, producing examples of C-terminal mosaicism.

4.5.4.3 Mosaic Set_64 and Set_32

Set_64 (Figure 4.15) came about by 64-A undergoing reparation of its pseudogenic CTD through interaction with one of a number of different 3' ends, and the N- and C-terminal pseudogene 64-B contributing various segments to the NTD. Set_32 (Figure 4.16) is representative of many of the mosaic sets identified, in that it was formed by a full-length *VSG* and a read sequence. The two clones were identified in the same infection and appear to have been formed by the activation of C-terminal pseudogene 32-A by 3' donation, followed by segmental conversion with 32-B in the NTD.

---

[64] The earlier appearance of Set_04 probably did not involve segmental gene conversion, as discussed in section 4.5.8.

[65] Marcello & Barry (2007) identified Tb11.v4.0074 as the main donor of their Set_10 mosaic VSG-22-07-02. However, it was found that the introduction of 10-A (Tb11.31.0001) as a putative donor made Tb11.v4.0074 redundant, as all set members could be explained without it.

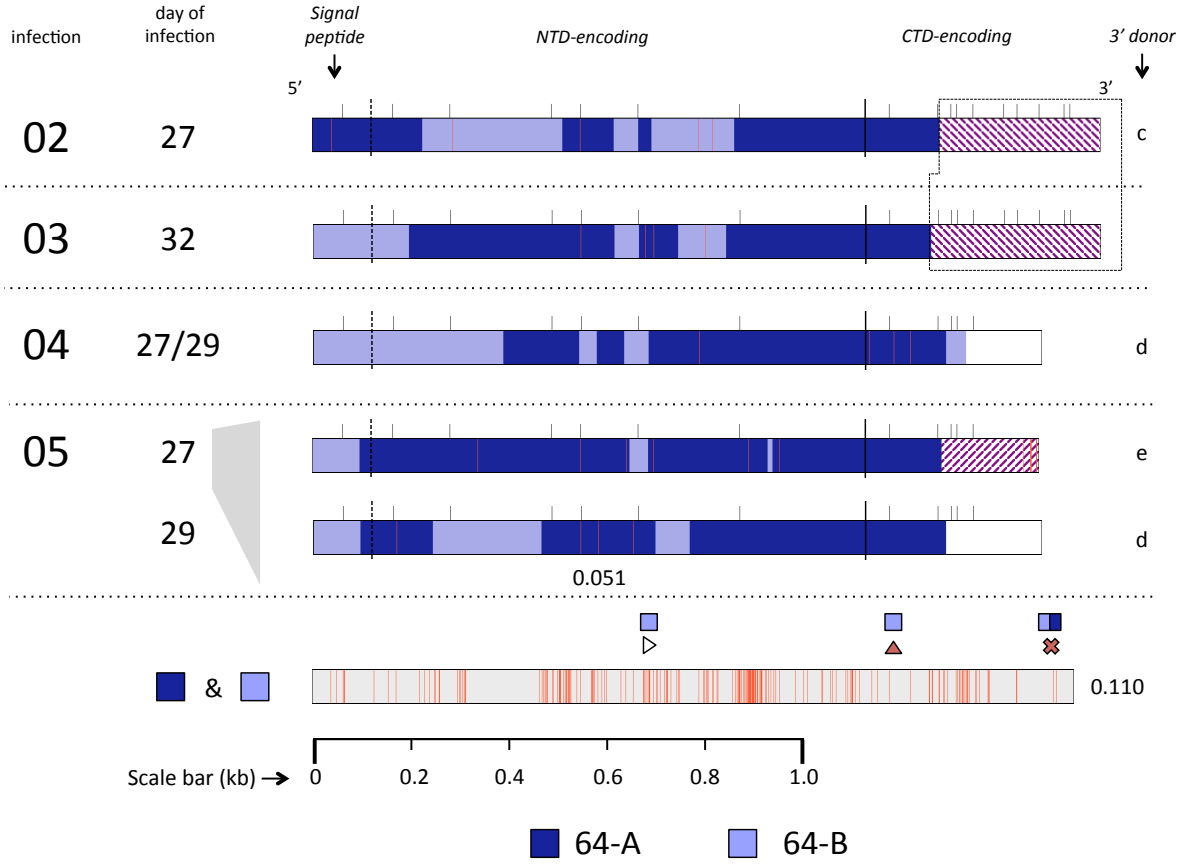


Figure 4.15 Set_64 mosaics. Diagram is arranged and annotated similarly to Figure 4.13. Both donors show degeneration towards their 3' end, the position in the alignment where degeneration begins is indicated by a red cross. 64-A is Tb927.8.420 and 64-B is Tb09.244.0360.

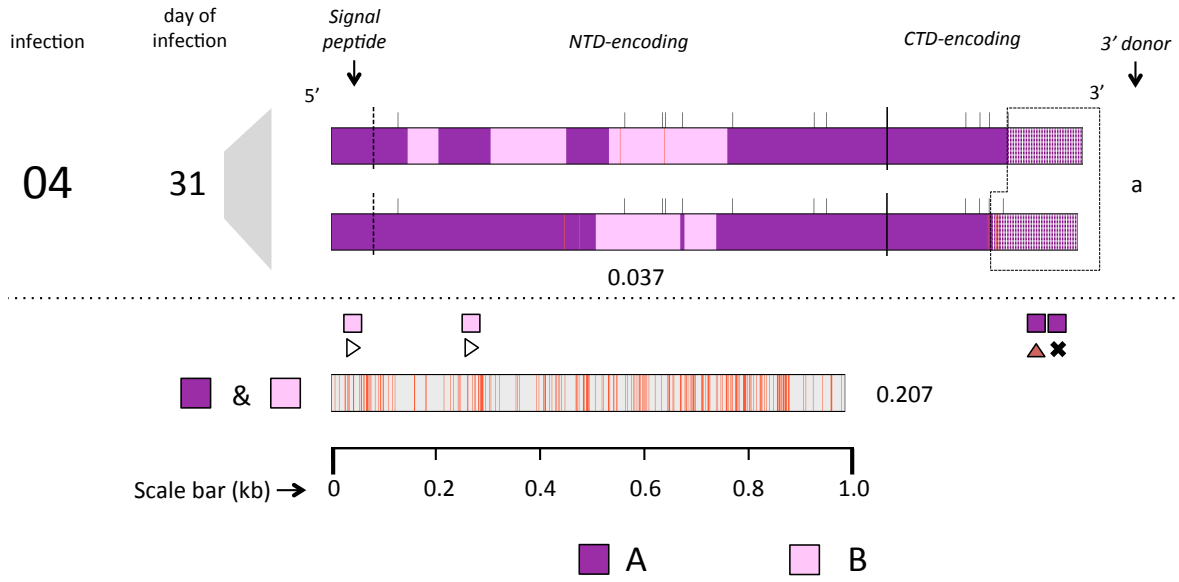


Figure 4.16 Set_32 mosaics. Diagram is arranged and annotated similarly to Figure 4.13. 32-A is Tb09.v2.0090 and 32-B is tryp_IXa-10f02_assembly. As donor B was incomplete, donor nucleotide comparisons could only be carried out for a portion of the sequence.

4.5.4.4 Mosaics summary

Table 4.5 shows details of all sets displaying apparent NTD mosaicism. Those sets discussed in the text and other diagrams are indicated by distinctive colours. The following sections describe some of the general patterns observed in mosaic *VSG*s.

4.5.5 Mosaic VSG*s were formed from similar donors*

All the mosaics detected in these infections shared high levels of identity with one another and with their donors, particularly in the NTD-encoding region. Each mosaic's donors were therefore members of the same *VSG* subfamily. Furthermore, the same few donors apparently contributed to form all of the set's expressed mosaic *VSG*s. As a consequence, all the mosaics of a set had the same NTD type as one another, and all were more similar to one another than they were to any *VSG* of another set. Mosaic *VSG* formation, therefore, did not appear to be a 'free-for-all'—in these comparatively short 4–5 week infections, there were apparently restrictions on the involvement of more variant donors in the NTD. This is in contrast with 3' donation, which seemed to be more promiscuous.[66]

[66] The exception to this is Set_12, which Marcello & Barry (2007a) showed involved the more variant donors 12-A (Tb11.30.0005) and 12-B (Tb11.v4.0021). However, the involvement of the additional donor was only in the very 3' most end of the NTD-encoding region (after position 980). As a consequence, this set was not considered here to be an example of mosaicism (instead an example of 3' donation). Although the two donors had 0.360 differences/nt when their sequence is considered as a whole, variability was concentrated in the more 5' end of the NTD encoding region where they did not segmentally combine (0.449 differences/nt before position 980; 0.203 differences/nt after position 980).

Set	Mosaicism		Max NTD variability		Donors		Donor properties	Variability amongst donors		
	With donors	Amongst clones	Set	Infection				with B	with C	with D
Set_22	*	*	0.076	0.052	A	**Tb05.5K5.330**	ψC	0.177	0.22	-
					B	Tb11.21.0004	atyp	-	0.14	-
					C	Tb927.5.4840	atyp	-	-	-
Set_29	*	*	0.081	0.078	A	*Tb09.244.0150*	ψC	[0.227]	-	-
					B	*927mc_IV-5e11_assembly*	[mc, ψC]	-	-	-
Set_31	*	*	0.074	0.003	A	**Tb09.354.0060**	atyp	0.103	-	-
					B	Tb927.6.5370	intact	-	-	-
Set_32	*	*	0.037	0.037	A	**Tb09.v2.0090**	ψC	[0.254]	-	-
					B	tryp_IXa-10f02_assembly	[read, ψN]	-	-	-
Set_04	*	*	0.116	0.057	A	***tryp_X-54b12.q1c-rev_assembly***	[read]	[0.223]	-	-
					B	*Tb10.v4.0061*	ψN ψC	-	-	-
Set_43	*	*	0.100	-	A	**Tb10.v4.0102**	atyp	0.304	-	-
					B	Tb10.v4.0081	ψN	-	-	-
Set_51	*	*	0.076	0.003	A	**tryp_IXa-20a05_assembly**	[read]	[0.271]	-	-
					B	Tb11.57.0023	ψN ψC	-	-	-

Table 4.5 Summary of mosaicism. For each set the means of identification (comparison with donors and/or comparison amongst clones) is indicated in the corresponding column by a '∗'. The maximum variability amongst clones of each set, clones of that set within an infection is given, as well as the pairwise variability between donors. Variability in this table is given as the number of differences between mature NTDs per amino acid: damaged donors were corrected conservatively to allow translation. Samples of the colours used to identify donors that feature in other Figures in this chapter are given alongside the donor name. Donor names in bold indicate the primary donor of a set, where one could be identified. Where donor names are italicized, these donors do not explain all the N-terminal variability of that set (the donor sequences were incomplete and/or there was the contribution of further unidentified donors). 'Donor properties' represents data from VSGdb supplemented by additional analyses; brackets indicate a partial read sequence that does not cover an entire VSG gene. Where the comparison is bracketed, it was carried out using one or more incomplete sequences. Table continues on following pages.

Set	Mosaicism		Max NTD variability		Donors		Donor properties	Variability amongst donors		
	With donors	Amongst clones	Set	Infection				with B	with C	with D
Set_40	*	*	0.111	0.111	A	**Tb11.1480**	atyp	0.169	-	-
					B	Tb10.v4.0063	ψN	-	-	-
Set_14	*	*	0.219	0.219	A	Tb11.09.0005	atyp	0.251	0.244	0.245
					B	Tb11.13.0003	ψN	-	0.257	0.268
					C	tryp_XI-1034g11_assembly	read, ψN ψC	-	-	0.214
					D	Tb10.v4.0009	ψN ψC	-	-	-
Set_17	*	*	0.196	0.087	A	***tryp_XI-1058d01_assembly***	[read, ψN]	[0.217]	[0.318]	-
					B	*Tb11.57.0032*	ψN ψC	-	0.332	-
					C	*Tb09.244.0090*	ψC	-	-	-
Set_10	*	*	0.110	0.096	A	**Tb11.31.0001**	intact	0.186	0.113	-
					B	Tb10.v4.0161	ψN	-	0.154	-
					C	Tb11.14.0001	ψN	-	-	-
Set_59	*	*	0.085	0.069	A	**Tb927.3.270**	ψN	[0.189]	-	-
					B	tryp_XI-1157a04_assembly	[read, ψN ψC]	-	-	-
Set_62	*	*	0.053	0.048	A	**Tb927.5.4670**	atyp	0.213	0.21	-
					B	Tb09.v4.0061	ψN	-	0.231	-
					C	Tb05.5K5.320	intact	-	-	-
Set_64	*	*	0.095	0.051	A	**Tb927.8.420**	ψC	0.193	-	-
					B	Tb09.244.0360	ψN ψC	-	-	-
Set_16	*	*	0.051	-	A	**Tb11.48.0003**	ψC	[0.074]	-	-
					B	tryp_XI-934d12.q1k_rev	[read, ψN]	-	-	-
Set_49	*	*	0.110	0.110	A	***Tb11.43.0002***	ψN ψC	[0.22]	-	-
					B	*tryp_XI-909g03.q1k-rev*	[read]	-	-	-

Table 4.5 (continued)

Set	Mosaicism		Max NTD variability		Donors		Donor properties	Variability amongst donors		
	With donors	Amongst clones	Set	Infection				with B	with C	with D
Set_24	*		-	-	A	**Tb08.27P2.460**	intact	[0.19]	-	-
					B	trypA24a5.p1p	[read]	-	-	-
Set_48	*		-	-	A	**Tb11.35.0001**	intact	0.22	-	-
					B	Tb09.244.1360	ψC	-	-	-
Set_46	*		-	-	A	**Tb11.16.0003**	atyp	0.347	-	-
					B	Tb10.v4.0095	intact	-	-	-
Set_42	*		-	-	A	**Tb927.8.170**	atyp	0.21	-	-
					B	Tb10.v4.0070	intact	-	-	-
Set_33	*		-	-	A	**Tb10.v4.0031**	intact	0.297	-	-
					B	Tb09.v4.0071	ψC	-	-	-
Set_20	*		0.024	0.000	A	*Tb927.7.6530*	ψC	-	-	-
Set_63		*	0.041	0.033	A	*Tb927.5.4930*	ψN	[0.134]	-	-
					B	*tryp_XI-1007d10*	[read, ψN]	-	-	-
Set_84		*	0.096	0.071	A	*927mc_VIII-14d09_assembly*	mc	-	-	-
Set_11		*	0.048	0.045	A	*Tb927.3.190*	intact	-	-	-
Set_54		*	0.155	0.155	A	***Tb11.v4.0029***	intact	[0.134]	[0.231]	-
					B	*tryp_IXa-29e12_assembly*	[read]	-	[0.26]	-
					C	*tryp_X-99f10_assembly*	[read, ψN]	-	-	-
Set_37		*	0.065	0.065	A	*Tb09.v4.0123*	ψC	-	-	-
Set_69		*	0.086	0.029	-	*none identified*	-	-	-	-

Table 4.5 (continued) See Appendix Figure 7.13 for details on Set_20.

The greatest degree of variation between donors was seen between Tb11.13.0003 (14-B) and the tryp_XI-1034g11 read assembly (14-C), which contribute to Set_14 mosaic *VSG*s; these two donors had 0.269 differences/nt across their entire length. The most similar donors, Tb09.354.0060 (31-A) and Tb927.6.5370 (31-B) had 0.059 differences/nt. In general, mosaic donors had between 0.1 and 0.2 differences/nt across their full-lengths, and between 0.1 and 0.35 differences/aa in their mature NTD.[67]

The apparent demand for identity between donors suggests that increasing divergence between donors inhibits their ability to interact and form functional mosaics. It might therefore be expected that mosaics with more similar donors would tend to appear earlier in an infection. However, a preliminary investigation into this relationship saw no obvious correlation between time of first appearance and identity between donors for the mosaics identified here (data not shown). It is possible that the differentiation-based control likely dominating in these infections complicated this analysis, as the expansion (and hence detection) of a particular mosaic lineage became separated in time from its construction, due to suppression of that subpopulation by differentiation.

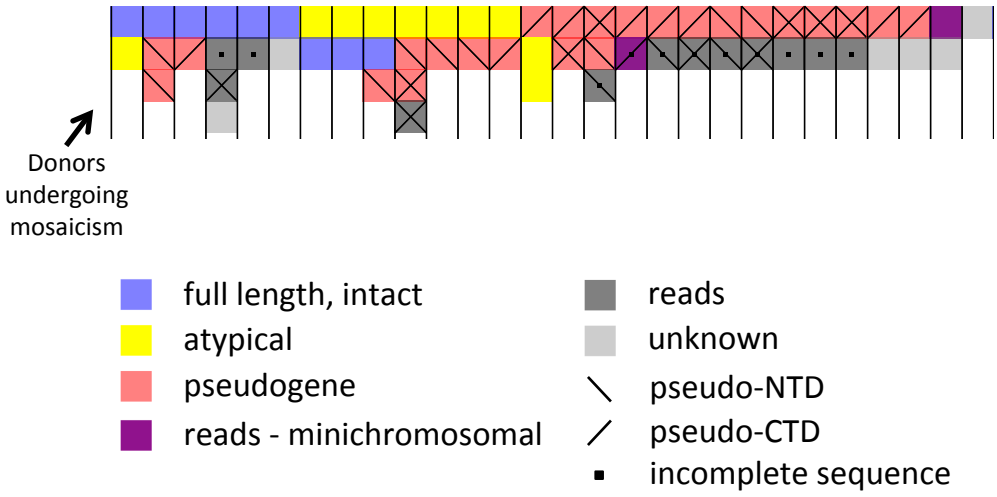


Figure 4.17 Mosaicism repairs pseudogenes. For the 28 sets undergoing mosaicism, each putative donor is indicated by a coloured square. The properties of the donors are indicated, according to the key. This figure is a cropped version Figure 3.7, with additional details on the pseudogenicity of the donors.

[67] Frameshifts in the NTD-encoding region of mosaic donors were corrected for this analysis by the addition or deletion of bases. In no cases did such modifications increase variation between donors.

As detailed in Figure 4.17, mosaicism allowed the partial expression of damaged silent *VSG*s. Of the 28 sets, 16 showed the repair of a pseudogenic donor by mosaicism.[68]

4.5.6 Segmental gene conversion did not demand perfect identity

The homology between donors apparently required for mosaic construction indicates that this process of segmental gene conversion was probably catalysed by a type of homologous recombination, whereby identities between donor sequences were required to guide introduction of the converted segment. To investigate the sequence requirements for segmental gene conversion, 435 boundaries of segmental gene conversion from Set_04, Set_10, Set_14, Set_22, Set_31, Set_32 and Set_40 were studied. Three properties associated with recombination—low GC-content, length of perfect identity between donors, and the presence in the donors of palindromic sequences—were investigated.

Beyond the general homology seen between mosaic donors, there were no outstanding features common to all the boundaries of segmental gene conversion. In some cases, a region of perfect identity between donors at the boundaries was not present, suggesting that perfect identity between donors is not required at the boundary of segmental gene conversion. It is possible that identity with the donors was ablated by the subsequent occurrence of point mutations at that position. However, the identification of little (less than 3 bp) or no identity at the boundaries of 33 apparent segmental gene conversion events, from five out of seven sets investigated, makes this explanation unconvincing.

4.5.7 Similar mosaic junctions were identified in different infections

In some cases, mosaic donors appeared to interact at the same position forming a 'conserved' mosaic junction. Conserved junctions were seen most strikingly with Set_14 mosaics, shown in Figure 4.18. Other examples were found in Set_04 and Set_17. In each of these cases, the boundary of recombination occurred in a region of complete identity between the donors 5–10 bp long.

[68] This figure does not include those sets involving the repair of CTD pseudogenes by 3' donation.

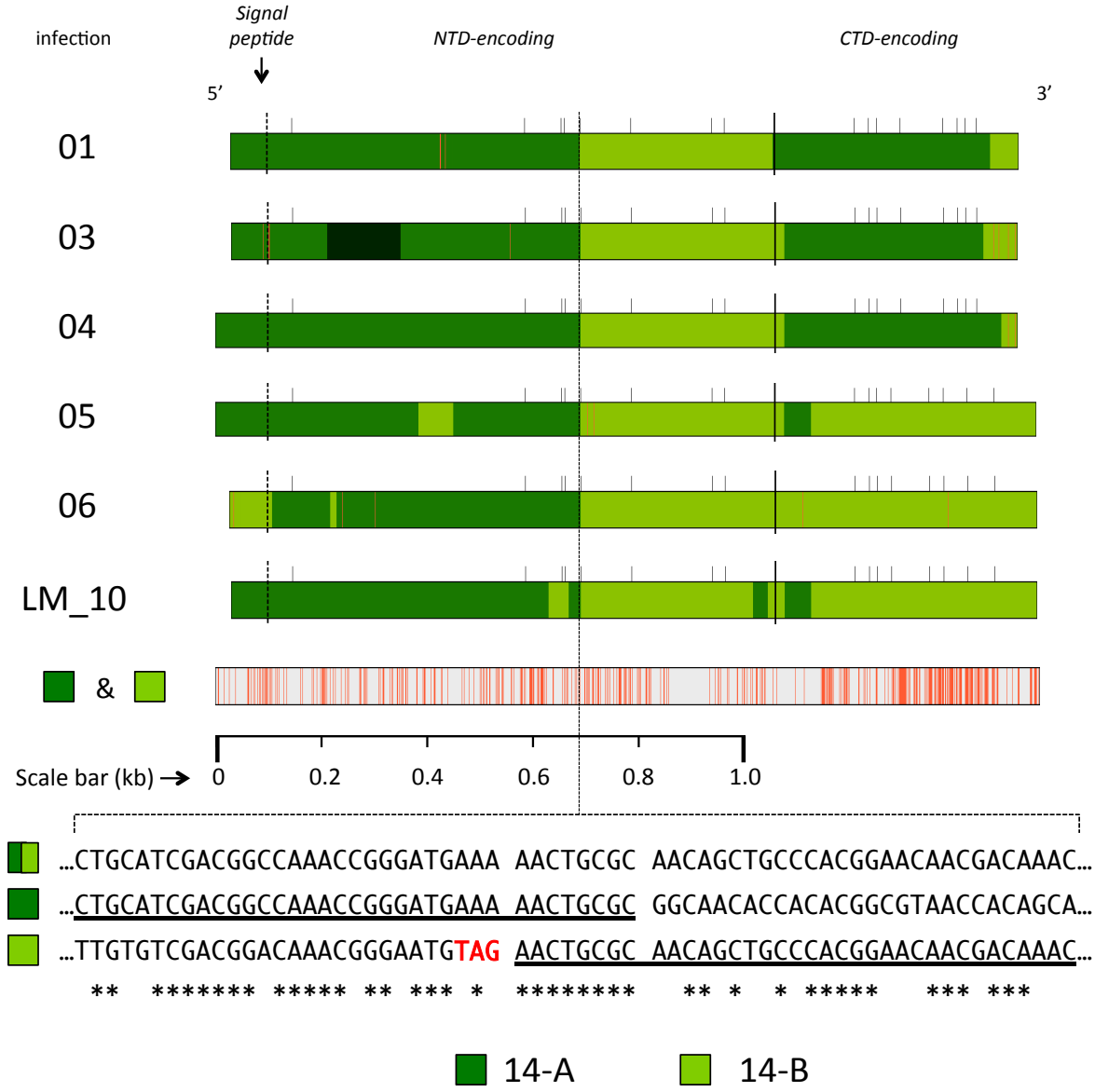


Figure 4.18 Conserved mosaic junction. At the top, diagrams representing expressed *VSG*s from five infections are shown. The conserved mosaic junction is indicated by a dotted line. At the bottom, a multiple sequence alignment of the region covering the junction is provided: the sequences, in order from top to bottom, are a cDNA clone sequence, donor 14-A, donor 14-B. The cDNA clone sequence matches all the clones at the top, except that from mouse 05 (two point mutations) and from LM_10 (another segmental involvement of 14-B to the 5' end). The underlined regions of 14-A and 14-B match the cDNA clone sequence exactly. Asterisks below the alignment indicate positions where both donors match. 14-B possesses an in-frame 'TAG' stop codon, coloured red. 14-A is Tb11.09.0005 and 14-B is Tb11.13.0003. Between the cDNA diagrams and the sequence alignment, a diagram representing the locations of nucleic acid differences between 14-A and 14-B is provided. Other diagrams of Set_14 mosaics are shown in section 4.5.10.

The most straightforward explanation for this finding is that there were silent genes containing these specific junctions present in the genome at the beginning of infection. However, the PCR analysis on the Set_14 junction in section 4.5.2

suggests that if this particular junction was present in the inoculum, it was present in just a small number of the infecting parasites, below detection levels.[69] Further experiments are required to resolve this question. It is interesting to note that Kamper & Barbet (1992) identified a 24 bp mosaic junction that was the same in three independently isolated mosaics (Watat 1.12, 1.13 and 1.14, see Chapter 1 section 1.6.9), although it is possible that this particular junction of theirs could be traced back to an initial rabbit infection. Were conserved mosaic junctions genuine, it would suggest that certain segmental conversion events are favoured over others, hinting at the mechanisms underlying mosaic *VSG* formation.

4.5.8 There was only limited evidence for progressive mosaicism

Mosaics are hypothesised to 'build up', in either an active or silent ES (Barry *et al.*, 2005). In this model, one donor is first copied into the ES. Other donors then interact with the first donor, sequentially contributing segments to build an increasingly complex string of mosaic *VSG*s over time.

The data presented in Figure 4.19 show some evidence for this process occurring. For example, Set_40 had two donors, 40-A (Tb11.1480, a *VSG* with an atypical GPI anchor sequence) and 40-B (Tb10.v4.0063, an N-terminal pseudogene). Barring a few point mutations and a short 3' donation correcting the GPI anchor sequence, a clone detected at day 27 in mouse 02 (02-27c01, indicated by a star in Figure 4.19) showed identity with 40-A across its whole length. Yet at the same timepoint in this infection, three other related variants were found, each of which had blocks that mismatch with 40-A and show greater identity at these positions with the N-terminally pseudogenic 40-B. These other

[69] An approximate calculation, where the mass of DNA present in a single parasite genome is 26 Mbp x 2.5 (to account for the fact that the genome project sequenced only haploid megabase chromosomes) x 610 = ~40 x 10^9 Da, finds that approximately 15,000 copies of a single-copy gene are present per nanogram of pure trypanosome gDNA. For the PCR reactions in section 4.5.2, 30 ng gDNA was used. A PCR titration experiment using limiting dilutions of isolated plasmid template found that the PCR reaction performed to detect the junction was sensitive enough to detect the mosaic junction when the template was present in as few as 1000 copies (data not shown), although it is likely that reactions using gDNA templates are in general less sensitive than using plasmid templates due to impurities in the preparation and non-specific interactions between template DNA and primers.

variants could have been formed by 40-B contributing segments to an expression-linked copy of 40-A (represented by the starred clone).

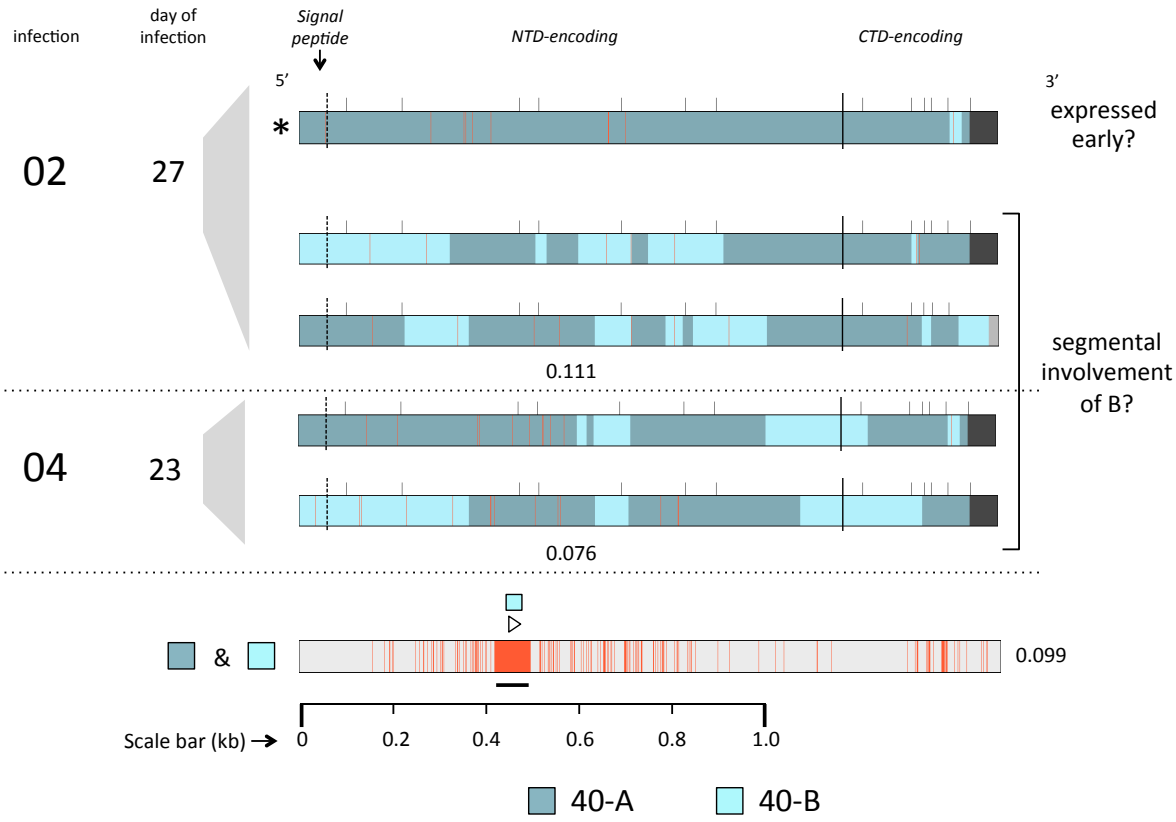


Figure 4.19 Set_40 progressive mosaicism. Diagram is arranged and annotated similarly to Figure 4.13. Mouse 04 variants were more variant from one another than they were from their putative predecessor (the 'early' mouse 02 variant). 40-A is Tb11.1480 and 40-B is Tb10.v4.0063. The black bar under the donor comparison indicates a region of sequence missing in 40-B.

Similarly with Set_04 (Figure 4.20): a clone retrieved from day 32 of mouse 01, and a clone identified by (Marcello & Barry, 2007a) (from infection LM_03), are almost identical (a diagram representing these clones, 01-32c11 and VSG-09-03-04, is indicated by a star in Figure 4.20). Unfortunately, no genomic copy donor could be found to match these sequences for the full length, however, both of these sequences match the 851 bp long tryp_X-54b12.q1c-rev read assembly (04-A) very well for its entire length. Across its length, 04-A contains no stop codons or frameshifts. The other clones of this set show the involvement of Tb10.v4.0061 (04-B), an N- and C-terminal *VSG* pseudogene that has apparently contributed one large segment in the case of the mouse 05 variant, and has undergone more complex segmental conversion events in the case of the mouse 06 variant, and the other variant from mouse 01.

In neither of these cases do we see increasing involvement of the additional donor in an infection over time: for Set_40 the clones from mouse 02 showing progressive mosaicism were all retrieved from the same point, and for Set_04 the more complex clone from mouse 01 was retrieved from the earlier timepoint. In other cases, patterns of progressive mosaicism were even more elusive. Many sets, for example Set_10, show similar or identical mosaics at sequential timepoints, as can be seen in Figure 4.14. Other sets show apparently more complex mosaics at earlier timepoints, for example Set_04 in mouse 01, Set_14 in mouse 05 or Set_10 in mouse 04 (Figure 4.20, Figure 4.23 and Figure 4.14 respectively). When considering all mosaics together there was not a clear general progression from simple to more complex mosaics across an infection as it progresses.

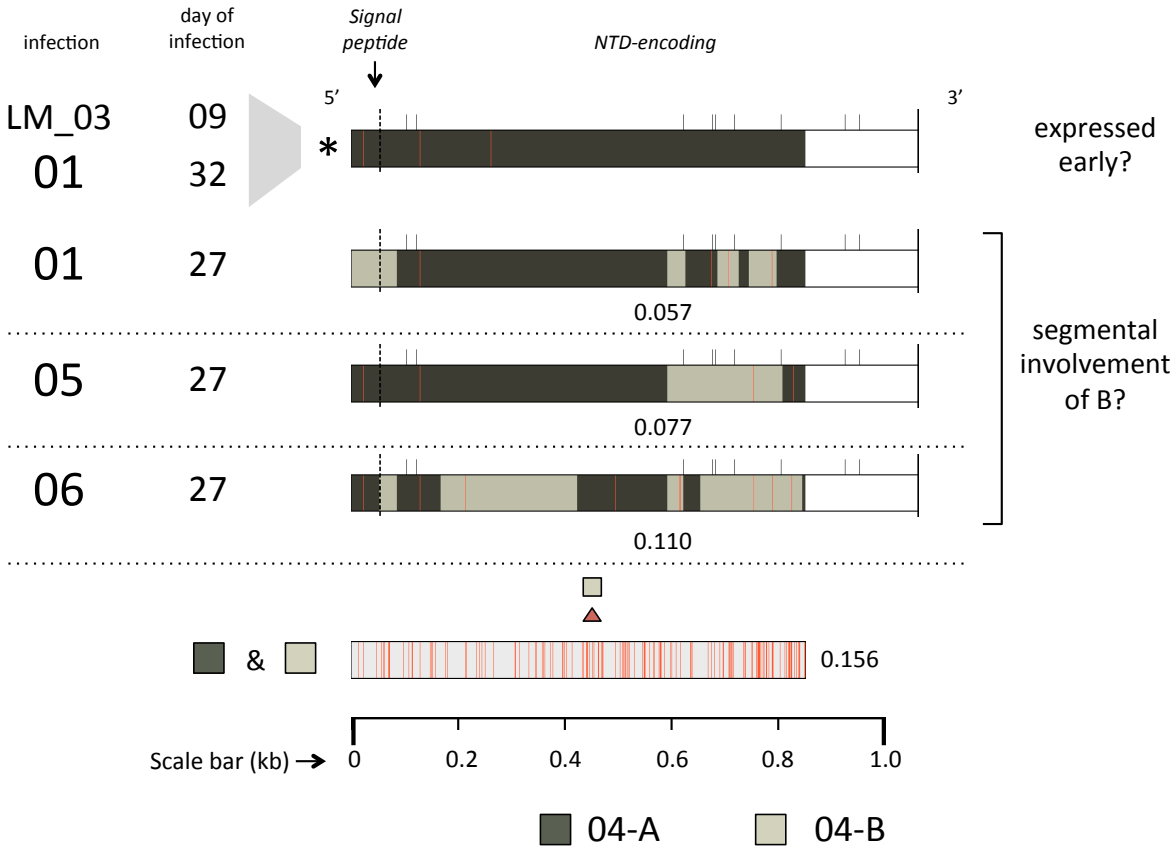


Figure 4.20 Set_04 progressive mosaicism. Diagram is arranged and annotated similarly to Figure 4.12 however maximum variability for each expressed sequence was calculated by comparison with the putative predecessor (the day 09 / day 32 variant present in LM_03 and mouse 01 and indicated by a star). 04-A is tryp_X-54b12.q1c-rev_assembly and 04-B is Tb10.v4.0061.

4.5.9 Diverse mosaics were constructed from the same donors

Mosaic donors combine in diverse ways to yield considerable expressed NTD variation amongst members of the set (see section 4.5.10). Within an infection, the mosaic sets have between 0.024 and 0.219 differences/aa across the expressed NTDs of their least related members.

Are the expressed mosaics of a mosaic set more diverse than the silent donors that create them? To investigate this question, the 'expressed' (that is, excluding the sequence encoding the signal peptide) NTD aa variability between mosaic donors[70] were compared with the expressed NTD aa variability between the most divergent mosaics that they form. The data is included in Table 4.5 above. For all sets, mosaic donors were themselves more divergent than the expressed mosaics they generate. This is perhaps not surprising, considering that more variant, degenerate parts of donors may be the regions preventing their full-length expression. The implications of this finding are discussed in Chapter 6 (section 6.4.5).

4.5.10 Diverse mosaics appeared within an infection

To examine more closely the mosaics present in an infection, Set_14 mosaics were studied in greater detail. This set was selected for in-depth analysis as it was detected in many infections (five here, plus one in the pilot study), and in three of these infections it was detected at many timepoints. It is therefore a regular participant in TREU 927/4 infections. Some of its clones were verified experimentally by PCR. Finally, Set_14 mosaics had the highest levels of within-set variation, and are thus an interesting example of the degree of variation that can be generated by segmental gene conversion.

Set_14 mosaics were apparently constructed from four different donors: Tb11.09.0005 (14-A), Tb11.13.0003 (14-B), a read assembly constructed from tryp_IXa-24h09.q1c-rev, tryp_XI-1034g11.p1k and tryp_XI-1034g11.q1k-rev (tryp_XI-1034g11_assembly, 14-C), and Tb10.v4.0009 (14-D). These donors have at most 0.269 differences/nt across their whole length. To investigate whether

[70] Donor damage was corrected for this analysis, see footnote 67.

there were further unknown genomic copies that might have contributed to the mosaic set, a probe made from the 14-A NTD-encoding region was hybridized to a dot blot of Set_14 mosaics and a Southern blot of ten samples of pre-infection TREU 927/4 genomic DNA digested with one of ten different restriction enzymes. The results of these hybridizations are shown in Figure 4.21 and Figure 4.22. The dot blot shows that the probe was able to bind to all of the Set_14 mosaics, making it likely that it would identify all related donors in the gDNA. On the Southern blot, bands corresponding to 14-A, 14-B and 14-D could be identified by reference to the location of restriction enzyme sites in the published genome sequence. There were two further bands identified in many of the reactions, one of which hybridized more strongly to the probe than the other. Given the results of the dot blot, the weaker-hybridizing band (indicated by a ball-and-stick) likely corresponds with the silent copy matching the read assembly, 14-C. This indicates that there was probably a further related genomic *VSG* that has not yet been sequenced. The involvement of a potential further donor in the construction of mosaics is unclear, however, since in general there are few regions of identical variation in different infections (with the exception of that referred to and experimentally tested in section 4.5.2 above), and even 14-D did not appear to contribute significantly to the mosaic set.[71]

Within an infection, these donors joined in varied ways to generate a wide variety of different mosaics. This was seen most extensively in the two infections studied in greater depth, mouse 04 and mouse 05. Set_14 variability is shown in Figure 4.23.

[71] Further Southern hybridizations were performed following submission of this thesis. These experiments resolved the question of the additional bands in Figure 4.22. See Figure 7.38 (Appendix 7.4.3) for further details.

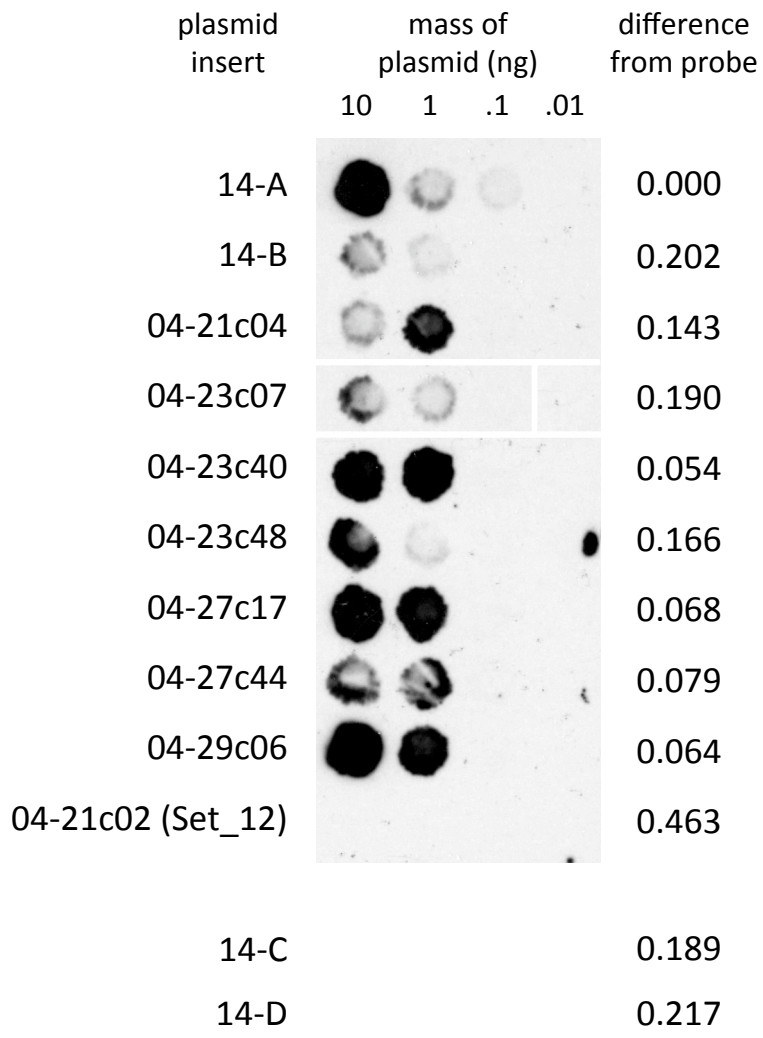


Figure 4.21 Dot blot of related mosaic DNA. The blot was prepared using isolated plasmid as described in Chapter 2, hybridized with a probe complementary to 14-A, and washed to high stringency (70ºC, buffer prepared according to kit instructions). For each sequence, the number of nucleic acid differences between the probe sequence and the sequences of the blotted DNA per base is indicated. Data are also included for donors 14-C and 14-D, but as these sequences were not cloned into a plasmid they could not be included on the blot.

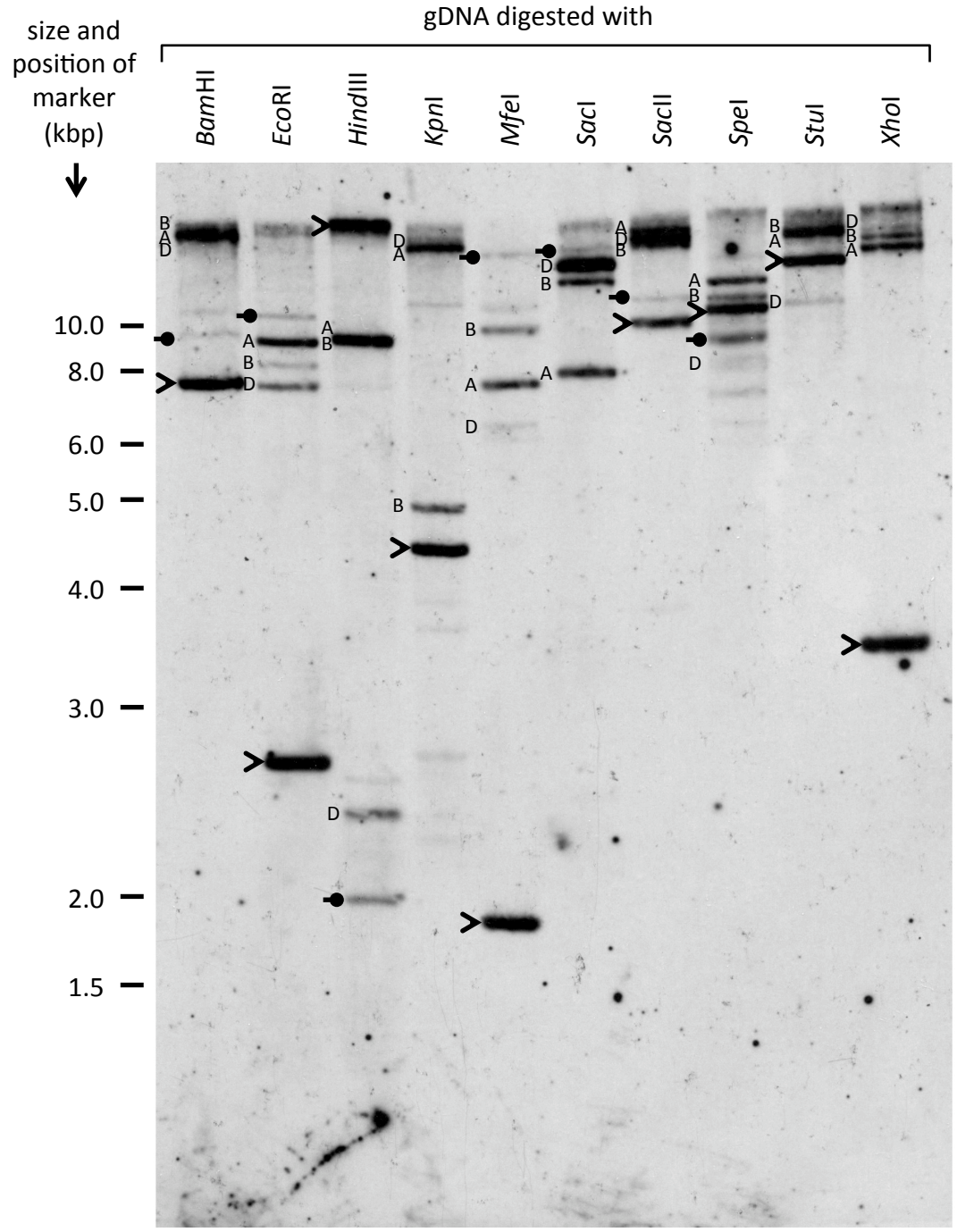


Figure 4.22 Southern blot of gDNA. The blot was prepared using pre-infection 927 genomic DNA digested with one of ten restriction enzymes as described in Chapter 2, hybridized with a probe complementary to 14-A and washed to high stringency (70ºC, buffer prepared according to kit instructions). Bands corresponding with the predicted sizes of each of the donors identified in the genome (14-A, 14-B and 14-D) are indicated. Two additional bands could be identified in many lanes, one binding the probe more strongly than the other (indicated by an arrowhead and a ball-and-stick respectively). Set_14 mosaics were not detected in pre-infection cDNA. Additional bands seen in the samples digested with *Kpn*I and *Spe*I could be due to star activity as the reaction was not carried out with high-fidelity versions of the enzyme. Note that this figure has been superseded by a technically superior one: Figure 7.38 (Appendix 7.4.3).

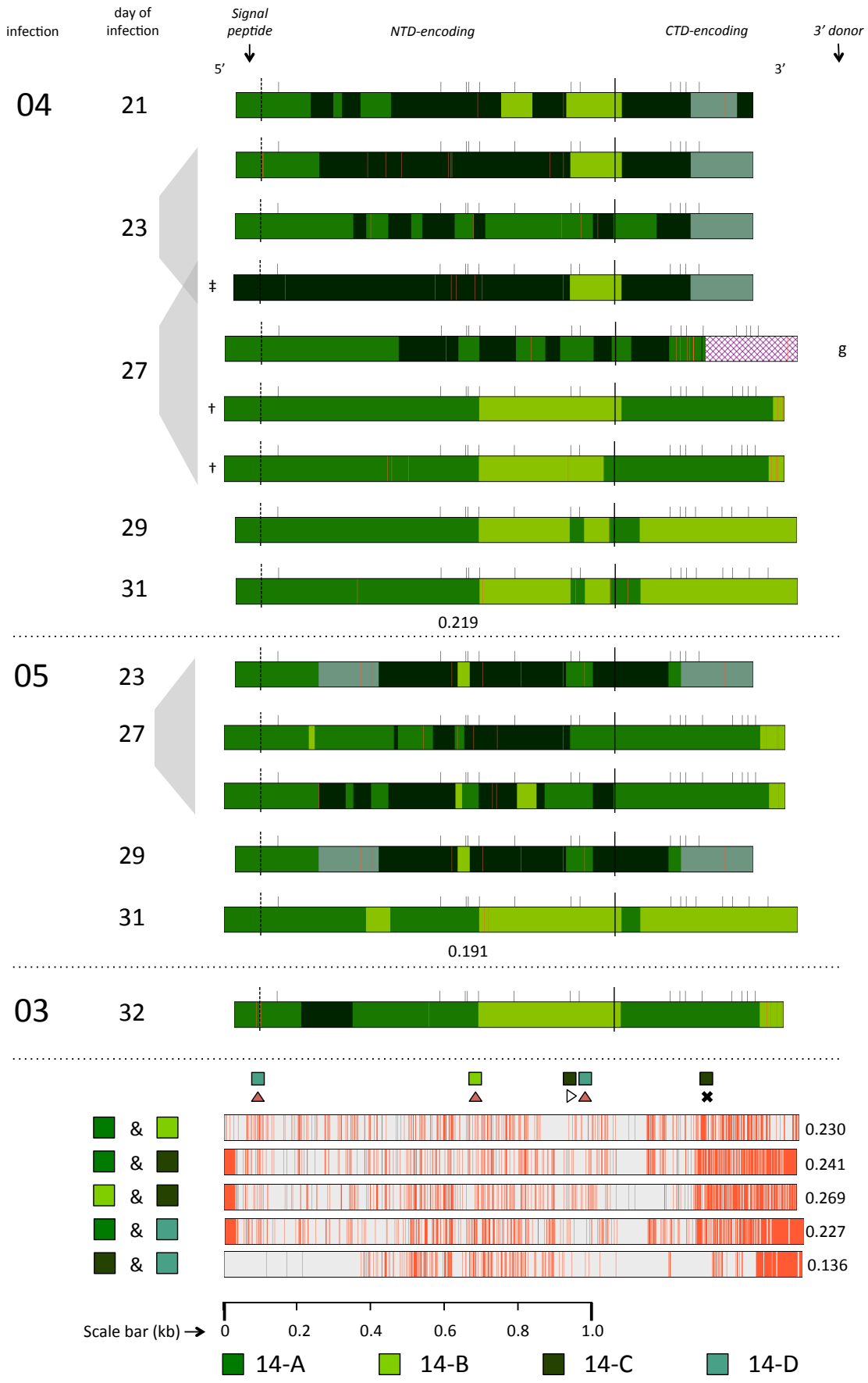


Figure 4.23 Set_14 mosaics. Diagram is arranged and annotated similar to Figure 4.13. Variation in the length of the CTD-encoding region means that this part of the diagram is not necessarily aligned with the donor comparisons. 14-B had no junctions with 14-D, so this comparison has not been included. 14-A is Tb11.09.0005, 14-B is Tb11.13.0003, 14-C is tryp_XI-1034g11_assembly and 14-D is Tb10.v4.0009.

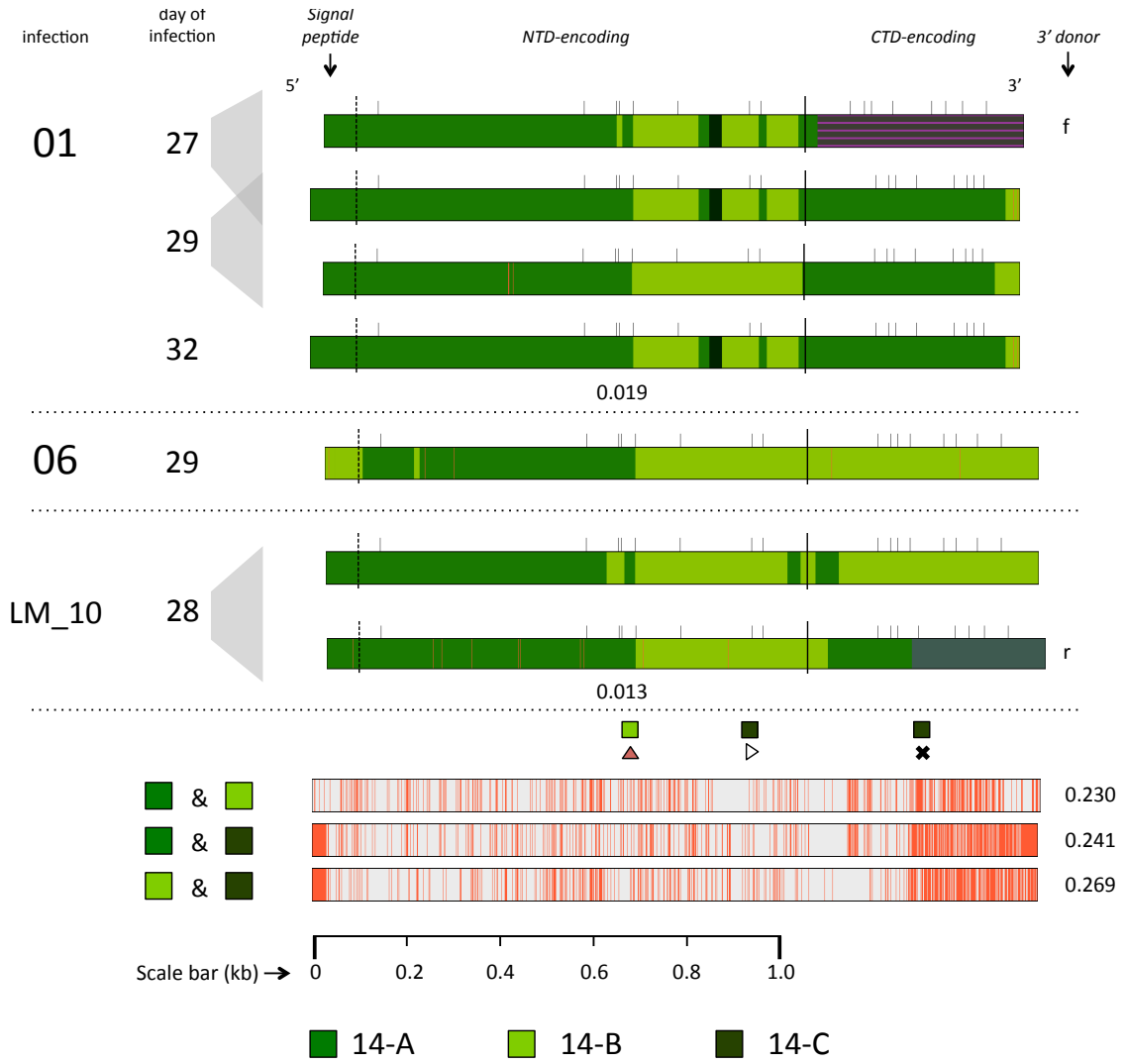


Figure 4.23 (continued). 14-D comparisons are not shown, as 14-D does not contribute to the NTD-encoding regions in these mosaic variants.

4.5.10.1 The development of Set_14 mosaics: later stages

First, consider the later-stage mosaics of this family: those identified from day 27 in six infections. These are the mosaics possessing the conserved mosaic junction identified in section 4.5.7. Marcello & Barry (2007a) speculated that the Set_14 mosaic lineage identified in that pilot study began as a full-length copy of 14-A, which acquired a segment from 14-B to form an intermediate mosaic. Such a 'mosaic intermediate' was found in two clones in mouse 04, indicated by '†' in Figure 4.23. As with the mosaics identified in the pilot study, further involvement by 14-B appears to develop this lineage further in mouse 04 to give those variants that are present at days 29 and 31. Similar mosaic intermediates may have occurred undetected in mouse 01, mouse 03, mouse 05 and mouse 06, and led to the Set_14 mosaics in those populations, although the

Set_14 mosaic identified in mouse 06 hints at a different route whereby the 3' boundary of conversion of the inserted 14-B segment is 3' of the coding sequence. The expression of full-length 14-A was not detected; indeed, it is possible that the atypical nature of its GPI anchor signal sequence precludes its full length expression (the 3' end of 14-A was not seen in any Set_14 *VSG* clones).

4.5.10.2 The development of Set_14 mosaics: early stages

In mouse 04 and mouse 05, an earlier stage in the development of Set_14 mosaics was found, examples of which can be seen at days 21 and 23. This stage probably began with an expressed gene similar to 14-C, such as that indicated by '‡' in Figure 4.23, which overcame its pseudogenicity by mosaicism in its CTD with 14-D and either 14-B (mouse 04) or 14-A (mouse 05). The expressed *VSG* probably underwent various gene conversions with all three of the other donors to yield the apparently more complex early Set_14 mosaic *VSG*s. It is not known whether early-stage Set_14 mosaics were present in the other Set_14-expressing infections, or whether early-stage Set_14 mosaic expression necessarily preceded the later form.

One explanation for the apparently more complex Set_14 mosaics occurring earlier in infection than the apparently more simple Set_14 mosaics is that the interactions between donors required to generate the later-appearing mosaics may occur at a region where the donors do not so readily recombine. As described in section 4.5.7 the mosaic junction leading to the later-stage Set_14 mosaics is strictly conserved between infections, suggesting that there may be restrictions on the ability of these two donors to directly interact. Therefore, whilst superficially appearing simpler, the likelihood of the 'intermediate mosaic' forming may be lower than the apparently more 'complex' mosaic interactions that give the earlier-stage Set_14 mosaics.

4.5.10.3 The development of Set_14 mosaics: summary

Bringing this all together, a more complex hypothesis for how Set_14 mosaics have been constructed is described in Figure 4.24 and Figure 4.25 for mouse 04 and mouse 05 respectively.

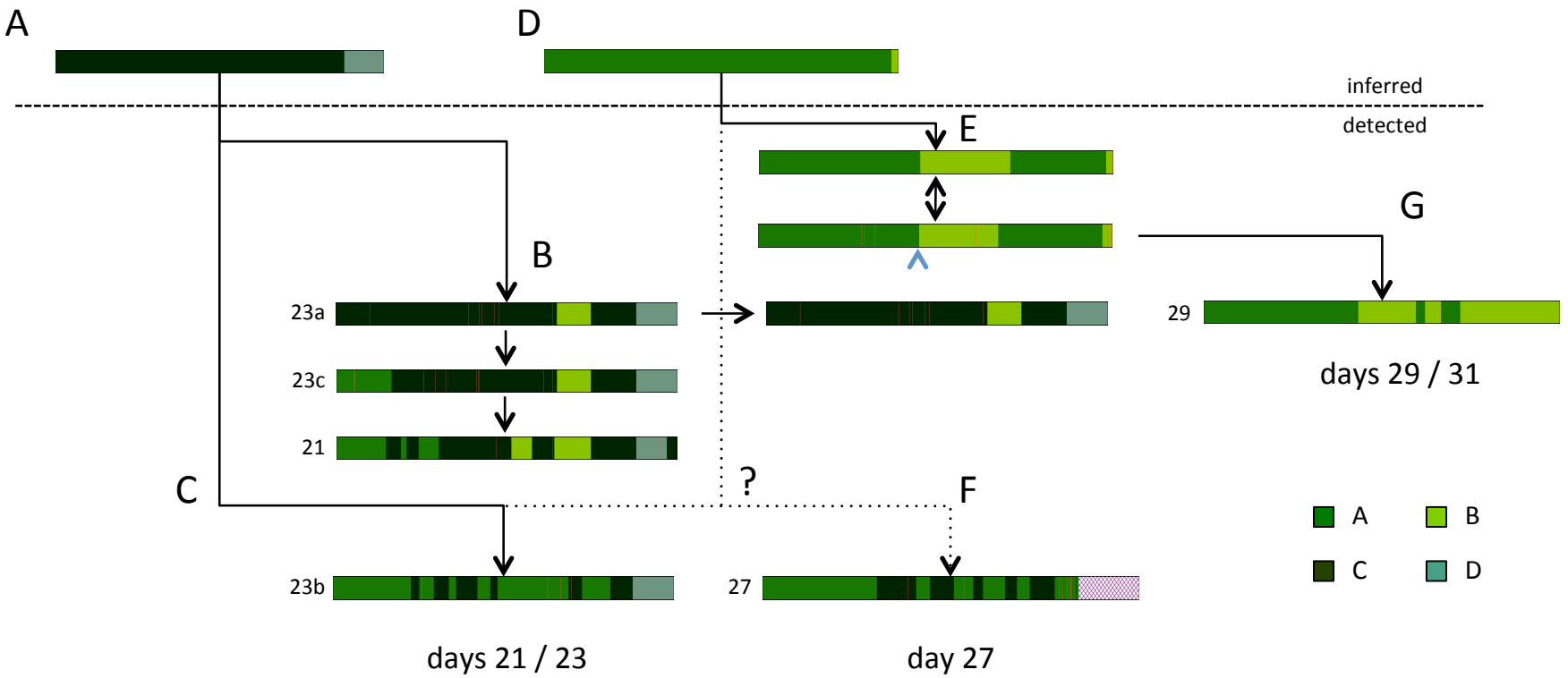


Figure 4.24 Development of Set_14 mosaics in mouse 04. Day of expression of selected variants is included on the left hand side of the variant. (A) The early stages were probably initiated by a copy of 14-C undergoing segmental conversion in its 3' end to overcome C-terminal pseudogenicity. This variant was not detected in any samples, but the common 14-C/14-D CTD was seen in many clones at day 23. (B) The inferred progenitor accumulated segments of 14-A and 14-B. The detection of the more complex variant at day 21 (before the others) indicates that this progression probably occurred rapidly in a subpopulation. The earlier variant was still detectable at day 27. (C) The inferred progenitor also acquired multiple segments from 14-A in a process that was probably independent from (B). (D) The later stage of the mosaic development was probably initiated by a full-length copy of 14-A, which acquired a segment from 14-B to produce the conserved mosaic junction (blue vertical arrowhead) and the variants present at day 27 (E). (F) Meanwhile, other interactions between 14-A, 14-C and a 3' donor produced another variant at day 27. This variant could have been produced either from development of the 14-D-like progenitor, the 14-A like progenitor, or possibly interactions between the two. (G) The 14-A/B variants at (E) undergo further interactions with 14-A and 14-B to produce the later-stage variant, which was the only form detected at days 29 and 31 of infection.

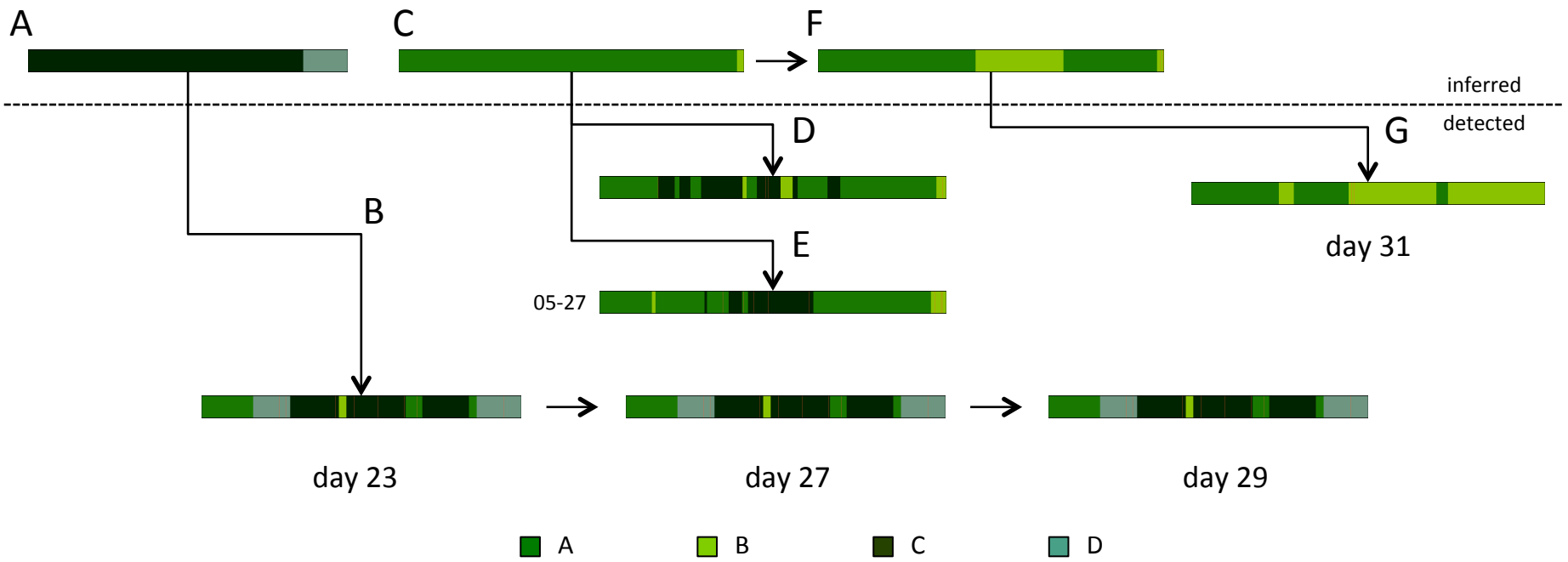


Figure 4.25 Development of Set_14 mosaics in mouse 05. Mouse 05 saw a smaller group of variants than mouse 04, but their progression can fit a similar pattern. (A) Again, the accumulation of many segments by a repaired 14-C could explain the mosaic variant (B) that persists between days 23 and 29 in this infection. (C) A repaired 14-A could have led to the two variants detected at day 27, each of which (D, E) could have each developed by the independent contribution of 14-C segments to the NTD of 14-A (and, to a lesser extent, contribution of 14-B segments). The repaired 14-A could have produced a mouse 04-like 'mosaic intermediate' undetected in this infection (F), which accumulated further segments from 14-B to become the later variant seen at day 31 (G).

4.5.11 Could mosaicism contribute to antigenic variation?

Mosaicism introduced considerable variation into the NTD of VSGs during an infection, by enabling partial expression of damaged *VSG*s and by combining donor sequences in different ways. Abundant variants produced from the same set of donors were identified over the course of infection. Could mosaic variation be sufficient to change epitope structure and therefore contribute to immune evasion directly? Addressing this question bioinformatically is complicated by the fact that approaches to assess positive selection are sensitive to the presence of recombination (Kosakovsky Pond *et al.*, 2006). The fact that the donors of identified mosaics shared substantial identity with one another could potentially limit the degree of antigenic variability they could introduce. Nevertheless, for some sets, the variation introduced rapidly into the NTD by mosaicism approaches levels associated with antigenic distinctiveness (Barbet *et al.*, 1989; Pays *et al.*, 1985a). Therefore, as described in the following chapter, the ability of segmental gene conversion to change VSG epitope structure was addressed experimentally.

4.6 Discussion

4.6.1 VSG*s have multiple levels of variation in an infection*

Chapter 3 revealed that, within an infection, a range of different antigen sets could be detected over time, representing the switching between different genomic *VSG*s typical of *T. brucei* antigenic variation. Looking at those sets of antigens in more detail, we uncover a further level of diversity amongst expressed *VSG*s. Related expressed *VSG*s often vary from one another, and from their best-matching genomic copy. This diversity takes the form of point mutation, 3' donation, and mosaicism. The bioinformatic and experimental analyses of the different patterns of diversity presented in this chapter allow us to draw a number of conclusions about the extent of these different processes, and their potential roles in infection.

4.6.2 Key findings of this chapter

4.6.2.1 Diversity amongst expressed *VSG*s is generated rapidly

T. brucei does not possess huge families of closely related *VSG* genes (Marcello & Barry, 2007a), so most of the variability amongst sequence-related, expressed *VSG*s appearing in each infection must have been produced in the period since the parasites were last cloned. For most of the infections, this period was ~72 generations. The extensive variability seen amongst mosaics in the clonal infections is consistent with a model whereby diverse related mosaics and 3' donation events routinely occur over the course of a 4–5 week infection.

4.6.2.2 The contribution of point mutations to antigenic variability within an infection is probably minor

Point mutations were detected in expressed *VSG* at a rate greater than that expected from experimental artefact. However, they did not vary expressed *VSG*s extensively, and their effects—changing single amino acids at a time—were minor when compared with other patterns of variation occurring alongside them in the parasite population (such as between-set variation covered in Chapter 3, and segmental gene conversion considered in sections 4.4 and 4.5). Therefore, in the context of a single infection, the selective advantages bestowed by point mutation are likely to be negligible. Their potential role in the biology of African trypanosome antigenic variation is likely to be longer-term.

4.6.2.3 The 3' end of an expressed *VSG* is often acquired from a different donor

Most sets showed some evidence of variation amongst their members at their 3' ends, which in many cases was considerable. Where they were identified, most of the 3' ends that were provided by a different donor came from likely telomere-proximal *VSG*s, indicating that the expressed *VSG* probably acquired its 3' end as it was copied into a telomere-proximal locus. 3' donation allowed the expression of damaged *VSG*s and could also recombine different types of CTD. The variation amongst sequence-related, expressed *VSG* in their 3' ends observed during an infection is consistent with the polyclonal activation of *VSG*s, in which each event generates a distinct expressed copy (Michels *et al.*, 1983;

Timmers *et al.*, 1987), although the possibility that variation accumulated in the 3' end of the expressed *VSG*s cannot be ruled out.

4.6.2.4 A diverse range of mosaic *VSG*s is frequently observed within an infection

Mosaicism was prevalent, particularly at later timepoints. Short of switching *VSG* NTDs completely, mosaicism was the most powerful means of introducing variation into the exposed part of the antigen. Complex mosaics developed rapidly, during an infection. A restricted palette of genomic *VSG*s could generate considerable diversity amongst expressed *VSG*s: from four silent donors eight distinct NTD variants were witnessed within an infection. Mosaicism allowed the contribution of damaged *VSG*s, but the variability amongst mosaics identified within an infection—even within a sample from that infection—hints at another possible role for mosaicism.

4.6.3 How might these findings have differed in other experimental models?

The mechanisms generating within-set variation are almost certainly parasite-intrinsic, so the patterns of variation observed during these infections are not limited to this experimental model *per se*. However, the strength of selection by the immune system and the total parasite population size could affect the degree of variation that is observed amongst expressed *VSG*s. Considering the results of Chapter 3 and previous studies (MacGregor *et al.*, 2011) differentiation was probably the dominant factor in these infections. Where the dynamics of infection are controlled primarily by the immune system, variation that does not change epitope structure will be disfavoured. As a consequence, variation produced by differences in 3' donation, as well as many point mutations and instances of mild mosaicism, may become harder to detect, while the greater selective advantage of more dramatic segmental gene conversion events in the NTD-encoding region might make such mosaics more pronounced.

Longer infections where there is greater selection for antigenic novelty and a larger parasite population may facilitate the development of more complex antigenically distinct mosaic *VSG*s. Such circumstances may also break down the apparent requirement for homologous mosaic donors, revealing rare variants

that have undergone non-homologous segmental gene conversion events in their NTD.

4.6.4 Concluding remarks

In many cases, the expression of *VSG*s appears to go hand-in-hand with the generation of diversity, as *VSG*s acquire point mutations, different 3' ends or segments from related genes. Can such variation vary epitope structure? Point mutations are unlikely to vary VSGs rapidly enough to enable immune evasion. 3' donation accesses genomically-encoded NTD variability, greatly increasing the antigenic variability available to the trypanosome, but does not itself vary the epitope-bearing NTD. Could mosaicism in the NTD introduce antigenic novelty, given that related mosaics utilize the same group of sequence-related *VSG* donors? To answer this question, the ability of mosaicism to change VSG epitope structure was addressed experimentally. The results are described in the following chapter.

Chapter 5

The epitope structure and antigenic function of mosaic VSGs

5 The epitope structure and antigenic function of mosaic VSGs

5.1 Introduction

5.1.1 Can the products of mosaic VSG*s contribute to antigenic variation during an infection?*

As seen in the previous chapter, a variety of related mosaic *VSG*s could be detected within a single infection. Were it the case that mosaic construction was a one-off event, occurring at low frequency and simply providing occasional access to damaged silent *VSG*s, we would not expect to see many manifestations of that repaired gene. The diversity and abundance of related mosaics existing in an infection suggest instead that segmental gene conversion leading to mosaic *VSG* expression is a dynamic, on-going process. However, it may simply be the case that once an expression-linked copy of a *VSG* is created it is especially vulnerable to recombination with other, highly related *VSG*s, resulting in little or no impact on its antigenic profile. Mosaic *VSG* donors have high levels of sequence identity with one another, suggesting that a restricted set of donor genes is favoured in the creation of mosaics. This narrow palette may not be sufficient to generate antigenic novelty in the VSG coat within the course of a 4-5 week infection.

Is mosaic *VSG* construction capable of generating new epitopes, or introducing new epitopes from pseudogenic donors, over the course of an infection? That is, does mosaic *VSG* construction play a direct role in antigenic variation during an infection?

5.1.2 Key hypothesis to be tested

The aim of this chapter was to examine the antigenic profiles of related mosaic *VSG*s occurring within a single infection. The (null) hypothesis to be tested was that an antibody response raised against a surface coat made from the product of any member of a mosaic *VSG* set found within an infection is able to recognize surface coats made from products of any other member of that mosaic *VSG* set found within that infection. That is, within an infection, the products of related mosaic *VSG*s are antigenically cross-reactive.

To test this hypothesis, a representative mosaic *VSG* set from one infection was selected and studied in depth. Of all the mosaic *VSG*s observed in chronic mouse infections to date, Set_14 mosaics are the best characterized, both experimentally and bioinformatically. Set_14 mosaic cDNA clones were retrieved from six out of 15 chronic mouse infections analysed[72] (see Chapters 3 and 4 for further details). From a single infection (mouse 04), eight distinct Set_14 mosaic cDNA clones were retrieved, representing a diverse range of related mosaics. The presence of these mosaics spanned ten days in this infection, which normally would exceed the time for specific antibodies to arise. The diverse and prevalent Set_14 mosaic set, summarized in Figure 5.1 and Table 5.1, was therefore deemed to be the best available to test the null hypothesis outlined above.

Of the Set_14 mosaic *VSG*s found in mouse 04, six were selected for testing: 04-21c04, 04-23c07, 04-23c40, 04-23c48, 04-27c44 and 04-29c06. For brevity and clarity, these clones are referred to as variants '21', '23a', '23b', '23c', '27' and '29' respectively. In addition, to investigate the epitope structures of Set_14 mosaics further, a related mosaic from mouse 05, was studied: 05-27c09, referred to as variant '05-27'. Diagrams representing each of these mosaics are indicated in Figure 5.1 by a '*'. All seven of these *VSG*s have been formed by the segmental combination of four putative donors. The mature NTD amino acid variability, and the full-length nucleic acid variability, between the different mosaics are indicated in Table 5.1.

Note added in proof: experiments conducted following submission of this thesis revise some of the results and analyses presented in this chapter. The text here has been left uncorrected. Although the overall conclusions of this chapter remain the same, the reader is urged to consult Appendix 7.4.4 for further details.

[72] Five of the 11 infections presented here, plus one out of four infections to progress into the chronic stage previously reported by Marcello & Barry (2007).

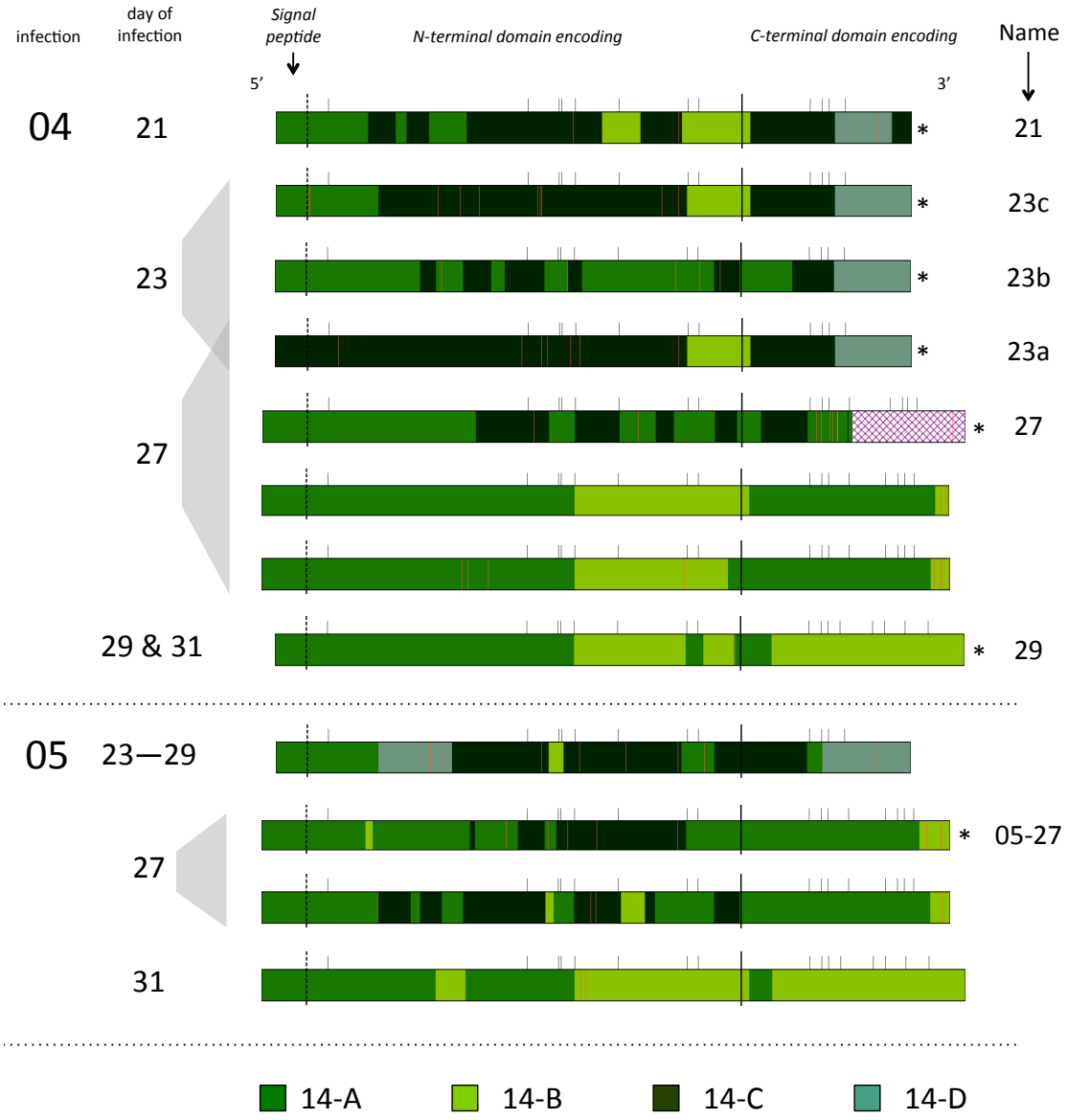


Figure 5.1 Twelve diverse Set_14 mosaic *VSG*s identified in mouse 04 and mouse 05. In diagrams representing cDNA sequences, a black bar indicates the boundary between N- and CTD-encoding regions. A dotted line indicates the division between the N-terminal signal peptide-encoding and mature NTD-encoding regions. Cysteine codons are represented by a bar projecting above the diagram. Segments donated by the putative donors 14-A, 14-B, 14-C and 14-D are coloured in different shades of green, according to the key at the bottom of the image. The mouse and day of infection where each of these mosaic variants was identified is shown on the left hand side of the Figure. '∗' indicates those mosaics investigated in this chapter. The names referring to each of these mosaics are referred to are given on the right hand side, next to each cDNA diagram.

		Set_14 mosaic variant							Set_14 mosaic donors			
		21	**23a**	**23b**	**23c**	**27**	**29**	**05-27**	**14-A**	**14-B**	**14-C**	**14-D**
Mosaics	**21**	-	0.087	0.153	0.072	0.100	0.125	0.100	0.197	0.207	**0.116**	0.217
	23a	0.061	-	0.197	0.047	0.150	0.209	0.138	0.253	0.223	**0.041**	0.239
	23b	0.091	0.115	-	0.181	0.078	0.131	0.109	**0.075**	0.260	0.181	0.233
	23c	0.042	0.030	0.096	-	0.134	0.191	0.122	0.231	0.238	**0.078**	0.258
	27	0.120	0.149	0.110	0.130	-	0.138	0.081	**0.109**	0.276	0.134	0.236
	29	0.141	0.192	0.146	0.177	0.146	-	0.128	**0.081**	0.169	0.231	0.230
	05-27	0.128	0.146	0.120	0.128	0.092	0.152	-	**0.125**	0.263	0.128	0.220
Donors	**14-A**	0.168	0.197	**0.094**	0.181	**0.106**	0.134	**0.065**	-	0.251	0.257	0.268
	14-B	0.190	0.196	0.229	0.212	0.226	**0.095**	0.227	0.229	-	0.244	0.245
	14-C	**0.083**	**0.036**	0.124	**0.060**	0.172	0.220	0.173	0.222	0.233	-	0.214
	14-D	0.135	0.131	0.150	0.154	0.207	0.210	0.205	0.218	0.222	0.133	-

Table 5.1 The Set_14 mosaic *VSG*s under investigation were compared with one another and with their putative donors. For each comparison, the mature NTD amino acid variability (top right half of the matrix) and full-length nucleic acid variability (bottom left half of the matrix) are given in differences/aa and differences/nt respectively. For the expressed mosaics, variabilities of less than 0.1 are highlighted in red. The best matching donor (that is, the lowest variability between an expressed mosaic and a donor) is indicated, for each mosaic, in bold.

5.2 Experimental approach

5.2.1 Overview of experimental design

VSGs possess a characteristic tertiary structure thought to be critical to their function (Blum *et al.*, 1993), are post-translationally modified by N-glycosylation (Mehlert *et al.*, 2002), and their close packing in the complete surface coat shields their membrane-proximal end from access by antibodies and other immune effectors (Schwede *et al.*, 2011). The intact surface coat may also induce different qualities of immune response compared with inoculation with purified protein coupled with adjuvant (Clarke *et al.*, 1987; Miller *et al.*, 1984b; Theodos *et al.*, 1990). Therefore, rather than exogenously expressing tagged versions of the products of mosaic *VSG*s in an established bacterial protein expression system, each of the *VSG*s under investigation was expressed as an intact surface coat by a laboratory-adapted strain of *Trypanosoma brucei*.

Transgenic trypanosomes expressing different Set_14 VSG coats were used to initiate infections of mice, which were cured by chemotherapy at the first parasitaemic peak. This elicited an antibody response against that specific mosaic variant. Antibody-containing plasma was harvested, adding to a panel of polyclonal antibody responses directed against each variant. The ability of each antibody response to bind and neutralize different Set_14 VSG surface coats was assayed on live trypanosomes by immunofluorescence, complement mediated lysis, and agglutination. In addition, several monoclonal antibody (mAb)-producing hybridoma lines were cloned, to further dissect the immune response, and specifically identify epitopes shared amongst related mosaics.

5.2.2 Mosaics were expressed using a two-plasmid system

To generate a trypanosome line stably expressing a mosaic VSG as its surface coat, a two-plasmid system was used. Previously, Smith et al. (2009) used plasmid p221_PUR117VSG_UTR to target VSG 427-4 (VSG117, see footnote 73) to

[73] Throughout this Chapter, the Lister 427 VSGs are referred to by their 'preferred name' (as documented by 'Lister 427 *VSG* Summary', G. Cross, pers. comm.), with the word 'Lister' omitted for brevity. Their 'local name' is also provided where they are introduced. For reference, the four Lister 427 VSGs referred to in this Chapter are Lister 427-2 (VSG221), Lister 427-4 (VSG117), Lister 427-6 (VSG121) and Lister 427-9 (VSGVO2 or 17.23).

the promoter-proximal end of bloodstream expression site (BES) 1 of the Lister 427 strain of *T. brucei*, yielding clones expressing two *VSG*s from the same expression site. A further round of transfection with plasmid pBS_VSG221KO replaced the endogenous telomere-proximal *VSG* at that expression site, VSG 427-2 (VSG221), with a selectable marker (G. Rudenko, Imperial College London, pers. comm.), resulting in clones expressing just the exogenous, promoter-proximal *VSG* under double drug selection.

Plasmid p221_PUR117VSG_UTR, kindly gifted by G. Rudenko, was modified to allow expression of the Set_14 mosaic *VSG*s. First, the VSG 427-4 gene was replaced with a multiple cloning site containing recognition sites for the restriction endonuclease *Sbf*I as described in Chapter 2 (section 2.4.5). This modified p221_PUR117VSG_UTR was relabelled 'pVSG' for brevity. Each of the Set_14 mosaic *VSG*s to be investigated was amplified from its respective cloning plasmid using primers that added *Sbf*I sites to the amplicon[74], was purified on a 1% agarose gel, and was cloned into the *Sbf*I site of pVSG. Clones were sequenced using primers pVSG_insert_F and pVSG_insert_R (see Appendix 7.1) to ensure correct orientation and that no errors had been introduced. Figure 5.2 shows a diagram of the modified plasmids and the transfection strategy. Plasmids were linearized using *Not*I and *Xho*I for transfection.

Lister 427, a laboratory-adapted trypanosome line, was used as the background for these experiments. Not only is it easily cultured and genetically modified, but it also has a repressed rate of *VSG* switching compared with more animal-adapted strains: measured to be as low as 10^{-7} per cell per generation (McCulloch *et al.*, 1997). In addition, the dual plasmid system outlined above had already been used successfully with this line (Smith *et al.*, 2009). Specifically, an RNAi-capable line (13-90) was used (Wirtz *et al.*, 1999), in case future experiments were to investigate the relationship between surface coats constructed from mosaic VSGs and other cellular processes.

[74] The entire *VSG* insert was amplified from the cloning plasmid and cloned into pVSG. Each cloned sequence therefore included 17 bp of the spliced leader sequence at its 5' end, and a portion of the 3' UTR (up to and including the 16-mer) at its 3' end (these portions were not present in the VSG 427-4 in the unmodified p221_PUR117VSG_UTR plasmid). The additional sequence at the 5' end of the *VSG* did not introduce any open reading frames. The Set_14 mosaic from day 27 (04-27c44) possessed an atypical GPI anchor signal sequence—possibly due to a frameshift at its extreme 3' end—that was not corrected here. The results described in the rest of this Chapter indicate that this atypical GPI anchor signal sequence was processed as normal.

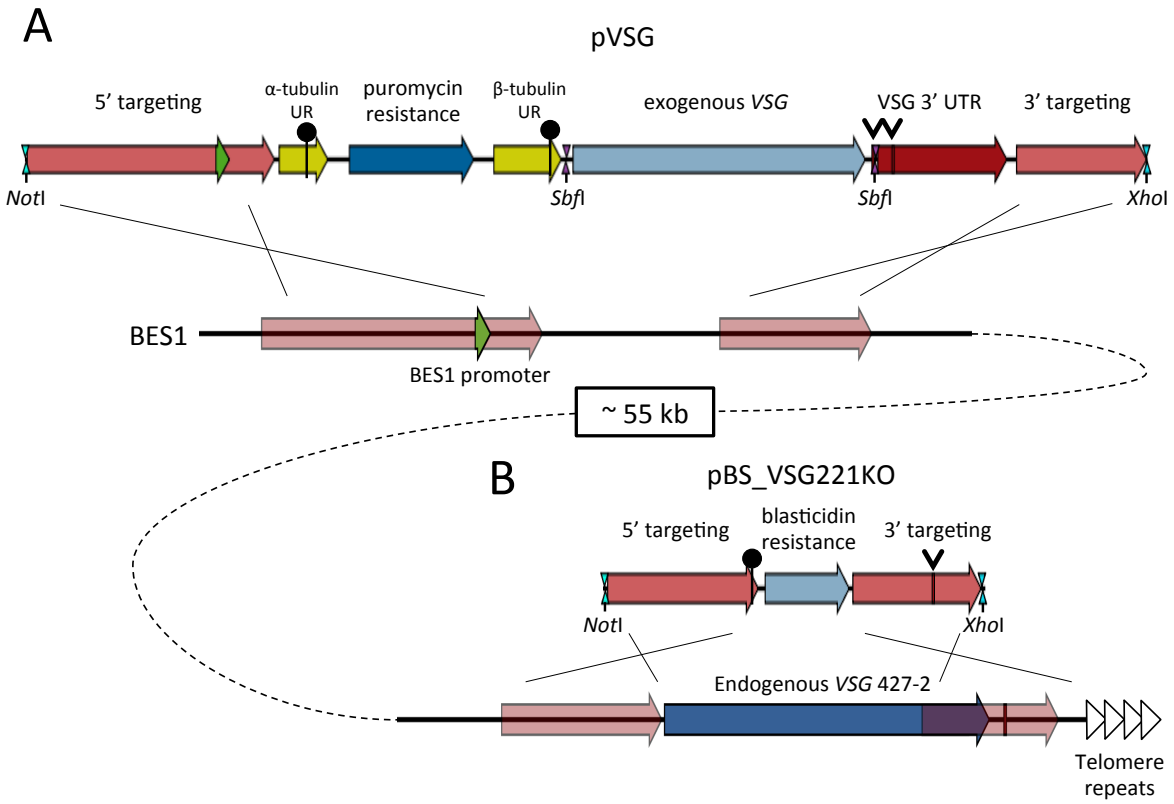


Figure 5.2 Plasmids and transfection strategy. Bloodstream expression site 1 (BES1) is shown. (A) First, transfection with pVSG linearized by *Not*I and *Xho*I causes integration of the exogenous *VSG* at the promoter proximal end of BES1. Crosses indicate the region in which homologous recombination occurs. Selection of clones is achieved with puromycin. (B) A second round of transfection with pBS_VSG221KO linearized by *Not*I and *Xho*I replaces the endogenous VSG 427-2 with a resistance marker, and selection of clones is achieved with blasticidin. Double resistant clones have therefore integrated the exogenous *VSG*, and lost the endogenous *VSG*, in BES1. The acceptor sites for the trans-spliced leader sequence are indicated by black circles, and copies of the 16-mer sequence are indicated by vertical arrowheads.

5.2.3 RT-PCR showed expression of the mosaic VSGs *in Lister 427*

The two-plasmid system was employed to generate trypanosomes expressing just the Set_14 mosaic *VSG*s under investigation. In addition, trypanosomes expressing just VSG 427-4 were generated by the same process (using the unmodified VSG 427-4-containing plasmid p221_PUR117VSG_UTR). The VSG 427-4-expressing trypanosomes were used as a control to assess whether any of the effects observed were due to the artificial nature of the expression system, rather than the expression of a mosaic *VSG*.

After the first round of transfection, the *VSG* expression of clones growing on the appropriate drugs was tested by PCR on cDNA (RT-PCR). All clones showed dual expression of the exogenous *VSG* and the endogenous *VSG* 427-2 (data not

shown). After both rounds of transfection, stabilates of the clones were prepared and *VSG* expression was again tested by RT-PCR. RT-PCR was carried out using primers directed against VSG 427-2, Set_14 mosaic *VSG*s, VSG 427-4, and two further *VSG*s commonly expressed by Lister 427, VSG 427-6 (VSG121) and VSG 427-9 (VSGVO2). The results are shown in Figure 5.3. As can be seen, trypanosomes subjected to transfection with both plasmids were expressing the exogenous *VSG*, and were no longer expressing VSG 427-2. In most cases, no expression of either VSG 427-6 or VSG 427-9 could be detected.[75]

5.2.3.1 Digestion of amplification products showed no other *VSG* transcripts

To exclude the possibility that *VSG*s other than the exogenous, promoter proximal *VSG* were being expressed, all sequences possessing the *VSG*-specific 16-mer in their 3' UTR were amplified from cDNA using *VSG*-specific primers (as in Chapter 3 section 3.3.3). The products were gel-purified, digested with *Eco*RV 'high fidelity' (HF) restriction endonuclease and separated on an agarose gel (Figure 5.4). All the Set_14 mosaic *VSG*s under investigation have an *Eco*RV recognition site at approximately 400 bp. Cut and uncut PCR product was run on a gel. In the cut lane, bands with an approximate size of 1000 bp and 400 bp can be seen. No residual uncut product is seen in the lanes containing product treated with the enzyme. This result indicates that all detectable products possess the *Eco*RV site predicted of the Set_14 mosaics. Three representative mosaic-expressing lines were tested by amplification-digestion: 21, 23a and 29.

[75] The exception to this is the 05-27-expressing trypanosomes, which show product from both the VSG 427-6 and VSG 427-9 reactions. The most likely explanation for this pattern is that theirs has become a meroclonal population: possibly trypanosomes in their population switched as the clone expanded. As the 05-27 clones were obtained only towards the end of the project it was not possible to re-clone them. The fact that the 05-27 trypanosome population may be expressing other *VSG*s alongside the Set_14 mosaic demands that the results obtained with this line must be interpreted carefully.

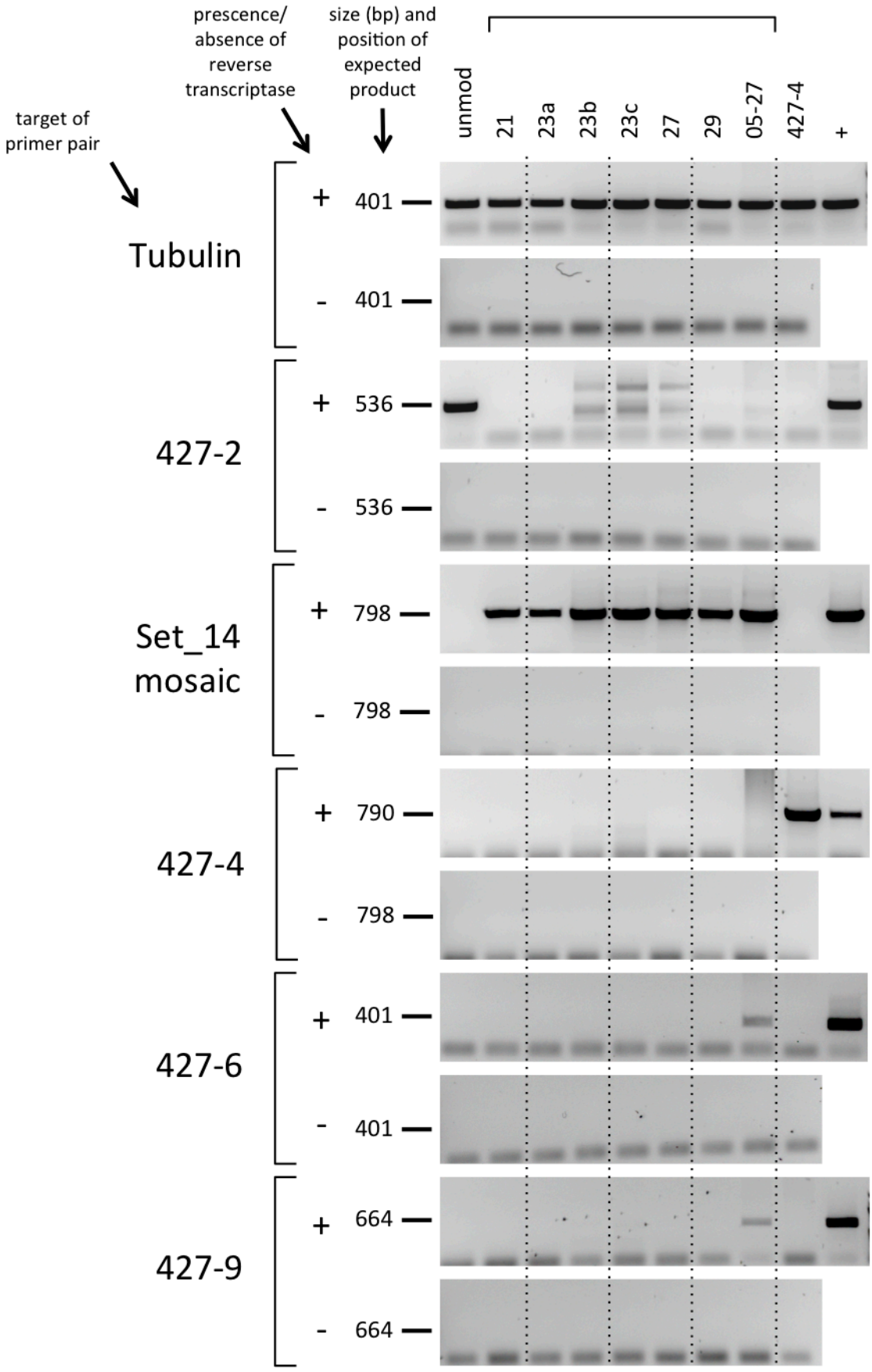


Figure 5.3 Post-cloning RT-PCR. Six primer pairs were used to test expression of different VSGs and tubulin (to control for equal template loading). cDNA from nine lines was tested. Reactions were carried out on total cDNA, as well as the product of a reaction from which reverse transcriptase was absent, as a control for gDNA contamination. The Set_14 mosaic expressors are indicated by a bar above the lane labels. Dotted lines have been placed over the image to aid identification of the cDNA template used for each reaction.

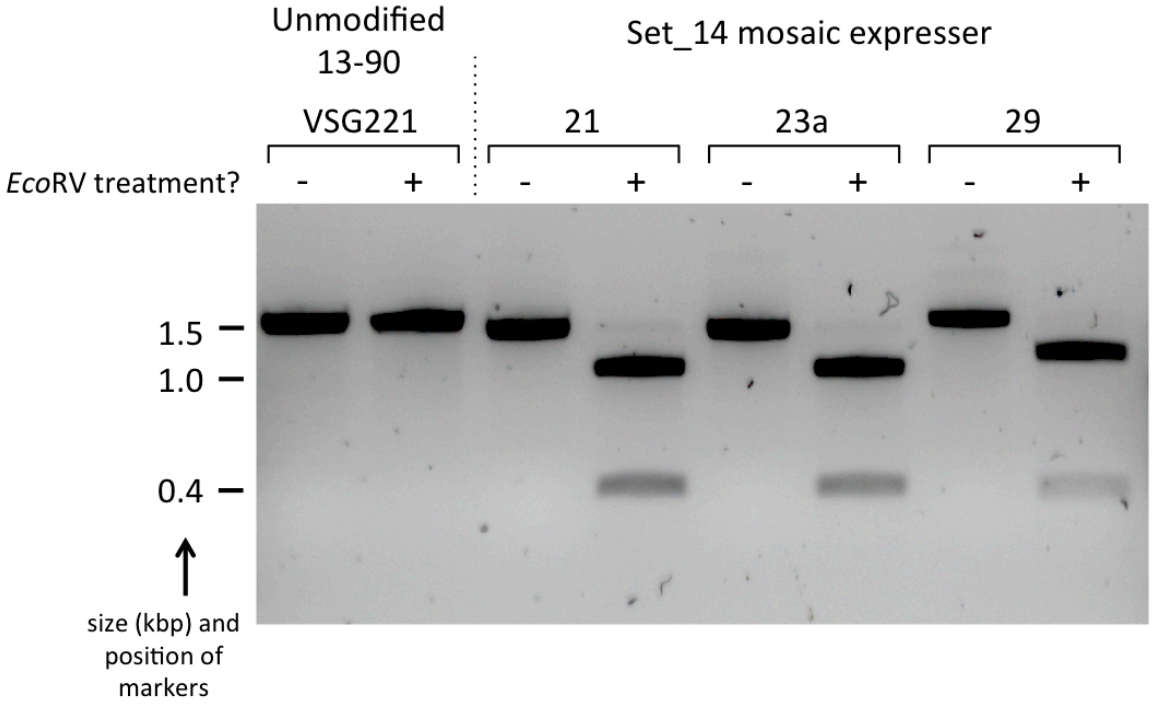


Figure 5.4 Amplification-*Eco*RV restriction digest. cDNA from unmodified trypanosomes, as well as those from lines expressing Set_14 variants 21, 23a and 29, was amplified using primers directed against the spliced leader and the *VSG*-specific 16-mer. The product was gel purified. 100 ng of the purified product was digested with *Eco*RV-HF for 10 hours and separated on a 1% agarose gel alongside 100 ng of undigested product.

5.2.4 SDS-PAGE showed a variant band that was verified by mass spectometry

To ensure that transcripts detected by RT-PCR were being translated into a protein coat, crude cell lysate was run on SDS-PAGE (Figure 5.5). As can be seen, the thick Coomassie-blue-stained band corresponding to VSG 427-2 seen in the untransfected cells was lost in all the modified lines, and replaced with a different-sized band. The distance that each of these variant bands had migrated broadly corresponds with the relative predicted molecular weights of the different VSGs under investigation (Table 5.2), in that 21, 23a, 23b and 23c are all expected to have similar molecular weights to that of each other, whereas 27 and 29 are expected to be ~5 kDa larger. Discrepancies between the predicted molecular weight and the distance that they have migrated, particularly when comparing the Set_14 VSGs with the VSG 427-2 and VSG 427-4 controls, could be due to the extent of glycosylation in these different VSGs: as shown in Table 5.2, the Set_14 VSGs under consideration all have more Asn-Xaa-Ser/Thr potential glycosylation sites than VSGs 427-2 or 427-4.

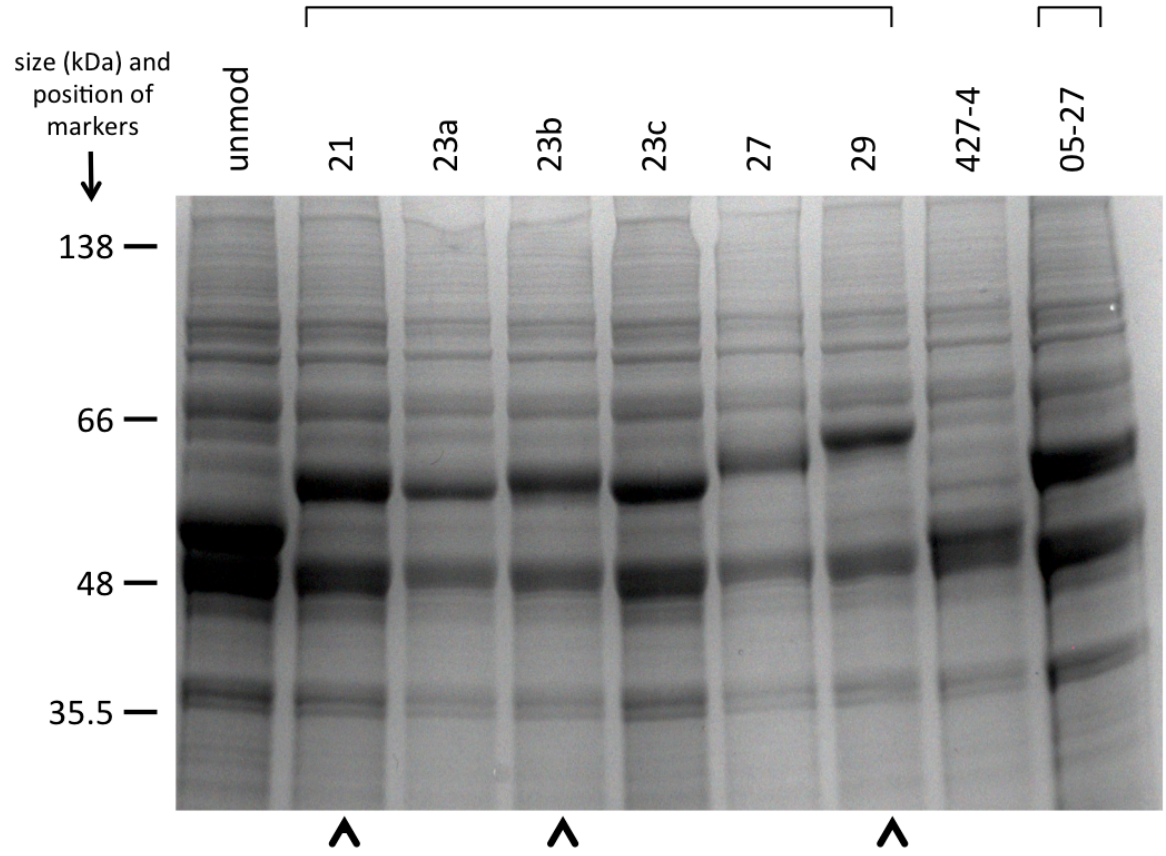


Figure 5.5 SDS-PAGE. Crude cell lysate from 5 x 10^6 cultured cells was separated on a 10% Bis-Tris acrylamide gel. Lysates from the Set_14 mosaic expressors are indicated by a bracket above the lane labels, 'unmod' indicates the unmodified 427-2-expressing line. The variant band lies between the weight markers corresponding with 48 and 66 kDa. Arrowheads mark those lanes for which the variant band was excised and examined by mass spectrometry.

	VSG	Mass of mature protein (kDa)	# of NXS/T sites
	427-2	46.2	2
Set_14 variants	**21**	45.6	4
	23a	45.7	4
	23b	45.3	4
	23c	45.7	4
	27	48.6	3
	29	49.3	4
	05-27	47.8	3
	427-4	50.1	1

Table 5.2 Expected mass of VSGs. The predicted molecular weights of the different mature VSGs were calculated by cropping the predicted N-terminal and GPI anchor signal sequences from the amino acid sequences, and calculating the molecular weight of the remaining sequence using CLC Genomics Workbench. The number of putative glycosylation sites was found by searching the sequence for asparagines followed by any amino acid, followed by either serine or threonine.

For three of the Set_14 mosaic expressors, 21, 23b and 29, this band was excised from the SDS-PAGE gel and analysed by tryptic digest mass spectrometry (R. Burchmore, University of Glasgow) to test whether they corresponded with the anticipated VSG. In all three cases, both of the Set_14 donors present in the standard MASCOT database, 14-A (Tb11.09.0005) and 14-B (Tb11.13.0003; for this study a custom database containing the complete mosaic sequences was not used) were represented, they were each identified by at least one significant peptide match, and they were the only VSGs identified.

5.2.5 *Transgenic trypanosomes were expressing the mosaic* VSG

In summary, analyses of both RNA and protein of the doubly-transfected Lister 427 lines indicate that the trypanosomes were expressing the exogenous, promoter-proximal *VSG*, and that this *VSG* mRNA was being translated to protein. With the exception of 05-27, there was no evidence that any other *VSG* was being expressed.

5.2.6 *Transgenic trypanosomes showed abnormal growth in culture*

Compared with wild-type Lister 427 cells, the doubly-transfected transgenic *VSG* expressors showed some differences in growth under standard *in vitro* culture conditions. The following qualitative observations were made casually as cells were being prepared for experiments. Compared with wild-type cells, the transgenic cell lines appeared to yield a greater frequency of 'monster' cells, with a swollen cytoplasm, multiple flagella, and occasionally detached flagella. They appeared to be more sensitive to high cell density (>2 x 10^6 ml^{-1}) than the unmodified cells, with recovery from a period of overgrowth often impossible. In addition, transgenic cell lines would occasionally form clumps of cells, even at relatively low cell density (~1 x 10^5 ml^{-1}), a phenomenon usually seen only in overgrown wild-type cell cultures. Finally, on occasion, the resistance of the transgenic cell lines to the two drugs used for selection was lost, with the cells showing greatly reduced motility and death. Loss of drug resistance could occur despite continued exogenous *VSG* expression (detected by RT-PCR, data not shown).

One possible mechanism accounting for poor growth and drug sensitivity in these lines is the architecture of the modified expression site. As seen in Figure 5.2, the exogenous *VSG* and drug selection markers introduced to the expression site contain additional untranslated regions. There is some recent evidence that the conserved 16-mer present in the *VSG* 3' untranslated region is involved in silencing of inactive *VSG* genes (S. Hutchison & D. Horn, pers. comm.; P. Batram, pers. comm.). The exact mechanism is unknown, although it may be through competition for an RNA binding protein. The modified expression site contains 16-mers linked both to the expressed *VSG* and to the blasticidin resistance gene.[76] Interference between these sequences could plausibly account for some of the behaviour of these lines.

Notwithstanding these differences from wild-type cell growth, the transgenic cell lines generally grew well in culture, and could be maintained for many weeks with continued transgenic *VSG* expression (confirmed by RT-PCR, data not shown).

5.2.7 Transgenic trypanosomes showed abnormal growth in vivo

To raise antibodies against the various surface coats under investigation, both untransfected and transgenic trypanosomes were used to infect mice. Mice rapidly mount an antibody response against the highly immunogenic VSG surface coat of African trypanosomes. Once mice have been exposed to a particular surface coat, the trypanosomes can be cleared by treatment with cymelarsen (20 $mg.kg^{-1}$). This is necessary as the Lister 427 line used as the background strain for the transgenic *VSG* expressors is monomorphic and hence highly virulent: without treatment the mice rapidly succumb to the infection. Post-treatment, the immune response against VSG continues to develop, and antibodies at high titre can be retrieved with the blood 7–12 days after infection (although they appear sooner, J. D. Barry, pers. comm.).

Preliminary infections of mice demonstrated that a substantially greater intraperitoneal inoculum of transgenic trypanosomes was required to establish

[76] In the case of the Set_14 mosaic-expressing trypanosomes, the promoter-proximal (exogenous) *VSG* possesses two 16-mers and the blasticidin resistance gene possesses one. In the case of the VSG 427-4-expressing trypanosomes, there is just one additional 16-mer, associated with the blasticidin resistance gene. This discrepancy was due to the means by which the Set_14 mosaics were cloned from their subcloning plasmids into pVSG (see footnote 74).

infection compared with untransfected cells. It was found that an intraperitoneal inoculum of 1 x 10^6 trypanosomes was required to reliably establish an infection (compared with 5000 or fewer unmodified Lister 427, G. Cross pers. comm.). With inoculum sizes of 5000 or 50,000, the trypanosomes could not always establish an infection; this was found to be true even in mice immunosuppressed with cyclophosphamide (data not shown). In addition, mice could, on occasion, control the infection, leading to decrease in parasitaemia (data not shown). Poor *in vivo* growth has previously been observed with similar *VSG* transgenic trypanosomes lines (G. Rudenko, pers. comm.), and was found to be the case both for Set_14 mosaic expressors, and the control VSG 427-4 expressor, which is known not to inhibit growth when expressed endogenously.

To test whether poor *in vivo* growth was due to a problem with the expression of the exogenous *VSG* or the loss of the telomere-proximal VSG 427-2, double expressors (which had undergone only one round of transfection to insert a promoter-proximal exogenous *VSG*, leaving the telomere-proximal VSG 427-2 intact) were also used to infect animals. Unlike the single expressors, the double expressors could initiate infections with a smaller inoculum size of 5000, in the three instances attempted (data not shown),[77] suggesting that it is the replacement of the telomere-proximal VSG 427-2 gene with a drug resistance gene that is largely responsible for poor *in vivo* growth.

Because growth in culture places selective pressures on trypanosomes that might be suboptimal for growth *in vivo*, attempts were made to increase the virulence of transgenic trypanosome lines by rapidly passaging through mice every 1-3 days. Such attempts were made with three trypanosome lines (expressing Set_14 variants 21, 23a and 29), before recloning and testing by RT-PCR to ensure continued exogenous *VSG* expression. However, no substantial increase in infectivity was observed (data not shown).

Despite the poor *in vivo* growth and survival compared with wild-type cells, the transgenic lines could still be used to raise antibodies, as an immune response develops even following inoculation with osmotically lysed trypanosome ghosts (Hsia *et al.*, 1996).

[77] *In vivo* growth of double-expressors was tested only for Set_14 mosaics 21, 23a and 29.

5.2.8 Recently cloned cells were used for experiments and expression was verified by RT-PCR

Considering the unpredictable growth in culture (see section 5.2.6), care was taken to ensure that those trypanosomes used to initiate infections and run experiments were indeed expressing the transgenic *VSG*s under investigation by using cells that had very recently been cloned under drug selection and tested by RT-PCR.[78] For the immunofluorescence experiments, trypanosomes were grown from RT-PCR-tested stabilates (see section 5.2.3) in the absence of drug selection for no longer than one week, to ensure healthy cells. During this time all immunofluorescence experiments were conducted (sections 5.3.1 and 5.3.5). A similar approach was used for the complement-mediated lysis and agglutination experiments (sections 5.3.2, 5.3.3 and 5.3.6). RT-PCR was carried out on the cultures after the experiments. The results of the RT-PCR following the complement-mediated lysis and agglutination experiments are shown in Figure 5.6; similar results were obtained following the immunofluorescence experiments (data not shown).

As can be seen, the cultures used for the experiments were still expressing the transgenic *VSG*s under investigation. Expression of other Lister 427 *VSG*s (VSG 427-6 and VSG 427-9) could also be detected, at varying levels in different cultures, suggesting that, in the absence of drug selection, there are present within the cultures trypanosomes that have come to express different VSGs (either through *in situ* switching to another expression site, or through duplication into BES 1). However, the abundance of the product amplified using transgenic *VSG*-specific primers suggests that there remained many trypanosomes in the culture that continued to express the exogenous *VSG*, lending support to the premise that the experiments were indeed conducted on the VSG surface coats under investigation.

[78] The results of this RT-PCR were the ones shown to illustrate transgenic *VSG* expression in Figure 5.3.

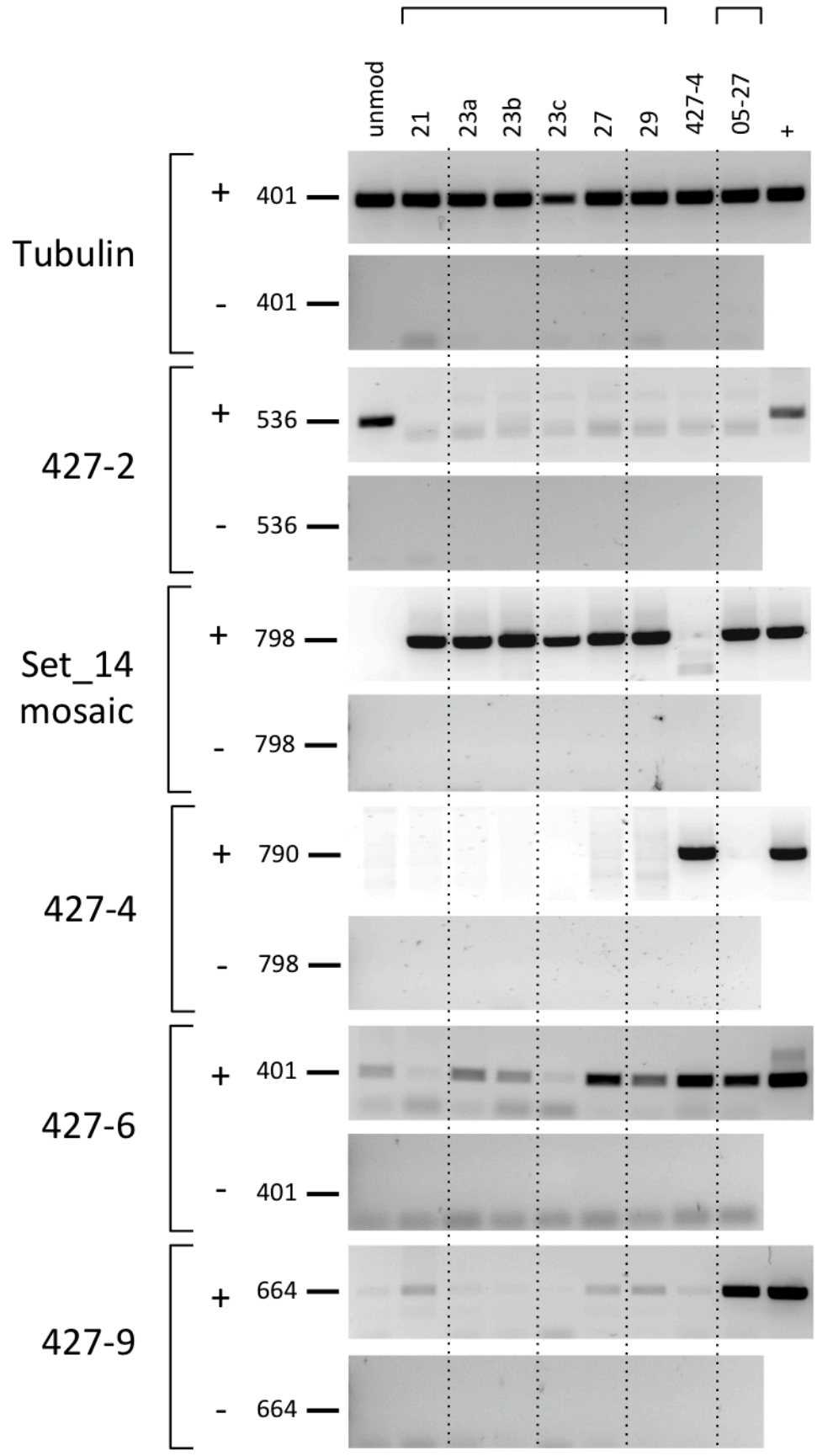


Figure 5.6 Post-complement-mediated lysis RT-PCR. An experiment identical to that described in Figure 5.3 was carried out using the same cultures as used for the complement-mediated lysis and agglutination assays. The Figure is annotated as Figure 5.3. Note that compared with Figure 5.3, VSG 427-4 and 05-27 expressor reactions were loaded into different lanes.

5.2.9 Antibodies against the mosaic coats were raised by infection and cure

The standard procedure for infection to raise antibodies was as follows: 1 x 10^6 trypanosomes were injected intraperitoneally into a Balb/c mouse. When the parasitaemia exceeded antilog 7.2 parasites.ml^{-1} (as scored by the method of Herbert & Lumsden, 1976) mice were cured by a dose of cymelarsen. Five to eight days after cure, blood was collected by cardiac puncture under terminal anaesthesia into CBSS containing 5% w/v sodium citrate. Plasma was retrieved by centrifuging the sample at 14,000 *g* for 10 minutes, removing the supernatant into a clean microfuge tube, centrifuging the supernatant again for 10 minutes, again removing the supernatant into a clean microfuge tube, and adding, as a preservative, sodium azide to a final concentration of 0.05%. At least two anti-plasma were raised against each trypanosome line under consideration. Anti-plasma were named 'anti-XX (Y)' where XX is the variant against which they were raised, and Y is a capital letter from A–D to distinguish the different replica anti-plasma. From four infections, splenocytes were also retrieved for the purposes of generating hybridoma lines (see section 5.3.4).

The question of whether those trypanosomes present in the infection were indeed continuing to express the exogenous *VSG* was tested for three of the Set_14 mosaic expressors (21, 23a and 29). Samples of blood taken by venesection at the peak of parasitaemia were tested for exogenous *VSG* expression by RT-PCR. An abundant product was seen when using the Set_14 mosaic *VSG*-specific primers, with weak or no product from primers directed towards the other Lister 427 VSGs (data not shown), lending support to the premise that the trypanosomes present at the time of cure were indeed expressing the exogenous *VSG* under investigation.[79]

With mosaic *VSG*-expressing trypanosomes and antibodies against them in hand, experiments could be carried out to assess their serological relationships.

[79] In addition, samples of blood taken from infections with three Set_14 mosaic expressors (21. 23a and 29) were tested by indirect immunofluorescence with the monoclonal antibodies mAb-21A and mAb-29B (see section 5.3.5). The overwhelming majority of trypanosomes (>99% of >300 observed) present in the 21 and 23a infections reacted with mAb-21A, and the overwhelming majority of trypanosomes present in the 29 infections reacted with mAb-29B (data not shown), again in support of the premise that the mosaic *VSG*s were being expressed by the infecting trypanosomes.

5.3 Serological analysis of mosaic VSGs

5.3.1 Indirect immunofluorescence showed mosaic expressors were antigenically distinct

5.3.1.1 Indirect immunofluorescence was performed on formaldehyde-fixed cells

To test whether polyclonal antibodies raised against one variant could bind to that and another variant surface coat, indirect immunofluorescence (immunofluorescence assay, IFA) was carried out on live cells. For these experiments, the polyclonal antibodies present in the harvested plasma were used as the primary antibody, and fluorescently labelled goat anti-mouse Ig antibodies were used as the secondary antibody. Preliminary studies (and previous work (J. D. Barry, pers. comm.)) indicated that, were live cells to be used, these experiments would need to be carried out and scored rapidly as surface bound antibodies are rapidly sorted to the posterior of the trypanosome, internalized, and degraded (Engstler *et al.*, 2007; O'Beirne *et al.*, 1998; Pal *et al.*, 2003). The number of antibody-antigen combinations to be examined was too many for such an approach to be feasible, therefore formaldehyde fixation was investigated. Formaldehyde kills and preserves cells and tissues by crosslinking exposed proteins (Fox *et al.*, 1985). Because formaldehyde fixation can result in the addition of antigen-masking groups to exposed proteins, it could change the structure of exposed epitopes.[80] The potential for antigen masking could have complicated interpretation of the immunofluorescence data, so the effects of fixation in the presence of primary antibody were investigated. Preliminary experiments identified no non-specific fixation of primary antibody if formaldehyde fixative was briefly applied in its presence, and substantial posterior accumulation of antibody was seen if a wash step was included between treatment with primary antibody and formaldehyde fixation. For the purposes of identifying exposed epitopes by immunofluorescence, it was therefore decided that cells would be briefly fixed in the presence of the primary antibody, ten minutes after addition.

[80] From the technical advice sheet: 'Fixation and Permeabilization for IHC and ICC', Abcam, Cambridge, UK.

All procedures were carried out at 4ºC, or on ice. Briefly, cultured cells were washed in trypanosome dilution buffer (TDB, (Cross, 1975)) + 1% w/v bovine serum albumin (BSA), and resuspended in anti-plasma at a 1:25 dilution in TDB + 1% w/v BSA. The suspensions were incubated for 10 minutes, after which formaldehyde in PBS was added to a final concentration of 4% w/v. Fixation was carried out for 5 minutes. Cells were washed twice with phosphate-buffered saline (PBS) and resuspended in secondary antibody[81] at a 1:500 dilution in PBS + 1% w/v BSA. After incubation for 15 minutes, the cells were washed twice again in PBS, mounted and visualized. This method is a modified version of that kindly provided by M. Engstler (pers. comm.).

5.3.1.2 Some mosaic VSGs cross-reacted, others were antigenically distinct

Indirect immunofluorescence was carried out as described above for nine different trypanosome lines and 22 different anti-plasma.[82] Figure 5.7 shows three representative cells from each antigen-antibody combination. This figure is summarized in Table 5.3.

Figure 5.7 →Indirect immunofluorescence with polyclonal antibodies. Indirect immunofluorescence was carried out on live cells using different antiplasma as described in the text. Samples of stained cells were mounted on glass slides and observed at 1000x magnification in the DIC, FITC and PE channels. The different antiplasma used are indicated at the top. For each antiplasma-VSG combination minimally 200 trypanosomes were observed microscopically in all channels. Images of three representative trypanosomes (i–iii) in the DIC and FITC channels are provided for each antibody-antigen combination. 'Unmod' refers to the 427-2-expressing unmodified 13-90 trypanosomes. Additional images are provided for the anti-427-4-427-4 combination. Note: following submission of the thesis, additional experiments were performed that revise some of the results presented in this figure. Please see Appendix 7.4.4 and 7.4.5 for details.

[81] The secondary antibodies used in all cases were Alexa 488-conjugated goat anti-mouse IgG H+L (Invitrogen, Paisley, UK) which reacts against IgG heavy chains and all classes of immunoglobulin light chains from mouse, and, where isotype differentiation was required, Alexa 594-conjugated goat anti-mouse IgM (μ chain) (Invitrogen, Paisley, UK) which is specific for the IgM isotype.

[82] For reasons of time, some of the replica anti-plasma could not be tested on all trypanosome lines (most notably VSG 427-4 expressors).

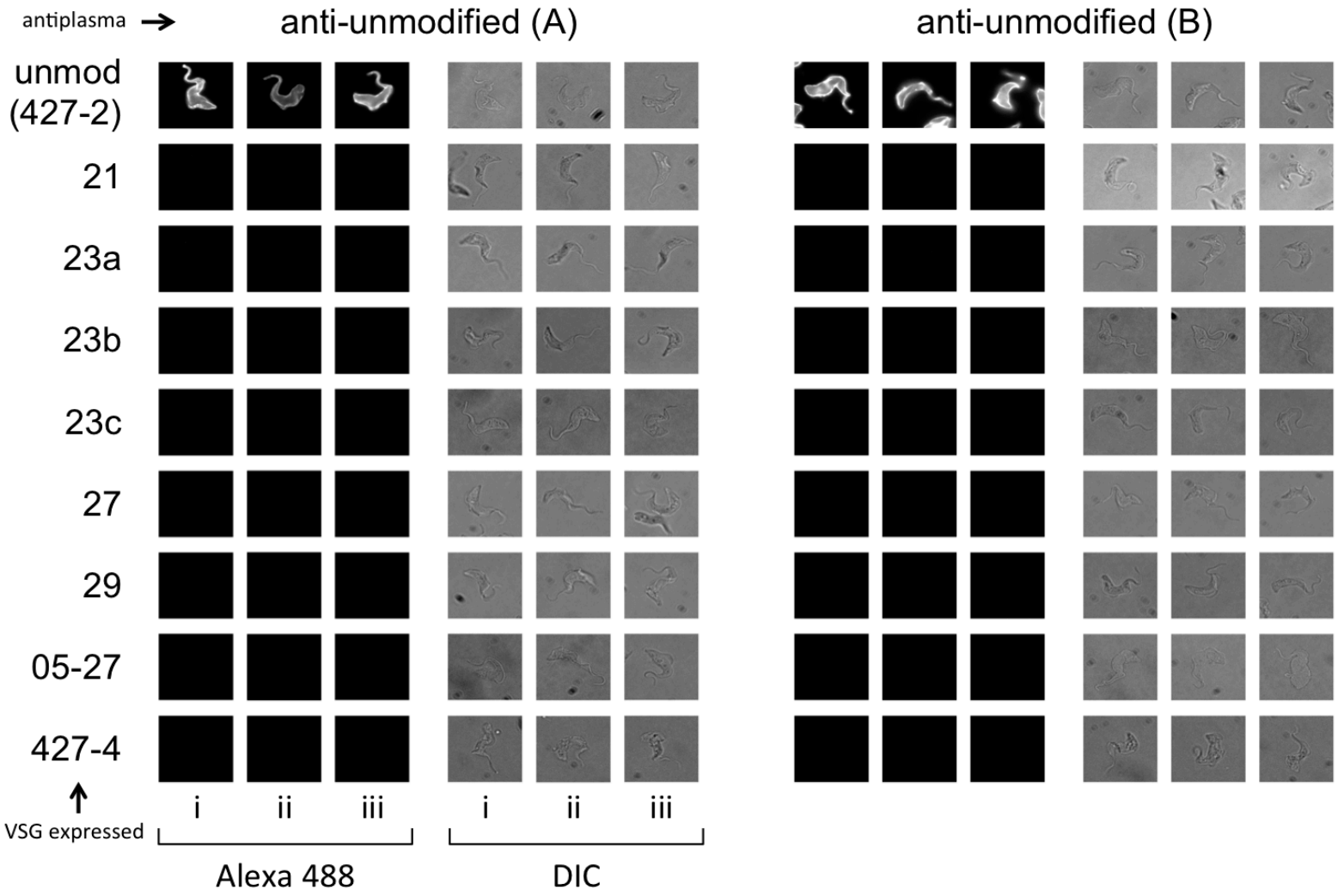


Figure 5.7 IFA with polyclonal antibodies. See previous page for details.

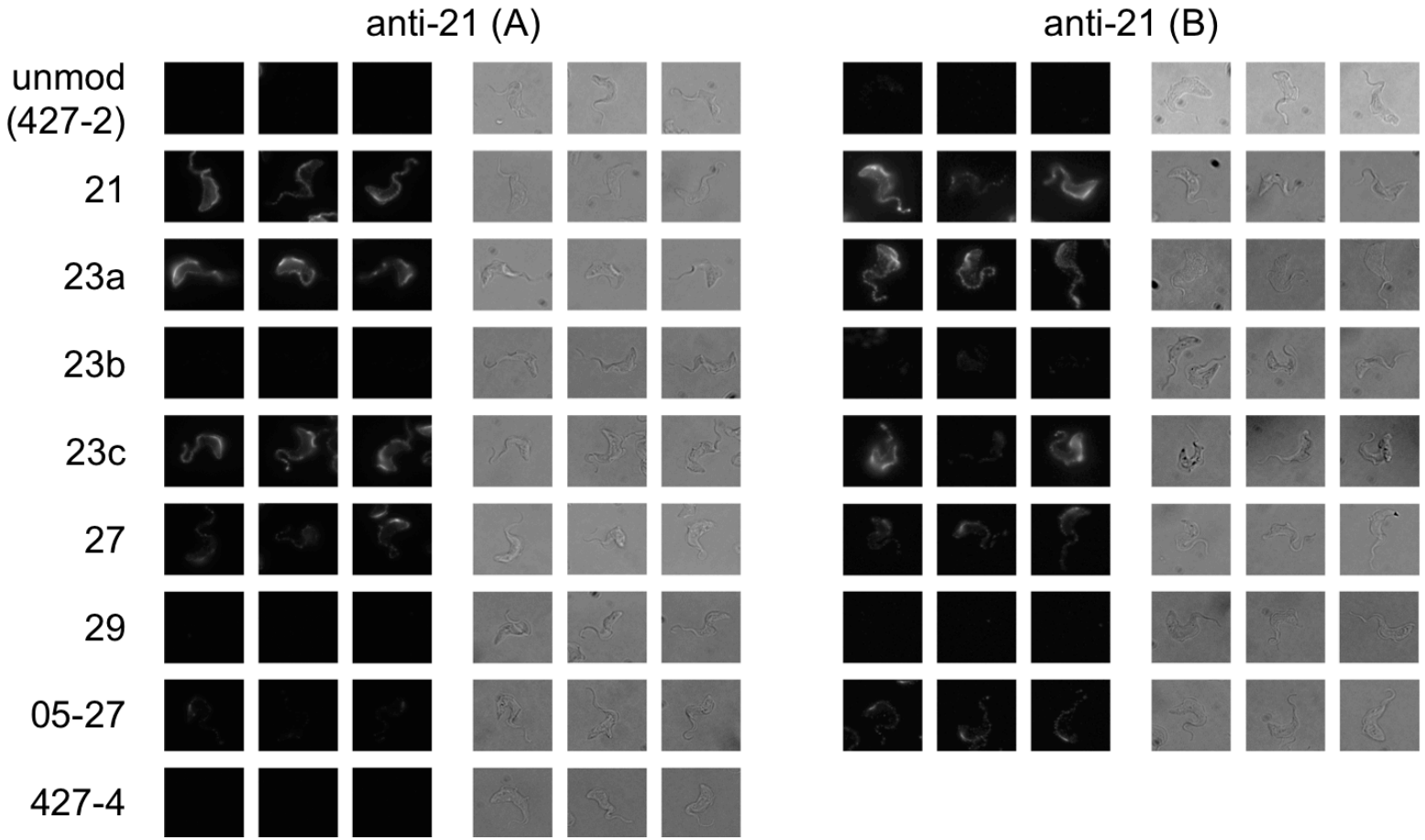


Figure 5.7 (continued)

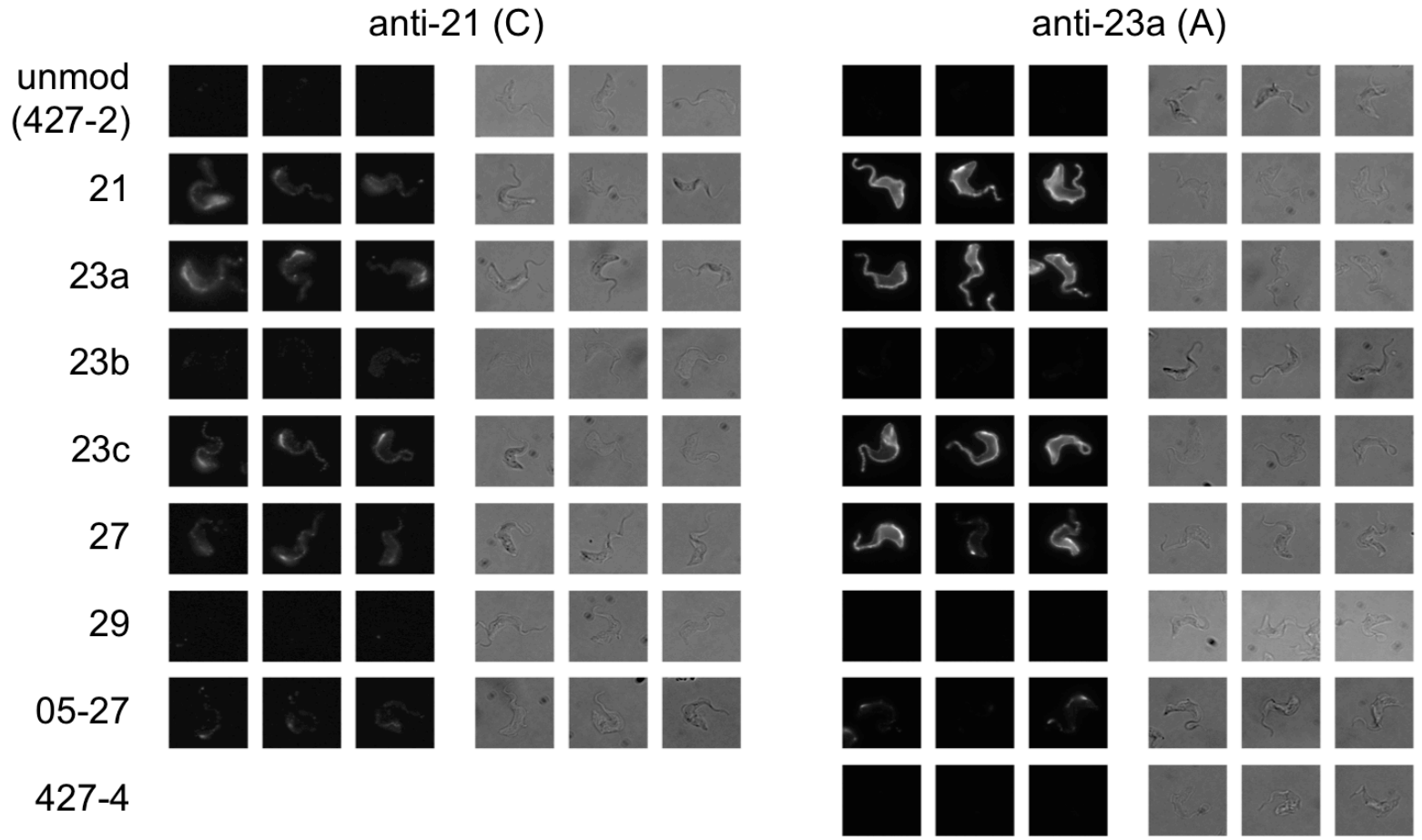


Figure 5.7 (continued)

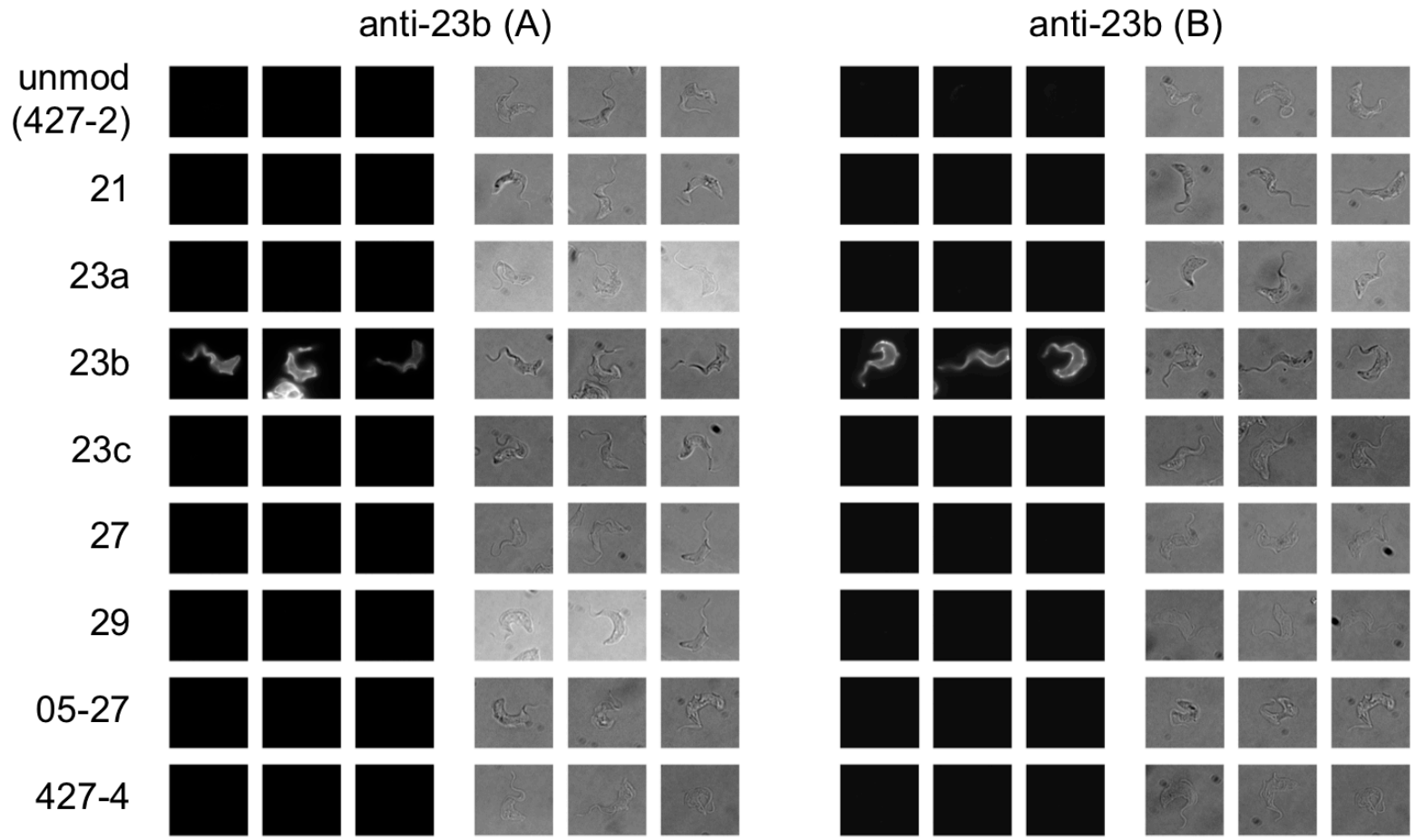


Figure 5.7 (continued)

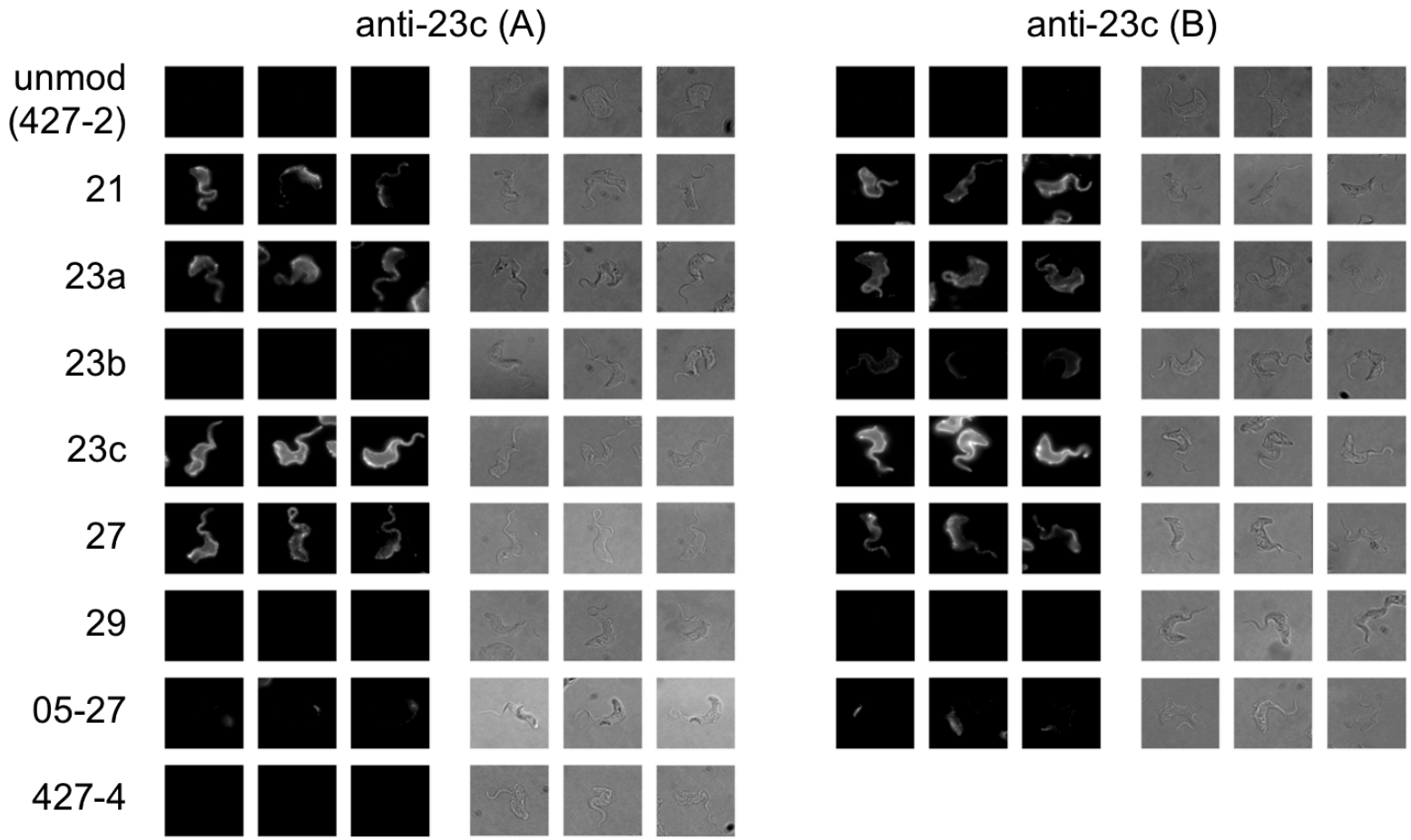


Figure 5.7 (continued)

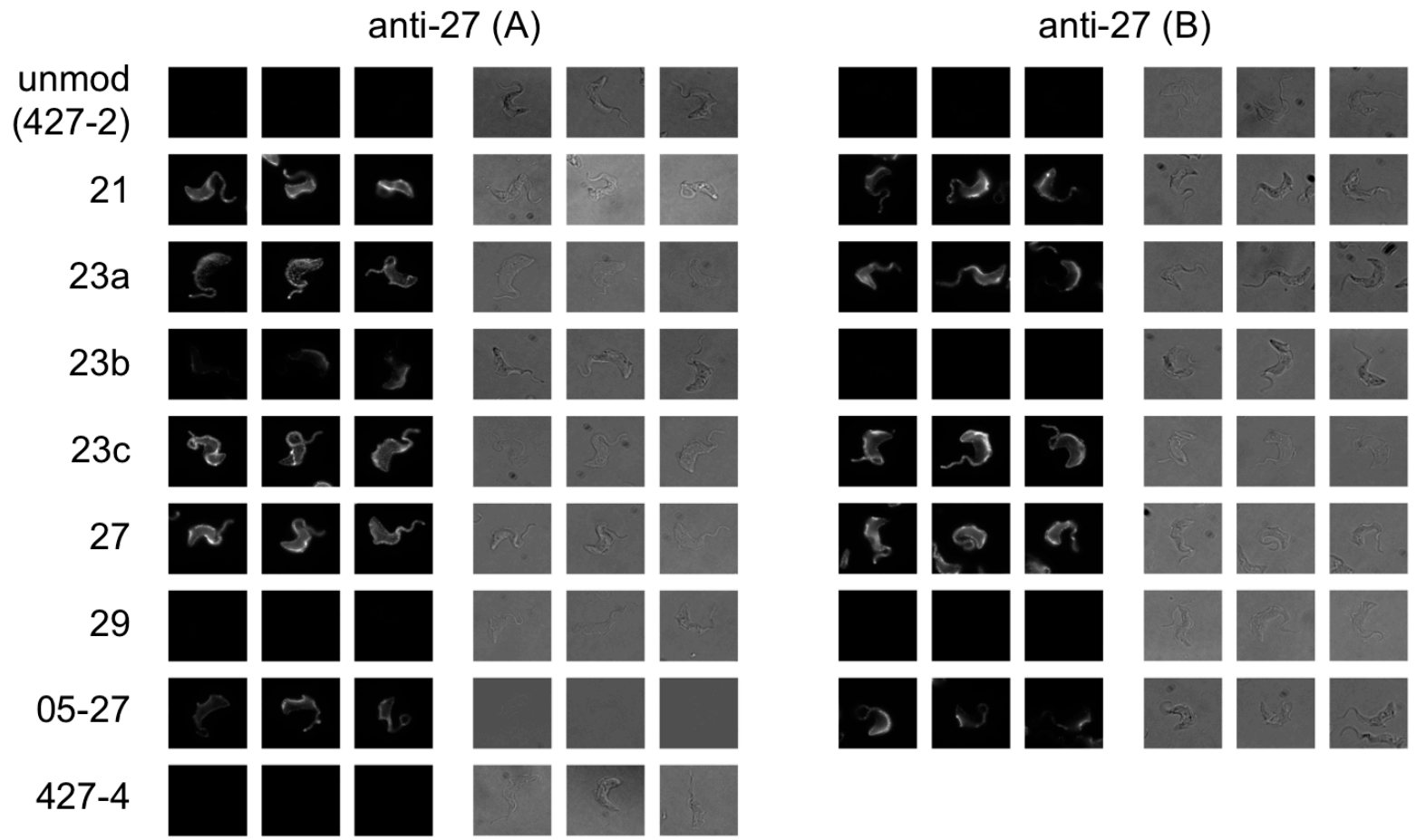


Figure 5.7 (continued)

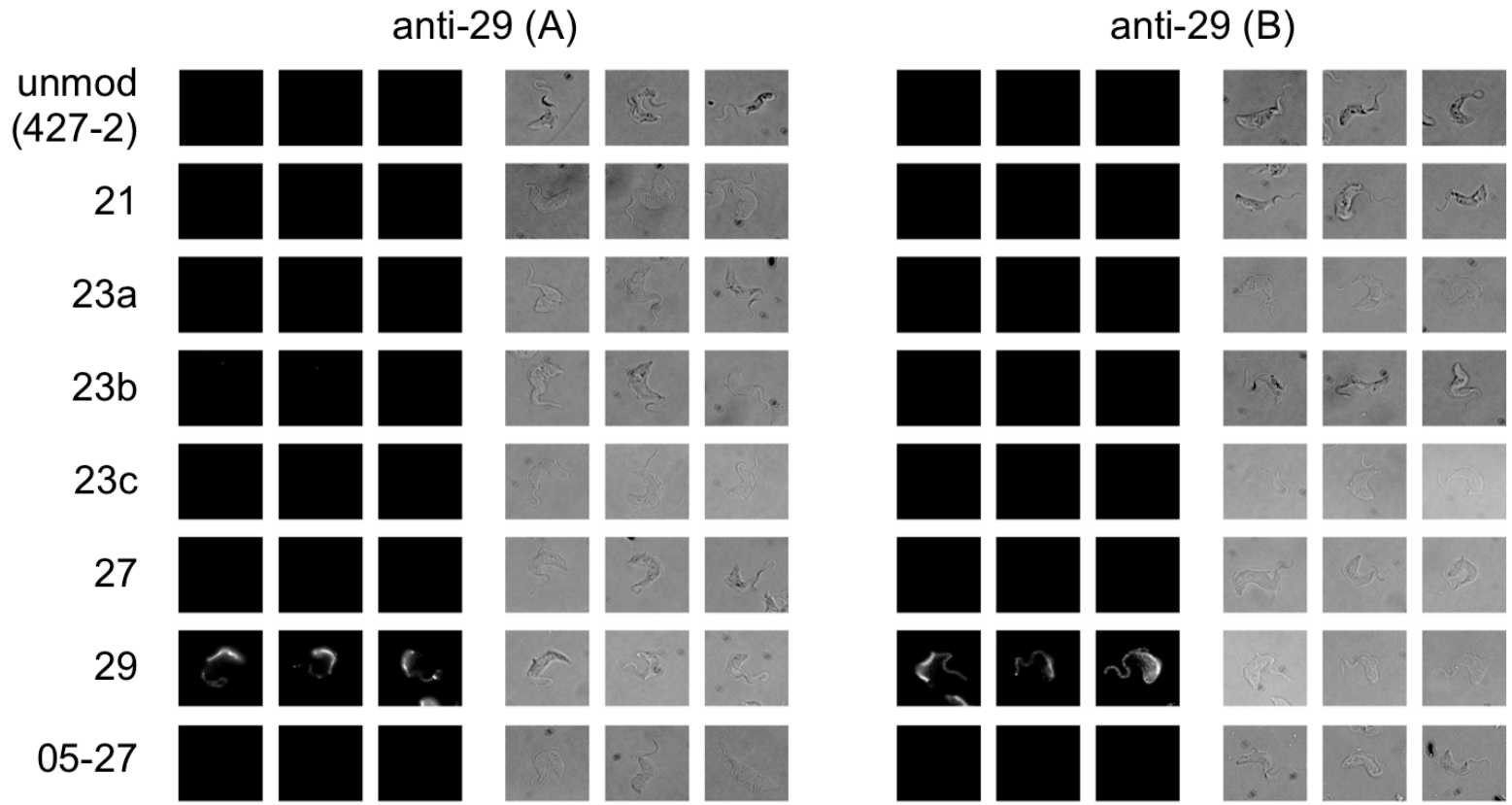


Figure 5.7 (continued)

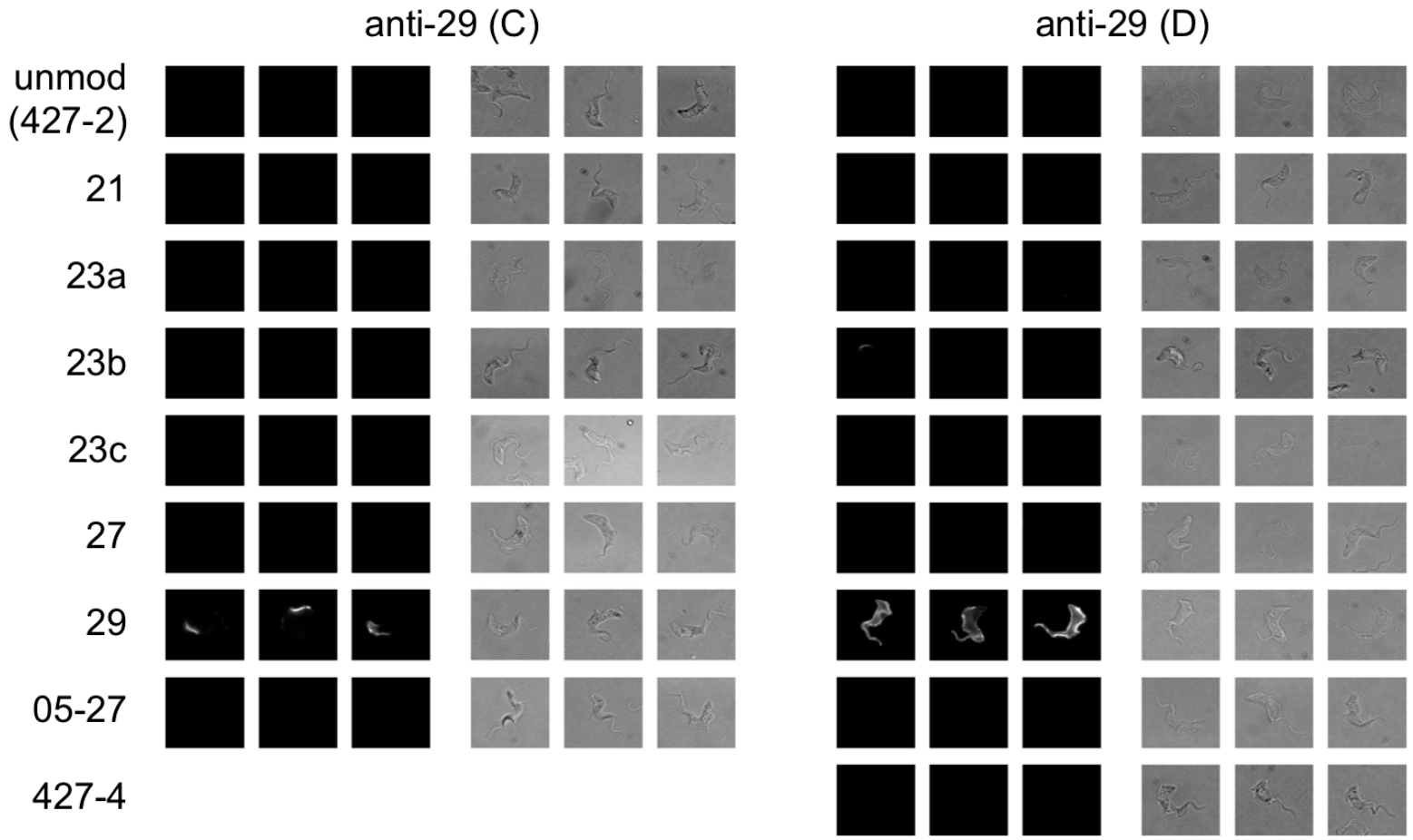


Figure 5.7 (continued)

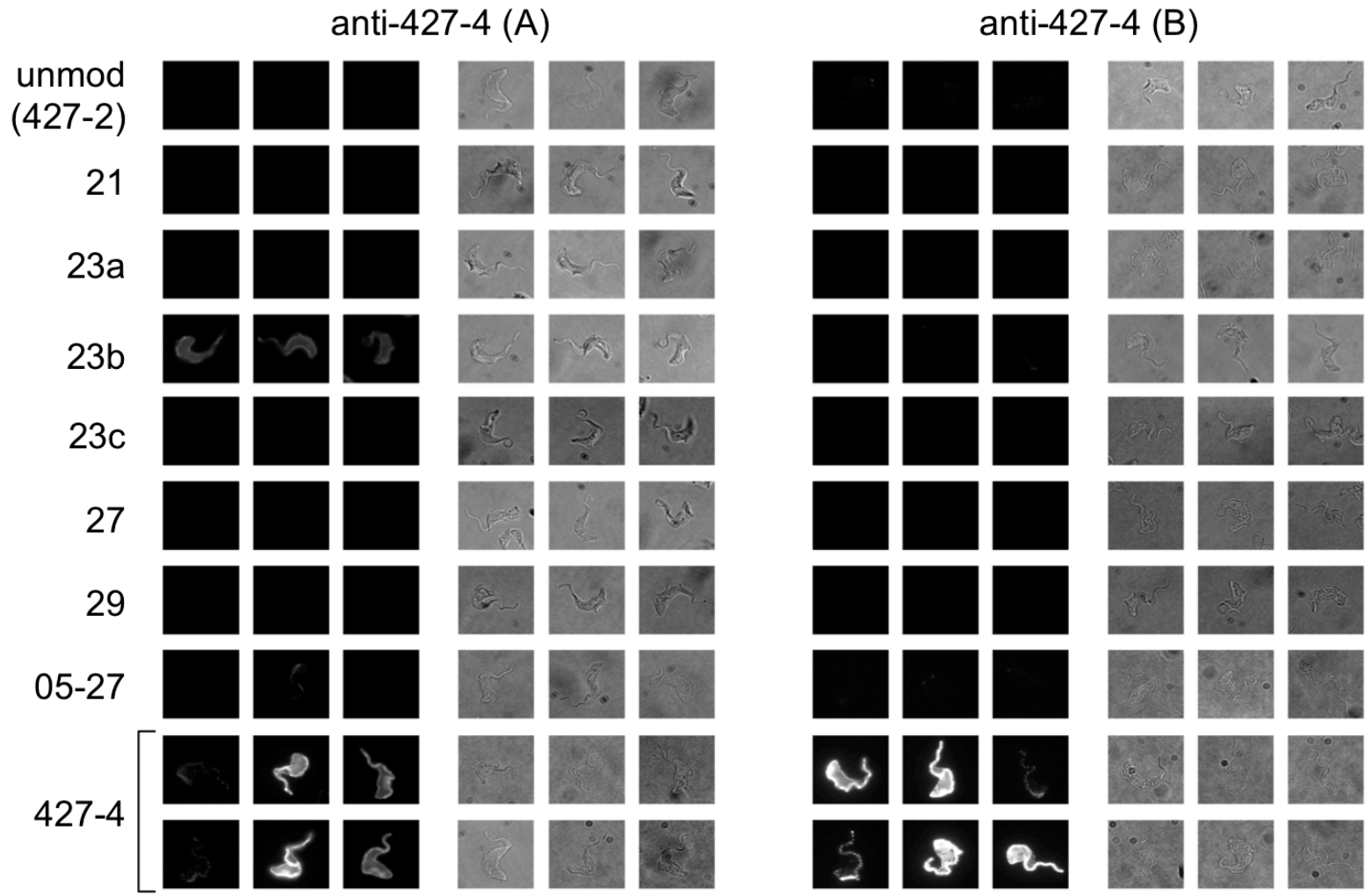


Figure 5.7 (continued)

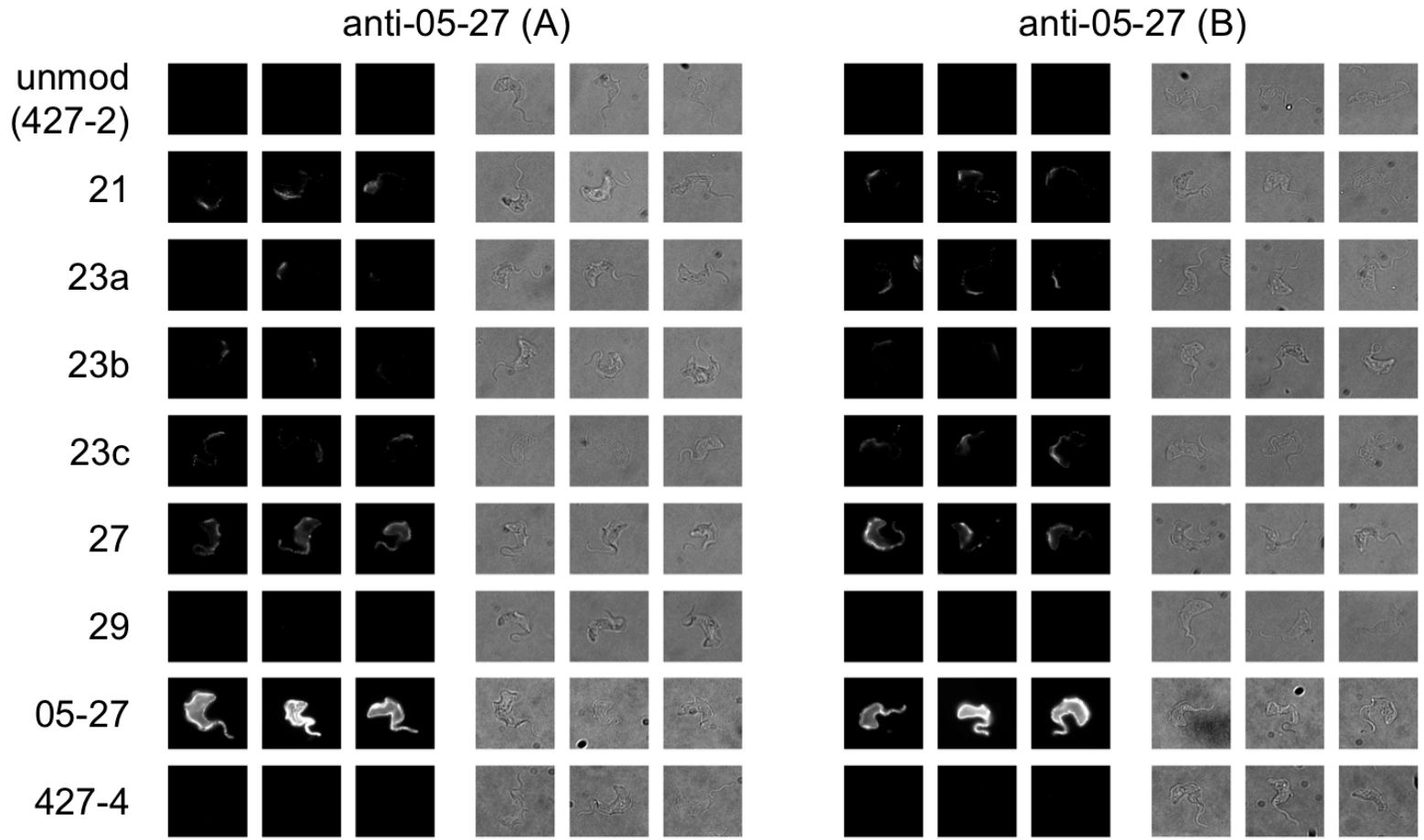


Figure 5.7 (continued)

	VSG expressed	VSG antiplasma raised against								
		427-2	21	23a	23b	23c	27	29	05-27	427-4
	427-2	+++	-	-	-	-	-	-	-	-
Mosaic expressors	21	-	++	++	-	++	+	-	+	-
	23a	-	++	++	-	++	++	-	+	-
	23b	-	-	-	+++	-	+/-	-	~	+/-
	23c	-	++	++	-	+++	++	-	+	-
	27	-	+	++	-	++	++	-	++	-
	29	-	-	-	-	-	-	++	-	-
	05-27	-	+	+	-	+	+	-	+++	-
	427-4	-	-	-	-	-	-	-	-	+/+++

Table 5.3 Indirect immunofluorescence with polyclonal antibodies. A summary of the results shown in Figure 5.7. '+++' indicates strong, eggshell-like fluorescence. '++' indicates strong fluorescence, but perhaps with a speckled appearance or posterior accumulation. '+' indicates patchy or weak fluorescence. '~' indicates very weak fluorescence that was seen on only a few trypanosomes. '-' indicates no signal. Cells are coloured in intensity according to the strength of their signal. Where two different scores are given, separated by a '/', the results varied within a sample, or between replica antiplasma. Note: following submission of the thesis, additional experiments were performed that revise some of the results presented in this table. Please see Appendix 7.4.4 and 7.4.5 for details.

First, the patterns of cross-reaction between antibodies and antigens will be considered. As expected, antibodies raised against unmodified, VSG 427-2-expressing trypanosomes did not crossreact with any of the transgenic cell lines. In a reciprocal manner, antibodies raised against the transgenic cell lines did not bind to unmodified VSG 427-2-expressing trypanosomes. Similarly, (albeit with one exception, discussed below), antibodies raised against transgenic VSG 427-4-expressing trypanosomes did not bind to the Set_14 mosaic-expressing trypanosomes, and neither did antibodies raised against any of the Set_14 mosaic-expressing trypanosomes bind to the VSG 427-4-expressing controls.

On the other hand, antibodies raised against several of the Set_14 variants were able to bind to other Set_14 VSG coats. This cross-reaction was exactly reciprocal. Specifically, polyclonal antibodies raised against any of Set_14 variants 21, 23a, 23c or 27 were able to bind to all four of these antigens. These antigens also cross-reacted reciprocally with a Set_14 variant identified in another infection, 05-27. Serological cross-reaction was not seen with the Set_14 variant retrieved from the later timepoint of the same infection (day 29), and neither did antibodies raised against variant 29 bind to any of the other Set_14 variants.

Taken together, these results suggest that of the Set_14 variants identified in a single infection, 21, 23a, 23c and 27 shared at least one epitope that was exposed on the surface of the live trypanosome, and none of these putative epitopes was possessed by 29.

The pattern of cross-reaction shown by the Set_14 variant 23b is harder to interpret. Antibodies raised against trypanosomes expressing this mosaic did not crossreact with any of the other antigens under consideration, including those formed from related Set_14 mosaic *VSG*s. This seems to suggest that this surface coat was also antigenically distinct from those of the other Set_14 variants. However, this lack of cross-reaction was not exactly reciprocal: in some cases, antibodies raised against other Set_14 mosaic expressors were able to bind to 23b-expressing trypanosomes (anti-23c (B) and anti-27 (A)). In one case a mouse exposed to VSG 427-4-expressing trypanosomes produced antibodies that could bind to 23b-expressing trypanosomes. This unexpected observation indicates that the results of experiments on the 23b-expressing line should be interpreted carefully, and is discussed in detail in light of other results (section 5.3.7).

Next, the qualitative staining patterns will be considered. As shown in Figure 5.7, in both cases unmodified VSG 427-2-expressing trypanosomes treated with homologous primary antibodies displayed a uniform, 'eggshell'-like pattern of fluorescence across their entire surface. This eggshell staining was occasionally seen in the case of the transgenic trypanosomes (for an example, see 05-27). However, a qualitatively different staining pattern was also observed, with a distinctly punctate, 'speckled' effect, and extensive accumulation of stain around the posterior and flagellar pocket of the trypanosome (for an example, see 05-27). Both eggshell and speckled trypanosomes were identified in the same preparation (the case of VSG 427-4-expressing trypanosomes treated with anti-427-4 (B)).

Initially it was considered that the speckled staining pattern was due to aberrant artefactual structure of the transgenic VSG surface coat, but this hypothesis is not consistent with the results of the mAb analysis described in section 5.3.5. More likely the poor growth of transgenic trypanosomes *in vivo* led to those mice having less exposure to those antigens than mice infected with unmodified VSG

427-2-expressing trypanosomes, resulting in a weaker immune response and lower plasma antibody titres. To test whether low antibody concentration could cause the speckled staining pattern in unmodified 427-2-expressing trypanosomes, immunofluorescence using limiting dilutions of primary antibody was carried out. Although a weaker signal was seen as the primary antibody was diluted, the characteristic punctate speckledness was not seen, suggesting that it was not simply antibody concentration that contributed to speckledness (data not shown). An attempt was made to boost the antibody titres by allowing the infections to progress further. This met with some success: anti-29 (D) shows a stronger, less specked staining pattern than the other anti-29 antibodies, which were produced from animals cured earlier in infection (data not shown).[83]

The accumulation of stain around the flagellar pocket was most likely due to the concentration of antibodies in this location by hydrodynamic drag, which, although reduced at lower incubation temperature, is not ablated (Engstler *et al.*, 2007). Preliminary studies in which a wash step was included between treatment with primary antibody and fixation exaggerated this accumulation further, as trypanosomes had more time to sort bound antibodies to this location (data not shown). As can be seen in Figure 5.7, there is a difference in degree in posterior antibody accumulation between different antigen-antibody combinations. This could be caused by variation in both titre and isotype composition of the different polyclonal antibody preparation, as antibody concentration and isotype both affect the rate and extent to which the antibodies are sorted on the trypanosome's surface (Engstler *et al.*, 2007).

5.3.2 Antigenic distinctness protected the later mosaic from complement mediated lysis

To test whether the apparent antigenic distinctness of some Set_14 variants had functional significance, allowing bound antibodies to activate complement and kill the trypanosomes, live cells were exposed to varying concentrations of the antibody in the presence of complement-competent plasma (the complement-mediated lysis, CML, assay). Briefly, complement-competent plasma was

[83] In this context it is interesting to note the results of Dubois *et al.* (2005) who reported reduced antibody responses against transgenic double-VSG expressing trypanosomes. It may be that impaired antibody responses are a general feature of trypanosomes manipulated to express transgenic *VSG*s rather than a phenomenon particular to double-VSG expression.

prepared from freshly collected guinea pig blood by centrifugation at 14,000 *g* for 10 minutes, retrieval of the supernatant into a fresh tube, removal of any remaining blood cells by an additional 10 minute 14,000 *g* centrifugation, and again retrival of the supernatant into a fresh tube. This purified guinea pig plasma (GPP) was divided into small volumes (1–2 ml), frozen, and stored at -80ºC. On the day of the experiment, GPP was thawed and centrifuged at 14,000 *g* for 10 minutes to separate any aggregates that may have formed. Representative polyclonal antibodies[84] directed against the variants under consideration were serially threefold diluted in the GPP, and 5 µl drops were distributed into the wells of a Terasaki plate. 5 µl of 1 x 10^7 ml^{-1} trypanosomes, prepared in advance by resuspending cultured cells in the correct volume of guinea pig plasma, were added to the antibody dilutions. The plate was covered and the reaction was incubated at room temperature for one hour before observation under phase contrast microscopy and scoring each well for dead cells. The maximum dilution of antiplasma that could still cause substantial cell death (>95%) was recorded. For each trypanosome line, cells were incubated at the same concentration in GPP to control for non-specific killing.

	VSG expressed	VSG antiplasma raised against: 427-2	21	23a	23b	23c	27	29	05-27	427-4	GPP
	427-2	1/486	-	-	-	-	-	-	-	-	-
Mosaic expressors	**21**	-	1/54	1/54	-	1/18	1/18	-	1/18	-	-
	23a	-	1/54	1/54	-	1/54	1/18	-	1/6	-	-
	23b	-	-	-	1/162	1/18	1/54	1/18	1/18	1/54	-
	23c	-	1/18	1/18	-	1/54	1/18	-	1/18	-	-
	27	-	1/18	1/18	-	1/18	1/54	-	1/18	-	-
	29	-	-	-	-	-	-	1/54	1/6	-	-
	05-27	X	X	X	X	X	X	X	1/18	X	-
	427-4	-	-	-	-	-	1/6	-	-	1/162	-

Table 5.4 Complement-mediated lysis with polyclonal antibodies. Representative antiplasma (anti-427-2 (B), anti-21 (B), anti-23a (A), anti-23b (A), anti-23c (A), anti-27 (A), anti-29 (D) and anti-427-4 (A)) were serially diluted threefold with guinea pig plasma and 50,000 trypanosomes in an equal volume of guinea pig plasma (GPP) were added. The reaction was incubated in a humid Terasaki plate for one hour at room temperature, and the maximum dilution of antiplasma causing >95% cell death was recorded for each antiplasma-VSG combination. '-' indicates no cell death up to 1/6 dilution. 'X' indicates experiment was not performed. As a negative control, cells were incubated in GPP without antiplasma. Note: following submission of the thesis, additional experiments were performed that revise some of the results presented in this table. Please see Appendix 7.4.4 and 7.4.5 for details.

[84] Representative anti-plasma were used for the CML and the agglutination assays, as there was insufficient time to perform the assay with all available antiplasma. For these experiments, anti-unmodified (B), anti-21 (B), anti-23a (A), anti-23b (A), anti-23c (A), anti-27 (A), anti-29 (D) and anti-427-4 (A) were used. Only homologous polyclonal antibody CML and agglutination assays were performed with 05-27.

The results of this experiment are shown in Table 5.4. Important to note is that the survival of the trypanosomes in GPP indicates their possession of an intact surface coat (Ferrante & Allison, 1983). Consistent with the immunofluorescence, the four cross-reacting Set_14 variants showed reciprocal complement-mediated lysis, whereas VSG 427-2-expressing trypanosomes and mosaic variant 29-expressing trypanosomes appeared to be antigenically distinct, both from the cross-reacting Set_14 variants and one another (time restrictions prevented this assay from being applied to 05-27-expressing trypanosomes).

Similar to the results of the immunofluorescence, the 23b-expressing cells showed cross-reaction with anti-VSG 427-4, as well as with some of the antibodies raised against other Set_14 variants. The non-reciprocal cross-reaction of 23b, and susceptibility of 23b to anti-427-4, may represent general fragility of the cultured cells in suboptimal conditions (GPP). It may also represent genuine cross-reaction, and will be discussed in section 5.3.7.

In general, the transgenic trypanosomes required higher concentrations of anti-plasma to cause lysis when compared with the anti-plasma raised against VSG 427-2. The reduced potency of the anti-plasma could be a manifestation of the speckled staining pattern seen in immunofluorescence, corresponding with a reduction in the total number of antibodies binding.

5.3.3 Agglutination assays are consistent with complement-mediated lysis and immunofluorescence

A third approach to investigating the antigenic relatedness of the Set_14 variants was the agglutination assay. Within a certain stoichiometric range, VSG-binding antibodies are able to characteristically agglutinate trypanosomes. Agglutination happens rapidly, and can be easily observed under phase contrast or bright field microscopy. Briefly, trypanosomes and antibodies were prepared as for the complement-mediated lysis assay, except trypanosome dilution buffer (TDB) was used as the diluent and the medium for trypanosome resuspension (rather than GPP). For this assay, trypanosomes were at a concentration of 2 x 10^7 parasites.ml^{-1}. Plates were incubated at room temperature for 30 minutes before the wells were examined and scored using phase contrast microscopy at

400x magnification. The maximum dilution of antiplasma able to agglutinate trypanosomes was recorded.

The results of the agglutination assays are shown in Table 5.5. These results are broadly consistent both with the immunofluorescence and with the complement-mediated lysis experiments. Some non-reciprocated reaction was seen at higher concentrations of the anti-29 antibodies. The reasons for this are unclear, but it may be that the transgenic cells were particularly prone to agglutination: witness the difference in titres seen between the CML and agglutination assays with antiplasma raised against variant 27, for example.

	VSG expressed	VSG antiplasma raised against 427-2	21	23a	23b	23c	27	29	05-27	427-4	TDB
	427-2	1/54	-	-	-	-	-	-	-	-	-
Mosaic expressors	21	-	1/54	1/54	-	1/54	1/162	-	1/18	-	-
	23a	-	1/54	1/54	-	1/54	1/162	-	1/18	-	-
	23b	-	-	1/6	1/54	1/54	1/54	1/54	-	1/18	-
	23c	-	1/18	1/18	-	1/54	1/162	1/6	1/18	-	-
	27	-	1/18	1/54	-	1/18	1/162	1/18	1/18	-	-
	29	-	-	-	-	-	-	1/54	-	-	-
	05-27	X	X	X	X	X	X	X	1/18	X	-
	427-4	X	X	X	X	X	X	X	-	X	-

Table 5.5 Agglutination with polyclonal antibodies. Representative antiplasma (the same as Table 5.4) were serially diluted threefold with TDB and 100,000 trypanosomes in an equal volume of TDB were added. The reaction was incubated in a humid Terasaki plate for one hour at room temperature, and the maximum dilution of antiplasma causing agglutination was recorded for each antiplasma-VSG combination. '-' indicates no agglutination up to 1/6 dilution. As a negative control, cells were incubated in TDB without antiplasma. 'X' = experiment not performed. Note: following submission of the thesis, additional experiments were performed that negate some of the results presented in this table. Please see Appendix 7.4.4 and 7.4.5 for details.

5.3.4 Monoclonal antibodies were generated to investigate specific epitopes

To dissect the epitope structure of related mosaic VSGs, splenocytes from infected mice were fused to form immortal, antibody-producing hybridomas. These were cloned, providing a pure source of monoclonal antibody (mAb).

Briefly, spleens were aseptically harvested from mice killed by exsanguination under terminal anaesthesia. These were the same mice from which polyclonal

antibodies were obtained.[85] Splenocytes were fused to cultured P3-X63-Ag8-653 myeloma cells using concentrated PEG solutions (as described in Chapter 2, section 2.5.3) distributed across the wells of five 96-well cloning plates, and incubated at 37ºC in 5% CO_2 in air. Twenty-four hours after fusion, cells were put under selection by adding HAT selection medium to each well of the plate to a final concentration of 1x. Cells were screened by indirect immunofluorescence for mAb production 10–12 days after cloning. Wells containing hybridoma producing antibodies of interest were recloned by limiting dilution two further times to ensure pure populations of hybridoma.

Supernatants from wells showing hybridoma growth were screened at each cloning step by differential indirect immunofluorescence. Undiluted supernatants were applied to slides containing acetone-fixed trypanosomes expressing the surface coat under investigation or unmodified trypanosomes expressing VSG 427-2 as a negative control. Slides were incubated in a humid chamber for 30 minutes at room temperature. They were then washed twice in PBS, before the secondary antibodies (see footnote 81) were applied at a 1:2000 dilution in PBS with 1% BSA. Slides were again incubated in a humid chamber for 30 minutes at room temperature, washed twice again, and mounted in Vectashield DAPI mounting reagent (Vector Laboratories, Peterborough, UK). Slides were examined using a 'Zeiss Axioscop 2 mot +' microscope at 1000x magnification. Comparison of fluorescence at 488 nm and 594 nm ('FITC' and 'PE' channels) allowed isotyping of the antibodies present, as IgM isotype antibodies binding the Alexa-594 labelled secondary antibodies would give a strong signal at 594 nm. Antibodies giving no signal at 594 nm were assumed to be IgG, although it is impossible to be sure as the secondary antibody used can also bind to other isotype light chains. Double-cloned cell lines were grown in 100 ml complete RPMI for approximately one week until the cells began to die, and the supernatant was harvested by pelleting the cells at 14,000 *g* for ten minutes. Sodium azide was added to a final concentration of 0.05% as a preservative. These stocks of culture supernatant were used as the source of mAbs for the experiments described in the next section.

[85] Ultimately monoclonal antibodies were obtained from animals that provided anti-21 (A) ('04-21c04-1.1H5a', named mAb-21B for brevity), anti-21 (C) ('04-21c04-2G1a3', named mAb-21A for brevity), anti-29 (A) ('04-29c06-1F10a', named mAb-29B for brevity) and anti-29 (C) ('04-29c06-3D8c', named mAb-29A for brevity).

The production of hybridomas producing mAb of interest was attempted for three different Set_14 mosaic variants, 21, 23a and 29. However, cloned, reactive mAb were obtained for only two of these variants, 21 and 29. Compared with previous work (J. S. Crowe, Ph.D. thesis, 1983, University of Glasgow), this is a relatively low number of mAbs. As shown in Figure 5.7 above, polyclonal responses against transgenic *VSG*-expressing trypanosomes often gave a weaker signal than those responses against unmodified VSG 427-2-expressing trypanosomes. A weaker signal may be due to lower concentration of these antibodies in the plasma, which in turn may be due to lower numbers of anti-VSG antibody-producing B-cells in the spleen, which could also account for the decreased number of hybridoma clones retrieved.

5.3.5 Monoclonal antibody cross-reaction broadly corresponded with that of the polyclonal antibodies

The four mAbs produced were tested on acetone-fixed trypanosomes, in the same manner as supernatant screening (briefly described in section 5.3.4 above), and the results are shown in Figure 5.8 and in Table 5.6. The results broadly corresponded with the results of the polyclonal antibodies. The two mAbs raised against Set_14 variant 29 detected just this mosaic variant, and did not bind to any of the other antigens tested. This demonstrated that 29 had at least one epitope, possibly two, which were not present on any of the other Set_14 mosaics. The two mAbs directed against variant 21 had different patterns of cross-reaction. The IgM isotype mAb (mAb-21B) bound to variant 23a as well as 21. The IgG isotype (mAb-21A) was more promiscuous, reacting with variants 23a, 23c, 27 and 05-27 as well as 21. This pattern of cross-reaction indicated that there was at least one epitope shared amongst all five of these Set_14 variants, and another epitope possessed by just two of them, (21 and 23a). None of the mAbs produced recognized 23b. [86]

The process of acetone fixation permeabilizes the trypanosome as it disrupts the surface membrane, also causing the disruption of the VSG surface coat, exposing epitopes that would otherwise be buried on a living trypanosome. To test whether any of the mAbs produced recognized epitopes exposed on a live

[86] Note: experiments performed following submission of this thesis revise some of the statements in this paragraph. Please see Appendix 7.4.4 and 7.4.5 for further details.

trypanosome's surface coat, indirect immunofluorescence was carried out on live cells using mAbs. The method was the same as that used for polyclonal antibodies, described in 5.3.1.2, except instead of a 1:25 dilution of mouse anti-plasma, undiluted hybridoma culture supernatant was used as the primary antibody.

The results for mAb-21A and mAb-29B are shown in Figure 5.9, and are summarized in Table 5.6. No fluorescence was seen from mAb-21B or mAb-29A (data not shown), indicating either that the epitopes recognized by these antibodies were buried in an intact surface coat, or that the antibodies were at such low concentration in the supernatant that they were unable to cause a reaction in live cell immunofluorescence. Given the clear staining seen using acetone-fixed cells, the first possibility—that they target buried epitopes—is the more likely one.

The epitope recognized by mAb-21A, on the other hand, was evidently exposed on live trypanosomes. The patterns of cross-reaction corresponded exactly with the immunofluorescence carried out on acetone-fixed trypanosomes, and with the immunofluorescence seen for the polyclonal antibodies raised against this VSG (section 5.3.1.2). However, unlike the polyclonal antibodies the staining pattern seen with this mAb is the classical eggshell-like staining, suggesting that the speckled pattern seen with the polyclonal antibody immunofluorescence was due to the quality and/or concentration of the antibody present, rather than the structure of the surface coat.

The mAb mAb-29B is harder to interpret. Although it definitely recognized an exposed epitope on live 29-expressing cells (compare with the result of mAb-29A), the staining pattern was quite different, showing very weak, speckled staining. This staining pattern was somewhat reminiscent of the polyclonal antibody staining pattern. Such a staining pattern could have been produced if the antibody was at low concentration in the supernatant (although the clear binding to acetone-fixed cells suggests that this was not the case). A more likely explanation is that the epitope targeted by this particular mAb was only partially exposed on the surface. Steric hindrance prevented many of these antibodies binding on a live surface coat, but once the cells were disrupted by acetone

fixation, the antibody was able to bind more easily, giving the more uniform staining pattern seen in Figure 5.8.

Figure 5.8 → Indirect immunofluorescence with monoclonal antibodies on fixed trypanosomes. Indirect immunofluorescence was carried out on acetone-fixed cells using different hybridoma supernatants as described in the text. Samples were observed at 1000x magnification in the DAPI, FITC (able to detect Alexa 488) and PE (able to detect Alexa 594) channels. The different source supernatants are indicated at the top. For each supernatant-VSG combination over 200 trypanosomes were observed microscopically in all channels. Images of three representative trypanosomes (i–iii) in the DAPI and FITC channels are provided for each antibody-antigen combination for mAb-21A, mAb-29A and mAb-29B; for mAb-21B images are provided in the DAPI and PE channels (as this mAb was able to bind to the Alexa 594-conjugated anti-mouse μ chain secondary antibody). Note: following submission of the thesis, additional experiments were performed that revise some of the results presented in this figure. Please see Appendix 7.4.4 and 7.4.5 for details.

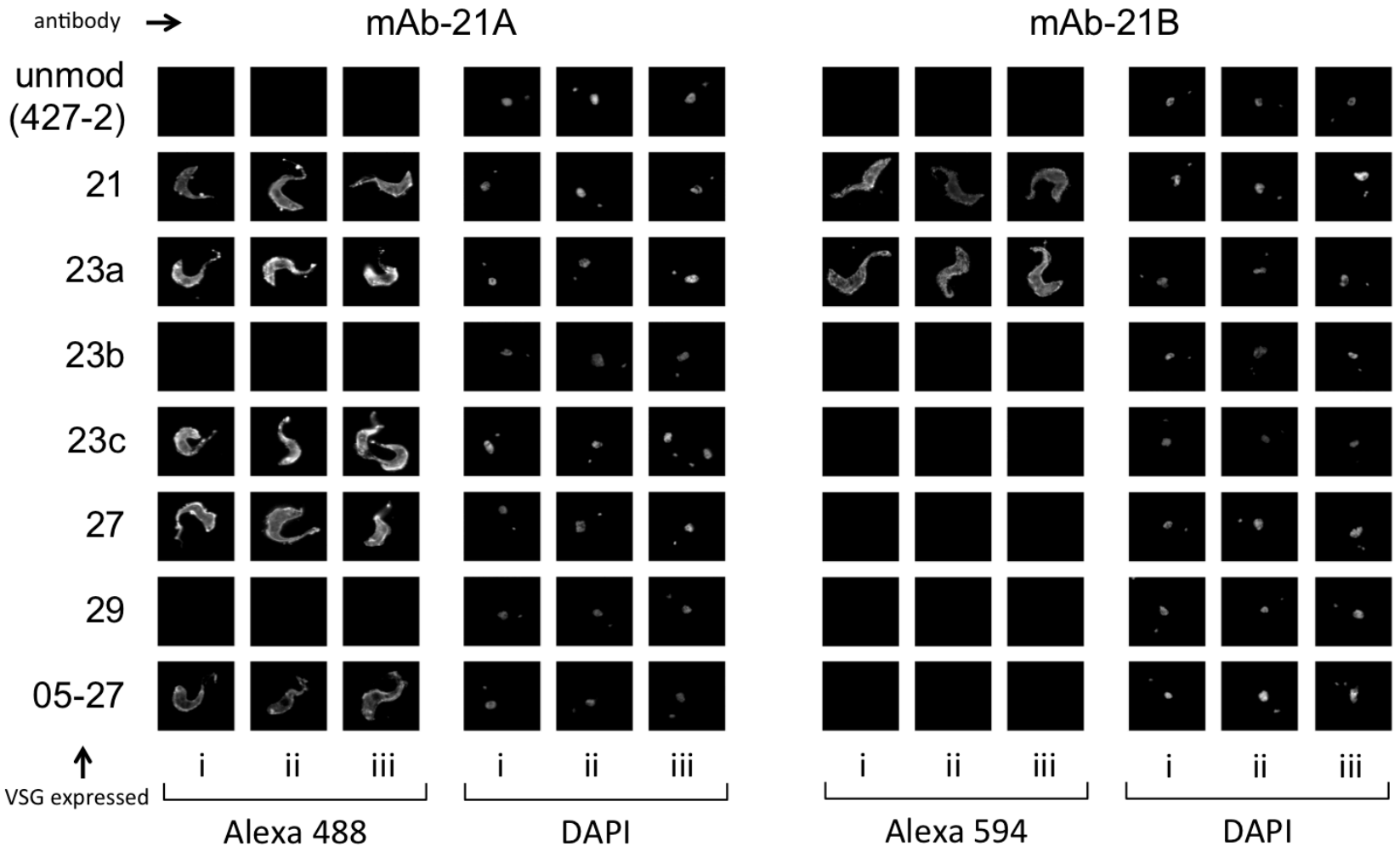


Figure 5.8 Indirect immunofluorescence with monoclonal antibodies on fixed parasites. See previous page for details.

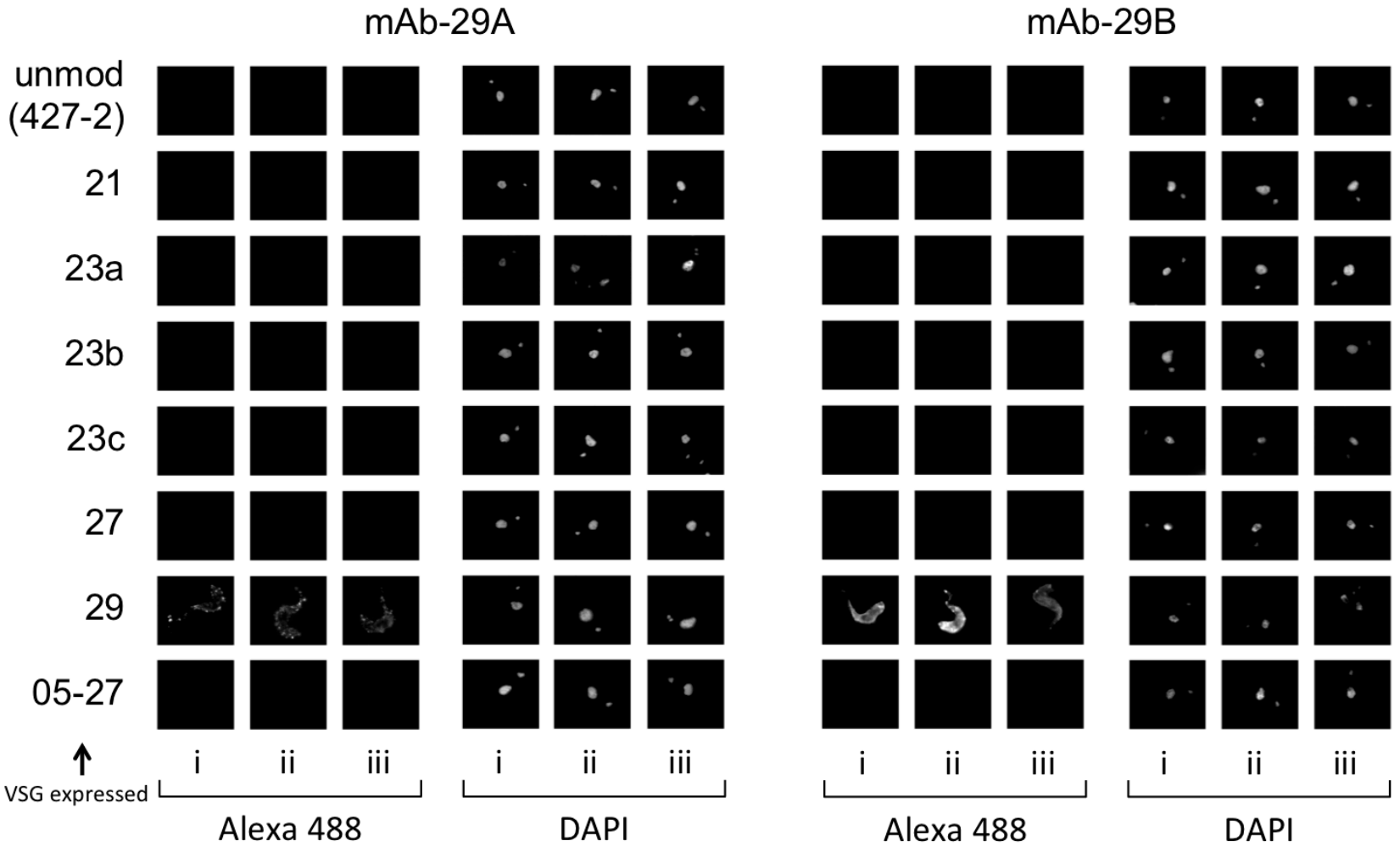


Figure 5.8 (continued)

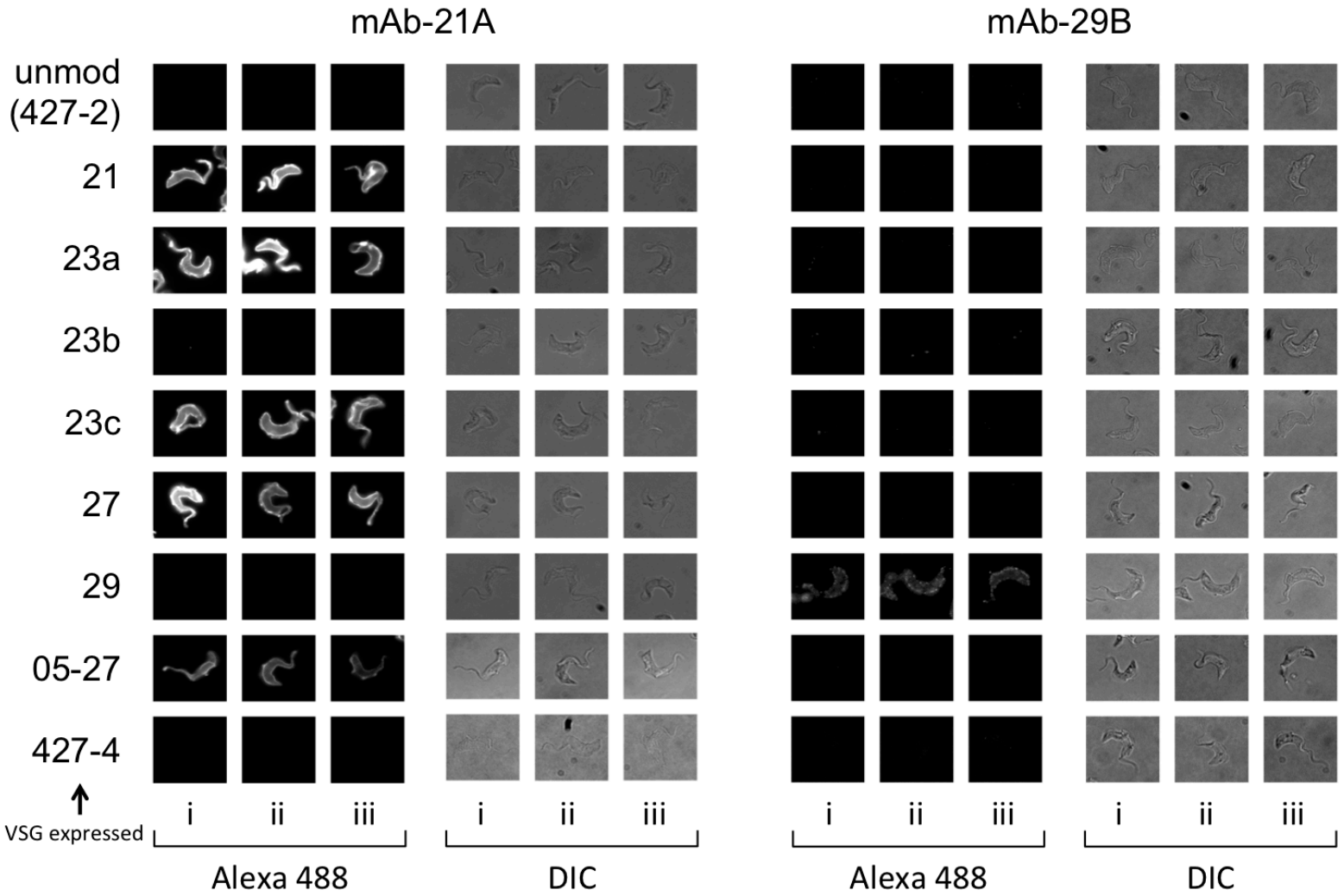


Figure 5.9 Indirect immunofluorescence with monoclonal antibodies on live parasites. See following page for details.

Figure 5.9 ←Indirect immunofluorescence with monoclonal antibodies on live trypanosomes. Indirect immunofluorescence was carried out on live cells using different hybridoma supernatants as described in the text. This image was prepared and is annotated as in Figure 5.7. Note: following submission of the thesis, additional experiments were performed that revise some of the results presented in this figure. Please see Appendix 7.4.4 and 7.4.5 for details.

VSG expressed		mAb-21A		mAb-21B		mAb-29A		mAb-29B	
		Fixed	Live	Fixed	Live	Fixed	Live	Fixed	Live
	427-2	-	-	-	-	-	-	-	-
Mosaic expressors	21	+++	+++	+++	-	-	-	-	-
	23a	+++	+++	+++	-	-	-	-	-
	23b	-	-	-	X	-	X	-	-
	23c	+++	+++	-	X	-	X	-	-
	27	+++	+++	-	X	-	X	-	-
	29	-	-	-	-	+++	-	+++	+
	05-27	+++	+++	-	X	-	X	-	-
	427-4	X	-	X	X	X	X	X	-

Table 5.6 Indirect immunofluorescence with monoclonal antibodies. A summary of the results shown in Figure 5.8 and Figure 5.9. '+++' indicates strong fluorescence, '+' indicates a weak or patchy fluorescence, '-' indicates no signal. 'X' indicates no data for that antibody-VSG combination. Note: following submission of the thesis, additional experiments were performed that revise some of the results presented in this figure. Please see Appendix 7.4.4 and 7.4.5 for details.

5.3.6 Monoclonal antibodies gave a consistent result with complement lysis and agglutination assays

To test whether the live-cell binding by mAb-21A and mAb-29B seen using immunofluorescence corresponded with functional activity, CML assays were set up. Broadly, the method was identical to that described in section 5.3.2 and 5.3.3. In this case, however, undiluted supernatant was used as the source of the primary antibody.

The results are shown in Table 5.7 and closely match the results of the immunofluorescence. Corresponding with its weak immunofluorescence staining pattern, mAb-29B showed no reaction on variant 29, and so was not tested on all antigens. Compared with the polyclonal antibodies, mAb-21A was able to activate complement and lyse the cells at a higher dilution. Agglutination assays were set up using mAb-21A, the results of which are shown alongside the CML results in Table 5.7.

VSG expressed		mAb-21A AggIn	mAb-21A CML	mAb-29B CML
	427-2	-	-	-
Mosaic expressors	21	1/162	1/162	-
	23a	1/162	1/162	-
	23b	-	-	X
	23c	1/162	1/162	-
	27	1/162	1/162	-
	29	-	-	-
	05-27	1/54	1/162	X
	427-4	-	-	X

Table 5.7 Complement-mediated lysis and agglutination assays with monoclonal antibodies. 'AggIn' = agglutination assay. Reactions were performed and scored as with polyclonal antibodies (Table 5.4 and Table 5.5). 'X' indicates no data for that antibody-VSG combination. Note: following submission of the thesis, additional experiments were performed that revise some of the results presented in this figure. Please see Appendix 7.4.4 and 7.4.5 for details.

In general, the mAbs gave a much more unambiguous result than the polyclonal antibodies, consistent with their specificity for a single epitope that is either present or absent.

5.3.7 Behaviour of 23b-expressing trypanosomes was consistent with double expression

The results of the 23b-expressing cell line superficially indicated that this *VSG* encoded a product antigenically distinct from those of the other Set_14 mosaic *VSG*s under investigation: antibodies raised against trypanosomes expressing this *VSG* showed no cross-reaction with any of the other cell lines by either immunofluorescence, complement-mediated lysis, or agglutination. None of the mAbs that reacted with the other variants recognized 23b in immunofluorescence assays. However, these patterns were not completely reciprocal: trypanosomes expressing 23b reacted to antibodies raised against other Set_14 VSGs, and, unexpectedly, to antibodies raised against VSG 427-4-expressing trypanosomes.

The fact that the antibodies raised against 23b were specific for 23b suggests this discrepancy was primarily due to the cells themselves, rather than a genuine example of VSG cross-reaction. To test whether the cells were expressing any other *VSG*s besides the Set_14 mosaic under consideration, PCR was attempted

on cDNA produced from cloned cells[87] using *VSG*-specific primers (as in section 5.2.3.1). As can be seen in Figure 5.10, the 23b-expressing cells clearly showed two distinct bands of between 1500 and 1700 bp. One interpretation of this observation is that individual 23b-expressing cells were simultaneously expressing two different *VSG*s. Besides increasing the number of epitopes exposed on the trypanosome surface, this may have made this trypanosome line especially fragile.

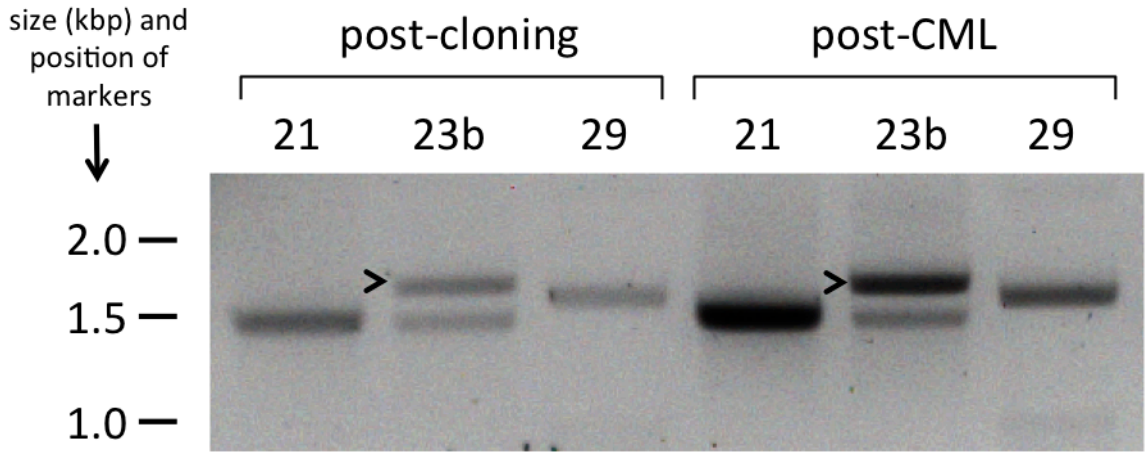


Figure 5.10 Variant 23b-expressing clones were probably expressing an additional *VSG*. cDNA from lines expressing Set_14 variants 21, 23b and 29 was amplified using primers directed against the spliced leader and the *VSG*-specific 16-mer and the products separated on a 1% agarose gel. The source of the template is shown at the top of each lane. Template cDNA made from cells immediately after cloning ('post-cloning') and from cells after the complement-mediated lysis ('post-CML') and agglutination assays was used. Arrowheads indicate the additional, unexpected bands in the 23b lanes.

One model for the unexpected behaviour of 23b-expressing cells, drawing together all available evidence, is that the plasmids did not both correctly integrate in this line, leaving doubly-transfected cells expressing both 23b and an unknown endogenous Lister 427 VSG, 'VSG X'. The SDS-PAGE result and mass spectrometry suggested that if two VSGs were present, 23b was probably at least equally represented. However, VSG X would still have been expressed on the trypanosome surface. During the course of the antibody-production infections, trypanosomes multiplying rapidly without on-going drug selection may have switched to express a different *VSG*. In one of the VSG 427-4-expressor infections, a trypanosome early on switched expression to VSG X, causing a

[87] The cDNA template was the same as that for the PCR analyses presented in Figure 5.3. There was not sufficient time to reclone these cells, as the original clones were only obtained towards the end of the project.

strong antibody response to be raised against both VSG 427-4 and VSG X, producing the patterns of cross-reaction seen in the previous sections. One way of testing this hypothesis would be to test all of the antibodies raised against a panel of endogenous Lister 427 VSGs. Observing some occasional cross-reaction would be consistent with this model. A complementary approach would be to show that antibodies verified to be specific to VSG 427-4 (perhaps mAbs) show no cross-reaction whatsoever with 23b.

However, without further evidence either way, it is not possible to say whether 23b is indeed antigenically distinct from the other Set_14 mosaics, as its lack of reaction may have been due to its co-expression alongside another VSG.

As described in section 5.2.3, variant 05-27 also showed expression of non-Set_14 mosaic *VSG*s. The positive reaction of 05-27 with mAb-21A (raised against the verified variant 21 expressors) indicates that 05-27 is being expressed by most of these trypanosomes, and is cross-reactive with other Set_14 mosaics.

5.3.8 Summary of experimental results: related mosaics can be antigenically distinct

Table 5.8 shows a summary of all the experimental results presented in this chapter. Figure 5.11 shows the Set_14 mosaics investigated in this chapter, arranged according to their antigenic relatedness. Polyclonal immunofluorescence, complement-mediated lysis, and agglutination assays showed that the Set_14 mosaic variant expressed on day 29 of the infection of mouse 04 shared none of its epitopes with earlier-expressed variants 21, 23a, 23c and 27. These earlier variants share at least one epitope that was exposed on the intact surface coat, an epitope that was also present on a related mosaic from a different infection (05-27).

From the same set of donors, mosaicism generated an antigenically novel variant (29) that did not cross-react serologically with those variants expressed earlier in the infection.

	VSG expressed	antiplasma raised against																						monoclonal antibodies mab-										
		427-2			21			23a			23c			27			29			05-27			427-4			21A				21B		29A		29B		
		I	C	A	I	C	A	I	C	A	I	C	A	I	C	A	I	C	A	I	C	A	I	C	A	I(f)	I	C	A	I(f)	I	I(f)	I	I(f)	I	C
	427-2	3	5	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Set_14 mosaics	21	-	-	-	2	3	3	2	3	3	2	2	3	1	2	4	-	-	-	1	2	2	-	-	-	3	3	4	4	3	-	-	-	-	-	-
	23a	-	-	-	2	3	3	2	3	3	2	3	3	2	2	4	-	-	-	1	1	2	-	-	-	3	3	4	4	3	-	-	-	-	-	-
	23c	-	-	-	2	2	2	2	2	2	3	3	3	2	2	3	-	-	1	1	2	2	-	-	-	3	3	4	4	-	X	-	X	-	-	-
	27	-	-	-	1	2	2	2	2	3	2	2	2	2	3	3	-	-	2	2	2	2	-	-	-	3	3	4	4	-	X	-	X	-	-	-
	29	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	3	3	-	1	-	-	-	-	-	-	-	-	-	-	1	-	3	1	-
	05-27	-	-	X	1	X	X	1	X	X	1	X	X	1	X	X	-	X	X	3	2	2	-	X	X	3	3	4	3	-	X	-	X	-	-	X
	427-4	-	-	X	-	-	X	-	-	X	-	-	X	-	1	X	-	-	X	-	-	-	3	4	X	X	-	-	-	X	X	X	X	X	-	X

Table 5.8 Summary of serological analyses of Set_14 mosaic VSGs. A summary of the results shown in Table 5.3, Table 5.4, Table 5.5, Table 5.6 and Table 5.7. 'I' indicates the results of IFA on live cells, 'C' indicates the results of CML, 'A' indicates the results of agglutination, 'I(f)' indicates the results of IFA on acetone fixed cells. For the immunofluorescence assays, '3' indicates strong, eggshell-like fluorescence. '2' indicates strong fluorescence, but perhaps with a speckled appearance or posterior accumulation. '1' indicates patchy or weak fluorescence. '-' indicates no signal. For CML and agglutination assays, the score is given as the number of threefold dilutions still able to give a signal (that is, 1 = 1/6, 2 = 1/18, 3 = 1/54 and so on). 'X' indicates that that particular test was not carried out with that antibody-VSG combination. 23b was omitted due to inconsistencies in VSG expression (section 5.3.7). The non-reciprocal cross-reaction seen in the agglutination test with variant 29 is discussed in section 5.3.3. Note: following submission of the thesis, additional experiments were performed that revise some of the results presented in this figure. Please see Table 7.10 for an updated version, and Appendix 7.4.4 and 7.4.5 for details.

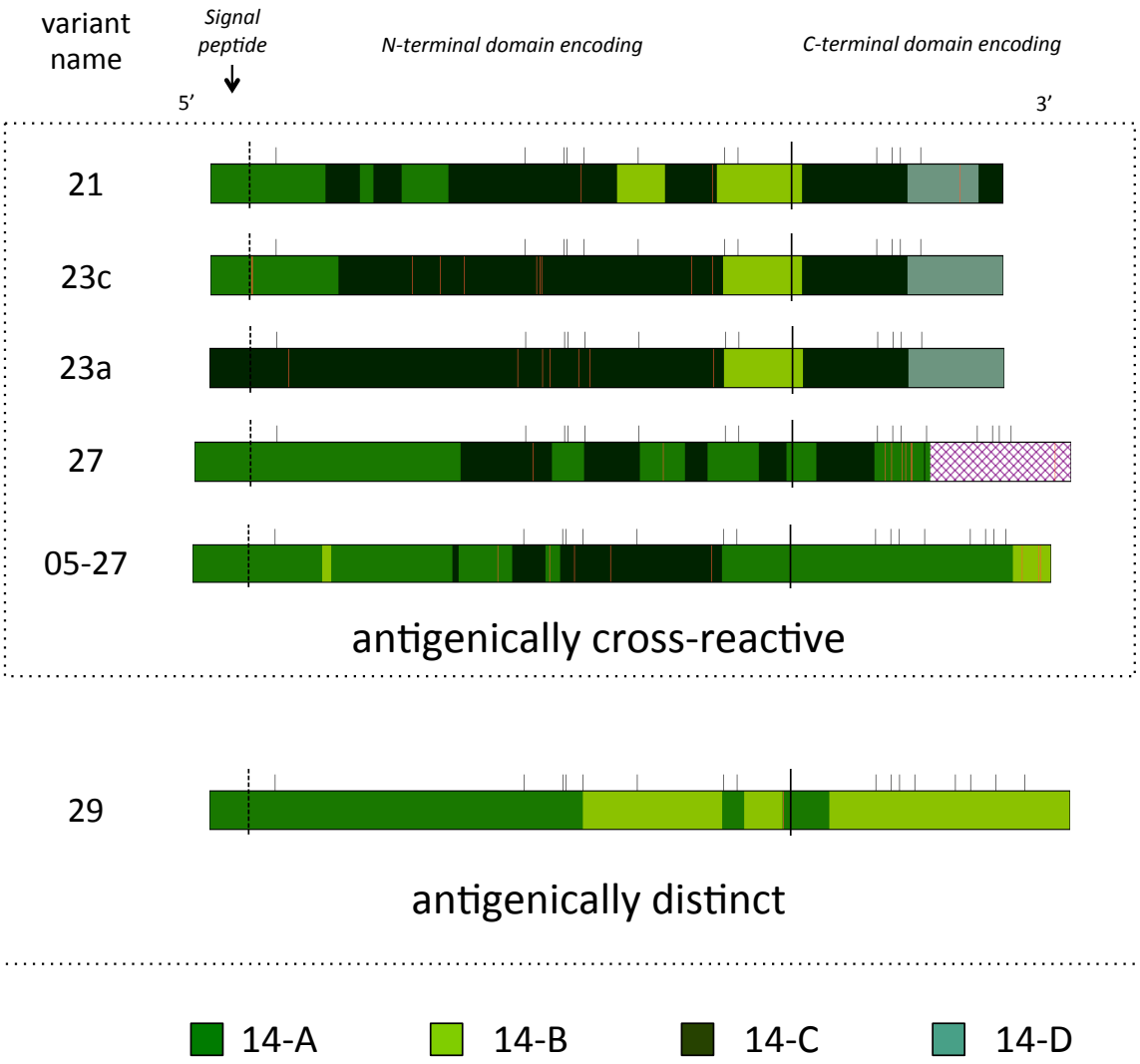


Figure 5.11 Variant 29 was antigenically distinct. The diagrams of the different mosaic *VSG*s examined in this chapter (shown in Figure 5.1) have been rearranged according to antigenic cross-reaction. The maximum variability amongst mature NTDs of the cross-reacting group was 0.150 differences/aa (23a and 27), whereas the minimum variability between variant 29's mature NTD and that of any member of the cross-reacting group was 0.125 differences/aa (29 and 21).

5.4 Structural analysis of distinct mosaics

5.4.1 What amino acid differences were required for antigenic novelty?

To study the relationship between amino acid sequence and antibody binding, the amino acid sequences of the mosaic VSGs were analysed in depth with reference to the experimental results, to ask the question: 'what changes at the amino acid level abrogate specific antibody binding?'.

Working from the premise that the amino acid chains probably fold in a similar manner for all of the related Set_14 VSGs under investigation, the aligned amino acid sequences of the different variants were compared. The alignment did not include any gaps. Differences in amino acid residues between the VSGs were considered for each position in the alignment. If a VSG that did not bind a particular antibody had a different amino acid from all of the antibody-binding VSGs at a given point in the alignment, the change could have contributed to the antibody's inability to bind.

The polyclonal antibodies showed reciprocal cross-reaction between variants 21, 23a, 23c, 27 and 05-27. Variant 29, which appeared in the same infection as variants 21, 23a, 23c and 27, did not bind antibodies raised against those variants; neither did those variants bind antibodies raised against variant 29. The differences contributing to variant 29's complete antigenic distinctness were investigated by recording those positions in the alignment where variant 29 was different from all of the cross-reacting VSGs. Twenty-six such differences were found. The positions of these differences in the amino acid sequence of the NTD are shown in Figure 5.12. Variation associated with the abrogation of antibody binding was present in the more C-terminal two-thirds of the NTD. The structural features of this region are considered further in section 5.4.2 below, and in Table 5.9.

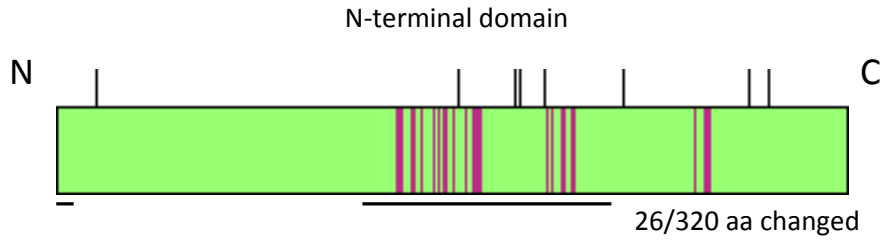


Figure 5.12 Residues contributing to variant 29 antigenic distinctness in the infection of mouse 04. In a multiple sequence alignment of mature NTD amino acid sequences, positions where residues of the mature variant 29 NTD were different from all of variants 21, 23a, 23c and 27 are shown in magenta. As with the images of cDNA clones, the position of conserved cysteine residues (cysteines were in the same position for all Set_14 mosaic NTDs) are shown as bars protruding from the top of the diagram. The underlined regions correspond with regions predicted to be exposed at the top of the molecule.

A similar approach was employed to identify the likely epitopes of the mAbs. Monoclonal antibody mAb-21A gave a pattern of cross-reaction similar to the polyclonal antibodies raised against variants 21, 23a, 23c and 27, and the variant

expressed in the mouse 05 infection (05-27). Because this mAb was able to bind to live trypanosomes, it is unlikely that the CTD contains its epitope (Schwede *et al.*, 2011). Positions in the alignment where all the antibody-binding VSGs had the same amino acid, but variant 29 (which did not bind the mAb) differed, were recorded. The results, if just the NTD is considered, are shown in Figure 5.13. Was the mAb-21A epitope a linear one, these regions would be the strongest candidate for its location. However, the epitope may be conformational, so it is impossible to draw any conclusions about its location without further experiments.

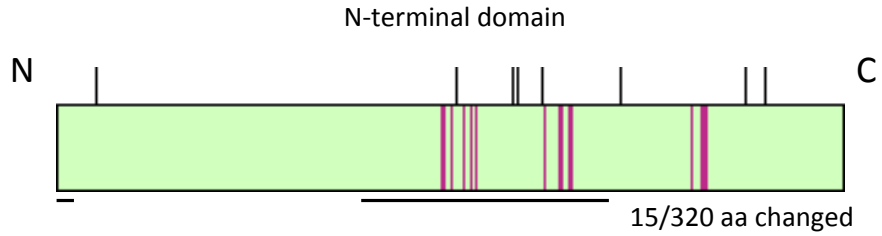


Figure 5.13 Residues contributing to the mAb-21A epitope structure. In a multiple sequence alignment of mature NTD amino acid sequences, positions where the variants that bound mAb-21A (21, 23a, 23c, 27 and the mouse 05 Set_14 variant 05-27) were identical to one another but different from the variant that did not bind mAb-21A (29) are shown in magenta. Figure is annotated as Figure 5.12. Note: analysis of additional experiments, performed following submission of this thesis, produces a different figure. Please see Figure 7.39 for an updated version, and Appendix 7.4.4 and 7.4.5 for details.

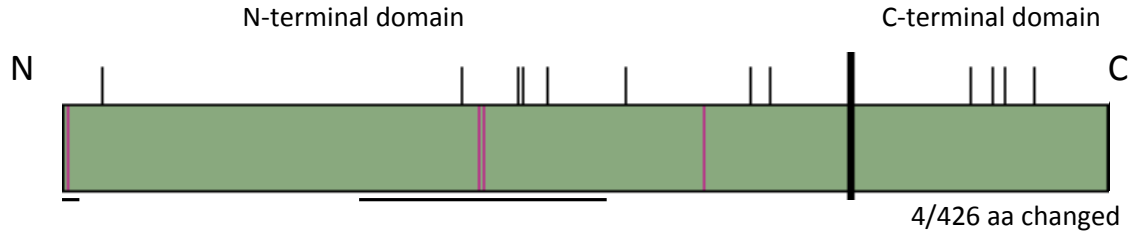


Figure 5.14 Residues contributing to the mAb-21B epitope structure. Positions where the variants that bound mAb-21B (21 and 23a) were identical to one another and different from the most similar variant that did not bind (23c) are shown in magenta. Figure is annotated as Figure 5.12, with the positions of cysteine residues conserved amongst variants 21, 23a and 23c shown. As mAb-21B could not bind to live trypanosomes, the entire mature VSG is shown, with a black bar separating the putative boundary between the N- and CTDs. Note: analysis of experiments performed following submission of this thesis produces a different figure. Please see Figure 7.40 for an updated version, and Appendix 7.4.4 and 7.4.5 for details.

29 escape		
Pos	X-rct	29
138	A	G
139	N	A
140	E	A
144	V	G
145	S	A
148	N	S
153	N	D
155	A	S
157	S	N
158	S	T
161	K	T
166	N	S
169	D	S
170	A/G	S
171	D	Q
172	A/G	S
199	G	A
201	T/K	A
205	K	T
206	P	T
209	D	Q
210	A	N
259	K	S
263	K	P
264	S	T
265	A	Q

mAb-21-A		
Pos	X-rct	29
157	S	N
158	S	T
161	K	T
166	N	S
169	D	S
171	D	Q
199	G	A
205	K	T
206	P	T
209	D	Q
210	A	N
259	K	S
263	K	P
264	S	T
265	A	Q

mAb-21-B		
Pos	X-rct	23-c
2	D	T
170	A	G
172	G	A
262	T	A

Acidic
Basic
Non-polar
Polar

Table 5.9 Amino acid substitutions associated with antibody escape. Three tables show the amino acid substitutions associated with variant 29 antigenic distinctness (left, see Figure 5.12), abrogation of mAb-21-A binding (middle, see Figure 5.13) and mAb-21-B binding (right, see Figure 5.14). 'Pos' = amino acid position in mature NTD; 'X-rct' = identities of amino acids at those positions in cross-reacting antigens; '29' or '23-c' = identities of amino acids at those positions in the antigenically distinct antigens. Emboldened rows correspond with positions predicted to be exposed on the top of the molecule. Note: analysis of experiments performed following submission of this thesis produces a different figure. Please see Table 7.11 for an updated version, and Appendix 7.4.4 and 7.4.5 for details.

Monoclonal antibody mAb-21B gave a different pattern, binding just 21 and 23a, and then only in an acetone-fixed state.[88] Comparing 23c, which didn't bind the antibody, with the two VSGs that did, it can be seen that just four differences at the amino acid level abrogated binding of this mAb. These are shown in Figure 5.14. All these differences were located in the generally more exposed NTD, but

[88] Note that the findings presented in Appendix 7.4.4 and 7.4.5 would cause revision of much of this paragraph.

evidently affected an epitope that was made accessible when the coat was disrupted by acetone. Differences at or near these positions were not consistently seen when comparing the antibody-binding 21 and 23a VSGs with any of the other non-antibody-binding variants 27 or 29, or 05-27 from mouse 05 (data not shown). That there is no specific region in which changes are necessary for antibody escape suggests that mAb-21B's epitope is conformational, as changes to different parts of the molecule all have the same effect of abrogating antibody binding. Although mAb-21B did not have any functional activity, the fact that just four amino acid changes could abrogate its binding is consistent with the finding that VSGs do not require many changes to their amino acid sequence to enable escape from individual antibodies (Baltz *et al.*, 1991). The amino acid substitutions associated with antibody escape are shown in Table 5.9. Non-conservative changes to charged residues in the region of variant 29 predicted to form the exposed loops may underlie that variant's antigenic distinctness.

Neither of the variant 29-binding mAbs bound any of the other Set_14 variants. Compared with the cross-reacting Set_14 VSGs presented here, 29 is generally more distinct at the amino acid level (see Table 5.1). The different results seen when performing immunofluorescence with the two different 29-specific mAbs on live cells suggests that there were two different epitopes, one of which was more exposed on an intact surface coat than the other.

Considering the patterns of segmental contribution by the donor sequences in Figure 5.11, the antigenically distinct variant 29 was the only mouse 04 Set_14 mosaic considered here that showed no contribution of donor 14-C (tryp_XI-1034g11_assembly). The region that 14-C contributed to the cross-reacting mosaics was replaced in variant 29 by two large segments from 14-A and 14-B. This difference may help explain the antigenic distinctness of variant 29.

5.4.2 Changes associated with antigenic distinctness were located in the predicted N-terminal loops

To speculate further on the effects that amino acid differences have on the structure of the mosaic VSG, structural prediction algorithms I-TASSER and PHYRE2 were used to predict the tertiary structure of the mosaic VSGs (Kelley & Sternberg, 2009; Zhang, 2008). In essence, I-TASSER predicts the secondary

structure of the query sequence and uses this pattern to search a database of solved crystal structures for similar secondary structure patterns. Each of the best matches from this search is used as a template to which the query sequence is fitted. PHYRE2 uses PSI-BLAST to find putative templates, and thus depends more immediately on the amino acid sequence itself. Predicted structures of Set_14 variants were attempted using both I-TASSER and PHYRE2. When predicting structures, I-TASSER frequently identified the two solved VSG structures (MITat 1.2 and ILTat 1.24: PDB identifiers 1vsg and 2vsg respectively) as the best matches. PHYRE2 was not able to predict a full structure based on automatic alignments alone. However, if a solved VSG structure was specified (in this case, 1vsg), PHYRE2 was able to thread the amino acid sequence to this template. The predicted structures of variant 29 are shown in Figure 5.15.

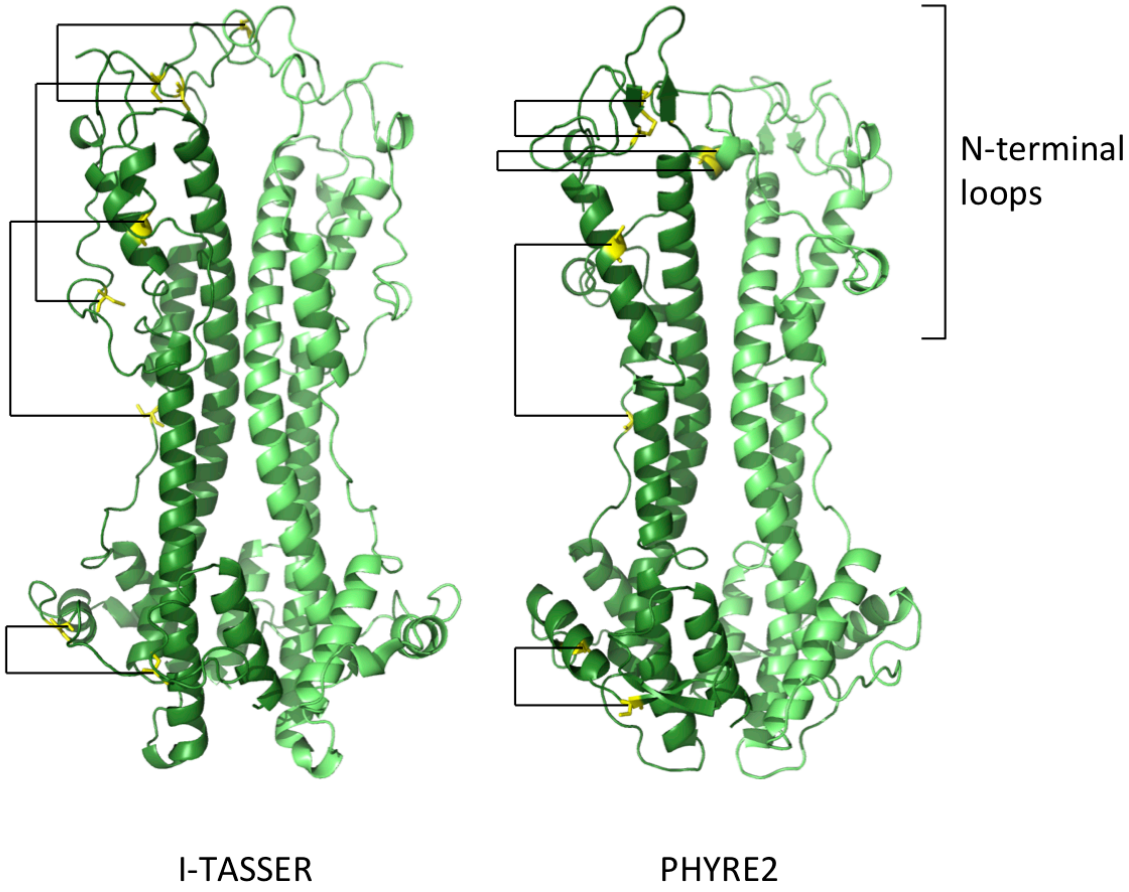


Figure 5.15 Variant 29 structure predictions. The three-dimensional protein structure for variant 29 dimers, as predicted by I-TASSER (Zhang, 2008) and PHYRE2 (Kelley & Sternberg, 2009), are shown. For each structure, the two constituent monomers are coloured a different shade of green, with the cysteines of one chain coloured in yellow. On each figure, each pair of cysteines that likely lie in close proximity to form disulphide bridges are indicated and connected by black lines. Images were generated using Pymol (Schrödinger, LLC).

There are considerable differences between the two predictions, particularly in the more disordered N-terminal loop region. Furthermore, it is likely that both

of these structures vary from the actual three-dimensional structure of a Set_14 VSG: while the query sequence was fitted to a type A VSG NTD (1VSG) for both I-TASSER and PHYRE2, Set_14 mosaic VSGs have B-type NTDs. No crystal structure has yet been solved for B-type NTDs, but the disulphide bridges formed by the conserved cysteine residues have been identified (Bussler *et al.*, 1998). Neither the I-TASSER nor the PHYRE2 models placed all of the disulphide-bridge-forming cysteines in close proximity, although PHYRE2 managed somewhat better than I-TASSER. The predicted structures must therefore be treated with caution. Nevertheless, even imprecise predictions are likely to be informative when it comes to the gross features of VSG structure: what parts of the sequence are likely to constitute the buried alpha helices and what parts of the sequences are likely to contribute to the exposed N-terminal loops (Blum *et al.*, 1993).

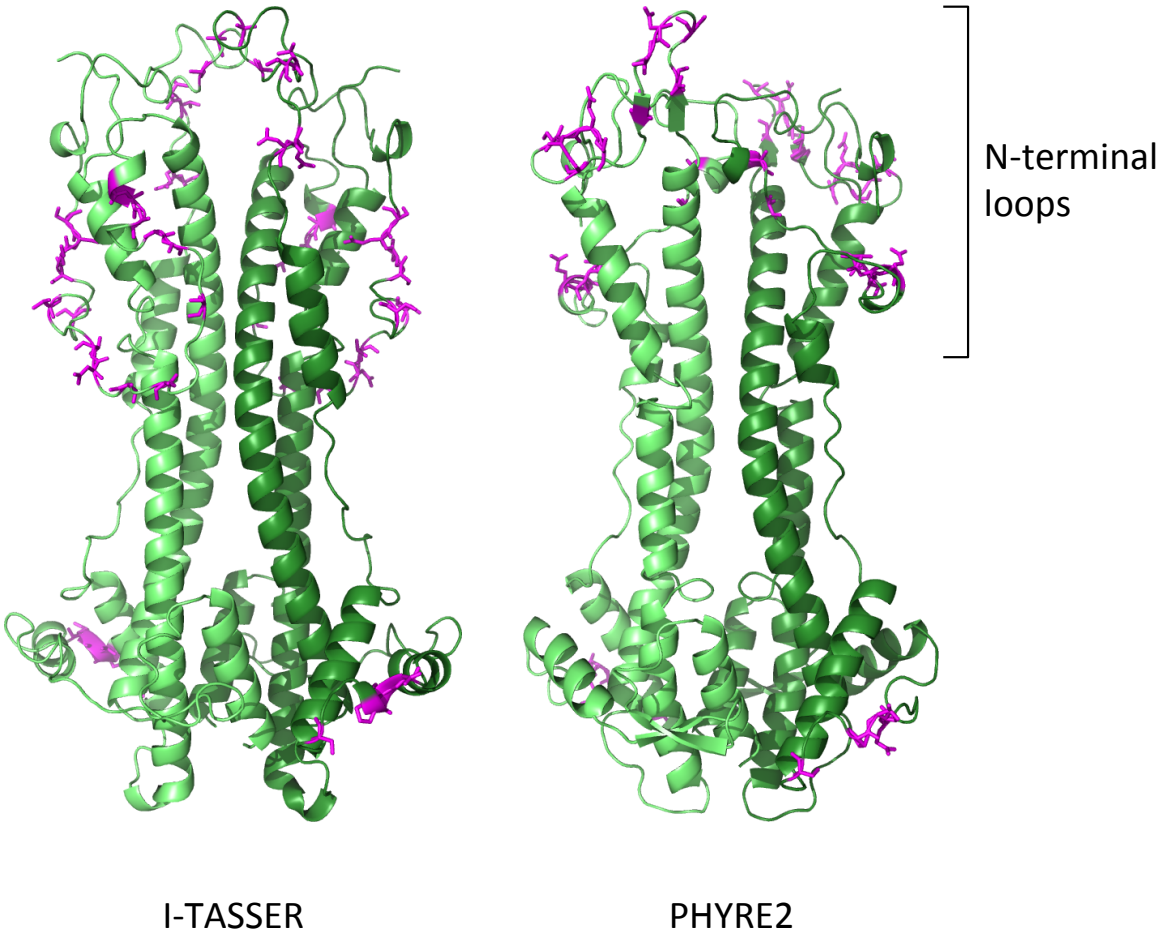


Figure 5.16 Predicted location of residues potentially contributing to variant 29 antigenic distinctness in the infection of mouse 04. The residues identified in Figure 5.12 were indicated on the predicted structures of variant 29 in magenta.

In what region of variant 29 are those changes associated with antigenic distinctness located? Figure 5.16 shows the predicted structures of variant 29 with those amino acids that are possibly associated with variant 29's antigenic

distinctness highlighted (that is, those differences shown in Figure 5.12). As can be seen in the figure, these differences are located primarily in the N-terminal loops of the predicted structure. It is in this part of VSG that forms the most immediate interface with the immune system, and where, to date, most epitopes have been identified (Hsia *et al.*, 1996). The differences associated with variant 29's antigenic distinctness therefore likely occur in a part of VSG that interacts directly with antibodies, consistent with their role in abrogating antibody binding.

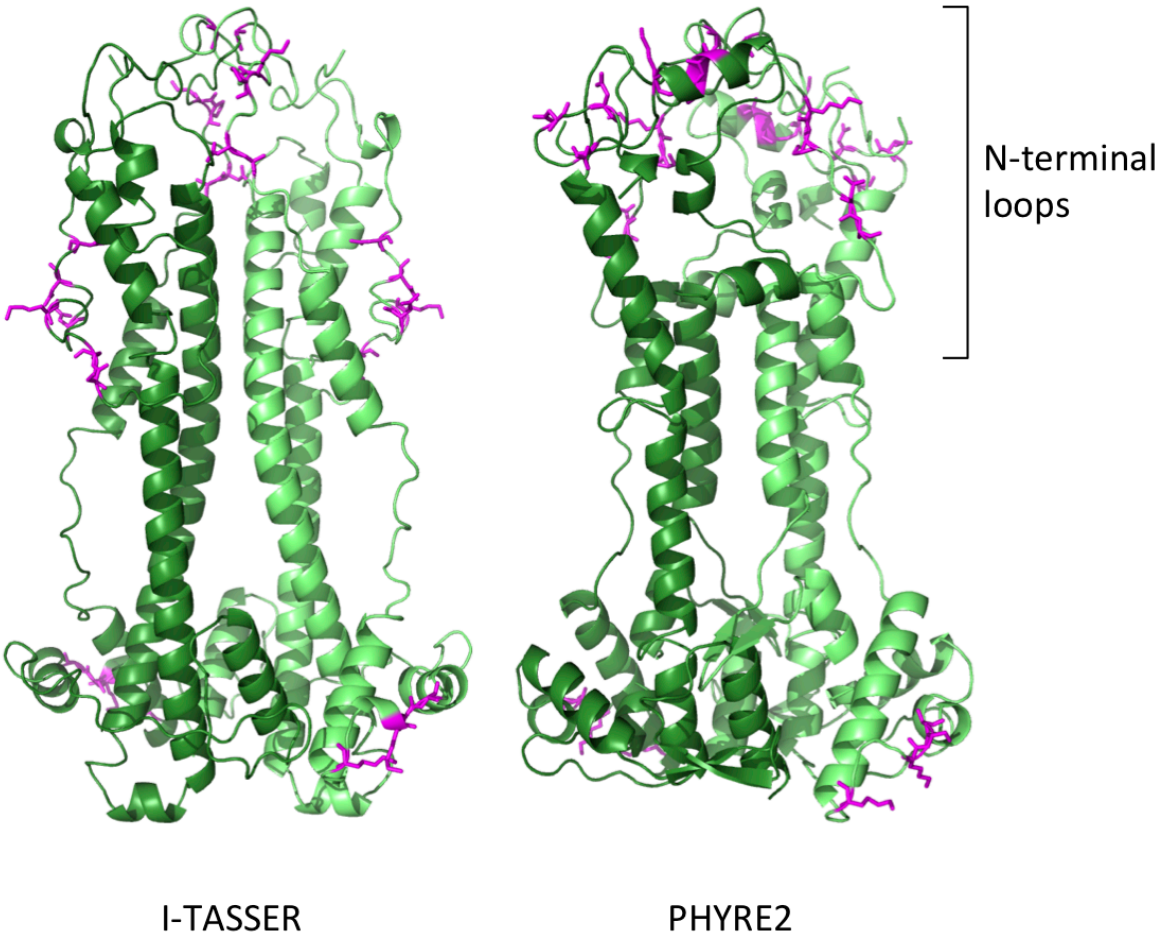


Figure 5.17 Predicted location of residues contributing to mAb-21A binding. Predictions of variant 21's three-dimensional structure were made using I-TASSER and PHYRE2 as with variant 29 (Figure 5.15). The residues identified in Figure 5.13 were indicated on the predicted structure in magenta. Note: analysis of experiments performed following submission of this thesis produces a different figure. Please see Figure 7.41 for an updated version, and Appendix 7.4.4 and 7.4.5 for details.

Predictions of variant 21's three-dimensional structure were also made, and the putative mAb-21A and mAb-21B epitopes described in section 5.4.1 were investigated. These are shown in Figure 5.17 and Figure 5.18. Consistent with its ability to bind live trypanosomes, differences associated with the abrogation of mAb-21A binding were predicted to occur predominantly at the membrane-

distal N-terminal loops of the VSG. For mAb-21B, which was unable to bind to live trypanosomes and likely binds a conformational epitope, three out of four differences associated with variant 23c being unable to bind the antibody occurred at positions expected to form exposed surface loops. The more membrane-proximal position,[89] which may be less accessible in a live surface coat (especially to an IgM antibody), probably makes a greater contribution to the epitope structure in this case.

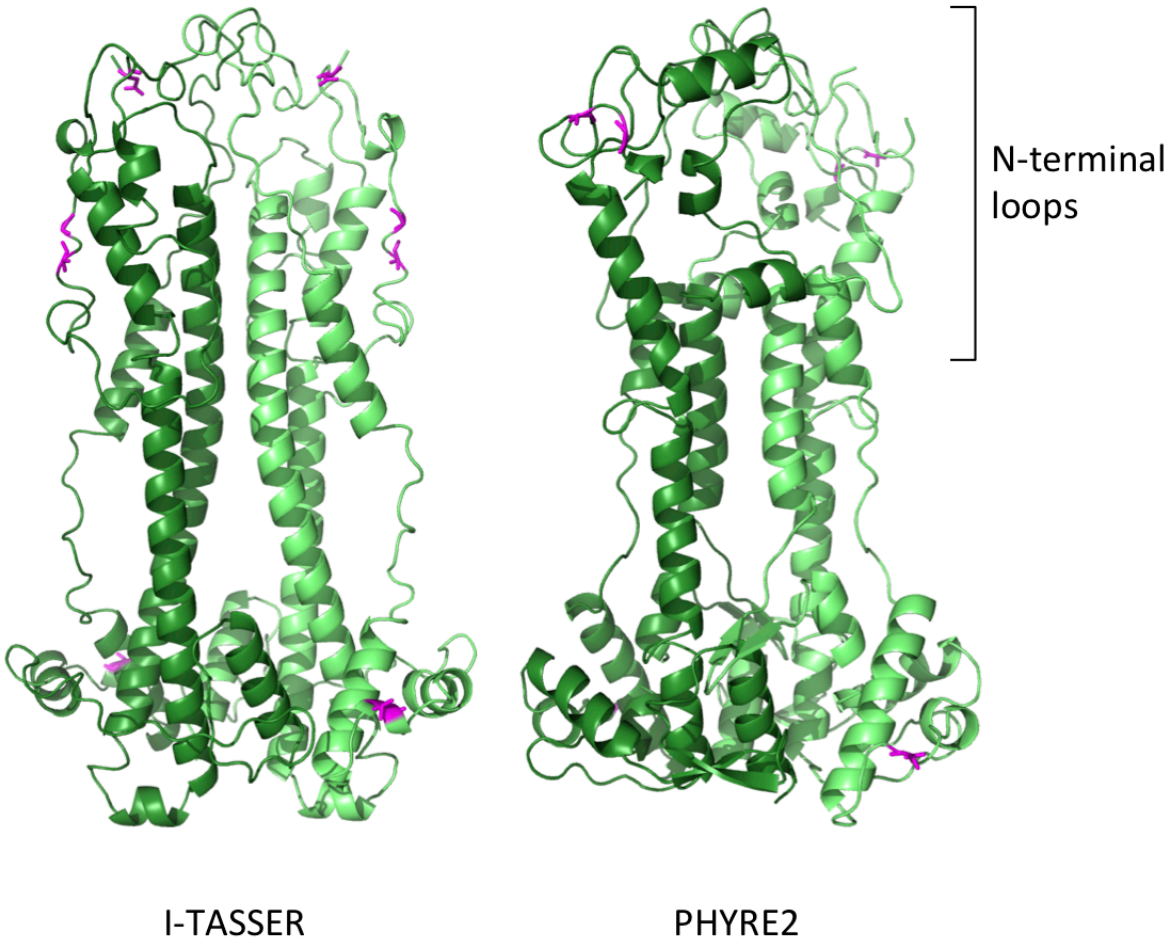


Figure 5.18 Predicted location of residues contributing to mAb-21B binding. The residues identified in Figure 5.14 were indicated on the predicted structures of variant 21 in magenta. Note: analysis of experiments performed following submission of this thesis produces a different figure. Please see Figure 7.42 for an updated version, and Appendix 7.4.4 and 7.4.5 for details.

5.5 Discussion

5.5.1 Key finding of this chapter

In this chapter, the products of related mosaic *VSG*s that had appeared in the same chronic mouse infection were analysed serologically to test the null

[89] For variant 23c, the change at this position was non-conservative (threonine to alanine).

hypothesis that they were antigenically indistinguishable. Immunofluorescence, complement-mediated lysis, and agglutination assays all showed that a mosaic variant that had appeared later in the chronic mouse infections was completely serologically distinct from related variants appearing earlier in that infection.

The same set of related donor *VSG*s was therefore able to combine in various ways through mosaicism to produce more than one antigenically distinct variant. Mosaicism is capable of contributing directly to antigenic variation during an infection.

5.5.2 Further experiments

This study has identified at least one epitope amongst a set of mosaic VSGs that changes over the course of a single infection. Additional experiments could be carried out to further validate these results. For example, a Western blot using the mAbs would verify that they do indeed bind to the variant band seen in the SDS-PAGE. Transmission electron microscopy of the transgenic cell lines would show whether these lines are in fact able to produce the dense layer of VSG that is characteristic of the surface coat. Immuno-gold labelling would conclusively demonstrate that the mAbs raised do indeed bind to the coat itself.

An expanded palette of mAbs would be invaluable in dissecting the epitope structure of mosaic VSGs. In particular, mAbs directed against other Set_14 variants (besides variants 21 and 29), and additional mAbs targeting exposed epitopes, could be used to indirectly map the number of different VSG epitopes and how they change over the course of an infection. If the screening approach used live cells (either through immunofluorescence as described in section 5.3.1.1, or by complement-mediated lysis), rather than acetone-fixed ones, generation of mAbs with functional relevance would be favoured.

In the case of the Set_14 mosaics, structural analyses could reveal the extent to which segmental gene conversion affects the form of the expressed antigen, and would throw more light on its effects on epitope structure. Co-crystallisation of the mosaic VSG with the mAb antigen-binding fragment could be carried out alongside, to specifically show the epitope(s) involved.

In the longer term, improving the reliability and throughput of the expression system would enable a variety of experiments. A one-step transfection approach (replacing the telomere-proximal VSG 427-2 with a defined exogenous VSG) would make the process of generating transgenic lines both quicker and easier. Linking exogenous *VSG* expression with a fluorescence gene like GFP would provide an easily detectable proxy for expression of the expected surface coat, providing a quick way of confirming correct *VSG* expression and increasing the confidence with which VSG-binding antibodies are identified. The scope of this analysis could then be easily expanded to include other mosaic sets.

In essence, a refined VSG expression system could ultimately be a way of creating a stable 'snapshot' of the antigenic profile of what is naturally a constantly changing trypanosome population. This would allow the researcher to experimentally define the parameters that affect the composition of a trypanosome population, that is, identify and quantify the forces that influence antigenic variation in the dynamic *in vivo* population.

5.5.3 Concluding remarks

The experimental results reported in this chapter show that in an infection, related mosaics—expressed *VSG*s formed from the same set of donor genes—can be antigenically distinct. These results demonstrate a functional role for segmental gene conversion during a chronic infection.

With such a large archive of silent *VSG* genes, what could mosaicism contribute to the broader survival strategy of African trypanosomes? In the following chapter, the findings reported here will be evaluated with regards to the key findings of Chapters 3 and 4, and previous work.

Chapter 6

Discussion

6 Discussion

6.1 Introduction

This research set out to assess whether segmental gene conversion plays a prominent role in changing the antigenic profile of the trypanosome population over the course of an infection. The general pattern of antigenic variation consists of three steps: generation of diversity, expression of diversity, and success of diversity under selection. For African trypanosomes, beginning each infection with a large, diverse antigen archive, we might naively have considered *Trypanosoma brucei* antigenic variation within a host as being solely a question of expression: what specific archive *VSG*s are expressed, and when they appear. The large degree of archive pseudogeny hinted that this may not be the case, and the results presented here suggest that *VSG* expression goes hand-in-hand with generation of *VSG* diversity.

The key findings were:

(i) There was a diverse range of antigens expressed during infection: Relatively short infections, lasting 4–5 weeks, can show the expression of a large number of different *VSG*s.

(ii) Expressed *VSG* diversity was generated during infection: Segmental gene conversions can introduce diversity that did not exist in the genome at the start of infection, with a limited number of donors forming large 'sets' of mosaic *VSG*s.

(iii) Mosaics were antigenically distinct: *VSG* diversity produced over the course of infection by segmental gene conversion has functional relevance in the evasion of adaptive immune responses.

These findings will be discussed in the broader context of African trypanosome biology and general pathogen survival strategies.

6.2 Population diversity during infection

6.2.1 'Diversity' comprises species richness and species evenness

The notion of 'diversity' is multifaceted. Diversity is an important aspect of ecological investigation, as scientists attempt to quantify the effects of various

environmental changes on wildlife populations. Broadly, species diversity breaks down into two independent components: species richness (the total number of different species present) and species evenness (the difference in abundance between the species present) (Hill, 1973). These concepts are further complicated by relatedness between species—many different but similar species might be considered a less diverse population than a population of fewer, more distinct species (Leinster & Cobbold, 2012). Both richness and evenness function in African trypanosome antigenic variation, and these will be considered first in sections 6.2.2 and 6.2.3, before moving on to discuss relatedness in section 6.3.

6.2.2 Antigen richness in trypanosome infections is probably related to weak immune selection and the dominance of differentiation-dependent control

There is much 'antigen richness' in the data presented here. Even in those samples where only one distinct variant was detected, it does not seem unlikely that many other sets were present in that host's blood at that time, beneath the limit of detection. How does this pattern come about, and how does it contribute to the African trypanosome's survival?

The observation of many different variants at the same timepoint could be due to a low rate of clearance of old variants, and/or a high rate of activation of new variants. With this host-parasite combination there is evidence for both these processes. The mouse sustains a large number of parasites with TREU927 relative to its size: if the parasitaemia is estimated at ~10^7 ml^{-1}, the total population is ~2 x 10^7 parasites, approximately 1 x 10^9 parasites per kg of body mass[90]. This varies between hosts: cattle, for example, may have a much lower parasitaemia, usually between 1 and 100 parasites per ml, suggesting a total population of up to 2 x 10^6 parasites, approximately 6 x 10^3 parasites per kg of body mass[91] (L. Morrison, PhD Thesis (2004)). All things being equal, mice have fewer immune effectors and less lymphoid tissue per parasite than is the case with most other hosts (La Greca & Magez, 2011; Morrison *et al.*, 2005). In addition, defined mechanisms of immune suppression mean that particularly in the later stages of infection, mice probably exert only very weak immune

[90] Assuming a mouse body mass of 20 g and a blood volume of 2 ml.

[91] Assuming a cow body mass of 320 kg and a blood volume of 20 L, and rounding to one significant figure.

selection, if at all (Radwanska *et al.*, 2008). That the parasitaemia did not generally peak and trough over the course of infection (although see section 6.2.5, below) is consistent with the premise that density-dependent differentiation is the main influence on the size of the parasite population in this mouse model (MacGregor *et al.*, 2011).

Maintaining a large parasite population also favours the occurrence of rarer *VSG* activation events in the population, and by keeping their numbers below the threshold for inducing immune responses, density-dependent differentiation promotes the persistence of these expressors (Gjini *et al.*, 2010). A back-of-the-envelope calculation, estimating a generic switch rate of 10^{-3} per parasite per generation (Turner & Barry, 1989), a generation time of 6 hours (Turner, 1990) and a population of 10^7 parasites for 30 days gives a figure of 1.2×10^6 switch events. This is certainly an overestimate (accounting neither for parasite death nor for non-proliferative non-switching stumpy form trypanosomes) but it does not seem unreasonable to reckon that the total number of switches in the parasite population within each of these infections is at least in the thousands. Antigen richness in the parasite population might therefore be expected, as might the persistence and reappearance of antigens.

How relevant is the observation of antigen richness to other host-parasite combinations? The later stages of infection are generally underrepresented in the data. The observations of Capbern et al., (1977), who identified antibodies against as many as 98 different VATs over 45 days of *T. equiperdum* rabbit infections, make the 31 sets observed in mouse 04 over 10 days not excessive.[92] Through its carrying capacity and rate of immune response, the host species could affect the antigen richness of the parasite population, as could parasite-intrinsic factors such as the activation probabilities of its *VSG* genes (Barry, 1986; Gjini *et al.*, 2010). Still, barring exceptional circumstances, the textbook figure showing sequential, homogeneous parasitic waves is certainly a misleading oversimplification.

A study conducted on *Borrelia burgdorferi* encountered similar non-saturating diversity amongst antigenic variants cloned from an infection (Coutte *et al.*,

[92] Rabbits also have a substantially lower total carrying capacity than do mice.

2009). Extreme antigenic richness may therefore be a factor common to many antigenically variable pathogens in their pre-emption of specific immune responses.

6.2.3 *Polyclonal* VSG *activation may play a role in infection dynamics*

'Antigen evenness' refers to the relative abundance of the different variants in the population. A variant's activation probability could affect its relative abundance both through the normal dynamics of a peak (when a variant is activated for the first time, the population expands until it is subsequently controlled and cleared), and through convergent switching (higher activation probability encourages different parasites to convergently switch to a variant). In the mouse model, where density-dependent factors may control the parasite population (MacGregor *et al.*, 2011), switching to a novel variant is often not advantageous: switchers will not be able to rapidly expand in a clonal manner, as the population is already saturated. Convergent switching may therefore play the more important role in determining the dominance of a particular antigen type. When the immune system eventually removes a dominant variant, space is left for a subpopulation to expand. A polyclonal variant subpopulation, made up of multiple convergent switchers, has a greater 'head-start' over other variant subpopulations (that are not so easily switched to), so even if those other variants have *appeared* in the population they are crowded out by the polyclonal expansion of the convergently-formed VAT (Lee & Van der Ploeg, 1987; Timmers *et al.*, 1987). The number of related *VSG*s with different 3' donations reported in Chapter 4 might indicate convergent switching, representing a specific *VSG* that has undergone different conversion events to become activated (Michels *et al.*, 1983)

In this context it is interesting to note that *Plasmodium falciparum var* genes have a convergent switching hierarchy. For each clone, switching is directional, and a particular *var* variant switches via one of several possible second 'intermediate' variants to a third variant (Recker *et al.*, 2011). This 'single-many-single' switching pattern is proposed to be an advantageous compromise between protecting the archive and maximising the switching options of an individual, and thus is better able to cope with encountering a non-naïve host. Such a highly structured pattern may be particularly necessary for *P. falciparum*,

as its *var* archive is substantially smaller than that of the African trypanosomes *VSG* archive (~60 vs. ~1600 (Kyes *et al.*, 2007; Marcello & Barry, 2007a))

For trypanosomes, convergent switching observed here is a population-level effect, and therefore it is unlikely that it represents anything more than an aspect of the switching hierarchy. Its ultimate effect on the outcome of infection would be inconsequential in most cases, but its occurrence may have implications for our understanding of activation probabilities (activation probabilities may be greater than appreciated) and infection dynamics (VATs may frequently be polyclonal).

6.2.4 Hierarchy is robust, but further knowledge of later switching mechanisms is required

Despite observed antigen richness, expression of diversity is evidently controlled: given that potentially thousands of switches could occur, 'only' 83 different sets were detected over all 22 infections analysed, and their appearance varied over time with both early and late variants distinguishable. At a gross level, the timing of appearance of a *VSG* set was associated with its putative genomic location, consistent with previous studies (Morrison *et al.*, 2005; Robinson *et al.*, 1999). On the other hand, that some sets represent damaged genomic copies while some intact genomic copies are not detected at all suggests that 'ease of activation' is not appreciably affected by intactness, discussed in further in sections 6.4.2 and 6.4.4.

The occasional re-expression of some early-appearing genes in the later stage of infection is consistent with the tendency of strains to revert to their original antigen type following passage (Gray, 1965), as these genes have apparently high activation probability. Turner (1999) made a distinction between 'real' and 'effective' switch rates to highlight this occurrence, predicting that 'effective' switch rates decrease as infections progress. If early *VSG*s do indeed have a greater activation probability, and the mouse is capable of only weak immune selection, why was this not observed more often? It could be that as array genes become involved in 'stage 2' antigenic variation (see section 6.2.5), *VSG* specific homologies become a more important determinant of activation probability (that is, the switch-off *VSG* affects the switch-on *VSG*). As a mechanism, switch rates changing relative to the *VSG* that is expressed has been proposed in the context

of mosaic *VSG*s by Pays et al. (1985a). This relativism could also include 3' donation or even *VSG*-specific UTR flanks (discussed further in section 6.4.4). Relativistic switching would minimise unproductive switching to 'early' VATs, and thus would have particular relevance later in infection. Detailed examination of 3' donation and flanking regions may increase our understanding of the genetic mechanisms underlying hierarchy, shedding more light on the events that come to the fore after the early stage of infection.

6.2.5 'Stage 2' antigenic variation may become the 'new normal' as early VSG*s expire*

The later stages of infection were associated with several characteristic features when compared with the earlier stages:

a) Greater levels of within-set variation: a lack of 3' donation and mosaics.
b) Greater numbers of identifiable array *VSG*s corresponding with the sets.
c) Disappearance of earlier-occurring sets (Set_08, for example).
d) Elevated level of parasitaemia.
e) Follows a deep trough in parasitaemia—a population crash.

These patterns have been referred to as 'stage 2' antigenic variation[93] (L. Marcello, Ph.D. thesis, 2006, University of Glasgow).

One explanation for these phenomena is that the host has finally raised productive immune responses against all the easily-accessible 'early' variants. As those trypanosomes are destroyed, subpopulations of variants activated less easily, previously suppressed by density-dependent differentiation, can proliferate and rise to pre-eminence. This shift in expression marks a distinct change in the use of the archive, and the appearance of array genes is coincident with an increase in within-set variation of mosaics and 3' donations.

Is 'stage 2' antigenic variation the 'new normal' as the infection progresses? The shift occurs approximately two weeks before the mouse mean time of death, so it is impossible here to separate 'stage 2' antigenic variation from the

[93] Not to be confused with second-stage sleeping sickness, when parasites invade the central nervous system.

terminal stages of trypanosomiasis. Conceptually, the parasitaemia up to the crash might represent the first 'peak', that is, the switching between and eventual exhaustion of a block of *VSG*s 'between which switching occurs rapidly' (Gjini *et al.*, 2010). In larger hosts, where the kinetics of antigenic variation are accelerated (Barry, 1986), the exhaustion of the initial 'block' would occur more rapidly, and 'stage 2' antigenic variation would assume prominence for the rest of infection. If the infection lasted long enough, there may be further stages, perhaps corresponding with groups of even less accessible archive *VSG*s. Investigating *VSG* expression with relation to the genome in a range of hosts may provide data to complement this study, helping to dissect these patterns.

6.2.6 *Easily-accessible* VSG*s may help an invading population form a 'bridgehead'*

If the later stage is the normal state of infection, why have an early stage of infection at all? The possible risks associated with segmental gene conversion (outlined in section 6.4.8) might make it advantageous for the trypanosome population to build up numbers in the new host, forming a 'bridgehead'. Analogous patterns have already been characterized in metacyclic trypanosomes present in a tsetse fly's salivary glands, which show heterogeneity in expressed *MVSG*s, pre-empting circulating immunity when they are deposited during tsetse fly feeding (Barry *et al.*, 1998). By coming equipped with a number of easily-accessible, functional *VSG*s, the initial bloodstream form population can proliferate and begin antigenic variation without having to rely on riskier recombination events. During this period, there are abundant opportunities for members of the population to attempt more risky, low probability segmental switching events, forming a population to initiate 'stage 2' antigenic variation (section 6.2.5).

Having easy access to a set of functional *VSG*s would also be more competitive in the early stages of mixed infections, particularly where differentiation controls parasitaemia, as described in section 6.4.8. Strain-specificity in this 'block' of *VSG*s could be easily developed, by direct exchange between ES and telomere for example, allowing reinfection or superinfection of partially-immune hosts. *VSG* expression and infection progression is summarized in Figure 6.1.

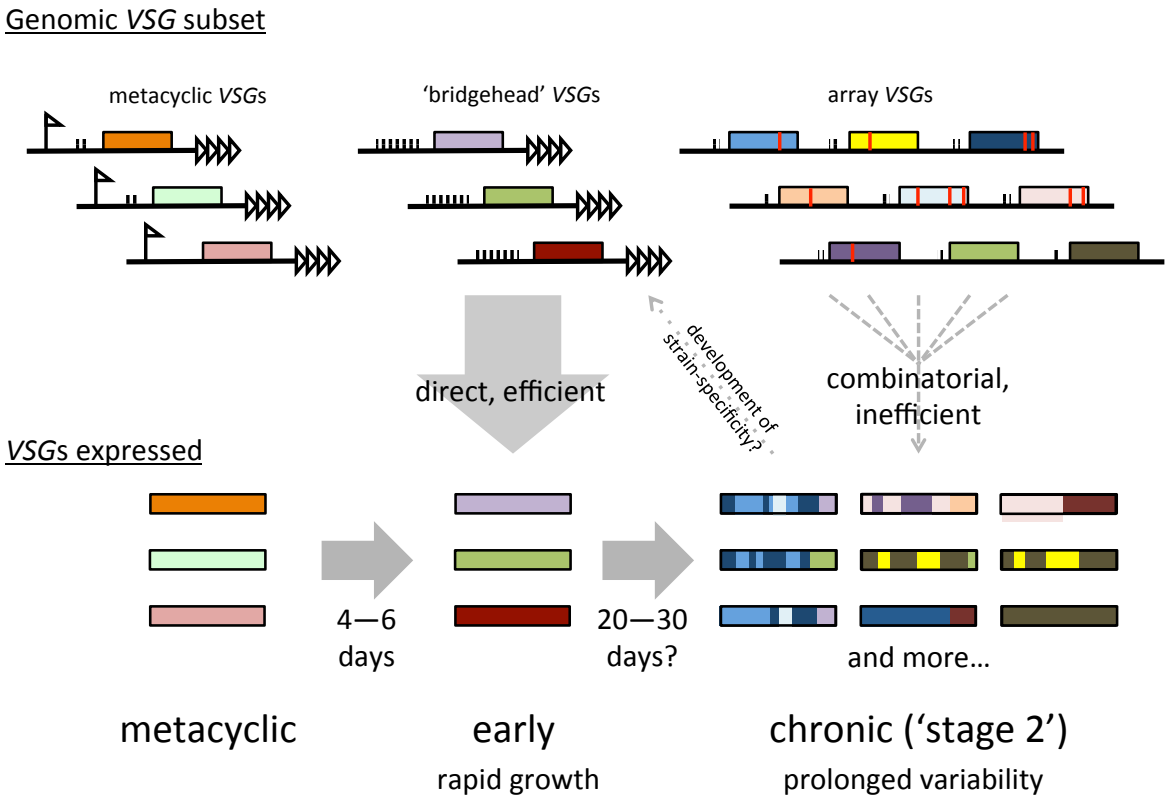


Figure 6.1 *VSG* expression and the progress of infection. In this simplified model, the three subsets of genomic VSGs—metacyclic, 'bridgehead' and archive—are shown at the top. As the infection progresses, different parts of this repertoire are used. The infecting metacyclic trypanosomes express a range of metacyclic *VSG*s, allowing them to overcome previous immunity. Over the next few days they switch to the telomere-proximal, 'bridgehead' *VSG*s. This easily-activated 'intermediate' subset allows the trypanosome population to outcompete strains that progress too rapidly to inefficiently-activated (and mostly pseudogenic) subtelomeric archive *VSG*s. Progression to array *VSG* expression has occurred by c. 20 days later (Morrison *et al.*, 2005) and is frequently accompanied by segmental gene conversion events that can both open up the wealth of archive diversity and generate variability in the process. The 'bridgehead' (and indeed the metacyclic) *VSG* population can be replaced, perhaps via gene conversion from the expression site as indicated, a process that could develop strain-specificity and facilitate re- or super-infection of subsequent hosts.

6.3 The patterns of generated diversity

6.3.1 Expressed VSGs *were frequently different from their donors and were often unique*

When comparing expressed *VSG*s between infections and with the genome, patterns of within-set variation were clear: point mutations, 3' donation, and mosaicism. Only five clones matched their corresponding genomic copy exactly. Each of these patterns will be discussed, before a broader discussion on the consequences of infection-generated diversity for African trypanosome antigenic variation.

6.3.2 Point mutations are probably a minor process in antigenic variation within a host

Point mutations were detected between related *VSG*s within an infection. Although there was not time to measure directly the degree to which they were artefactual, calculations and comparisons indicated that most of them were genuine. Despite this, there was no evidence that point mutations were contributing to antigenic variation directly: non-synonymous mutations were not especially favoured, and they did not tend to accumulate in parts of *VSG* associated with antigenic variation. Partly, this could be due to convergent switching alongside weak immune selection on the part of the host, and the crowding out of point mutation variation by more substantial between-set variation and mosaicism. It is possible that point mutation has a more important role when access to diversity is limited: when more substantial switch events are disallowed, point mutations in *VSG* are demonstrably able to abrogate the binding of specific monoclonal antibodies (Baltz *et al.*, 1991; Lu *et al.*, 1994). Would this often occur in real infections? Few pathogens generate sufficient antigenic variability to maintain a chronic infection from point mutations alone: those that do are viruses, able to exploit a latent phase and accumulate mutations in the absence of constant immune selection (Bowen & Walker, 2005; Wolinsky *et al.*, 1996). In *T. brucei*, point mutations have not been able to confer resistance to physiologically relevant polyclonal antibodies, as antibody responses against *VSG* are too broad (Graham & Barry, 1996). Furthermore, given the apparent ease with which segmental gene conversion occurs (section 6.4.2) it is implausible that African trypanosome antigen variability routinely becomes so restricted that they come to rely on point mutations for persistence within a host.

Nevertheless, point mutations could play an important role in the long-term diversification and therefore survival of African trypanosomes in a host *population*, similar to the process of antigenic drift by influenzavirus (Boni, 2008), and may be a useful mechanism to build up strain specificity (Hutchinson *et al.*, 2007).

6.3.3 Segmental gene conversion is a frequent occurrence in antigenic variation

As a means of generating an expressed *VSG*, segmental gene conversion is sometimes characterized as an ancillary process, only undertaken by 'desperate' parasites to make the best of a 'damaged' archive or to drag out an infection when the host has developed immunity to the products of intact archival genes (Barbet & Kamper, 1993; Barry, 1997). The chronic mouse infections described in Chapters 3 and 4 produced an abundance and variety of segmental gene conversion events. If these events are indeed genuine (for reasons of time they could not all be verified by PCR) this pattern is inconsistent with both of these models. A subsidiary role for segmental gene conversion does not explain the rapid appearance of mosaics (within 3 weeks of infection), the range of mosaic diversity seen within a single infection (15 distinct variants) or the evidence for mosaicism in 28 out of 83 sets analysed, including between 'functional intact' donor genes. That the finer-grain 'within-set' variation of segmental gene conversion is clear amidst complete 'between-set' variation suggests segmental gene conversion is a principal aspect of *T. brucei* antigenic variation. How do related-but-different mosaics and 3' donations observed here affect our understanding of the mechanics of *T. brucei* segmental gene conversion? What are their roles in antigenic variation, and how might they contribute to the trypanosome's survival and continuing evolution?

6.3.4 3' donation is akin to variable region cassette exchange

The 3' donation events for which donors could be found are consistent with a model whereby the 3' end of recombination occurs within the CTD-encoding region of the resident *VSG*, as many of the 3' donors identified were probably telomere-proximal. Therefore, just the variable region of the telomere-proximal *VSG* is replaced, and the region encoding the structural CTD is (partially) conserved. In this way, 3' donation is akin to the use of variable region cassettes by numerous antigenically variant bacteria: *N. gonorrhoeae pilE* and *A. marginale msp2*, for example (Vink *et al.*, 2011). Possessing just a single copy of a conserved region is an efficient use of genomic space, and prevents deleterious mutations, fixed by genetic drift, from precluding expression of a variable region. If 3' donation is able to efficiently exchange domains in *T. brucei*, why are there so many *VSG* CTDs in the archive? The mechanism of gene

conversion may demand a region of homology in the 3' flank that is routinely provided by the CTD-encoding region, and it is possible that variability in the 3' flank of recombination, which overlaps part of the *VSG* CTD-encoding region, plays a role in determining activation probabilities of *VSG*s, as it does in *B. hermsii* Vsp/Vlp (Barbour *et al.*, 2006). Comparisons between *T. brucei*, *T. congolense*, and *T. vivax* suggest that CTD exchange evolved only in the *T. brucei* line, whereas *VSG* diversity is ancestral (Jackson *et al.*, 2012). It is therefore possible that archive CTDs are degraded relics that have freely accumulated mutations in the absence of their expression (Taylor & Rudenko, 2006).

6.3.5 Are mosaics built up progressively?

What do the observed patterns of mosaicism tell us about the development of mosaics? The current model holds that sequential segmental gene conversion events progressively build up in an expression site *VSG* to form a 'string' of related mosaics (Barry *et al.*, 2005). This is analogous to the mechanism seen in *Anaplasma*, where later *msp2* variants showed more complex patterns of segmental conversions (Futse *et al.*, 2005; Palmer *et al.*, 2007). Evidence for such a model of 'progressive mosaicism' was found in two sets where it appears that some mosaics have accumulated more segments than others (section 4.5.8). However, other mosaic lineages are less clear, showing either no direct relationship between sequentially-expressed mosaics, or an apparent decrease in complexity (the apparent number of segmental gene conversion events required to produce the mosaic) over time.

Given the antigenic richness of the large parasite population alongside weak immune selection (discussed in section 6.2.2), it is likely that the lack of evidence for progressive mosaicism reflects difficulties in isolating an individual mosaic string. Perhaps it is only once a mosaic has stably formed that the parasites expressing it grow to great enough numbers to be detected, and once there, these mosaic expressors dominate and persist in a similar way to non-mosaic sets described in Chapter 3. Furthermore, a mosaic string may indeed become increasingly complex as an infection progresses, but without a potent immune response rapidly removing those variants that have not changed, the more complex variants will remain rare and hard to detect, particularly in an

infection controlled by differentiation. Meanwhile, within the parasite population, other parasites may reactivate the first donor, which then embarks on a new process of segmental conversion and mosaicism, forming a parallel lineage. Deeper, more frequent sequencing may be necessary to better witness longitudinal development of mosaic strings in an infection. It may also be that 'complexity' is a poor metric—for some sets apparently simple mosaics might require the interaction of donor sequences that do not so readily combine.

6.3.6 Mosaics appear quickly

How quickly are mosaics made? Some expressed mosaics are complex, with as many as eight distinct segments, and are detected as early as day 21. It is therefore unlikely that the addition of a segment occurs at the background switch rate of maximally 0.01 per cell per division (Turner & Barry, 1989), since on average only one segment would be added every ~1.5 days (assuming a doubling time of six hours and that all segmental conversion events are successful) and require a large number of predecessors in the lineage to provide sufficient substrate for segmental conversion to occur. The number of diverse mosaics identified in a given infection, coupled with the lack of predecessors detected in previous samples of that infection, suggests that segmental conversion occurs much more rapidly. One possibility is that mosaics can be 'nucleated': once an initiating donor *VSG* is present in an expression site, other donors can contribute easily thanks to extensive sequence homologies (Pays *et al.*, 1985a). It is also possible that the RAD51-independent recombination pathway that is the most promising candidate for driving segmental recombination occurs at a higher rate in pleomorphic cells than in the monomorphic lines in which it was characterized (Conway *et al.*, 2002).

In *Borrelia burgdorferi* antigenic variation, 'intermittent recombination' has been reported, where replication apparently 'skips' back-and-forth between templates to produce a multiple-segmented mosaic, possibly in one step (Coutte *et al.*, 2009). Interestingly, intermittent recombination was found to occur more readily in a mutant lacking a component of the replication machinery, RuvA helicase, involved in Holliday junction resolution (Lin *et al.*, 2009a). It is possible that a similar process could occur in *T. brucei*, helping to produce some of the multiply-segmented *VSG* mosaics here.

6.3.7 Mosaics are formed from similar donors but junctions do not require much identity

Regardless of the speed at which complex mosaics are formed, there is a clear preference for using homologous donors. The maximum divergence between donors observed here was 0.269 differences/nt (that is, 73.1% identity), so it remains the case that all mosaics identified to date are formed from donors of the same *VSG* 'family'. Homology between donors may reflect structural demands—perhaps combinations of more variant donors do not form a stable VSG fold. Alternatively, and more likely, it might reflect the mechanics of recombination: segmental gene conversion between more diverse donors is disfavoured by the recombination machinery, even though it could produce a functional *VSG*. It would be interesting to test these options, by artificially producing non-homologous mosaics and attempting to express them in trypanosomes.

Despite general homology across their whole length, mosaic gene conversion does not occur in regions of extensive identity. For some junctions, there was no region of complete identity between sequences, and the size of some converted segments was very short. This is in common with *Neisseria gonorrhoeae* and *B. burgdorferi*, where regions of identity at a junction can be short and segments small (Coutte *et al.*, 2009; Criss *et al.*, 2005). Nonetheless, in these pathogens too, donors show homology with one another. In *A. marginale*, segmental gene conversion occurs in a variable region between conserved flanks (with a boundary of recombination 'anchored' in one, the other, or both flanks) (Futse *et al.*, 2005; Meeus *et al.*, 2003), and in *B. burgdorferi*, invariant regions in the expressed antigen are shared amongst donors (Zhang *et al.*, 1997). Some homology between donors may be necessary, aligning them and preventing disorderly and inexact segmental gene conversion from introducing radical changes into the expressed *VSG* (such as large additions or contractions, or changes to the number or spacing of critical cysteine residues), or from mis-directing recombination elsewhere. Shorter conversion events in parts of the gene that varies between donors would generate greater diversity than longer conversion events that require identical flanks. That segmental conversion events do not seem to concentrate in one part of the *VSG*

gene is also consistent with the premise that segmental gene conversion is broadly an undirected process.

The appearance of identical mosaic junctions in different infections is intriguing, but further work is required to confirm it is more than an artefact or an exception. It would be interesting to examine whether there are indeed specific segmental conversion events that occur more readily between donors, as this might show that there is more structure to *T. brucei* segmental gene conversion events than is currently reckoned.

The need for homology has some important consequences. As only 40% of *VSG*s are in families (Marcello & Barry, 2007a), does this mean that mosaicism is restricted for most archive *VSG*s? Certainly, if non-family *VSG*s are able to interact it is at a much lower rate than that between family *VSG*s, and so would appear much later on, if at all. 3' donation hints that lower-homology interactions are still possible, although these may not involve the same recombination machineries. If homologous donors are favoured in mosaic interactions, does this restrict the diversity that mosaicism can introduce? This question is addressed in sections 6.4.5 and 6.4.6 below.

6.4 The role of generated diversity

6.4.1 With such a large archive, why does T. brucei need segmental gene conversion?

If segmental gene conversion is indeed an integral mechanism, it means that expressed *VSG*s are frequently different from all of the silent archive *VSG*s. This is not exceptional amongst antigenically variant pathogens—indeed, only *Plasmodium* and *Giardia* appear to have mechanisms whereby antigens do not routinely segmentally recombine during an infection (Prucca & Lujan, 2009; Scherf *et al.*, 1998). However, most segmentally-recombining pathogens have only small archives,[94] and therefore depend on combinatorial diversity for antigenic novelty. For African trypanosomes, which possess an archive of silent genes unequalled in size, what might be the advantages of a dysjunction between archive and expression? This question is all the more important given

[94] *Babesia bovis*, with ~350 silent *ves1-α* and ~80 *ves1-β* constitutes another exception (Brayton *et al.*, 2007).

the potential for error that segmental gene conversion brings to the process: the dangers of introducing frameshifts and stop codons must be overcome or accommodated for it to be a stable strategy.

There are four possible advantages of a dysjunction between archive and expression, which can broadly be described as Reparation, Relaxation, Hierarchy and Variation. These are discussed in the following sections.

6.4.2 Segmental gene conversion can repair damaged VSG*s*

Evidently, with an archive where the majority of silent *VSG*s are otherwise unusable, segmental gene conversion provides a means for their reparation. Sixteen of the 28 mosaic sets utilized segments from a total of 24 damaged genomic copies, and 26 of the 47 sets showing 3' donation corresponded with genomic copies with damaged C-terminal domains.[95] Altogether 53 out of the 91 genomic *VSG*s used were damaged.[96] Utilization of pseudogenic donors was previously shown for the mosaics and 3' donations described by previous studies (Aline *et al.*, 1994; Barbet & Kamper, 1993; Roth *et al.*, 1989; Thon *et al.*, 1989). It would be interesting to see whether *T. congolense* and *T. vivax*, both of which have a greater proportion of non-pseudogenic archive *VSG*s (Jackson *et al.*, 2012), show a similar degree of segmental gene conversion. If reparation is the main advantage of a disjunction, they might be expected not to.

Reparation of damaged *VSG*s can be clearly observed as a consequence of segmental gene conversion. Yet segmental gene conversion also occurs to undamaged *VSG*s, or to parts of a damaged *VSG* that are not damaged, and it generates extensive variability in the process. Segmental gene conversion may also generate damage, through out-of-frame recombination (in this study, one 3' donation event rendered an atypical *VSG* dysfunctional). The supposed advantages of archive reparation as a factor driving segmental gene conversion also begs the question: if the expression of damaged archive genes is so important, why not have a wholly intact archive in the first place? Is it simply that selection on each individual *VSG* for functionality is so low that they drift

[95] This figures includes 11 'atypical' *VSG*s, whose functionality is uncertain.

[96] This total relates to those genomic *VSG*s for which *any* donor sequence could be found, thus includes incomplete reads and read assemblies. See Appendix 7.2.9 for further details.

towards degeneracy, and segmental gene conversion has developed as an imprecise compensatory mechanism? If that is the case, how is it that *T. congolense* and *T. vivax* have managed to maintain equally large archives with a lower proportion of pseudogenes? Differences in their lifecycles may provide some answer: *T. brucei* goes through tighter population bottlenecks in the tsetse fly than *congolense* and *vivax* do (Oberle *et al.*, 2010; Peacock *et al.*, 2012), which may help to fix pseudogenizing mutations.

6.4.3 Segmental gene conversion may relax demands for functionality, unencumbering archive evolution

Perhaps it's the other way around, and that the ability to be repaired has relaxed the demand for silent *VSG*s to be completely functional. By forming a further level of separation between an archival *VSG* and its expression, 'reparation' permits silent *VSG*s to mutate—but to mutate within limits. Imagine 'expression ability' was instead a binary functional/non-functional for all archival *VSG*s. A new *VSG* gene appears—is 'born'—perhaps by duplication. For this gene to be a useful addition to the archive, it must encode both a functional coat and a divergent antigen. These are opposing pressures, and the likelihood is that as mutations accumulate, it would become dysfunctional and before it became usefully divergent. Once it has become a dysfunctional gene and is completely released from the selective pressure of expression it is more than likely to degrade and 'die' completely: the framework alpha helices and the conserved cysteine residues necessary for *VSG*'s function as a barrier would not be maintained. By permitting 'shades of grey' in functionality, segmental gene conversion might relax this tension, encouraging more radical divergence in archival genes whilst retaining a mild level of selection for function, as described in Figure 6.2.

There are some problems with this theory: the homology apparently required between segmental gene conversion donors means that the 'mild selection' of partial expression is limited in extent and couldn't sustain useful divergence beyond the threshold of recombination. In addition, the greatest 'relaxation' appears to be on the 3' end of the *VSG* gene, where divergence is not clearly favoured in antigenic variation. Nevertheless, re-evaluating selective pressures on the antigen archive in light of the role of segmental gene conversion in *VSG*

expression will undoubtedly be important in understanding how *VSG*s evolve. This is considered further in section 6.4.7 below.

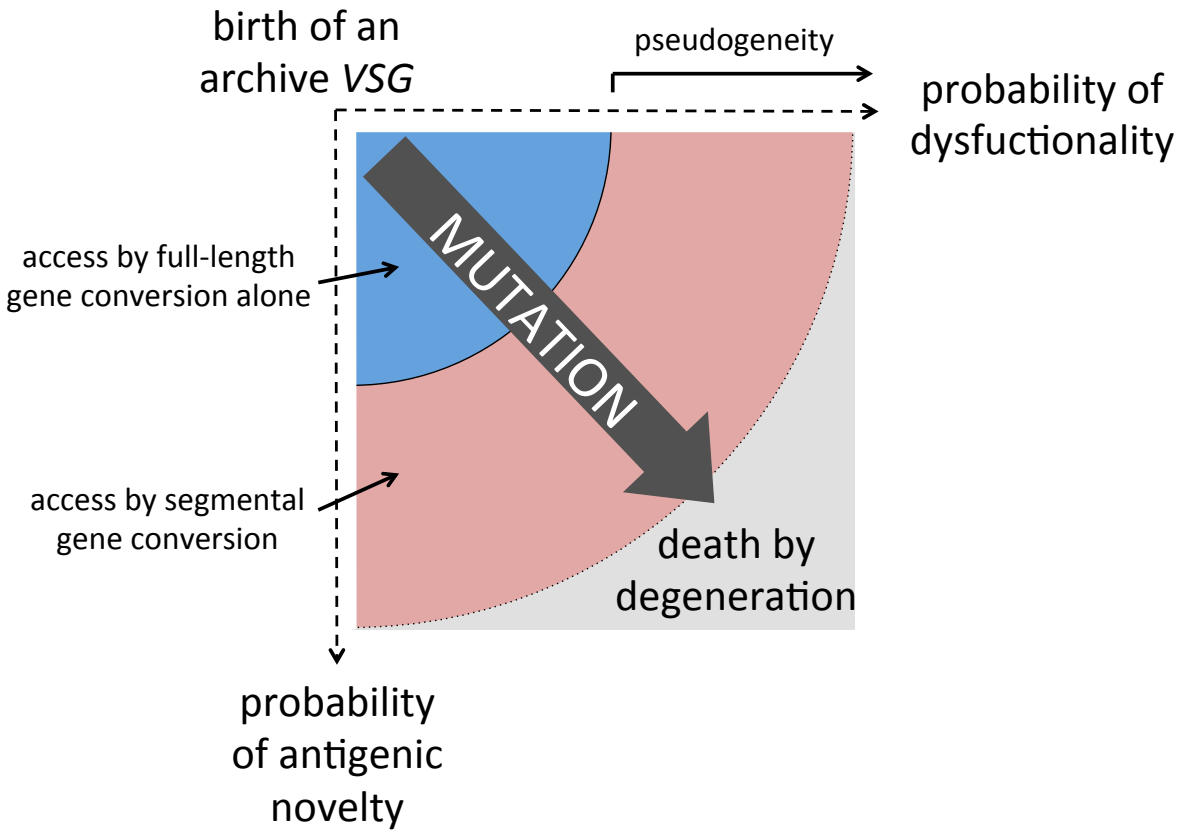


Figure 6.2 Segmental gene conversion could harness the diversity produced by archive evolution. In this simplified scheme, a *VSG* is 'born' by duplication of an intact *VSG*, either from elsewhere in the archive or by retrograde copying from the expression site. By accumulating mutations, the probability of becoming antigenically novel increases, but so does the probability of accumulating pseudogenizing mutations (and thus the 'death' of the gene). The probability of evolving an useable, antigenically novel VSG is increased if a degree of pseudogeneity is permitted, such as through use of segmental gene conversion.

6.4.4 Segmental gene conversion staggers expression and contributes to hierarchy

Mosaic *VSG*s generally appear later in infection. Indeed, in the data presented here, unambiguous mosaics do not appear until day 17 of infection. The requirement for segmental gene conversion to activate certain archival *VSG*s might be a mechanism for retarding their expression. By 'insisting' on mosaicism or 3' donation, the roll-out of the archive can be staggered (Thon *et al.*, 1990). Yet there are other mechanisms by which the expression of a particular genomic *VSG* could be delayed without compromising its integrity, and it is notable that some *VSG* sets corresponding to intact genomic copies (Set_21 for example) consistently appeared late in infection with little within-set

variation. Such a pattern indicates that hierarchy could be determined by other means beyond imprecise segmental gene conversion. *Borrelia hermsii*, for example, uses the positions and homologies of flanking sequences to exquisitely modulate *vsp*/*vlp* gene activation rates (Barbour *et al.*, 2006).

Segmental gene conversion could relate to hierarchy more subtly. Progressive mosaicism is implicitly hierarchical, as successive events accumulate in an expressed gene (Futse *et al.*, 2005; Pays *et al.*, 1985a). More broadly, the mechanisms behind segmental gene conversion could play a role in hierarchy beyond mosaicism or reparation. Segmental gene conversion is a mechanism that is clearly promoted by homology between donors, but these homologies do not necessarily need to be inside the NTD-encoding region of the activated genes. The formation of complex mosaics may be one aspect of a looser, more promiscuous, less efficient homology-driven recombination mechanism. The increasing prevalence of mosaics and 3' donations during an infection might signal the late-stage pre-eminence of this switching mechanism: a mechanism that causes *VSG*-specific homologies to determine switching hierarchy, rather than non-specific homologies between the 70-bp and telomere repeats. Such a mechanism would clearly have relevance to the 'block' structures of mathematical models (Gjini *et al.*, 2010).

6.4.5 Segmental gene conversion directly contributes to antigenic variation

Segmental gene conversion in trypanosomes may be an important, or even the only, way of generating and expressing antigenic variability, as it is in *Anaplasma* MSP2 (Zhuang *et al.*, 2007), *B. burgdorferi* VlsE (Zhang & Norris, 1998) and *N. gonorrhoeae* PilE (Hamrick *et al.*, 2001). In these pathogens, a small number of donor genes can combine in various ways to produce a wide range of epitope structures, and therefore dysjunction between archive and expression can be a source of diversity over the course of infection. Could segmental conversion in trypanosomes generate a further layer of antigen variability beyond that directly encoded in the archive? This question was addressed directly by the present work. Here, a set of related mosaics, Set_14, constructed from the same subset of donor genes, were identified both within a timepoint and from different timepoints of the same infection (Chapter 4). The products of these Set_14 mosaics were tested serologically, and it was found

that while antibodies raised against the earlier variants cross-reacted with one another, the later-appearing variant was antigenically distinct. This finding is consistent with previous studies on African trypanosome mosaicism, which have also shown some antigenic variability amongst related mosaics using monoclonal antibodies, although not from immediately longitudinal samples (Barbet *et al.*, 1989). Segmental gene conversion therefore contributes directly to antigenic variation.

Alongside 'between-set' variation, how important is this more subtle kind of variability to the overall process of antigenic variation? In other antigenically variable pathogens, accumulating segmental conversion events in the expressed antigen is the main source of antigen diversity, with *B. burgdorferi* and *A. marginale* both able to persist for years with only a handful of silent antigen genes in their repertoires (Coutte *et al.*, 2009; Zhuang *et al.*, 2007). In *T. brucei*, despite numerous genetic differences between related mosaics of Set_14, the yield in terms of (experimentally verified) antigenic novelty was not great. Out of five mosaics generated from the same palette of four donors, only one was antigenically distinct from the others. This seems not to be a particularly good use of the archive: combining four related silent genes to produce only two distinct expressed variants. Moreover, the antigenically distinct variant was divergent to 0.125 differences/aa in the NTD: unusually high for mosaic sets, with most other sets divergent to only 0.05–0.10 differences/aa in the NTD. It is therefore reasonable to presume that many of the members of these other sets are also incompletely antigenically distinct. It is possible that in longer infections and with stronger immune selection, the parasite population would reveal a greater number of distinct variants readily made from the same set of donor genes, or the segmental involvement of *VSG*s that do not share overall homology. Still, as far as these easily-made mosaics here are concerned, antigenic novelty does not come readily.

6.4.6 Complete antigenic distinctness may not be necessary in the context of similar mosaics

In the context of a complex infection, complete antigenic novelty may not be essential for each step in a progressive mosaic string. A slight change to epitope structure may decrease an antibody's binding affinity sufficiently to enable escape, particularly if those antibodies are at low titre and their clearance can

be enhanced by hydrodynamic flow-mediated antibody clearance (Engstler *et al.*, 2007). As the immune response develops with the assistance of antigen presenting helper T-cells, antibodies increase their affinity for their specific epitope and lose valency as they switch from IgM to IgG isotypes. If the antigens being presented and targeted are those of previous mosaics, increasing affinity for an out-of-date epitope and decreasing the avidity of the antibodies could assist in immune escape for the later mosaic expressor. Reinfection studies may help to resolve this question. In addition, mathematical models of infection indicate that incomplete cross-reaction can benefit a later mosaic by keeping that later variant's population below the threshold for specific antibody induction (Gjini *et al.*, 2010; Johnson *et al.*, 2012). If partial antigenic distinctness were sufficient for immune escape, segmental gene conversion may be a mechanism for expressing *epitopes* separately, rather than generating completely distinct variants.

Viruses can induce a phenomenon known as 'original antigenic sin' whereby the immune system mounts a non-neutralizing response against a later variant thanks to its immunological memory of a related variant (Kim *et al.*, 2009). Similar mechanisms may be relevant to *VSG* mosaicism: in a field population, memory may assist mosaic expressors in reinfection or superinfection.

6.4.7 Would mosaic 'saving' be an efficacious way of increasing archive diversity?

If introducing antigenic novelty was a key advantage of segmental gene conversion we might expect parasites to have means of circumventing the difficulties of creating useful mosaics. An obvious one would be the ability to retain the new information: a retrograde pathway for the expressed mosaic *VSG* to become part of the archive. Such *VSG* saving has not yet been clearly observed, but the presence of abundant minichromosomal telomere ends makes the mechanism of telomere exchange a possible pathway (Taylor & Rudenko, 2006). The advantage of saving this information for future infections might make intuitive sense from the point of view of an individual parasite: it has gained another easily accessible variant to deploy during its next infection. However, if 'damage' to archival genes can be frequently repaired during an infection as *VSG*s are activated and segmentally converted across the

population, as suggested by the results presented here, saving a repaired gene may not be exceptionally useful from this perspective.

What about archive diversity? For retrograde transfer to *increase* the antigenic diversity of the archive, it would have to occur only when the expressed *VSG* is antigenically novel (otherwise it would essentially only duplicate existing information, which could be detrimental where there is selection for limiting archive size). Given that four out of five mosaics tested here had some antigenic cross-reaction, it is likely that of all possible mosaics, the majority are not antigenically distinct from all of their donors or from one another. While it is possible to envisage mechanisms encouraging the saving of a mosaic that is antigenically distinct to that infection[97] it is unlikely that a mechanism could assess archival antigenic diversity, and save only those mosaics that increase it. More likely would be a mechanism that operates at a constantly low background rate, and thus would more frequently duplicate, rather than expand, antigen diversity.

Nevertheless, this form of gene 'duplication' is probably an important pathway for replacing those archival *VSG*s that have succumbed to degeneration, producing intact members of the *VSG* subfamily that can accumulate mutations and diverge. In addition, acquiring a new telomeric *VSG* would shuffle hierarchy, providing an accessible 'bridgehead' *VSG* that could be beneficial in the early stages of superinfection or reinfection (see section 6.2.6).

6.4.8 Are the disadvantages of segmental gene conversion indeed serious?

This section has been premised on the assertion that segmental gene conversion is accompanied by particularly deleterious disadvantages: the risk of frameshifts and errors, the loss of a surface coat, and so on. Are these disadvantages really that great?

[97] For example, perhaps retrograde transfer is a stochastic process whose likelihood of occurring increases with the amount of time that has passed since the most recent segmental gene conversion event, thus favouring *VSG*s produced by those segmental gene conversions that allow the parasite to survive for longer.

If segmental gene conversion has evolved as a mutator phenotype, it seems likely that mechanisms have evolved to control the immediate problems that segmental recombination brings to an individual parasite. Feedback loops to promote reparation of a damaged expression site-occupying *VSG* (hinted at by the response to lack of *VSG* mRNA (Smith *et al.*, 2009)) backup *VSG*s in other expression sites, and demands for homology to disfavour clearly incompatible donors: all of these may act to restrict the damage that segmental gene conversion could cause. Other mutator phenotypes systems have evolved similar mechanisms to restrict the potential for damage that enhanced mutation brings, for example by targeting mutation to specific loci (Pavri & Nussenzweig, 2011) or activating mutator phenotypes only in times of environmental stress (Caporale, 2003).

It is clear that many African trypanosome strategies evolve through second-order selection, so to understand the risks of segmental gene conversion we must also look beyond individual parasites. If individual parasites are frequently killed as a result of imprudent segmental gene conversions, what is the consequence for the infecting population? Particularly where density-dependent differentiation is dominant, parasite numbers are not restricted by replication rate or the availability of nutrients (Seed & Sechelski, 1988), so replacing those parasites that have fallen is not a great burden on the population. In clonal infections the benefits of segmental gene conversion could easily outweigh the costs. Attrition is commonplace in B-cell lineages, and is easily accommodated at the population level (Melchers *et al.*, 2000).

In mixed infections, which are likely to be commonplace (Balmer & Caccone, 2008; Macleod *et al.*, 2001), there is inter-clone competition for host resources or carrying capacity (Balmer *et al.*, 2009). There is thus more of a disadvantage of risky segmental conversion, as parasite strains that access only functional genes will be able to outcompete less efficient segmental converters in the early stage of infection. It is only later on that the potential for mosaic variation becomes advantageous. This might explain why mosaics do not appear in the earlier stage of infection: a strategy has evolved, whereby parasites strike a balance between efficacy of activation and scope of variability, with each strategy becoming deployed at different points in the infection. This proposition

could be tested with mathematical models of mixed infections, and analysis of archive use by different field isolates.

6.4.9 The role of segmental gene conversion in T. brucei infection

Segmental gene conversion is probably a mechanism under complex selection pressure, and has multiple functions and outcomes. It is unlikely that trypanosomes are able to restrain archive mutation, as selection on the archive is so indirect, and diversity is necessary for its function. Like other eukaryotic multi-gene families, such as olfactory receptors (Young *et al.*, 2008a), alpha-defensins (Das *et al.*, 2010) and immunoglobulin heavy chain variable regions (Ota & Nei, 1994), the *VSG* archive is probably subject to a process of birth-and-death evolution: genes are born by duplication, and while some 'die' by the accumulation of pseudogenizing mutations, others diversify and persist (Nei & Rooney, 2005). Segmental gene conversion has clear utility in accessing epitopes encoded by damaged *VSG*s, 'extending the life' of silent *VSG*s and harnessing the divergence introduced by mutation. In accommodating pseudogenes, segmental gene conversion introduces an additional layer of variability, extending the function of the archive even further. I propose that the key role of segmental gene conversion in African trypanosomes is as a mechanism of generating increased diversity from a degenerating and diversifying archive.

6.5 Concluding remarks

6.5.1 An holistic view of antigenic variation demands consideration of mechanisms

African trypanosomes possess archives of antigen genes unparalleled in size. But antigenic variation, even in trypanosomes, is driven not by the repertoire of silent genes, but by the genetic and epigenetic mechanisms that facilitate their expression, maintenance and evolution. It is these adaptive mechanisms that are likely subject to the greatest selection pressures. Trypanosomes must be prepared for uncertainty: they are at the whim of their vectors and have no control over where they end up. The size, immuno-competence, immunological memory and infection status of their next host cannot be anticipated. Archive use reflects this need for flexibility. In *T. brucei* at least, the archive should be

considered a resource that can be interpreted creatively, rather than a set script that is dogmatically adhered to. In this sense, the term 'archive'—with connotations of fixed records preserved for future reference—is an inadequate metaphor. Silent subtelomeric *VSG*s are more akin to a toolkit, albeit one that is highly dynamic. How this resource is managed, and utilized to such remarkable effect, will be a fascinating subject of research.

To what degree are these patterns shared with other, non-*brucei* African trypanosomes? There are clearly differences between species with regards to archive intactness and *VSG* relatedness (Jackson *et al.*, 2012), and it is likely that some of the genes involved in the maintenance and expression of the archive vary amongst African trypanosomes. Are the different species subject to different selective pressures? Have *T. congolense* and *T. vivax* developed different mechanisms for their archive's expression? It would be interesting to examine the relationship between genotype and phenotype in these species, to see whether their archives reflect differences in expression, and whether differing strategies between the three species have different evolutionary, epidemiological and ecological consequences.

6.5.2 What can trypanosomes tell us about the evolution of adaptability?

African trypanosome antigenic variation is a model system and a classic textbook example of a chronic host-parasite interaction. Many might consider deeper analysis of the system unwarranted: simply a case of tidying up loose ends, particularly as further studies of antigenic variation do not hold much promise for immediately useful therapeutic application. But the importance of our understanding of antigenic variation extends beyond disease control to diverse aspects of basic and applied biological science. Many fundamental mysteries remain.

The scope of our current understanding now puts us in a position to examine antigenic variation beyond the one-dimensional 'arms race', and consider it in an ecological and evolutionary framework. By examining how adaptability itself evolves and changes under differing selective pressures and host environments, competition from peers, and broader host and vector population structures, there is much we can learn from African trypanosomes.

Appendices

7 Appendices

7.1 Oligonucleotide sequences

Ch	Experiment	Primer name	Sequence
3	Amplification of *VSG*s from cDNA	Spliced Leader	GTTTCTGTACTATATTG
		16-mer	GTGTTAAAATATATCA
	Testing antigenic variation by PCR	Set_01_F	CGCAAACGCAACTATTAGCC
		Set_01_R	CAAGTGCAACAGCCTTCTTG
		Set_08_F	GCTCTCAGCCTAGGAAGCAA
		Set_08_R	TCAGTGTTGGCGCTATTGTT
		Set_12_F	AGTAACAGACAGGGCCAACG
		Set_12_R	CTGCTGTTTCGATCTCTTCT
		Set_14_F	AAAGGAAGGCACAGAAGG
		Set_14_R	GCAGTAATGTCTGATTCG
		Set_21_F	AGCCGCTGTTTTGCGTATAG
		Set_21_R	CGGAGTTACTCAGGAAAGCC
		Set_22_F	CCAACGACAAAGCACTCA
		Set_22_R	TGGCTTTGTTGATCGTGT
		Tubulin_F	CCAAGCTCGGCTACACGGTGT
		Tubulin_R	GGCTCAAACACAGCGTTCGA
	Screening colonies for insert	M13F	TGTAAAACGACGGCCAGT
		M13R	CAGGAAACAGCTATGACC
4	Testing mosaicism by PCR	14-AF	ACGGCCAAACCGGGATGAAA
		14-AR	ATCGCTTCGCGGACTTCCTC
		14-BF	CCATTGTCGACGCAACCGTT
		14-BR	GGCCGTCGACACAAAGACAG
		10-AF	CATAACCGCAATAGCCAGG
		10-AR	TTGTTTGGGCTATGCCTG
		10-BF	CGTAACCGCAATAGCCCAC
		10-BR	CGGCCTGTGCTATTTGTG
		PolI_F	GTGACGGTTCAGGGAACACT
		PolI_R	GCACCGAAATTTGACTTGGT
	Preparing probe for Southern blot	Tb11.09.0005_N_F	CAAAGAGCAAAAGAGATGACC
		Tb11.09.0005_N_R	TGATTCGGTTGTTGCTGTC

Table 7.1 The sequences of custom oligonucleotide sequences used in this work. Table continues on the following page.

Ch	Experiment	Primer name	Sequence
5	Primers for preparation of MCS	FseI/SbfI_5'	ATCAAGCTTGGCCGGCCAGTCCTGCAGGAATTCGAT
		FseI/SbfI_3'	ATCGAATTCCTGCAGGACTGGCCGGCCAAGCTTGAT
	Primers for sequencing insert	pVSG_insert_F	GAGGAGAAAGAATAGTAACC
		pVSG insert_R	GAAATTTGAGGGGGGAAAG
	Testing exogenous *VSG* expression	Tubulin_F	CCAAGCTCGGCTACACGGTGT
		Tubulin_R	GGCTCAAACACAGCGTTCGA
		427-2_F	ATGCCTTCCAATCAGGAGGC
		427-2_R	TGTATCGGCGACAACTGCAG
		Set_14_F	AAAGGAAGGCACAGAAGG
		Set_14_R	GCAGTAATGTCTGATTCG
		427-4_F	TAAAAGGAGACGGAGTGG
		427-4_R	GCTCTAGTTTGTTGTTGTTG
		427-6_F	ACCTGACATCGGACGGTAAC
		427-6_R	GTCGGTTATGTCGGCAAGTT
		427-9_F	GTAAACGGTCCGGAGTTCAA
		427-9_R	CATTTCCGCGTTGTCTTGTA
	Amplification of *VSG*s from cDNA	Spliced Leader	GTTTCTGTACTATATTG
		16-mer	GTGTTAAAATATATCA
	Amplification of *VSG*s from plasmid	L-S_SbfI	GATCCCTGCAGGTTTCTGTACTATATTG
		T-M_SbfI	GATCCCTGCAGGTGTTAAAATATATCA

Table 7.1 (continued)

7.2 Supplementary information to Chapters 3 and 4

7.2.1 Pedigree of trypanosomes used for chronic mouse infections

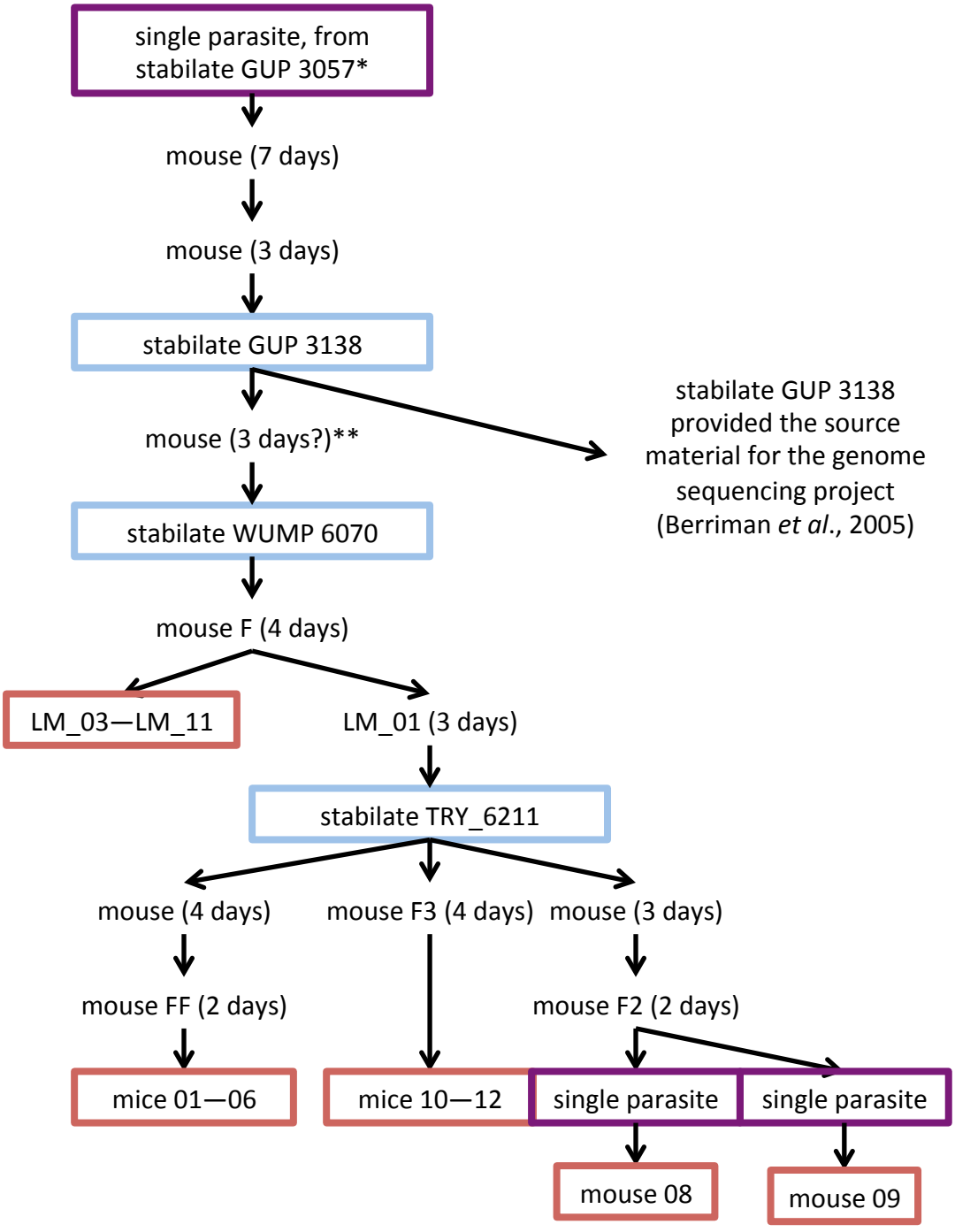


Figure 7.1 Pedigree of TREU 927/4 GUTat 10.1 trypanosomes used for infections. Purple boxes indicate verified cloning events. Blue boxes indicate stabilates. Red boxes indicate experimental mice. The identity of the trypanosomes was confirmed with genome markers specific for 927 (L. Marcello, Ph.D. Thesis, 2006, University of Glasgow). *The original source of TREU 927 was a tsetse fly in Kiboko, Kenya, in 1969/70 (G. Cross, pers. comm., http://tryps.rockefeller.edu/DocumentsGlobal/lineage_TREU927.pdf) . **The exact timing of the event(s) connecting stabilate GUP 3138 with WUMP 6070 is not known.

7.2.2 Parasitaemias of chronic mouse infections

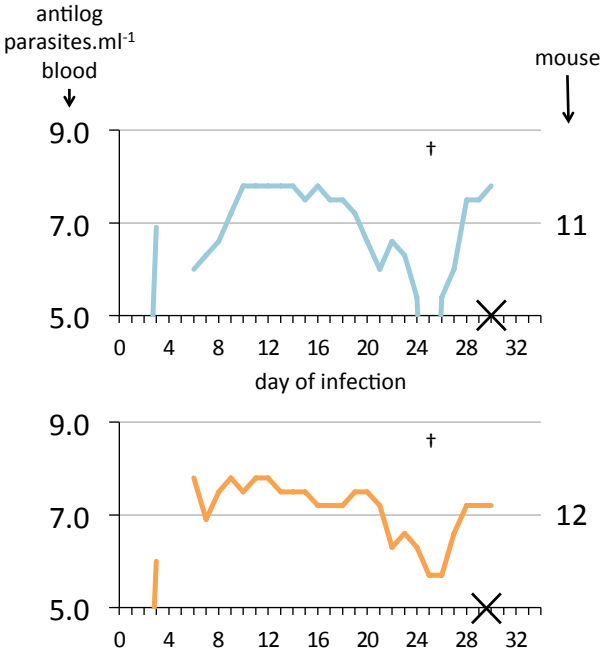


Figure 7.2 Parasitaemias of chronic mouse infections. Termination of the infection is indicated by a cross on the x-axis. A dagger indicates the transient dip in parasitaemia, described in Chapter 3. Figure continues on subsequent page.

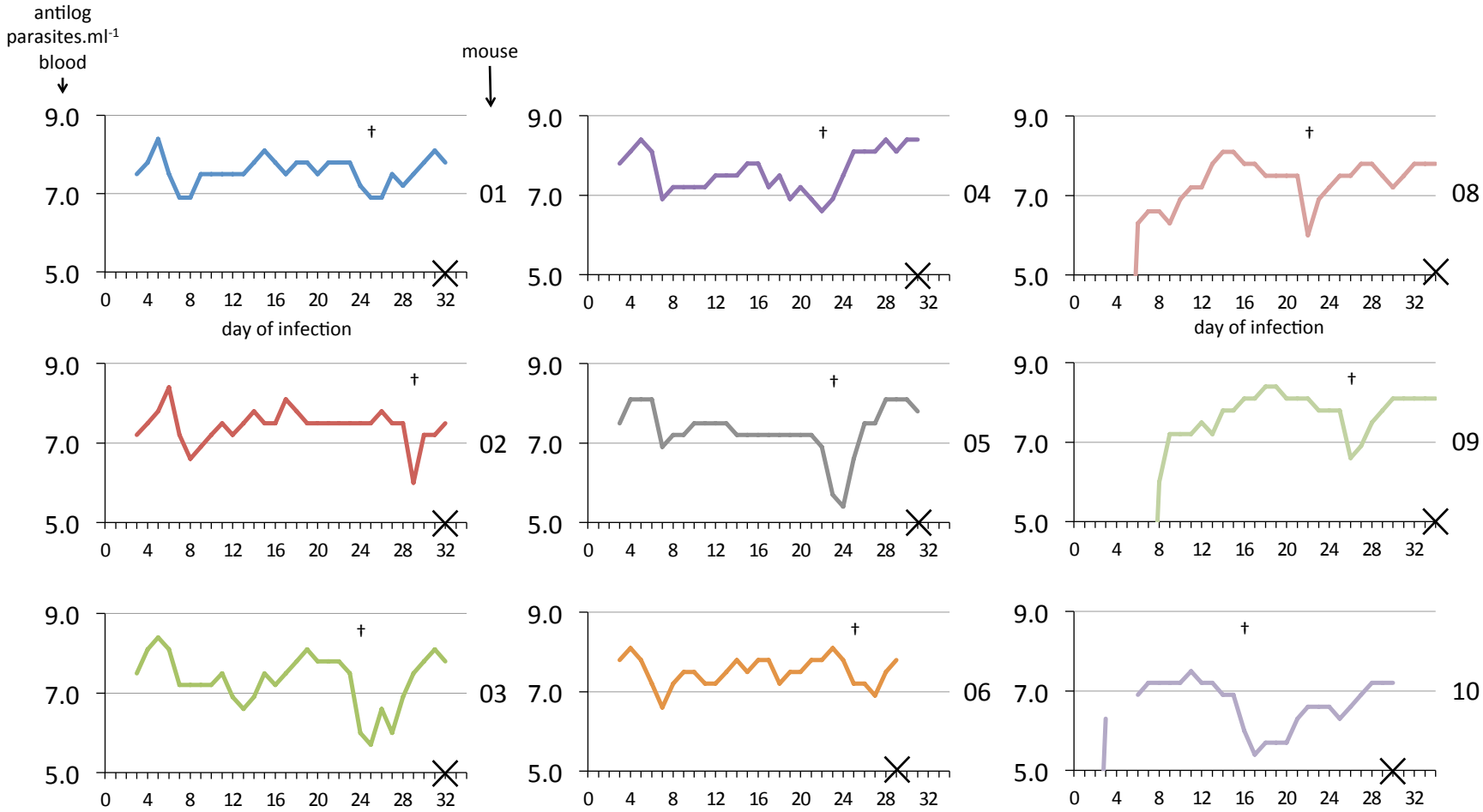


Figure 7.2 (continued)

7.2.3 VSG *clone details*

Clone name	Infection	Time point	Clone #	int. seq?	Mods	Set	type		dons for NTD?	mosaic from dons?	3' don. bound.	3' don
							N	C				
FF-00c02	FF	0	02			Set_01	A	2?a	y	n	0	—
FF-00c03	FF	0	03			Set_01	A	2?a	y	n	0	—
FF-00c04	FF	0	04			Set_01	A	2?a	y	n	0	—
FF-00c05	FF	0	05			Set_01	A	2?a	y	n	0	—
FF-00c06	FF	0	06			Set_01	A	2?a	y	n	0	—
FF-00c07	FF	0	07			Set_01	A	2?a	y	n	0	—
FF-00c08	FF	0	08			Set_01	A	2?a	y	n	0	—
FF-00c09	FF	0	09			Set_01	A	2?a	y	n	0	—
FF-00c10	FF	0	10			Set_01	A	2?a	y	n	0	—
FF-00c11	FF	0	11			Set_01	A	2?a	y	n	0	—
FF-00c12	FF	0	12			Set_01	A	2?a	y	n	0	—
06-20c05-del	6	20	05		T450del~	Set_01	A	2?a	y	n	0	—
VSG-03-01-01	LM_01	3	01	Y		Set_01	A	2?a	y	n	0	—
VSG-03-01-02	LM_01	3	02	Y		Set_01	A	2?a	y	n	0	—
VSG-03-02-01	LM_02	3	01	Y		Set_01	A	2?a	y	n	0	—
VSG-03-02-02	LM_02	3	02	Y		Set_01	A	2?a	y	n	0	—
VSG-09-03-01	LM_03	9	01	Y		Set_02	A	6	n	n	0	—
10-10c02	10	10	02			Set_03	A	6	p	n	0	—
10-10c03	10	10	03			Set_03	A	6	p	n	0	—
10-10c04	10	10	04			Set_03	A	6	p	n	0	—
10-10c05_x	10	10	05		failed R	Set_03	A	6	p	n	0	—
11-10c05	11	10	05			Set_03	A	6	p	n	0	—
11-10c07	11	10	07			Set_03	A	6	p	n	0	—
11-10c10	11	10	10			Set_03	A	6	p	n	0	—
11-17c04	11	17	04			Set_03	A	6	p	n	0	—
11-17c05	11	17	05			Set_03	A	6	p	n	0	—
11-17c06	11	17	06			Set_03	A	6	p	n	0	—
11-17c09	11	17	09			Set_03	A	6	p	n	0	—
11-17c12	11	17	12			Set_03	A	6	p	n	0	—
12-10c10-add	12	10	10		A105add	Set_03	A	6	p	n	0	—
12-10c12-add	12	10	12		A105add	Set_03	A	6	p	n	0	—
12-17c01	12	17	01			Set_03	A	6	p	n	0	—
12-17c02	12	17	02			Set_03	A	6	p	n	0	—
12-17c03	12	17	03			Set_03	A	6	p	n	0	—
12-17c05	12	17	05			Set_03	A	6	p	n	0	—
VSG-09-03-03_x	LM_03	9	03	Y		Set_03	A	6	p	n	0	—
VSG-09-04-01	LM_04	9	01	Y		Set_03	A	6	p	n	0	—
01-27c09	1	27	09			Set_04	B	1	p	y	{4}	—
01-32c11	1	32	11			Set_04	B	3	p	y	0	—
05-27c03	5	27	03			Set_04	B	3	p	y	0	—
06-27c12	6	27	12			Set_04	B	3	p	y	0	—
VSG-09-03-04	LM_03	9	04	Y		Set_04	B	3	p	y	0	—
04-31c16	4	31	16			Set_05	A	5b	n	n	0	—
11-10c01-del	11	10	01		A790del~	Set_05	A	5b	n	n	0	—
12-10c05	12	10	05			Set_05	A	5b	n	n	0	—
12-10c11-del	12	10	11		T754del	Set_05	A	5b	n	n	0	—
VSG-09-03-07	LM_03	9	07	Y		Set_05	A	5b	n	n	0	—
VSG-09-04-02	LM_04	9	02	Y		Set_05	A	5b	n	n	0	—
VSG-09-04-10_x	LM_04	9	10	Y		Set_06	A/C	1	n	n	0	—
11-07c02	11	7	02		stop	Set_07	A	2	n	n	0	—
11-07c04	11	7	04			Set_07	A	2	n	n	0	—
11-07c07	11	7	07			Set_07	A	2	n	n	0	—
11-07c08	11	7	08		atyp~	Set_07	A	2	n	n	0	—
11-10c03	11	10	03			Set_07	A	2	n	n	0	—
11-10c09	11	10	09			Set_07	A	2	n	n	0	—
11-14c01	11	14	01			Set_07	A	2	n	n	0	—
12-17c04	12	17	04			Set_07	A	2	n	n	0	—
12-17c08	12	17	08			Set_07	A	2	n	n	0	—
12-17c09	12	17	09			Set_07	A	2	n	n	0	—
12-24c02	12	24	02			Set_07	A	2	n	n	{2}	—
VSG-14-05-01	LM_05	14	01	Y		Set_07	A	2	n	n	0	—
01-20c02	1	20	02			Set_08	B	2	y	n	2+	a*

Table 7.2 Clone sequence details. See the end of the Table for details.

Clone name	Infection	Time point	Clone #	int. seq?	Mods	Set	type		dons for NTD?	mosaic from dons?	3' don. bound.	3' don
							N	C				
01-20c05	1	20	05		atyp~	Set_08	B	2	y	n	2+	a*
01-20c06	1	20	06			Set_08	B	2	y	n	2+	a*
01-20c09	1	20	09			Set_08	B	2	y	n	2	—
01-23c01	1	23	01			Set_08	B	2	y	n	2	—
01-23c02-del	1	23	02		A1318del~	Set_08	B	2	y	n	2	—
01-23c03	1	23	03			Set_08	B	2	y	n	2+	a*
01-23c04	1	23	04			Set_08	B	2	y	n	2	—
01-23c05	1	23	05		atyp~	Set_08	B	2	y	n	1	—
01-23c06-del	1	23	06		A1318del~	Set_08	B	2	y	n	2	—
01-23c08	1	23	08			Set_08	B	2	y	n	1	—
01-23c09	1	23	09			Set_08	B	2	y	n	2	—
01-23c10	1	23	10			Set_08	B	2	y	n	2+	a*
01-23c11	1	23	11			Set_08	B	2	y	n	2	—
01-23c12	1	23	12			Set_08	B	2	y	n	2+	a*
01-32c01-del	1	32	01		A1388del~	Set_08	B	2	y	n	2	—
01-32c02	1	32	02			Set_08	B	2	y	n	2	—
01-32c03	1	32	03			Set_08	B	2	y	n	1	—
01-32c04	1	32	04			Set_08	B	2	y	n	2	—
01-32c06	1	32	06			Set_08	B	2	y	n	1	—
01-32c07	1	32	07			Set_08	B	2	y	n	2	—
01-32c09	1	32	09			Set_08	B	2	y	n	1	—
03-20c01	3	20	01		stop	Set_08	B	2	y	n	2	—
03-20c02	3	20	02			Set_08	B	2	y	n	2	—
03-20c03	3	20	03			Set_08	B	2	y	n	2	—
03-20c04	3	20	04			Set_08	B	2	y	n	2	—
03-20c05	3	20	05			Set_08	B	2	y	n	2+	a*
03-20c07	3	20	07			Set_08	B	2	y	n	2	—
03-20c08	3	20	08			Set_08	B	2	y	n	2	—
03-20c09	3	20	09		stop	Set_08	B	2	y	n	2	—
03-20c10	3	20	10			Set_08	B	2	y	n	2	—
03-20c11	3	20	11			Set_08	B	2	y	n	1	—
03-20c12	3	20	12			Set_08	B	2	y	n	2	—
03-23c01	3	23	01			Set_08	B	2	y	n	1	—
03-23c02	3	23	02			Set_08	B	2	y	n	2+	a*
03-23c03_x	3	23	03		failed F	Set_08	B	2	y	n	2+	a*
03-23c04	3	23	04			Set_08	B	2	y	n	2+	a*
03-23c05	3	23	05			Set_08	B	2	y	n	2+	?
03-23c06	3	23	06			Set_08	B	2	y	n	2+	a*
03-23c07	3	23	07		atyp~	Set_08	B	2	y	n	1	—
03-23c09	3	23	09		atyp~	Set_08	B	2	y	n	1	—
03-23c11	3	23	11			Set_08	B	2	y	n	2	—
03-23c12	3	23	12			Set_08	B	2	y	n	1	—
05-21c01	5	21	01			Set_08	B	2	y	n	2+	a*
05-21c02	5	21	02			Set_08	B	2	y	n	2+	a*
05-21c03	5	21	03			Set_08	B	2	y	n	1	—
05-21c04	5	21	04			Set_08	B	2	y	n	2	—
05-21c06	5	21	06			Set_08	B	2	y	n	2+	a*
05-21c07	5	21	07			Set_08	B	2	y	n	2+	a*
05-21c08	5	21	08		atyp|	Set_08	B	2	y	n	2+	a*
05-21c09_x	5	21	09		failed F	Set_08	B	2	y	n	2+	a*
05-21c10	5	21	10			Set_08	B	2	y	n	2+	a*
05-21c11	5	21	11			Set_08	B	2	y	n	2+	a*
05-21c12	5	21	12			Set_08	B	2	y	n	1	—
05-21c13	5	21	13			Set_08	B	2	y	n	2+	a*
05-21c14	5	21	14			Set_08	B	2	y	n	2+	a*
05-21c15	5	21	15			Set_08	B	2	y	n	2+	a*
05-21c16	5	21	16		atyp~	Set_08	B	2	y	n	2+	a*
05-21c17	5	21	17			Set_08	B	2	y	n	2+	a*
05-21c18	5	21	18			Set_08	B	2	y	n	2+	a*
05-21c19	5	21	19			Set_08	B	2	y	n	2+	a*
05-21c20	5	21	20			Set_08	B	2	y	n	2+	a*
05-21c21	5	21	21			Set_08	B	2	y	n	2+	1*

Table 7.2 (continued)

Clone name	Infection	Time point	Clone #	int. seq?	Mods	Set	type		dons for NTD?	mosaic from dons?	3' don. bound.	3' don
							N	C				
05-21c22	5	21	22			Set_08	B	2	y	n	1	—
05-21c23	5	21	23			Set_08	B	2	y	n	2+	a*
05-21c24	5	21	24		stop	Set_08	B	2	y	n	2+	?
06-20c01	6	20	01			Set_08	B	2	y	n	2+	a*
06-20c02	6	20	02			Set_08	B	2	y	n	1	—
06-20c03	6	20	03			Set_08	B	2	y	n	2+	a*
06-20c04	6	20	04			Set_08	B	2	y	n	2+	a*
06-20c06-add	6	20	06		C252add	Set_08	B	2	y	n	2+	a*
06-20c07	6	20	07			Set_08	B	2	y	n	2+	a*
06-20c08	6	20	08			Set_08	B	2	y	n	2+	a*
06-20c09	6	20	09			Set_08	B	2	y	n	2+	a*
06-20c10	6	20	10			Set_08	B	2	y	n	2+	a*
06-20c11	6	20	11			Set_08	B	2	y	n	2+	a*
06-20c12	6	20	12			Set_08	B	2	y	n	2+	a*
06-23c01	6	23	01			Set_08	B	2	y	n	1	—
06-23c03	6	23	03			Set_08	B	2	y	n	1	—
06-23c04	6	23	04			Set_08	B	2	y	n	1	—
06-23c05	6	23	05			Set_08	B	2	y	n	1	—
06-23c06-del	6	23	06		A1460del~	Set_08	B	2	y	n	1	—
06-23c07	6	23	07			Set_08	B	2	y	n	1	—
06-23c08	6	23	08			Set_08	B	2	y	n	1	—
06-23c09-del	6	23	09		A1318del~	Set_08	B	2	y	n	2+	a*
06-23c10	6	23	10			Set_08	B	2	y	n	1	—
06-23c11	6	23	11		stop	Set_08	B	2	y	n	1	—
08-20c01	8	20	01			Set_08	B	2	y	n	1	—
08-20c02	8	20	02			Set_08	B	2	y	n	2+	a*
08-20c04	8	20	04			Set_08	B	2	y	n	1	—
08-20c05	8	20	05			Set_08	B	2	y	n	1	—
08-20c06	8	20	06			Set_08	B	2	y	n	2+	a*
08-20c07	8	20	07			Set_08	B	2	y	n	1	—
08-20c08	8	20	08			Set_08	B	2	y	n	2+	a*
08-20c09	8	20	09			Set_08	B	2	y	n	2+	a*
08-20c10	8	20	10			Set_08	B	2	y	n	2+	a*
08-20c11	8	20	11			Set_08	B	2	y	n	2+	a*
08-20c12	8	20	12			Set_08	B	2	y	n	1	—
09-20c01	9	20	01			Set_08	B	2	y	n	1	—
09-20c02	9	20	02			Set_08	B	2	y	n	1	—
09-20c03	9	20	03			Set_08	B	2	y	n	2+	a*
09-20c04	9	20	04			Set_08	B	2	y	n	1	—
09-20c05	9	20	05			Set_08	B	2	y	n	2+	a*
09-20c06	9	20	06			Set_08	B	2	y	n	2+	a*
09-20c07_x	9	20	07		failed F	Set_08	B	2	y	n	2+	a*
09-20c08	9	20	08			Set_08	B	2	y	n	1	—
09-20c10	9	20	10			Set_08	B	2	y	n	2+	a*
09-20c11	9	20	11			Set_08	B	2	y	n	1	—
11-07c05_x	11	7	05		failed R	Set_08	B	X	y	n	X	—
11-10c04	11	10	04			Set_08	B	2	y	n	1	—
11-17c02	11	17	02			Set_08	B	2	y	n	2	—
11-17c03	11	17	03			Set_08	B	2	y	n	2	—
11-17c07	11	17	07			Set_08	B	2	y	n	2	—
11-17c08	11	17	08			Set_08	B	2	y	n	2	—
11-17c11	11	17	11			Set_08	B	2	y	n	2	—
11-21c04_x	11	21	04		failed R	Set_08	B	X	y	n	X	—
11-21c07_x	11	21	07		failed F	Set_08	B	2	y	n	2+	a*
12-17c06	12	17	06			Set_08	B	2	y	n	2+	a*
12-17c07	12	17	07			Set_08	B	2	y	n	2	—
12-17c10	12	17	10			Set_08	B	2	y	n	2	—
12-17c11	12	17	11			Set_08	B	2	y	n	2+	a*
12-24c01	12	24	01			Set_08	B	2	y	n	2+	a*
12-24c03	12	24	03			Set_08	B	2	y	n	2	—
12-24c04	12	24	04			Set_08	B	2	y	n	1	?
12-24c08-add	12	24	08		G139add	Set_08	B	2	y	n	2	?

Table 7.2 (continued)

Clone name	Infection	Time point	Clone #	int. seq?	Mods	Set	type		dons for NTD?	mosaic from dons?	3' don. bound.	3' don
							N	C				
12-24c09-add	12	24	09		G139add	Set_08	B	2	y	n	2	?
12-24c11	12	24	11			Set_08	B	2	y	n	2	—
12-24c12	12	24	12			Set_08	B	2	y	n	2+	a*
12-24c14	12	24	14			Set_08	B	2	y	n	2+	a*
VSG-14-05-02	LM_05	14	02	Y		Set_08	B	2	y	n	2+	a*
VSG-14-06-01	LM_06	14	01	Y		Set_08	B	2	y	n	2+	a*
VSG-21-08-02	LM_08	21	02	Y		Set_08	B	2	y	n	2+	a*
VSG-21-08-03	LM_08	21	03	Y		Set_08	B	2	y	n	1	—
01-32c10	1	32	10			Set_09	A	6	y	n	4	?
04-31c37-del	4	31	37		T1517del~	Set_09	A	1	y	n	4	?
VSG-22-07-01	LM_07	22	01	Y		Set_09	A	3	y	n	0	—
01-27c08	1	27	08			Set_10	A	6	y	y	1	—
01-27c10	1	27	10			Set_10	A	6	y	y	1	—
01-27c12	1	27	12			Set_10	A	6	y	y	1	—
01-29c02	1	29	02			Set_10	A	6	y	y	1	—
01-29c04	1	29	04			Set_10	A	6	y	y	0	—
01-29c10	1	29	10			Set_10	A	6	y	y	0	—
01-29c11	1	29	11			Set_10	A	6	y	y	3	?
04-21c39	4	21	39			Set_10	A	6	y	y	0	—
04-21c40	4	21	40			Set_10	A	6	y	y	0	—
04-23c15	4	23	15			Set_10	A	6	y	y	0	—
04-23c19	4	23	19			Set_10	A	6	y	y	0	—
04-23c21	4	23	21			Set_10	A	6	y	y	0	—
04-23c36	4	23	36			Set_10	A	6	y	y	0	—
04-23c43	4	23	43			Set_10	A	6	y	y	0	—
05-23c27	5	23	27			Set_10	A	6	y	y	1	—
05-27c01	5	27	01			Set_10	A	6	y	y	0	—
05-27c04	5	27	04			Set_10	A	6	y	y	0	—
05-27c16	5	27	16			Set_10	A	6	y	y	0	—
05-27c23	5	27	23			Set_10	A	6	y	y	1	—
05-27c29	5	27	29			Set_10	A	6	y	y	1	—
05-29c05	5	29	05			Set_10	A	6	y	y	0	—
05-29c13-del	5	29	13		A1194del~	Set_10	A	6	y	y	0	—
05-29c15	5	29	15			Set_10	A	6	y	y	0	—
05-29c17	5	29	17			Set_10	A	6	y	y	2	?
05-29c21	5	29	21			Set_10	A	6	y	y	0	—
05-29c24	5	29	24			Set_10	A	6	y	y	0	—
11-27c04	11	27	04			Set_10	A	6	y	y	1	—
11-27c10	11	27	10			Set_10	A	6	y	y	1	—
11-30c02	11	30	02			Set_10	A	6	y	y	0	—
VSG-22-07-02-del	LM_07	22	02	Y	G135del	Set_10	A	6	y	y	0	—
04-21c06	4	21	06			Set_11	A	2	y	n	0	—
04-21c44	4	21	44			Set_11	A	2	y	n	0	—
05-23c04	5	23	04			Set_11	A	2	y	n	1	—
05-23c10	5	23	10			Set_11	A	2	y	n	1	—
05-23c15	5	23	15			Set_11	A	2	y	n	1	—
05-23c18	5	23	18			Set_11	A	2	y	n	1	—
05-23c24	5	23	24			Set_11	A	2	y	n	2	?
05-23c26	5	23	26			Set_11	A	2	y	n	1	—
11-30c06	11	30	06			Set_11	A	2	y	n	2	?
11-30c12	11	30	12			Set_11	A	2	y	n	1	—
11-30c14	11	30	14			Set_11	A	2	y	n	1	—
11-30c16	11	30	16			Set_11	A	2	y	n	1	—
11-30c17	11	30	17			Set_11	A	2	y	n	0	—
11-30c25	11	30	25			Set_11	A	2	y	n	1	—
11-30c28	11	30	28			Set_11	A	2	y	n	0	—
11-30c30	11	30	30			Set_11	A	2	y	n	1	—
11-30c34	11	30	34			Set_11	A	2	y	n	1	—
11-30c35	11	30	35			Set_11	A	2	y	n	1	—
VSG-22-07-03	LM_07	22	03	Y		Set_11	A	2	y	n	1	—
01-27c06	1	27	06			Set_12	A	5a	y	n	6	—
04-21c02	4	21	02			Set_12	A	4	y	n	6	—

Table 7.2 (continued)

Clone name	Infection	Time point	Clone #	int. seq?	Mods	Set	type		dons for NTD?	mosaic from dons?	3' don. bound.	3' don
							N	C				
04-23c14	4	23	14			Set_12	A	4	y	n	6	—
04-23c26	4	23	26			Set_12	A	4	y	n	6	—
04-23c29	4	23	29			Set_12	A	4	y	n	6	—
04-23c30	4	23	30			Set_12	A	4	y	n	6	—
04-23c32	4	23	32			Set_12	A	4	y	n	6	—
04-27c08	4	27	08			Set_12	A	5a	y	n	6	—
04-27c09	4	27	09			Set_12	A	5a	y	n	6	—
04-27c11	4	27	11			Set_12	A	4	y	n	6	—
04-27c16	4	27	16		atyp~	Set_12	A	4	y	n	2	—
04-27c19	4	27	19			Set_12	A	5a	y	n	6	—
04-27c27	4	27	27			Set_12	A	5a	y	n	6	—
04-27c28	4	27	28			Set_12	A	4	y	n	6	—
04-27c34	4	27	34			Set_12	A	4	y	n	6	—
04-27c36	4	27	36			Set_12	A	5a	y	n	6	—
04-27c37	4	27	37			Set_12	A	4	y	n	2	—
04-27c38	4	27	38			Set_12	A	4	y	n	6	—
04-27c46	4	27	46			Set_12	A	4	y	n	6	—
04-27c47	4	27	47			Set_12	A	4	y	n	6	—
04-29c02	4	29	02			Set_12	A	4	y	n	1	—
04-29c10	4	29	10			Set_12	A	5a	y	n	6	—
04-29c11	4	29	11			Set_12	A	5a	y	n	6	—
04-29c16	4	29	16			Set_12	A	5a	y	n	6	—
04-29c28	4	29	28			Set_12	A	4	y	n	6	—
04-29c31	4	29	31			Set_12	A	4	y	n	1	—
04-29c33	4	29	33			Set_12	A	5a	y	n	6	—
04-29c35	4	29	35			Set_12	A	4	y	n	1	—
04-31c05	4	31	05			Set_12	A	5a	y	n	6	—
04-31c07	4	31	07			Set_12	A	4	y	n	6	—
04-31c08	4	31	08			Set_12	A	5a	y	n	6	—
04-31c10	4	31	10			Set_12	A	5a	y	n	6	—
04-31c15	4	31	15			Set_12	A	5a	y	n	6	—
04-31c36-del	4	31	36		T1225del~	Set_12	A	5a	y	n	6	—
04-31c38	4	31	38			Set_12	A	5a	y	n	6	—
04-31c39	4	31	39			Set_12	A	5a	y	n	6	—
04-31c43	4	31	43			Set_12	A	4	y	n	1	—
04-31c44	4	31	44			Set_12	A	5a	y	n	6	—
04-31c46	4	31	46			Set_12	A	5a	y	n	6	—
05-27c08	5	27	08			Set_12	A	4	y	n	2	—
05-27c12	5	27	12			Set_12	A	5a	y	n	6	—
05-27c13	5	27	13			Set_12	A	5a	y	n	6	—
05-27c17-del	5	27	17		G577del	Set_12	A	5a	y	n	6	—
05-27c18	5	27	18			Set_12	A	5a	y	n	6	—
05-27c25	5	27	25			Set_12	A	4	y	n	6	—
05-29c01	5	29	01			Set_12	A	5a	y	n	6	—
05-29c04	5	29	04			Set_12	A	5a	y	n	6	—
05-29c06	5	29	06			Set_12	A	4	y	n	2	—
05-29c07	5	29	07			Set_12	A	5a	y	n	6	—
05-29c08	5	29	08			Set_12	A	5a	y	n	6	—
05-29c10	5	29	10			Set_12	A	5a	y	n	6	—
05-29c11	5	29	11			Set_12	A	5a	y	n	6	—
05-29c18	5	29	18			Set_12	A	5a	y	n	6	—
05-29c20	5	29	20			Set_12	A	5a	y	n	6	—
05-31c06	5	31	06			Set_12	A	5a	y	n	6	—
09-34c02	9	34	02			Set_12	A	5a	y	n	6	—
09-34c04-del	9	34	04		A1086del~	Set_12	A	5a	y	n	6	—
09-34c05	9	34	05			Set_12	A	5a	y	n	6	—
09-34c06	9	34	06			Set_12	A	5a	y	n	6	—
09-34c16	9	34	16			Set_12	A	5a	y	n	6	—
09-34c17	9	34	17			Set_12	A	5a	y	n	6	—
09-34c18	9	34	18			Set_12	A	5a	y	n	6	—
11-27c03	11	27	03			Set_12	A	5a	y	n	6	—
11-27c08	11	27	08			Set_12	A	5a	y	n	6	—

Table 7.2 (continued)

Clone name	Infection	Time point	Clone #	int. seq?	Mods	Set	type		dons for NTD?	mosaic from dons?	3' don. bound.	3' don
							N	C				
11-27c09	11	27	09			Set_12	A	4	y	n	6	—
11-30c10-del	11	30	10		G576del	Set_12	A	5a	y	n	6	—
VSG-22-07-04	LM_07	22	04	Y		Set_12	A	4	y	n	2	—
05-21c05	5	21	05			Set_13	B	2	y	n	2+	q
05-23c03	5	23	03			Set_13	B	2	y	n	2+	a
05-23c08	5	23	08			Set_13	B	2	y	n	2+	a
05-23c16	5	23	16			Set_13	B	2	y	n	2+	a
05-23c20	5	23	20			Set_13	B	2	y	n	2+	a
05-27c27	5	27	27		stop	Set_13	B	2	y	n	2+	a
06-29c05	6	29	05			Set_13	B	2	y	n	2+	a
VSG-24-09-01	LM_09	24	01	Y		Set_13	B	2	y	n	2+	a
01-27c07-del	1	27	07		A1057del	Set_14	B	3	y	y	6	f
01-27c11-del	1	27	11		A1396del~	Set_14	B	1	y	y	1	—
01-29c05-del	1	29	05		A1366del~	Set_14	B	1	y	y	1	—
01-29c09	1	29	09	Y		Set_14	B	1	y	y	1	—
01-32c08	1	32	08			Set_14	B	1	y	y	1	—
03-32c07	3	32	07			Set_14	B	1	y	y	1	—
04-21c04	4	21	04			Set_14	B	2	y	y	0	—
04-23c07	4	23	07			Set_14	B	2	y	y	0	—
04-23c40	4	23	40	Y		Set_14	B	2	y	y	0	—
04-23c47	4	23	47	Y		Set_14	B	2	y	y	0	—
04-23c48	4	23	48	Y		Set_14	B	2	y	y	0	—
04-27c07-del	4	27	07		A854del~	Set_14	B	2	y	y	0	—
04-27c17	4	27	17	Y		Set_14	B	1	y	y	1	—
04-27c21	4	27	21	Y		Set_14	B	1	y	y	1	—
04-27c25	4	27	25	Y	atyp~	Set_14	B	1	y	y	5	g
04-27c44	4	27	44	Y	atyp~	Set_14	B	1	y	y	5	g
04-29c06	4	29	06			Set_14	B	3	y	y	0	—
04-31c01	4	31	01			Set_14	B	3	y	y	0	—
04-31c06	4	31	06			Set_14	B	3	y	y	0	—
04-31c31	4	31	31	Y		Set_14	B	3	y	y	0	—
04-31c33	4	31	33	Y		Set_14	B	3	y	y	0	—
04-31c34	4	31	34			Set_14	B	3	y	y	0	—
04-31c35	4	31	35	Y		Set_14	B	3	y	y	0	—
05-23c12	5	23	12			Set_14	B	2	y	y	0	—
05-27c09	5	27	09			Set_14	B	1	y	y	1	—
05-27c19	5	27	19			Set_14	B	2	y	y	0	—
05-27c20	5	27	20			Set_14	B	2	y	y	0	—
05-27c21	5	27	21			Set_14	B	2	y	y	0	—
05-27c28	5	27	28			Set_14	B	1	y	y	1	—
05-27c32	5	27	32			Set_14	B	2	y	y	0	—
05-27c33	5	27	33			Set_14	B	2	y	y	0	—
05-29c23	5	29	23			Set_14	B	2	y	y	0	—
05-31c05	5	31	05			Set_14	B	3	y	y	0	—
06-29c07	6	29	07			Set_14	B	3	y	y	0	—
06-29c10	6	29	10			Set_14	B	3	y	y	0	—
VSG-28-10-02	LM_10	28	02	Y		Set_14	B	3	y	y	0	—
VSG-28-10-03	LM_10	28	03	Y		Set_14	B	3	y	y	5	r
VSG-28-10-04	LM_10	28	04	Y		Set_14	B	3	y	y	5	r
03-27c12	3	27	12			Set_16	A	3	p	y	5	?*
VSG-28-10-07	LM_10	28	07	Y		Set_16	A	3	p	y	5	?*
03-29c03	3	29	03			Set_17	B	2	y	y	{5}	a
03-29c19	3	29	19			Set_17	B	2	y	y	{5}	a
03-29c22	3	29	22			Set_17	B	2	y	y	{2+}	?
03-32c11	3	32	11			Set_17	B	2	y	y	{5}	a
08-34c01	8	34	01			Set_17	B	2	y	y	{2+}	a*
08-34c02	8	34	02			Set_17	B	2	y	y	{2+}	a*
08-34c03	8	34	03			Set_17	B	2	y	y	{2+}	a*
08-34c04	8	34	04			Set_17	B	2	y	y	{2+}	a*
08-34c05	8	34	05			Set_17	B	2	y	y	{2+}	a*
08-34c07-add	8	34	07		A982add~	Set_17	B	2	y	y	{2+}	a*
08-34c08	8	34	08			Set_17	B	2	y	y	{2+}	a*

Table 7.2 (continued)

Clone name	Infection	Time point	Clone #	int. seq?	Mods	Set	type		dons for NTD?	mosaic from dons?	3' don. bound.	3' don
							N	C				
08-34c09	8	34	09			Set_17	B	2	y	y	{2+}	a*
08-34c10	8	34	10			Set_17	B	2	y	y	{2+}	a*
08-34c11	8	34	11			Set_17	B	2	y	y	{2+}	a*
08-34c12	8	34	12			Set_17	B	2	y	y	{2+}	a*
08-34c14	8	34	14			Set_17	B	2	y	y	{2+}	a*
08-34c15	8	34	15			Set_17	B	2	y	y	{2+}	a*
08-34c20	8	34	20			Set_17	B	2	y	y	{2+}	a*
08-34c21	8	34	21			Set_17	B	2	y	y	{2+}	?
08-34c22	8	34	22			Set_17	B	2	y	y	{2+}	a*
08-34c23	8	34	23			Set_17	B	2	y	y	{2+}	a*
09-34c01	9	34	01			Set_17	B	2	y	y	{2+}	a*
09-34c03	9	34	03			Set_17	B	2	y	y	{2+}	a*
09-34c07	9	34	07			Set_17	B	2	y	y	{2+}	a*
09-34c08	9	34	08			Set_17	B	2	y	y	{2+}	a*
09-34c09	9	34	09			Set_17	B	2	y	y	{2+}	a*
09-34c10	9	34	10			Set_17	B	2	y	y	{2+}	a*
09-34c11	9	34	11			Set_17	B	2	y	y	{2+}	a*
09-34c12	9	34	12			Set_17	B	2	y	y	{2+}	a*
09-34c14	9	34	14			Set_17	B	2	y	y	{2+}	a*
10-30c05	10	30	05			Set_17	B	2	y	y	{2+}	a*
10-30c07	10	30	07			Set_17	B	2	y	y	{2+}	a*
12-30c09	12	30	09			Set_17	B	2	y	y	{2+}	a*
VSG-28-10-10	LM_10	28	10	Y		Set_17	B	2	y	y	{2+}	a*
01-27c05	1	27	05			Set_18	B	1	n	n	{3}	—
03-27c02	3	27	02			Set_18	B	1	n	n	0	—
03-27c08	3	27	08			Set_18	B	1	n	n	0	—
03-27c09	3	27	09			Set_18	B	1	n	n	0	—
03-29c05	3	29	05			Set_18	B	1	n	n	0	—
03-29c18	3	29	18			Set_18	B	1	n	n	0	—
03-32c01	3	32	01			Set_18	B	1	n	n	0	—
03-32c02	3	32	02			Set_18	B	1	n	n	0	—
03-32c03	3	32	03			Set_18	B	1	n	n	0	—
03-32c04	3	32	04			Set_18	B	1	n	n	0	—
03-32c05	3	32	05			Set_18	B	1	n	n	0	—
03-32c06	3	32	06			Set_18	B	1	n	n	0	—
03-32c09	3	32	09			Set_18	B	1	n	n	0	—
03-32c10	3	32	10			Set_18	B	1	n	n	0	—
04-27c01	4	27	01			Set_18	B	1	n	n	0	—
04-29c17	4	29	17			Set_18	B	1	n	n	0	—
VSG-28-11-01	LM_11	28	01	Y		Set_18	B	1	n	n	0	—
VSG-28-11-05	LM_11	28	05	Y		Set_18	B	1	n	n	0	—
10-10c10	10	10	10			Set_19	A	1	y	n	0	—
10-10c11	10	10	11			Set_19	A	1	y	n	0	—
VSG-28-11-02	LM_11	28	02	Y		Set_19	A	1	y	n	0	—
VSG-28-11-03	LM_11	28	03	Y		Set_19	A	1	y	n	0	—
VSG-28-11-04	LM_11	28	04	Y		Set_19	A	1	y	n	0	—
VSG-28-11-07	LM_11	28	07	Y		Set_19	A	1	y	n	0	—
VSG-28-11-08	LM_11	28	08	Y		Set_19	A	1	y	n	0	—
VSG-28-11-11	LM_11	28	11	Y		Set_19	A	1	y	n	0	—
05-27c22	5	27	22			Set_20	B	3	n	n	5	—
05-27c24	5	27	24			Set_20	B	3	n	n	5	—
VSG-28-11-09	LM_11	28	09	Y		Set_20	B	3	n	n	3	—
02-32c01	2	32	01	Y		Set_21	A	2	y	n	1	—
02-32c02	2	32	02	Y		Set_21	A	2	y	n	1	—
02-32c03	2	32	03	Y		Set_21	A	2	y	n	1	—
02-32c04	2	32	04	Y		Set_21	A	2	y	n	1	—
02-32c05	2	32	05	Y		Set_21	A	2	y	n	1	—
02-32c07	2	32	07	Y		Set_21	A	2	y	n	1	—
02-32c09	2	32	09	Y		Set_21	A	2	y	n	1	—
02-32c10-del	2	32	10		A1073del~	Set_21	A	2	y	n	1	—
02-32c11	2	32	11	Y		Set_21	A	2	y	n	1	—
02-32c12	2	32	12	Y		Set_21	A	2	y	n	1	—

Table 7.2 (continued)

Clone name	Infection	Time point	Clone #	int. seq?	Mods	Set	type		dons for NTD?	mosaic from dons?	3' don. bound.	3' don
							N	C				
04-31c27	4	31	27			Set_21	A	2	y	n	1	—
04-31c30-del	4	31	30		T1490del~	Set_21	A	2	y	n	1	—
04-31c47	4	31	47			Set_21	A	2	y	n	1	—
05-27c30	5	27	30			Set_21	A	2	y	n	1	—
05-29c12	5	29	12			Set_21	A	2	y	n	1	—
05-29c19_x	5	29	19			Set_21	A	2	y	n	1	—
05-29c30	5	29	30			Set_21	A	2	y	n	1	—
05-29c33	5	29	33			Set_21	A	2	y	n	1	—
05-31c01	5	31	01			Set_21	A	2	y	n	2	—
05-31c02	5	31	02			Set_21	A	2	y	n	2	—
05-31c07	5	31	07			Set_21	A	2	y	n	2	—
05-31c08_x	5	31	08			Set_21	A	2	y	n	2	—
05-31c09	5	31	09			Set_21	A	2	y	n	2	—
05-31c10	5	31	10			Set_21	A	2	y	n	2	—
05-31c11	5	31	11			Set_21	A	2	y	n	2	—
05-31c12	5	31	12			Set_21	A	2	y	n	2	—
05-31c13	5	31	13			Set_21	A	2	y	n	2	—
05-31c16	5	31	16			Set_21	A	2	y	n	2	—
05-31c19	5	31	19		atyp	Set_21	A	2	y	n	2	—
05-31c20	5	31	20			Set_21	A	2	y	n	2	—
05-31c22	5	31	22			Set_21	A	2	y	n	2	—
05-31c23	5	31	23			Set_21	A	2	y	n	2	—
05-31c24	5	31	24			Set_21	A	2	y	n	2	—
05-31c26	5	31	26		atyp	Set_21	A	2	y	n	2	—
05-31c30	5	31	30			Set_21	A	2	y	n	2	—
05-31c32	5	31	32			Set_21	A	2	y	n	2	—
VSG-28-11-13	LM_11	28	13	Y		Set_21	A	2	y	n	2	—
VSG-28-11-14	LM_11	28	14	Y		Set_21	A	2	y	n	2	—
02-20c01-add	2	20	01		G404add	Set_22	A	6	y	y	5	?
02-20c06	2	20	06	Y		Set_22	A	3	y	y	4	b*
02-20c07	2	20	07	Y		Set_22	A	3	y	y	4	b*
02-20c08	2	20	08	Y		Set_22	A	3	y	y	4	k
02-20c09	2	20	09	Y		Set_22	A	3	y	y	4	b*
02-20c11	2	20	11	Y		Set_22	A	3	y	y	4	k
02-23c01	2	23	01	Y		Set_22	A	3	y	y	4	b*
02-23c02	2	23	02	Y		Set_22	A	3	y	y	4	b*
02-23c04	2	23	04	Y		Set_22	A	3	y	y	4	b*
02-23c10	2	23	10	Y	atyp~	Set_22	A	3	y	y	4	k
02-23c11	2	23	11	Y		Set_22	A	3	y	y	4	b*
02-23c12	2	23	12	Y		Set_22	A	3	y	y	4	b*
04-21c03	4	21	03			Set_22	A	2	y	y	2	l
04-21c05	4	21	05			Set_22	A	2	y	y	2	l
04-21c09	4	21	09			Set_22	A	3	y	y	4	b*
04-21c10	4	21	10			Set_22	A	3	y	y	4	b*
04-21c11	4	21	11			Set_22	A	2	y	y	2	l
04-21c13	4	21	13			Set_22	A	3	y	y	3	b
04-21c14-del	4	21	14		C403del	Set_22	A	3	y	y	4	b*
04-21c17-add	4	21	17		C770add	Set_22	A	3	y	y	4	b*
04-21c19	4	21	19			Set_22	A	3	y	y	4	b*
04-21c22	4	21	22			Set_22	A	3	y	y	3	b
04-21c23	4	21	23		stop	Set_22	A	2	y	y	2	l
04-21c24	4	21	24			Set_22	A	2	y	y	2	l
04-21c27	4	21	27			Set_22	A	3	y	y	4	k
04-21c28	4	21	28			Set_22	A	3	y	y	3	b
04-21c30-del	4	21	30		A1003del~	Set_22	A	2	y	y	2	l
04-21c32	4	21	32			Set_22	A	3	y	y	4	k
04-21c36	4	21	36			Set_22	A	3	y	y	4	b*
04-21c37	4	21	37			Set_22	A	3	y	y	4	b*
04-21c38-del	4	21	38		A1003del~	Set_22	A	2	y	y	2	l
04-21c41	4	21	41			Set_22	A	3	y	y	4	k
04-21c43	4	21	43			Set_22	A	3	y	y	4	b*
11-17c01	11	17	01			Set_22	A	3	y	y	4	b*

Table 7.2 (continued)

Clone name	Infection	Time point	Clone #	int. seq?	Mods	Set	type		dons for NTD?	mosaic from dons?	3' don. bound.	3' don
							N	C				
11-17c10	11	17	10			Set_22	A	3	y	y	4	b*
05-29c14	5	29	14			Set_23	A	3	n	n	0	—
06-29c11	6	29	11			Set_24	A	2	p	y	2+	—
04-21c07	4	21	07			Set_25	A	2	p	n	2+	?
04-21c16	4	21	16			Set_25	A	3	p	n	4	?
03-29c01	3	29	01			Set_26	A	2	n	n	0	—
03-29c23	3	29	23			Set_26	A	2	n	n	0	—
02-23c09	2	23	09	Y		Set_27	B	3	y	n	6	j
04-23c27	4	23	27			Set_27	B	1	y	n	6	i*
04-23c41_x	4	23	41		failed R	Set_27	B	X	y	n	X	—
04-23c49_x	4	23	49		loss N	Set_27	B	1	y	n	6	?
04-27c33	4	27	33			Set_27	B	5b	y	n	6	h
04-27c48	4	27	48			Set_27	B	5b	y	n	6	h
05-29c25	5	29	25			Set_27	B	5b	y	n	6	h
05-29c29	5	29	29			Set_27	B	1	y	n	6	i*
03-32c12	3	32	12			Set_28	B	1	n	n	0	—
01-27c04	1	27	04			Set_29	B	3	p	y	0	—
01-29c08	1	29	08			Set_29	B	3	p	y	0	—
02-23c06	2	23	06	Y		Set_29	B	3	p	y	0	—
02-27c07	2	27	07	Y		Set_29	B	3	p	y	0	—
02-27c10_x	2	27	10			Set_29	B	3	p	y	0	—
04-23c33	4	23	33			Set_29	B	3	p	y	0	—
04-27c03	4	27	03			Set_29	B	3	p	y	0	—
04-27c26	4	27	26			Set_29	B	3	p	y	0	—
04-21c26-draft	4	21	26			Set_30	A	1	n	n	0	—
04-23c16	4	23	16			Set_31	A	3	y	y	5	n
04-23c50	4	23	50			Set_31	A	3	y	y	5	n
04-27c06	4	27	06			Set_31	A	3	y	y	5	n
05-27c06	5	27	06			Set_31	A	3	y	y	4	b
05-29c32	5	29	32			Set_31	A	3	y	y	4	b
05-31c14	5	31	14			Set_31	A	3	y	y	4	b
05-31c18	5	31	18			Set_31	A	3	y	y	4	b
05-31c21	5	31	21			Set_31	A	3	y	y	4	b
05-31c31_x	5	31	31		failed F	Set_31	A	3	y	y	4	b
06-27c06	6	27	06			Set_31	A	2	y	y	2	—
04-31c42	4	31	42			Set_32	B	2	p	y	2+	a
04-31c45	4	31	45			Set_32	B	2	p	y	5	a
11-30c15	11	30	15			Set_33	B	2	y	y	0	—
11-30c21	11	30	21			Set_33	B	2	y	y	0	—
11-30c27	11	30	27			Set_33	B	2	y	y	0	—
11-30c36	11	30	36			Set_33	B	2	y	y	0	—
03-27c10	3	27	10			Set_34	B	5b	y	n	5	h
05-31c25	5	31	25			Set_35	B	1	y	n	6	?
05-31c29	5	31	29			Set_35	B	1	y	n	6	?
09-20c09	9	20	09			Set_36	C	2?	y	n	2	—
04-23c01	4	23	01			Set_37	B	1	n	n	4	o
04-23c06	4	23	06			Set_37	B	6	n	n	4	p
04-23c08	4	23	08			Set_37	B	6	n	n	4	p
02-23c07	2	23	07	Y		Set_38	B	3	p	n	0	—
02-29c03	2	29	03	Y		Set_39	A	1	y	n	0	—
02-29c04_x	2	29	04		failed F	Set_39	A	1	y	n	0	—
05-29c02	5	29	02			Set_39	A	1	y	n	0	—
05-31c04	5	31	04			Set_39	A	1	y	n	0	—
05-31c15	5	31	15			Set_39	A	1	y	n	0	—
06-23c02	6	23	02			Set_39	A	1	y	n	0	—
06-23c12	6	23	12			Set_39	A	1	y	n	0	—
02-27c01	2	27	01	Y		Set_40	A	2	y	y	2	—
02-27c06	2	27	06	Y		Set_40	A	2	y	y	1	—
02-27c08	2	27	08	Y		Set_40	A	2	y	y	1	—
02-27c11	2	27	11	Y		Set_40	A	2	y	y	1	—
02-27c12	2	27	12	Y		Set_40	A	2	y	y	1	—
04-23c10	4	23	10			Set_40	A	2	y	y	2	—

Table 7.2 (continued)

Clone name	Infection	Time point	Clone #	int. seq?	Mods	Set	type		dons for NTD?	mosaic from dons?	3' don. bound.	3' don
							N	C				
04-23c25-del	4	23	25		A823del~	Set_40	A	2	y	y	1	—
05-23c19	5	23	19			Set_41	A	2	n	n	0	—
02-23c08	2	23	08	Y		Set_42	C	1	y	y	0	—
02-29c01	2	29	01	Y		Set_43	A	2	y	y	2	—
12-30c04	12	30	04			Set_43	A	2	y	y	2	—
05-23c13	5	23	13			Set_44	B	2	y	n	0	—
04-27c02	4	27	02			Set_45	B	2	n	n	0	—
04-23c20	4	23	20			Set_46	A	2	y	y	1	—
04-23c04	4	23	04			Set_47	A	2	n	n	0	—
05-23c02	5	23	02			Set_48	A	2	y	y	6	?
03-29c04	3	29	04			Set_49	A	2	p	y	{1}	—
03-29c24	3	29	24			Set_49	A	2	p	y	0	—
04-23c03	4	23	03			Set_49	A	2	p	y	0	—
04-27c10	4	27	10			Set_49	A	2	p	y	0	—
04-29c01	4	29	01			Set_49	A	2	p	y	0	—
04-29c03	4	29	03			Set_49	A	2	p	y	0	—
04-29c05	4	29	05			Set_49	A	2	p	y	0	—
04-29c12	4	29	12			Set_49	A	2	p	y	0	—
04-29c36	4	29	36			Set_49	A	2	p	y	0	—
04-31c32-del	4	31	32		A1099del~	Set_49	A	2	p	y	0	—
04-31c41	4	31	41			Set_49	A	2	p	y	0	—
04-21c21	4	21	21			Set_50	A	2	n	n	1	—
05-31c03_x	5	31	03		failed F	Set_50	A	2	n	n	6	m
05-31c17	5	31	17			Set_50	A	2	n	n	6	m
05-31c33	5	31	33			Set_50	A	2	n	n	6	m
03-29c02	3	29	02			Set_51	A	6	p	y	0	—
03-29c12	3	29	12			Set_51	A	6	p	y	0	—
05-23c05-add	5	23	05		A1084add	Set_51	A	6	p	y	0	—
05-23c22	5	23	22			Set_51	A	6	p	y	0	—
05-23c23	5	23	23			Set_51	A	6	p	y	0	—
05-23c25	5	23	25			Set_51	A	6	p	y	0	—
02-29c09	2	29	09	Y		Set_52	A	2	n	n	0	—
05-23c14	5	23	14			Set_53	A	5a	n	n	0	—
03-29c20	3	29	20			Set_54	A	5a	n	n	6	?
04-31c02	4	31	02			Set_54	A	5a	n	n	6	?
04-31c03	4	31	03			Set_54	A	5a	n	n	6	?
04-31c04	4	31	04			Set_54	A	5a	n	n	5	?
04-31c09	4	31	09			Set_54	A	5a	n	n	2+	?
04-31c11	4	31	11			Set_54	A	5a	n	n	6	?
04-31c12	4	31	12			Set_54	A	5a	n	n	6	?
04-31c13_x	4	31	13		failed R	Set_54	A	X	n	n	X	—
04-31c14	4	31	14			Set_54	A	5a	n	n	6	?
04-31c29	4	31	29			Set_54	A	5a	n	n	5	?
03-27c04	3	27	04			Set_55	A	3	y	n	4	f
03-27c06	3	27	06			Set_55	A	3	y	n	4	f
03-27c07	3	27	07			Set_55	A	3	y	n	4	f
04-23c35-del	4	23	35		A1154del~	Set_56	B	2	y	n	5	a
05-27c14	5	27	14			Set_56	B	3	y	n	4	b
05-29c03	5	29	03			Set_56	B	3	y	n	4	b
02-27c02	2	27	02			Set_57	B	6	y	n	4	p
02-29c02	2	29	02			Set_57	B	6	y	n	4	p
02-29c06	2	29	06			Set_57	B	6	y	n	4	p
02-29c11	2	29	11			Set_57	B	6	y	n	4	p
04-21c25	4	21	25		stop	Set_58	A	1d	y	n	6	s
01-29c06	1	29	06			Set_59	A	2	y	y	1	—
02-29c07	2	29	07	Y		Set_59	A	2	y	y	1	—
03-27c03	3	27	03			Set_59	A	2	y	y	1	—
03-29c21	3	29	21			Set_59	A	2	y	y	1	—
05-23c17-del	5	23	17		A566del~	Set_59	A	2	y	y	2+	a*
05-23c21-del	5	23	21		A566del~	Set_59	A	2	y	y	2+	a*
05-27c10	5	27	10			Set_59	A	2	y	y	2	—
10-30c06-del	10	30	06		A1087add~	Set_59	A	2	y	y	2+	a*

Table 7.2 (continued)

Clone name	Infection	Time point	Clone #	int. seq?	Mods	Set	type		dons for NTD?	mosaic from dons?	3' don. bound.	3' don
							N	C				
08-34c06	8	34	06			Set_60	A	1	y	n	5	g
08-34c16	8	34	16			Set_60	A	1	y	n	5	g
05-29c28	5	29	28			Set_61	A	2	p	n	0	—
04-21c35	4	21	35			Set_62	A	5a	y	y	5	—
04-23c09	4	23	09			Set_62	A	5a	y	y	1	—
11-27c05	11	27	05			Set_62	A	5a	y	y	5	—
04-23c02	4	23	02			Set_63	B	2	p	y	1	—
04-23c05	4	23	05			Set_63	B	2	p	y	1	—
11-30c04	11	30	04			Set_63	B	2	p	y	2	—
11-30c19	11	30	19			Set_63	B	2	p	y	2	—
02-27c03	2	27	03	Y		Set_64	A	3	y	y	5	c*
02-27c04	2	27	04	Y		Set_64	A	3	y	y	5	c*
02-27c05	2	27	05	Y		Set_64	A	3	y	y	5	c*
03-32c08	3	32	08			Set_64	A	3	y	y	5	c*
04-27c22	4	27	22			Set_64	A	2	y	y	5	d
04-29c07	4	29	07			Set_64	A	2	y	y	5	d
05-27c02	5	27	02			Set_64	A	2	y	y	5	d
05-27c05	5	27	05			Set_64	A	2	y	y	2+	e
05-27c26	5	27	26			Set_64	A	2	y	y	5	d
05-29c26	5	29	26			Set_64	A	2	y	y	2+	e
01-32c05	1	32	05		stop	Set_65	A	5a	n	n	0	—
02-29c08	2	29	08	Y		Set_66	B	3	n	n	0	—
02-32c08	2	32	08	Y		Set_67	A	2	p	n	5	?
04-21c01	4	21	01			Set_68	B	3	n	n	0	—
04-21c31	4	21	31			Set_68	B	3	n	n	0	—
05-23c11	5	23	11			Set_68	B	5b	n	n	{4}	?
11-30c13	11	30	13			Set_68	B	3	n	n	0	—
11-30c31	11	30	31			Set_68	B	3	n	n	0	—
11-30c32	11	30	32			Set_68	B	3	n	n	0	—
11-30c37	11	30	37			Set_68	B	3	n	n	0	—
04-27c12	4	27	12			Set_69	A	3	n	n	0	—
04-27c13-add	4	27	13		G720add	Set_69	A	3	n	n	0	—
04-27c15	4	27	15			Set_69	A	3	n	n	0	—
04-29c08	4	29	08			Set_69	A	3	n	n	0	—
04-29c09	4	29	09			Set_69	A	3	n	n	0	—
04-29c14	4	29	14			Set_69	A	3	n	n	0	—
04-29c15	4	29	15			Set_69	A	3	n	n	0	—
05-27c11	5	27	11			Set_69	A	3	n	n	{4}	?
05-29c09	5	29	09			Set_69	A	3	n	n	{4}	?
05-23c06	5	23	06		stop	Set_70	B	2	p	n	0	—
05-23c09	5	23	09			Set_71	A	2	p	n	0	—
04-23c13-add	4	23	13		A714add	Set_72	A	3	p	n	0	—
11-14c02	11	14	02			Set_73	B	2	n	n	0	—
11-14c04	11	14	04			Set_73	B	2	n	n	0	—
11-14c05	11	14	05			Set_73	B	2	n	n	0	—
11-14c06	11	14	06			Set_73	B	2	n	n	0	—
11-14c07	11	14	07			Set_73	B	2	n	n	0	—
11-14c08	11	14	08			Set_73	B	2	n	n	0	—
12-10c02	12	10	02			Set_73	B	2	n	n	0	—
12-10c01	12	10	01			Set_74	A	2	y	n	5	?
10-10c09_x	10	10	09		failed F	Set_75		2	n	n	0	—
11-14c03	11	14	03			Set_75	A	2	n	n	0	—
11-21c03	11	21	03			Set_75	A	2	n	n	0	—
11-21c06	11	21	06			Set_75	A	2	n	n	0	—
11-21c08	11	21	08			Set_75	A	2	n	n	0	—
12-10c03	12	10	03			Set_75	A	2	n	n	0	—
12-10c06	12	10	06			Set_75	A	2	n	n	0	—
11-21c01_x	11	21	01		failed F	Set_76		6	y	n	0	—
12-10c07	12	10	07			Set_76	A	6	y	n	0	—
12-10c08	12	10	08			Set_76	A	6	y	n	0	—
12-10c09	12	10	09			Set_77	A	2	y	n	0	—

Table 7.2 (continued)

Clone name	Infection	Time point	Clone #	int. seq?	Mods	Set	type		dons for NTD?	mosaic from dons?	3' don. bound.	3' don
							N	C				
05-27c15	5	27	15			Set_78	A	1	n	n	0	—
05-29c16_x	5	29	16		failed F	Set_78	A	1	n	n	0	—
05-29c22	5	29	22			Set_78	A	1	n	n	0	—
03-29c17	3	29	17			Set_79	B	3	p	n	6	j
05-27c31	5	27	31			Set_80	B	1	y	n	5	—
10-10c06	10	10	06			Set_81	A	1	y	n	0	—
10-10c08	10	10	08			Set_81	A	1	y	n	0	—
10-10c12	10	10	12			Set_81	A	1	y	n	0	—
03-27c01	3	27	01			Set_82	B	3	y	n	0	—
03-27c11	3	27	11			Set_82	B	3	y	n	0	—
10-10c01	10	10	01			Set_82	B	3	y	n	0	—
10-10c07	10	10	07			Set_83	A	2	y	n	0	—
01-29c03	1	29	03			Set_84	B	2	n	n	1	—
06-27c01	6	27	01			Set_84	B	2	n	n	1	—
06-27c02-edit	6	27	02		stop	Set_84	B	2	n	n	1	—
06-27c03	6	27	03			Set_84	B	2	n	n	1	—
06-27c04	6	27	04			Set_84	B	2	n	n	1	—
06-27c05	6	27	05			Set_84	B	2	n	n	1	—
06-27c07-edit	6	27	07		A939add	Set_84	B	2	n	n	1	—
06-27c08	6	27	08			Set_84	B	2	n	n	1	—
06-27c09	6	27	09			Set_84	B	2	n	n	1	—
06-27c10	6	27	10			Set_84	B	2	n	n	1	—
06-27c11-edit	6	27	11		stop A132del	Set_84	B	2	n	n	1	—
06-29c01	6	29	01			Set_84	B	2	n	n	1	—
06-29c02	6	29	02			Set_84	B	2	n	n	1	—
06-29c03	6	29	03			Set_84	B	2	n	n	1	—
06-29c04	6	29	04			Set_84	B	2	n	n	1	—
06-29c06	6	29	06			Set_84	B	2	n	n	1	—
06-29c09	6	29	09			Set_84	B	2	n	n	1	—
06-29c12	6	29	12			Set_84	B	2	n	n	1	—
10-30c01	10	30	01		loss C	Set_84	B	-	n	n	—	—
10-30c04	10	30	04			Set_84	B	2	n	n	2	—
10-30c08	10	30	08			Set_84	B	2	n	n	2	—
10-30c09	10	30	09			Set_84	B	2	n	n	2	—
10-30c12	10	30	12			Set_84	B	2	n	n	2	—
12-30c02	12	30	02			Set_84	B	2	n	n	2	—
12-30c03	12	30	03			Set_84	B	2	n	n	2	—
12-30c10	12	30	10			Set_84	B	2	n	n	2	—
12-30c11	12	30	11			Set_84	B	2	n	n	2	—

Table 7.2. Clone sequence details. Column 'int seq?' indicates the clones that were sequenced with an internal primer. 'Mods' indicates manual modifications made to the sequences (e.g. 'A132del' indicates that an 'A' at 132 was deleted, removing a frameshift and providing a sequence amenable to analysis) and/or abnormalities associated with the sequences. Abnormalities included in-frame stop codons ('stop'), failed forward or reverse reads, or in two cases, the loss of a part of the sequence. 'Atyp' indicates an atypical GPI anchor as a result of a frameshift towards the very 3'-most end of the sequence. A tilde ('~') indicates that the associated abnormality was associated with a tract of A or T nucleotides, and thus may have come about by polymerase slippage, and a '|' indicates it was due to possible sequence loss. 'Dons for NTD?' indicates whether a suitable (set of) donor(s) has been found for the NTD-encoding part of the gene (described in Table 7.2.8). 'Mosaic from dons?' indicates whether the expressed clone is an apparent mosaic from comparisons with the donor sequences. '3' don bound.' indicates whether the expressed *VSG* has undergone 3' donation, and where, the expressed sequence diverges from the NTD donors. The code is as follows: '1' = within GPI anchor sequence, '2' = between GPI anchor sequence and C-terminal-most cysteine, '3' within cysteines of C-terminal-most subdomain, '4' = between subdomains, '5' = within cysteines of N-terminal-most subdomain, '6' = upstream of N-terminal most subdomain. For one-subdomain CTDs, categories '3' and '4' were removed, whereas for type 2 CTDs, which have a more extensive region between the C-terminal-most cysteine and the GPI anchor, '2+' indicates that the boundary was closer to the cysteine than to the GPI anchor. Where the code is in curly brackets, identification was by comparing clones. '3' don' indicates the source of the 3' donor sequence, if known (see Chapter 4 section 4.4.4). '—' = donors not sought, '?' = donor could not be found, '*' = identical patterns in multiple infections.

7.2.4 'Genomic VSG *database'*

The 'genomic *VSG* database' was obtained in May 2010 by collecting all available *VSG* sequences from both TriTrypDB (Aslett *et al.*, 2010) (using the text search query 'vsg' and applying a filter to TREU 927/4 genes) and from VSGdb (all entries) (Marcello *et al.*, 2007). The list was made non-redundant by removing duplicate sequences. Where sequences in the two databases differed, the TriTrypDB sequence was favoured, unless the VSGdb sequence better matched the expressed clone sequence(s).

7.2.5 'Reads database'

The 'reads database' was obtained in December 2011 by collecting the following files from ftp://ftp.sanger.ac.uk/pub/databases/T.brucei_sequences/ :

T_brucei_chrIXa_reads.20Oct2001, T_brucei_chrIXa_reads.29May03, T_brucei_chrIXb_reads.20Oct2001, T_brucei_chrIXb_reads.29May03, T_brucei_chrI_reads.03Feb2000, T_brucei_chrXI_reads.V1_17Oct2002, T_brucei_chrXI_reads.V2_14Mar2003.fas, T_brucei_chrXI_reads.V3_11Jul03.fas, T_brucei_chrX_reads.03Mar2003, T_brucei_chrX_reads.20Oct2001, T_brucei_chrX_reads.30Apr2001, T_brucei_reads.21Dec1999, TbchrIXa_reads_26May04.fas, TbchrIXb_reads_26May04.fas, Tb927_IC_II_reads.fas, Tb927_IC_I_reads.fas, 177bp_repeat.dnaTb927_MC_cons.dna.crunch, 177bp_repeat.dnaTb927_MC_test_reads.dna.crunch, 927MC_readsII.fas, 927MC_readsIII.fas, 927MC_readsIV.fas, 927MC_readsV.fas, 927MC_readsVI.fas, 927MC_readsVII.fas, 927mcIII_v1.cons, 927mcII_v1.cons, 927mcIV_v1.cons, 927mcIX_v1.cons, 927mcVIII_v1.cons, 927mcVII_v1.cons, 927mcVI_v1.cons, 927mcV_v1.cons, Tb927MC_telo1_contigs.fas, Tb927MC_telo2_contigs.fas, Tb927MC_telo2_v2contigs.fas, Tb927MC_telo9_contigs.fas, Tb927MC_telo9_v2contigs.fas, Tb927MCtelo_1_test.fas, Tb927MCtelo_2_test.fas, Tb927MCtelo_9_test.fas, Tb927_MC.cons, Tb927_MCIX_contigs_6Mar08.fas, Tb927_MC_III_reads.fas, Tb927_MC_II_reads.fas, Tb927_MC_IV_reads.fas, Tb927_MC_IX_reads.fas, Tb927_MC_VIII_reads.fas, Tb927_MC_VII_reads.fas, Tb927_MC_VI_reads.fas, Tb927_MC_V_reads.fas, Tb927_MC_test_reads.fas, Tb927_MC_testreads_newlib.fas, uniprot-1-xTb927_MC_cons.dna.crunch, uniprot-1-xTb927_MC_test_reads.dna.crunch, TPunknown_May06.cons, Tbrucei_extrachr_testreads_Dec05.fas, chr_unknown_test_reads_Jan06.fas

7.2.6 Read assemblies

Assembly name	Set	Reads	Source
927mc_III-1b07_assembly	Set_03	927mc_III-1b07.p1k	Tb927_MC_testreads_newlib.fas
		927mc_III-1b07.q1k-rev	Tb927_MC_testreads_newlib.fas
		927mc_III-01b01.q1k-rev	Tb927_MCIII_reads.fas
tryp_X-54b12.q1c-rev_assembly	Set_04	tryp_X-54b12.q1c-rev	T_brucei_chrX_reads.30Apr2001
		tryp_XI-910b10.q1k	T_brucei_chrXI_reads.V3_11Jul03.fas
927mc_VI-6b02_assembly	Set_05	927mc_VI-6b02.p1k	Tb927_MC_VI_reads.fas
		927mc_VI-6b02.q1k-rev	Tb927_MC_VI_reads.fas
		927MC_V-01a09.p1k	927MC_readsV.fas
		927MC_VI-01c10.q1k-	927MC_readsVI.fas
		927MC_V-01a09.q1k-rev	927MC_readsV.fas
tryp_XI-1034g11_assembly	Set_14	tryp_IXa-24h09.q1c-rev	T_brucei_chrIXa_reads.20Oct2001
		tryp_XI-1034g11.p1k	T_brucei_chrXI_reads.V3_11Jul03.fas
		tryp_XI-1034g11.q1k-rev	T_brucei_chrXI_reads.V3_11Jul03.fas
tryp_XI-934d12_assembly	Set_16	tryp_XI-934d12.q1k-rev	T_brucei_chrXI_reads.V3_11Jul03.fas
		tryp_X-334d06.p1c	T_brucei_chrX_reads.03Mar2003
tryp_IXb-234c03_assembly	Set_17	tryp_IXb-234c03.q1c-rev	T_brucei_chrIXb_reads.29May03
		tryp_IXb-151d05-p1c-rev	TbchrIXb_reads_26May04.fas
tryp_XI-1058d01_assembly	Set_17	tryp_XI-1058d01.q1k	T_brucei_chrXI_reads.V3_11Jul03.fas
		tryp_XI-1121e11.p1k	T_brucei_chrXI_reads.V3_11Jul03.fas
		927mc_VI-8a07.p1k-rev	Tb927_MC_VI_reads.fas
tryp_IXa-29b02.q1c_assembly	Set_25	tryp_IXa-29b02.q1c	T_brucei_chrIXa_reads.20Oct2001
		tryp_IXb-301g08.p1c	T_brucei_chrIXb_reads.29May03
927mc_IV-5e11_assembly	Set_29	927mc_IV-5e11.p1k	Tb927_MC_IV_reads.fas
		927mc_IV-5e11.q1k	Tb927_MC_IV_reads.fas
tryp_IXa-10f02_assembly	Set_32	927mc_IX-10d02.q1k	Tb927_MC_IX_reads.fas
		tryp_IXa-10f02.p1c	T_brucei_chrIXa_reads.20Oct2001
tryp_IXa-20a05_assembly	Set_51	tryp_IXa-20a05.p1c	T_brucei_chrIXa_reads.20Oct2001
		tryp_IXa-6d11.q1c	T_brucei_chrIXa_reads.20Oct2001
		tryp_IXb-153c03.p1c-rev	T_brucei_chrIXb_reads.29May03
tryp_IXa-29e12_assembly	Set_54	tryp_IXa-29e12.p1c	T_brucei_chrIXa_reads.20Oct2001
		tryp_IXa-4a12.p1c	T_brucei_chrIXa_reads.20Oct2001
		tryp_IXb-361b07.p1c	T_brucei_chrIXb_reads.29May03

Table 7.3 Read assemblies. 'Source' refers to the file in which the read sequence was found (see section 7.2.5), 'Set' refers to the set for which the read assembly is putative donor. Table continues on following pages.

Assembly name	Set	Reads	Source
tryp_X-99f10_assembly	Set_54	927mc_IV-3d11.p1k-rev	Tb927_MC_IV_reads.fas
		927mc_VI-10d07.p1k	Tb927_MC_VI_reads.fas
		tryp_X-190c10.q1c-rev	T_brucei_chrX_reads.30Apr2001
		tryp_X-214a05.q1c-rev	T_brucei_chrX_reads.30Apr2001
		tryp_X-268e08.q1c	T_brucei_chrX_reads.30Apr2001
		tryp_X-277b10.q1c	T_brucei_chrX_reads.30Apr2001
		tryp_X-338f01.q1c-rev	T_brucei_chrX_reads.20Oct2001
		tryp_X-36c03.q1c	T_brucei_chrX_reads.30Apr2001
		tryp_X-99f10.p1c	T_brucei_chrX_reads.30Apr2001
		tryp_XI-143d03.q1ca-rev	T_brucei_chrXI_reads.V2_14Mar200
tryp_XI-1157a04_assembly	Set_59	tryp_XI-1157a04.p1k	T_brucei_chrXI_reads.V3_11Jul03.fas
		tryp_XI-1157a04.q1k-rev	T_brucei_chrXI_reads.V3_11Jul03.fas
tryp_IXa-10d01_assembly	Set_71	tryp_IXa-25g04.q1c-rev	T_brucei_chrIXa_reads.20Oct2001
		tryp_IXa-10d01.q1c-rev	T_brucei_chrIXa_reads.20Oct2001
tryp_XI-1009a05_assembly	Set_72	tryp_IXa-7a11.p1c-rev	T_brucei_chrIXa_reads.20Oct2001
		tryp_XI-1009a05.p1k-rev	T_brucei_chrXI_reads.V3_11Jul03.fas
		tryp_XI-1055f01.p1k-rev	T_brucei_chrXI_reads.V3_11Jul03.fas
tryp_XI-393f08_assembly	Set_74	tryp_XI-393f08.q1k	T_brucei_chrXI_reads.V2_14Mar200
		tryp_XI-949b02.q1k-rev	T_brucei_chrXI_reads.V3_11Jul03.fas
		tryp_XI-1098g01.q1k-rev	T_brucei_chrXI_reads.V3_11Jul03.fas
		trypA31g9.q1t-rev	T_brucei_chrI_reads.03Feb2000
		trypA29a10.p1p-rev	T_brucei_chrI_reads.03Feb2000
		tryp_XI-204f04.p1c-rev	T_brucei_chrXI_reads.V2_14Mar200
		tryp_X-48f04.p1c-rev	T_brucei_chrX_reads.30Apr2001
		tryp_XI-393f08.p1k-rev	T_brucei_chrXI_reads.V2_14Mar200
tryp_IXb-375c10_assembly	Set_76	tryp_IXb-375c10.p1c	T_brucei_chrIXb_reads.29May03
		tryp_IXb-62e09.q1c-rev	T_brucei_chrIXb_reads.29May03
		tryp_IXb-334g12.p1c-rev	T_brucei_chrIXb_reads.29May03
		tryp_XI-1117h04.q1k	T_brucei_chrXI_reads.V3_11Jul03.fas
		tryp_IXb-274b10.q1c-rev	T_brucei_chrIXb_reads.29May03
		tryp_IXb-1e12.q1c-rev	T_brucei_chrIXb_reads.29May03
		tryp_IXb-282d11.q1c	T_brucei_chrIXb_reads.29May03
		tryp_IXb-342g05.p1c-rev	T_brucei_chrIXb_reads.29May03
		tryp_IXb-375c10.q1c-rev	T_brucei_chrIXb_reads.29May03
		tryp_XI-1117h04.p1k-rev	T_brucei_chrXI_reads.V3_11Jul03.fas
		tryp_IXb-125h06.q1c-rev	T_brucei_chrIXb_reads.29May03
		tryp_IXb-282d11.p1c-rev	T_brucei_chrIXb_reads.29May03

Table 7.3 (continued)

Assembly name	Set	Reads	Source
tryp_XI-325h02_assembly	Set_77	tryp_XI-1024b10.p1k-rev	T_brucei_chrXI_reads.V3_11Jul03.fas
		tryp_XI-1128e06.p1kw-	T_brucei_chrXI_reads.V3_11Jul03.fas
		tryp_XI-1139e06.q1k	T_brucei_chrXI_reads.V3_11Jul03.fas
		tryp_XI-238f01.p1c-rev	T_brucei_chrXI_reads.V2_14Mar200
		tryp_XI-325h02.q1k-rev	T_brucei_chrXI_reads.V2_14Mar200
		tryp_XI-325h02.q1ka-rev	T_brucei_chrXI_reads.V2_14Mar200
		tryp_XI-330f02.q1k	T_brucei_chrXI_reads.V2_14Mar200
		tryp_XI-330f02.q1ka	T_brucei_chrXI_reads.V2_14Mar200
		tryp_XI-912a01.p1k-rev	T_brucei_chrXI_reads.V3_11Jul03.fas
927mc_III-7e05_assembly	Set_79	927mc_III-7e05.p1k	Tb927_MC_III_reads.fas
		927mc_III-7e05.q1k	Tb927_MC_III_reads.fas
927mc_VI-4a12_assembly	Set_80	927mc_VI-4a12.p1k	Tb927_MC_VI_reads.fas
		927mc_VI-5d03.p1k	Tb927_MC_VI_reads.fas
		927mc_V-8d11.p1k	Tb927_MC_V_reads.fas
		927mc_V-1d03.q1k-rev	927mcV_v1.cons
		927mc_VI-12c01.q1k-rev	Tb927_MC_VI_reads.fas
		927mc_VI-9c10.q1k	Tb927_MC_VI_reads.fas
		927mc_VI-11g04.p1k-rev	Tb927_MC_VI_reads.fas
927mc_VI-4g03_assembly	Set_81	927mc_VI-4g03.q1k	Tb927_MC_VI_reads.fas
		927mc_V-6d03.q1k	Tb927_MC_V_reads.fas
		927mc_VI-2g12.q1k	Tb927_MC_testreads_newlib.fas
		927mc_V-6d05.p1k	Tb927_MC_V_reads.fas
		927mc_VI-10e07.q1k	Tb927_MC_VI_reads.fas
927mc_VI-6c05_assembly	Set_82	927mc_VI-6c05.p1k-rev	Tb927_MC_VI_reads.fas
		927mc_VI-6c05.q1k-rev	Tb927_MC_VI_reads.fas
		927mc_VI-4a05.p1k-rev	Tb927_MC_VI_reads.fas
		927mc_VI-5e11.p1k	Tb927_MC_VI_reads.fas
		927mc_VI-5e11.q1k-rev	Tb927_MC_VI_reads.fas
		927mc_VI-5h02.p1k	Tb927_MC_VI_reads.fas
		927mc_VI-7e06.q1k-rev	Tb927_MC_VI_reads.fas
		927mc_VI-13c03.q1k	Tb927_MC_VI_reads.fas
927mc_VII-13h07_assembly	Set_83	927mc_VII-13h07.p1k-	Tb927_MC_VII_reads.fas
		927mc_VII-13h07.q1k-	Tb927_MC_VII_reads.fas
		927mc_VII-5h01.p1k-rev	Tb927_MC_VII_reads.fas
		927mc_VII-7g03.q1k-rev	Tb927_MC_VII_reads.fas
		927mc_VIII-1f08.q1k-rev	Tb927_MC_testreads_newlib.fas
		927mc_VII-11c05.p1k	Tb927_MC_VII_reads.fas
		927mc_VI-11f11.p1k-rev	Tb927_MC_VI_reads.fas
927mc_VIII-14d09_assembly	Set_84	927mc_VIII-14d09.p1k	Tb927_MC_VIII_reads.fas
		927mc_VIII-14d09.q1k	Tb927_MC_VIII_reads.fas
		927mc_IV-5h11.p1k	Tb927_MC_IV_reads.fas
		927mc_VII-5a05.q1k-rev	Tb927_MC_VII_reads.fas

Table 7.3 (continued)

Assembly name	Set	Reads	Source
d927mc_VIII-3d10_assembly	Set_85	927mc_VIII-3d10.p1k-rev	Tb927_MC_VIII_reads.fas
		927mc_VIII-3d10.q1k-rev	927mcVIII_v1.cons
		927mc_VIII-13b05.q1k	Tb927_MC_VIII_reads.fas
		927mc_VIII-14d10.q1k-	Tb927_MC_VIII_reads.fas

Table 7.3 (continued)

7.2.7 Number of sets detected across infection

Infection	Day of sampling																		sets / infection
	2	3	7	9	10	14	17	20	21	22	23	24	27	28	29	30	31	32	34	
FF	1																			1
1								1			1		6		5			5		11
2								1			5		4		6			2		15
3								1			1		6		8			5		16
4									11		15		10		6		8			31
5									2		13		14		13		7			30
6								2			2		3		4					9
8								1											2	3
9								2											2	4
10					6											3				9
11			2		4	3	3		3				3			6				15
12					7		3					2				3				12
LM_01		1																		1
LM_02		1																		1
LM_03				4																4
LM_04				3																2
LM_05						2														2
LM_06						1														1
LM_07										4										4
LM_08									1											1
LM_09												1								1
LM_10														3						3
LM_11														4						4
Sets / timepoint	1	1	2	5	13	4	4	4	15	4	31	3	27	7	29	10	12	11	3	**83**

Table 7.4 Number of sets detected at different timepoints. The table shows, for the 22 infections under consideration, the number of sets detected at each timepoint. The row and column totals do not equal the sums of the row or column as some sets were detected in multiple infections and timepoints.

7.2.8 Genomic copies

Set	Class	Donor(s)	Donor class	N-type	C-type
Set_01	single	Tb10.v4.0001	a	A	2a
Set_02	unknown	Unknown	?		
Set_03	mc	927mc_III-1b07_assembly	i,m	A	6
Set_04	mosaic	tryp_X-54b12.q1c-rev_assembly	i,r	B	-
		Tb10.v4.0061	nc,p	B	3
Set_05	mc	927mc_VI-6b02_assembly	i,m	A	5
Set_06	unknown	Unknown	?		
Set_07	unknown	Unknown	?		
Set_08	single	Tb09.v4.0077	f	B	2
Set_09	single	Tb09.244.1580	f	A	3
Set_10	mosaic	Tb11.31.0001	f	A	6
		Tb10.v4.0161	n,p	A	6
		Tb11.14.0001	n,p	A	6
Set_11	mosaic	Tb927.3.190	f	A	2
		Unknown	?		
Set_12	single	Tb11.30.0005	c,p	A	4
Set_13	single	Tb09.v4.0102	a	B	3u
Set_14	mosaic	Tb11.09.0005	a	B	1u
		Tb11.13.0003	n,p	B	3
		tryp_XI-1034g11_assembly	nc,r	B	d
		Tb10.v4.0009	n,p	B	2
Set_16	mosaic	Tb11.48.0003	c,p	A	3
		tryp_XI-934d12.q1k_assembly	nci,r	A	d
Set_17	mosaic	Tb11.57.0032	nc,p	B	d
		Tb09.244.0090	n,p	B	1
		tryp_XI-1058d01_assembly	ni,r	B	-
Set_18	unknown	Unknown	?		
Set_19	single	Tb927.5.5080	f	A	2
Set_20	mosaic	Tb927.7.6530	c,p	B	3
		Unknown	?		
Set_21	single	Tb10.v4.0088	f	A	2
Set_22	mosaic	Tb05.5K5.330	c,p	A	2
		Tb11.21.0004	a	A	2u
		Tb927.5.4840	a	A	2u
Set_23	unknown	Unknown	?		
Set_24	mosaic	Tb08.27P2.460	f	A	2
		trypA24a5.p1p	i,r		
Set_25	read	tryp_IXa-29b02.q1c_assembly	ci,r	A	1
Set_26	unknown	Unknown	?		
Set_27	single	Tb09.142.0470	c,p	B	3
Set_28	unknown	Unknown	?		
Set_29	mosaic	Tb09.244.0150	c,p	B	3
		927mc_IV-5e11_assembly	ci,m	B	-

Table 7.5 Genomic copies. See the end of the Table for details.

Set	Class	Donor(s)	Donor	N-type	C-type
Set_30	unknown	Unknown	?	-	-
Set_31	mosaic	Tb09.354.0060	a	A	2u
		Tb927.6.5370	f	A	2
Set_32	mosaic	Tb09.v2.0090	c,p	B	d
		tryp_IXa-10f02_assemblyz	ni,r	B	-
Set_33	mosaic	Tb09.v4.0071	c,p	B	d
		Tb10.v4.0031	f	B	2
Set_34	single	Tb09.v4.0073	c,p	B	1
Set_35	single	Tb09.v4.0075	n,p	B	1
Set_36	single	Tb09.v4.0088	a	C	2u
Set_37	mosaic	Tb09.v4.0123	c,p	B	1
		Unknown	?		
Set_38	unknown	Unknown	?		
Set_39	single	Tb09.v4.0177	f	A	1
Set_40	mosaic	Tb11.1480	a	A	2u
		Tb10.v4.0063	n,p	A	2u
Set_41	unknown	Unknown	?		
Set_42	mosaic	Tb10.v4.0070	f	C	1
		Tb927.8.170	a	C	1u
Set_43	mosaic	Tb10.v4.0102	a	A	2u
		Tb10.v4.0081	n,p	A	2u
Set_44	single	Tb10.v4.0145	f	Bu	2
Set_45	unknown	Unknown	?		
Set_46	mosaic	Tb11.16.0003	a	A	2u
		Tb10.v4.0095	f	A	2
Set_47	unknown	Unknown	?		
Set_48	mosaic	Tb11.35.0001	a	A	1u
		Tb09.244.1390	c,p	A	1
Set_49	mosaic	Tb11.43.0002	nc,p	A	2
		tryp_XI-909g03.q1k-rev	i,r	-	-
Set_50	single	Tb11.49.0005	c,p	A	2
Set_51	mosaic	Tb11.57.0023	nc,p	A	1
		tryp_IXa-20a05_assembly	i,r	A	-
Set_52	unknown	Unknown	?		
Set_53	unknown	Unknown	?		
Set_54	mosaic	Tb11.v4.0029	f	A	5
		tryp_IXa-29e12_assembly	i,r	-	-
		tryp_X-99f10_assembly	nc,r	A	5d
		Unknown	?		
Set_55	single	Tb11.v4.0031	c,p	A	d
Set_56	single	Tb11.v4.0035	c,p	B	1
Set_57	single	Tb11.v4.0038	f	B	6
Set_58	single	Tb927.3.180	a	A	1u
Set_59	mosaic	Tb927.3.270	n,p	A	2
		tryp_XI-1157a04_assembly	nci,r	d	d

Table 7.5 (continued)

Set	Class	Donor(s)	Donor class	N-type	C-type
Set_60	single	Tb927.3.490	a	A	2u
Set_61	mc	927mc_VIII-3d10_assembly	i,m	A	-
Set_62	mosaic	Tb927.5.4670	a	A	5u
		Tb05.5K5.320	f	A	5u
		Tb09.v4.0061	n,p	A	5u
Set_63	mosaic	Tb927.5.4930	n,p	B	2
		tryp_XI-1007d10.q1k	ni,r	-	-
Set_64	mosaic	Tb927.8.420	c,p	A	1
		Tb09.244.0360	nc,p	A	1
Set_65	unknown	Unknown	?		
Set_66	unknown	Unknown	?		
Set_67	read	tryp_XI-1084f09.p1k	nci,r	A	-
Set_68	unknown	Unknown	?		
Set_69	mosaic	Unknown	?		
Set_70	unknown	Unknown	?		
Set_71	read	tryp_IXa-10d01_assembly	i,r	A	-
Set_72	read	tryp_XI-1009a05_assembly	i,r	A	-
Set_73	unknown	Unknown	?		
Set_74	read	tryp_XI-393f08_assembly	i,r	A	2
Set_75	unknown	Unknown	?		
Set_76	read	tryp_IXb-375c10_assembly	r	A	6
Set_77	read	tryp_XI-325h02_assembly	i,r	A	2
Set_78	unknown	Unknown	?		
Set_79	mc	927mc_III-7e05_assembly	ci,m	B	-
Set_80	mc	927mc_VI-4a12_assembly	c,m	B	d
Set_81	mc	927mc_VI-4g03_assembly	m	A	1
Set_82	mc	927mc_VI-6c05_assembly	m	B	3
Set_83	mc	927mc_VII-13h07_assembly	i,m	A	-
Set_84	mosaic	927mc_VIII-14d09_assembly	m	A	2
		Unknown	?		

Table 7.5 Genomic copies. Table shows, for each set, the putative donor sequence(s) ('Donor(s)'). Details provided by L. Marcello (Ph.D. Thesis, 2006, University of Glasgow) were confirmed and supplemented by additional analyses to type and class sequences. 'Donor class': 'f' = full-length intact from arrays, 'a' = atypical, 'p' = pseudogene, 'n' = NTD frameshifts or stop codons, 'c' = CTD frameshifts or premature stop codons, 'i' = incomplete sequence, 'm' = minichromosomal read or read assembly, 'r' = read or read assembly, '?' = unknown, 'u' = uncertain functionality, 'd' = degenerate.

7.2.9 Summary of genomic copies

Class		Count
full length intact		**16**
atypical		**15**
array pseudogene		**31**
of which	ψN	*11*
	ψC	*15*
	ψN&C	*5*
minichromosomal read		**10**
of which	ψN	*0*
	ψC	*3*
	ψN&C	*0*
	incom.	*6*
read		**19**
of which	ψN	*3*
	ψC	*1*
	ψN&C	*5*
	incom.	*16*
inferred unknown		**27**
Total genes		**91**
including unknowns		***118***

Table 7.6 Summary of classes of donor genes. 'incom.' = incomplete sequence.

7.3 Supplementary information to Chapter 4

7.3.1 Examples of point mutation

Two examples are provided (Figure 7.3, Figure 7.4) to illustrate the presence of point mutations in expressed *VSG*s.

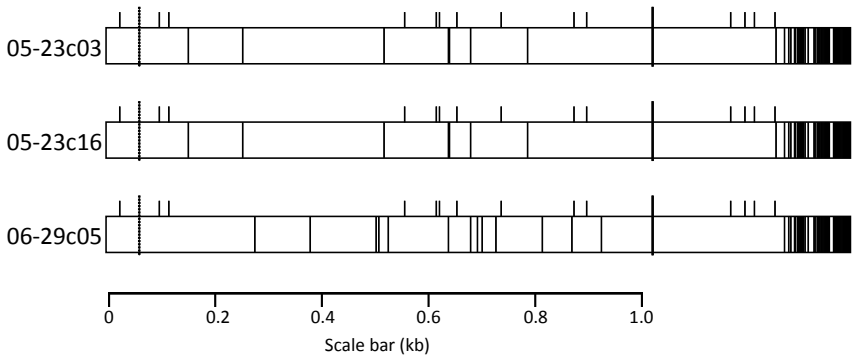


Figure 7.3 Point mutations in Set_13. Primary donor: Tb09.v4.0102. Images have been drawn similarly to those in Chapter 4 (Figure 4.13), except in monochrome: the position of cysteine codons are indicated by a thick bar projecting from the top of each diagram and the boundaries of the mature NTD are indicated by dashed and solid lines at the N- and C-terminal ends respectively. Black lines spanning each image, denoting point mutations, have been drawn in triple thickness for clarity. 05-23c03 and 05-23c16 were from independent reverse transcription-amplification reactions from the same mouse. They show a consistent pattern of point mutation, which is not shared with the clone from a different infection, 06-29c05. Each of these clones has also undergone 3' donation. Further clones from mouse 05 show differing point mutation patterns (data not shown).

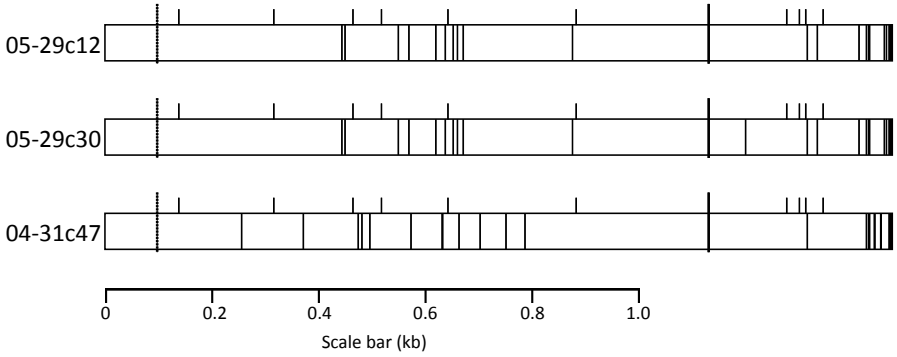


Figure 7.4 Point mutations in Set_21. Primary donor: Tb10.v4.0088. Images have been drawn as in Figure 7.3. 05-29c12 and 05-29c30 were from independent reverse transcription-amplification reactions from the same mouse. They show a consistent pattern of point mutation, which is not shared with the clone from mouse 04 (04-31c47). Each of these clones has also undergone mild 3' donation. Different patterns of point mutation in the expressed copy of Tb10.v4.0088 were also found in mouse 02 and mouse LM_11 (data not shown).

7.3.2 Sets undergoing 3' donation

Set	Identified from	
	don	cln
Set_08	*	*
Set_09	*	*
Set_10	*	*
Set_11	*	*
Set_12	*	*
Set_13	*	*
Set_14	*	*
Set_20	*	*
Set_21	*	*
Set_22	*	*
Set_27	*	*
Set_31	*	*
Set_32	*	*
Set_37	*	*
Set_40	*	*
Set_43	*	*
Set_50	*	*
Set_54	*	*
Set_56	*	*
Set_59	*	*
Set_62	*	*
Set_64	*	*
Set_84	*	*

Set	Identified from	
	don	cln
Set_25	~	*
Set_63	~	*
Set_16	*	
Set_24	*	
Set_34	*	
Set_35	*	
Set_36	*	
Set_46	*	
Set_48	*	
Set_55	*	
Set_57	*	
Set_58	*	
Set_60	*	
Set_67	*	
Set_74	*	
Set_79	~	
Set_80	~	
Set_04		*
Set_17		*
Set_07		*
Set_18		*
Set_49		*
Set_68		*
Set_69		*

Table 7.7 Sets undergoing 3' donation. A '∗' in indicates whether 3' donation was identified from the donor ('don') and/or by comparing clones ('cln'). The left hand table shows only those 23 sets that fulfil the conservative definition of 3' donation (both from donor and between clones) whereas the right hand table shows the additional 24 sets that are included in the more inclusive definition of 3' donation (from donor *or* between clones). A tilde ('~') indicates that comparison with the donor was being performed with an incomplete donor sequence.

7.3.3 Analysed boundaries of 3' donation

Clone name	Set	1º donor	3' donor	Donor NA variability	Length of boundary	
					perfect id	>90% id
09-20c10	08	Tb09.v4.0077	a	0.447	24	48
05-23c03	13	Tb09.v4.0102	a	0.431	27	107
04-31c42	32	Tb09.v2.0090	a	0.473	11	37
04-23c35-del	56	Tb11.v4.0035	a	0.45	6	9
05-23c17-del	59	Tb927.3.270	a	0.459	13	37
10-30c06-del	59	Tb927.3.270	a	0.459	11	37
02-20c06	22	Tb05.5K5.330	b	0.366	9	16
05-27c06	31	Tb05.5K5.330	b	0.441	15	39
05-27c14	56	Tb05.5K5.330	b	0.331	27	52
05-27c05	64	Tb927.8.420	e	0.353	11	19
04-23c27	27	Tb09.142.0470	i	0.454	14	22
02-20c08	22	Tb05.5K5.330	k	0.413	16	38
04-21c03	22	Tb05.5K5.330	l	0.487	18	34
04-23c16	31	Tb927.6.5370	n	0.47	8	11
04-23c01	37	Tb09.v4.0123	o	0.427	8	11
05-21c05	13	Tb09.v4.0102	q	0.307	7	10

Table 7.8 Analysed boundaries of 3' donation. '1º donor' = donor that was repaired by 3' donation; '3' donor' = 3' donor variant (see Chapter 4, section 4.4.4); 'Donor NA variability' = full length nucleic acid variability, in differences/nt; 'Length of boundary' = length of region of perfect identity, or greater than 90% identity (i.e. fewer than 0.1 differences/nt). Lengths are given in nt.

7.3.4 Diagrams of all mosaics

In this section, images representing all expressed mosaics are provided. Two diagrams are provided for each unique[98] clone. The upper is drawn as described in Chapter 4: the position of cysteine codons are indicated by a thick bar projecting from the top of each diagram and the boundaries of the mature NTD are indicated by dashed and solid lines at the N- and C-terminal ends respectively. Segments best matching a donor are coloured white (donor A), dark grey (donor B), light grey (donor C) or stippled (donor D) accordingly. A dark line spanning the upper diagram indicates points in the alignment where the expressed clone matched none of the putative donors; for clarity this line extends slightly below the diagram for clones for which more than one donor could be identified. The lower image provides a track for each donor. A black line indicates the location of a mismatch between the donor and the clone. Extensive regions coloured black indicate missing donor sequence. As imperfect automated scripts produced these images (see Chapter 2 sections 2.7.2.4 and 2.7.2.5), care must be taken in interpretation, as point mutations are occasionally misdrawn as segmental gene conversion events, particularly if many donors are being considered.

Data for each mouse-timepoint was produced from different reverse transcription-amplification reactions. Additionally, clones from mouse 05 with identifiers c13–c50 were produced from different reverse-transcription reactions from those with identifiers c01–c12.

Note that clones sequenced by this work have the format ‘XX-YYcZZ’ where ‘XX’ is the infection number (01–12), ‘YY’ is the day of infections and ‘ZZ’ is a two digit number unique to that mouse-timepoint. Sequences from Marcello & Barry, (2007a) have the format ‘VSG-AA-BB-CC’, where ‘AA is the day of infection, ‘BB’ is the infection number (01–11, refers to LM_01–LM_11) and ‘CC’ is a two digit number unique to that mouse-timepoint.

[98] Clones with duplicate patterns of segmental gene conversion, or differing by only a couple of nucleotides, have not been included, and are referred to in the captions.

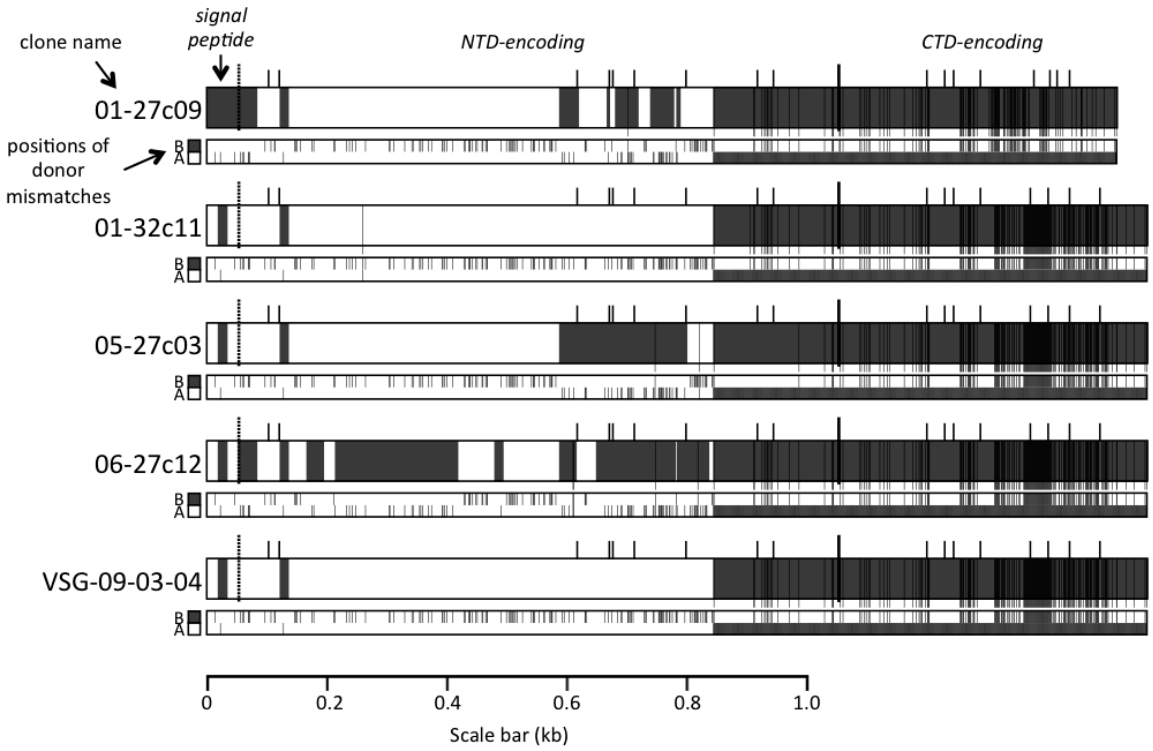


Figure 7.5 Set_04 mosaics. Identified mosaicism from donors and amongst clones. Donor A = tryp_X-54b12.q1c-rev_assembly, donor B = Tb10.v4.0061. Donor A is an incomplete partial sequence.

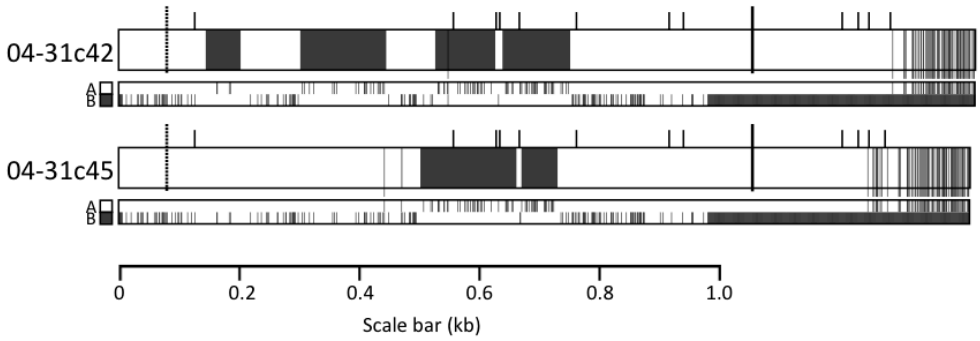


Figure 7.6 Set_32 mosaics. Identified mosaicism from donors and amongst clones. Donor A = Tb09.v2.0090, donor B = tryp_IXa-10f02_assembly. Donor B was is an incomplete sequence.

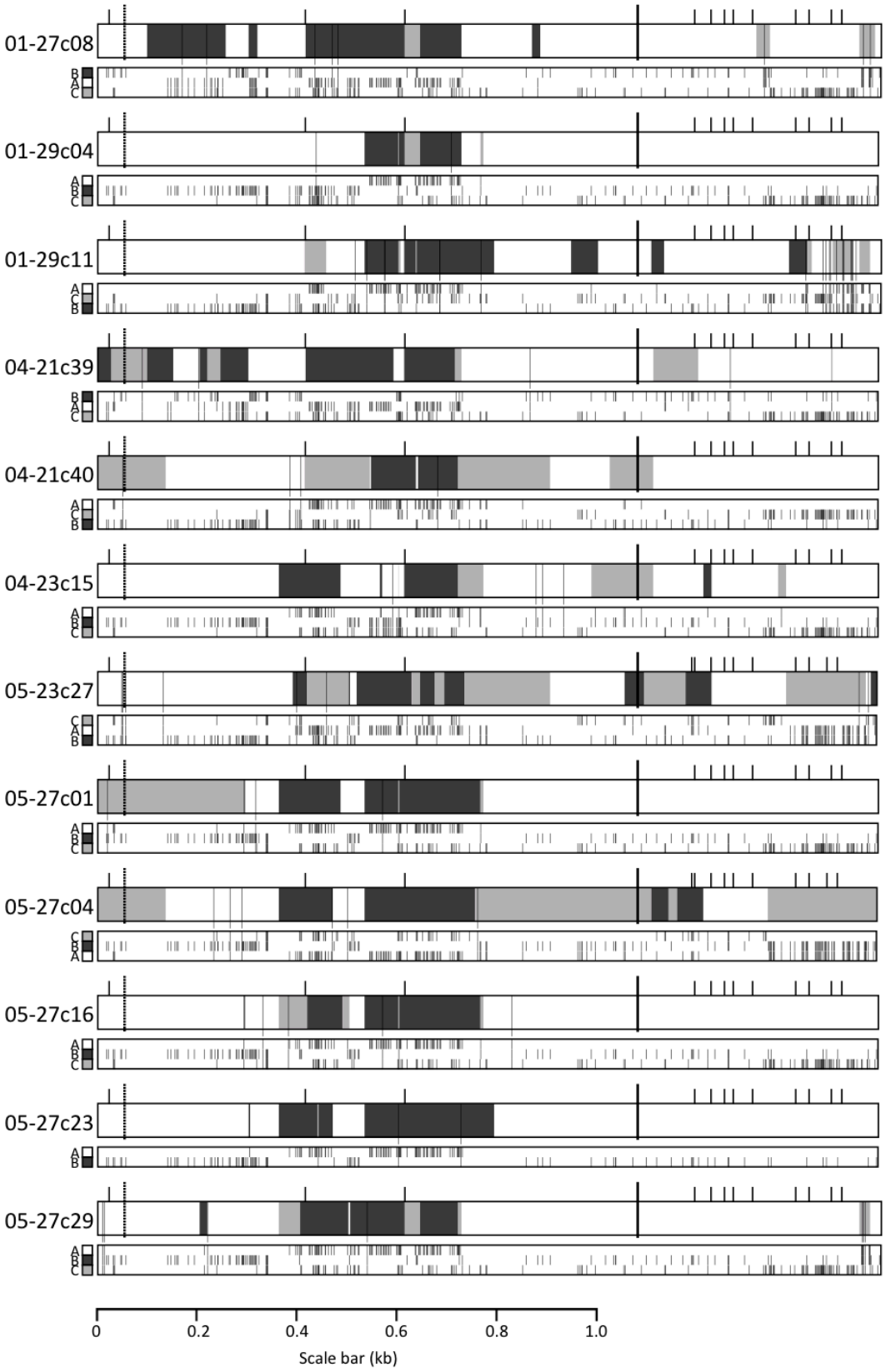


Figure 7.7 Set_10 mosaics. Identified mosaicism from donors and amongst clones. Donor A = Tb11.31.0001, donor B = Tb10.v4.0161, donor C = Tb11.14.0001. Clones not drawn: 05-29c15 matches 05-27c01; 05-29c13, 05-29c21, 05-29c24 match 05-29c05; 01-27c10, 01-27c12, 01-29c02 match 01-27c08; 01-29c10 matches 01-29c04; 04-23c19, 04-23c21, 04-23c36, 04-23c43 match 04-21c40.

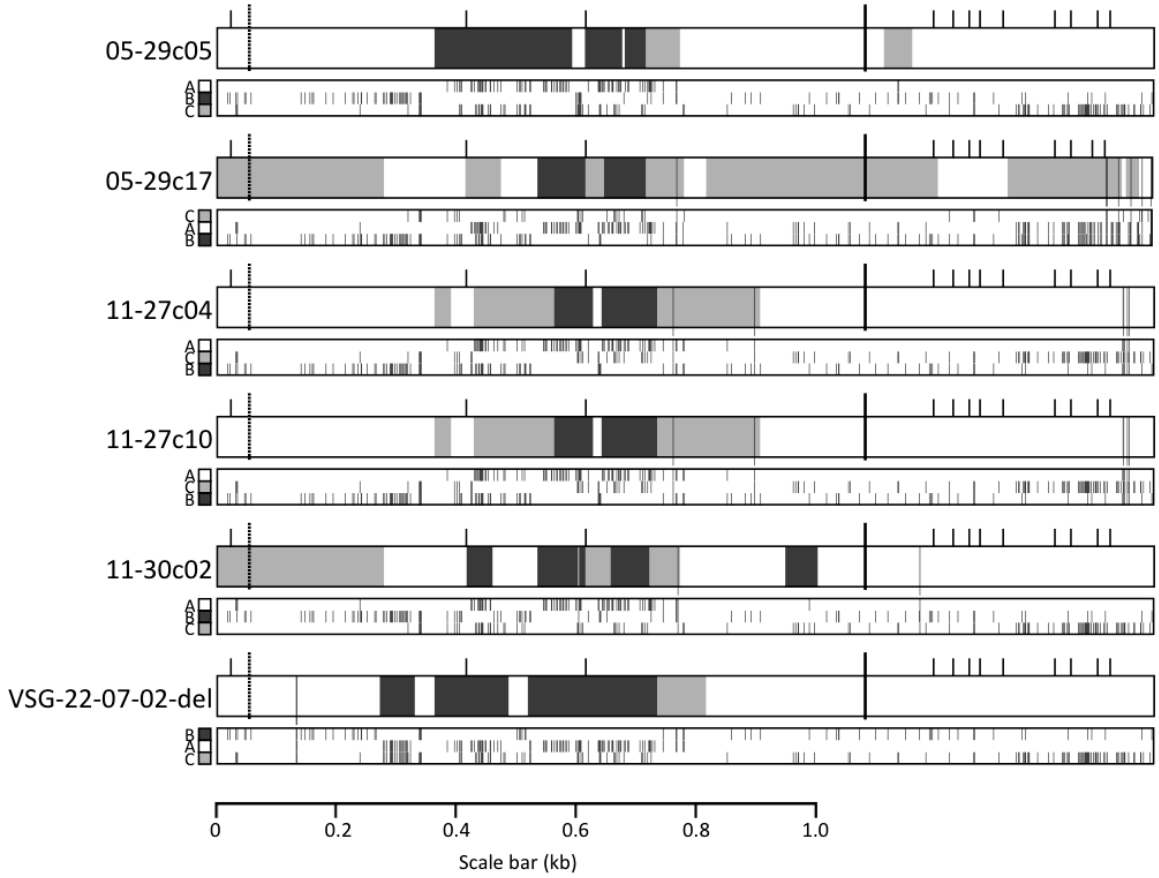


Figure 7.7 (continued)

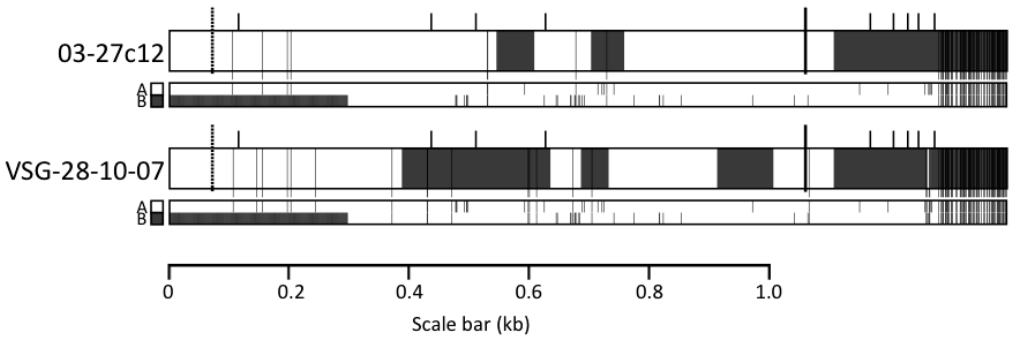


Figure 7.8 Set_16 mosaics. Identified mosaicism from donors and amongst clones. Donor A = Tb11.48.0003, donor B = tryp_XI-934d12_assembly. Donor B is an incomplete sequence.

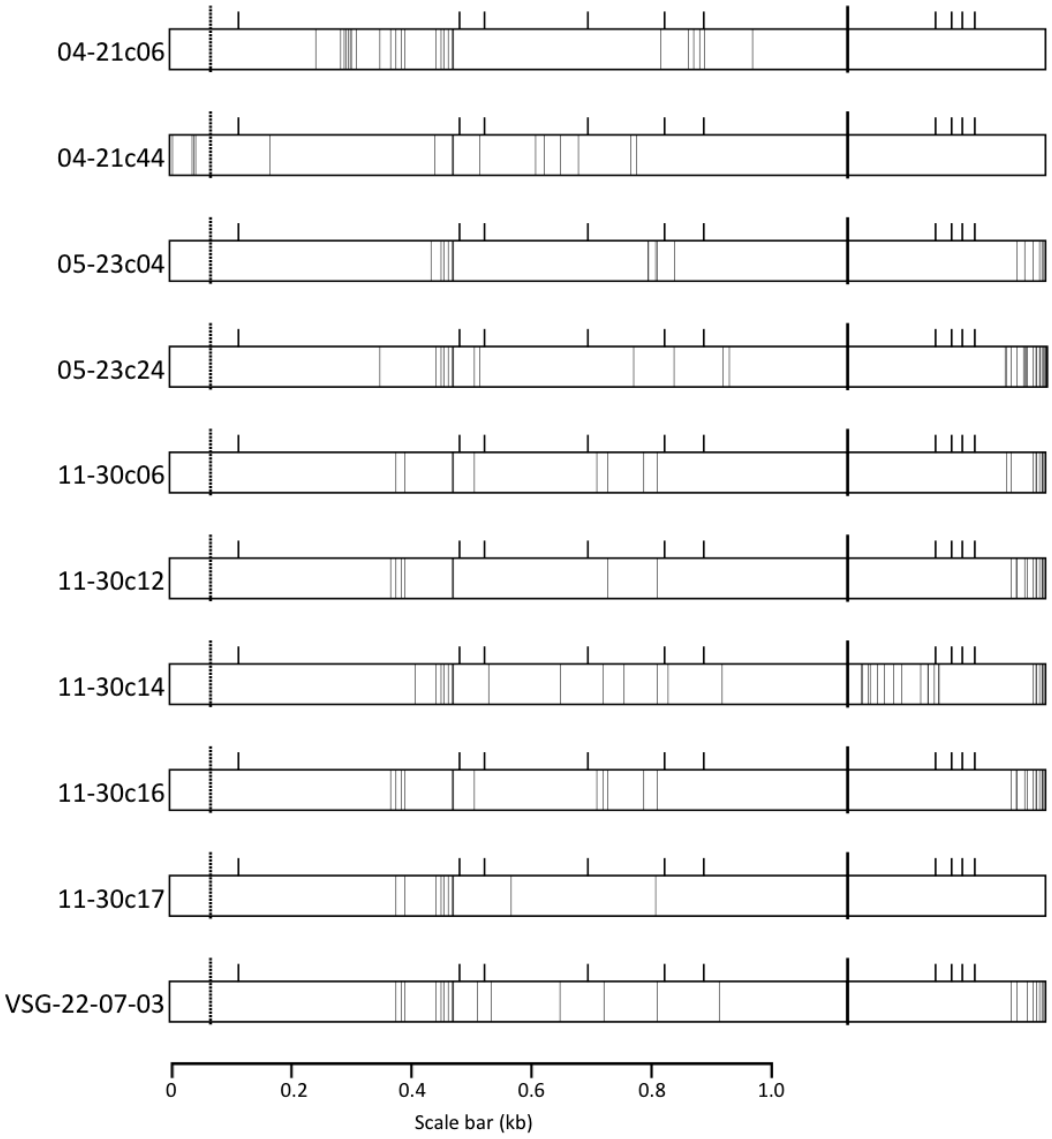


Figure 7.9 Set_11 mosaics. Identified mosaicism amongst clones only. Identified donor: Tb927.3.190. Clones not drawn: 05-23c10, 05-23c15, 05-23c18 and 05-23c26 match 05-23c04; 11-30c25, 11-30c30, 11-30c34 match 11-30c12; 11-30c35 matches 11-30c16; 11-30c28 matches 11-30c17.

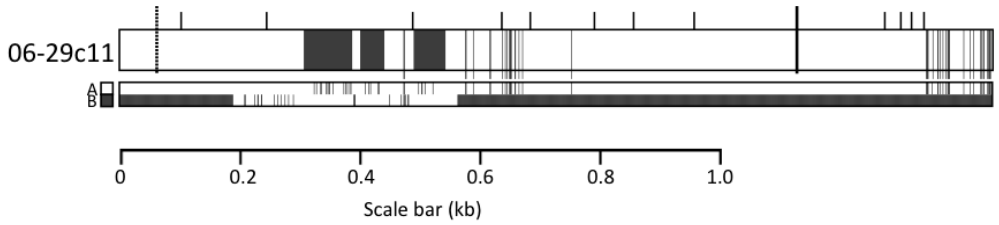


Figure 7.10 Set_24 mosaics. Identified mosaicism from donors only. Donor A = Tb08.27P2.460, donor B = trypA24a5.p1p. Donor B is an incomplete sequence.

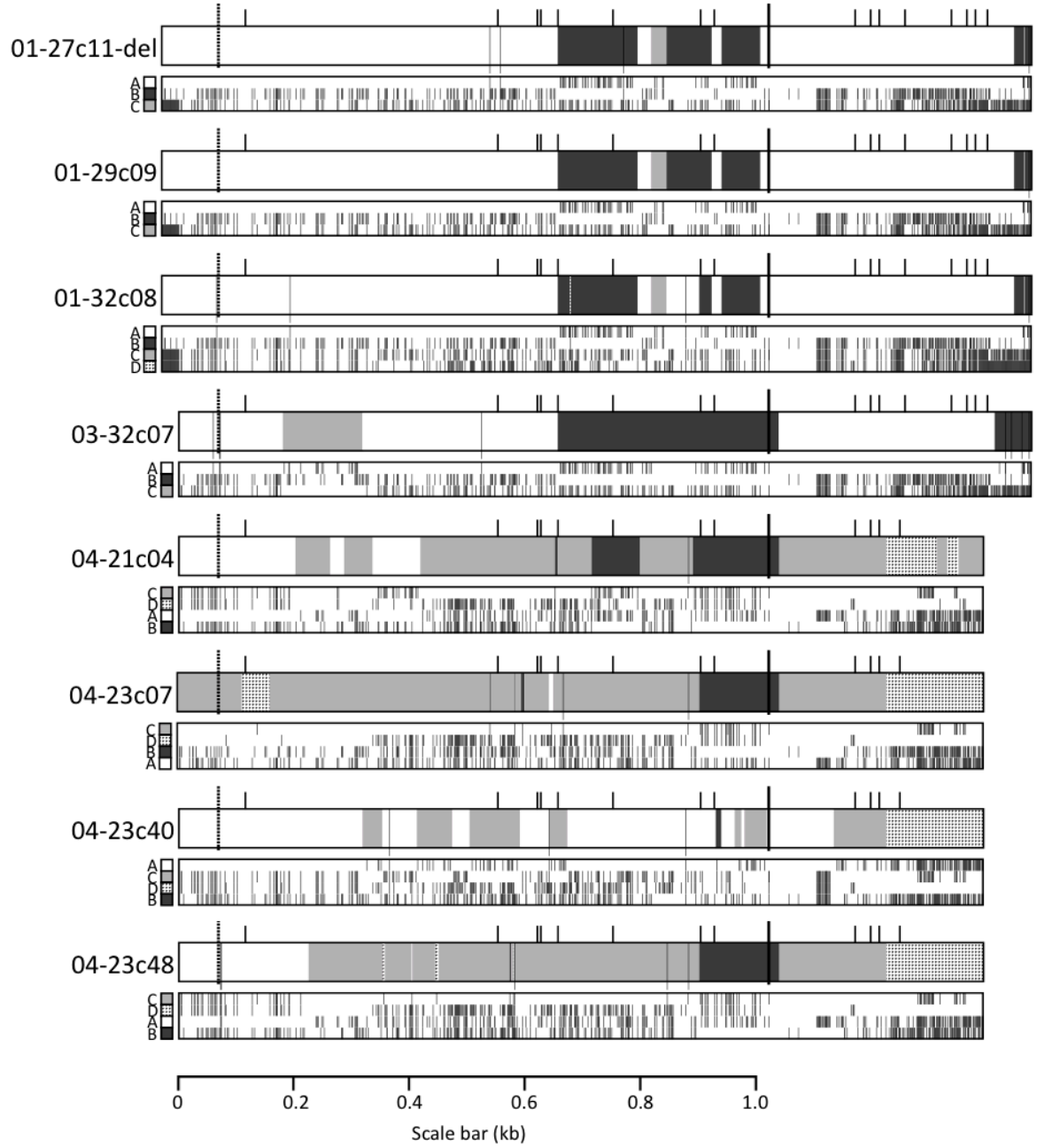


Figure 7.11 Set_14 mosaics. Identified mosaicism from donors and amongst clones. Donor A = Tb11.09.0005, donor B = Tb11.13.0003, donor C = tryp_XI-1034g11_assembly, donor D = Tb10.v4.0009. Clones not drawn: 01-29c05 matches 01-27c11-del; 04-27c07-del matches 04-23c07; 04-23c47 matches 04-23c40; 04-27c44 matches 04-27c25; 04-31c01, 04-31c06, 04-31c31, 04-31c33, 04-31c34 and 04-31c35 match from 04-29c06; 05-27c19, 05-27c20, 05-27c21, 05-27c32, 05-27c33, 05-29c23 match 05-23c12.

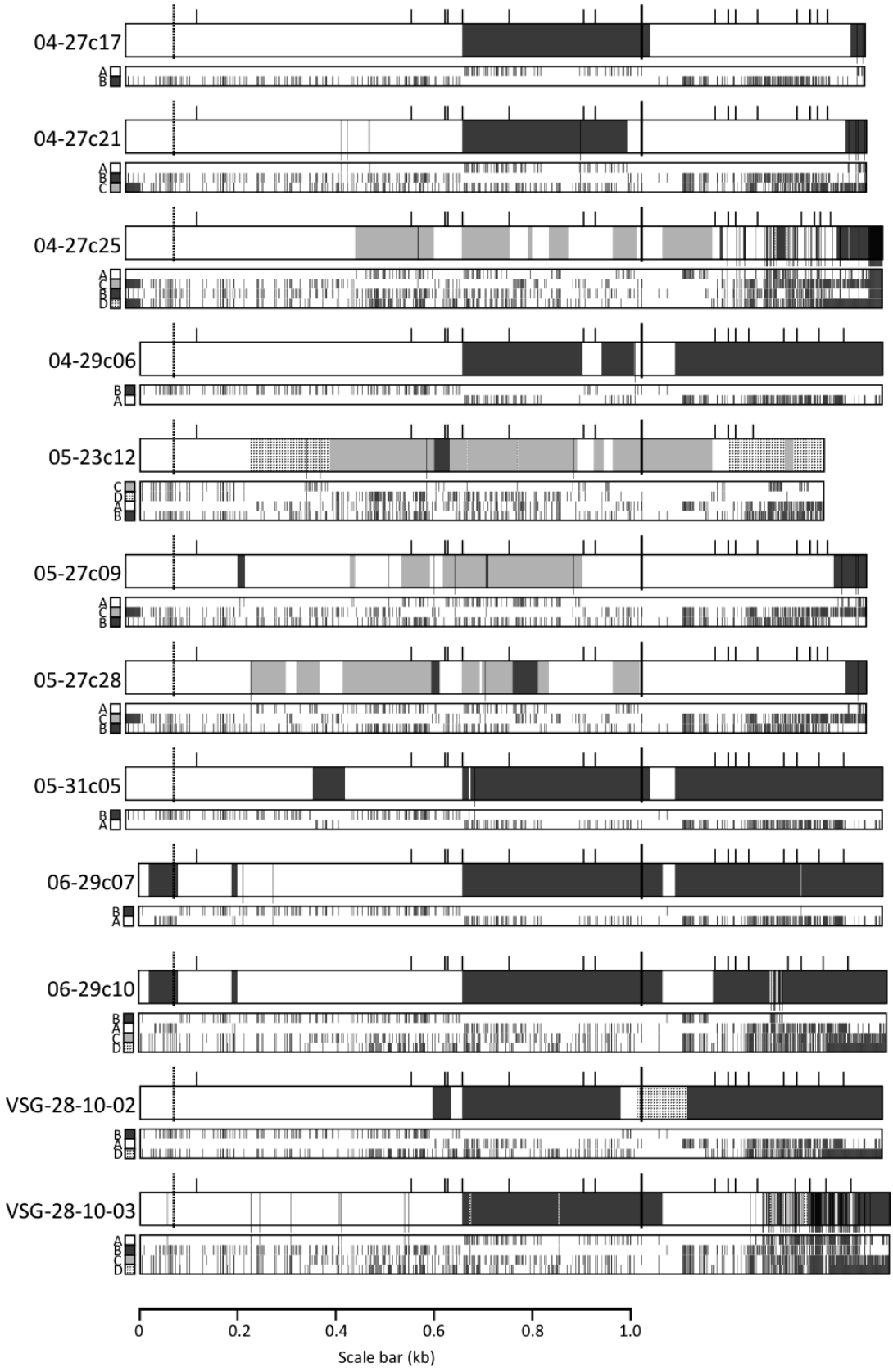


Figure 7.11 (continued)

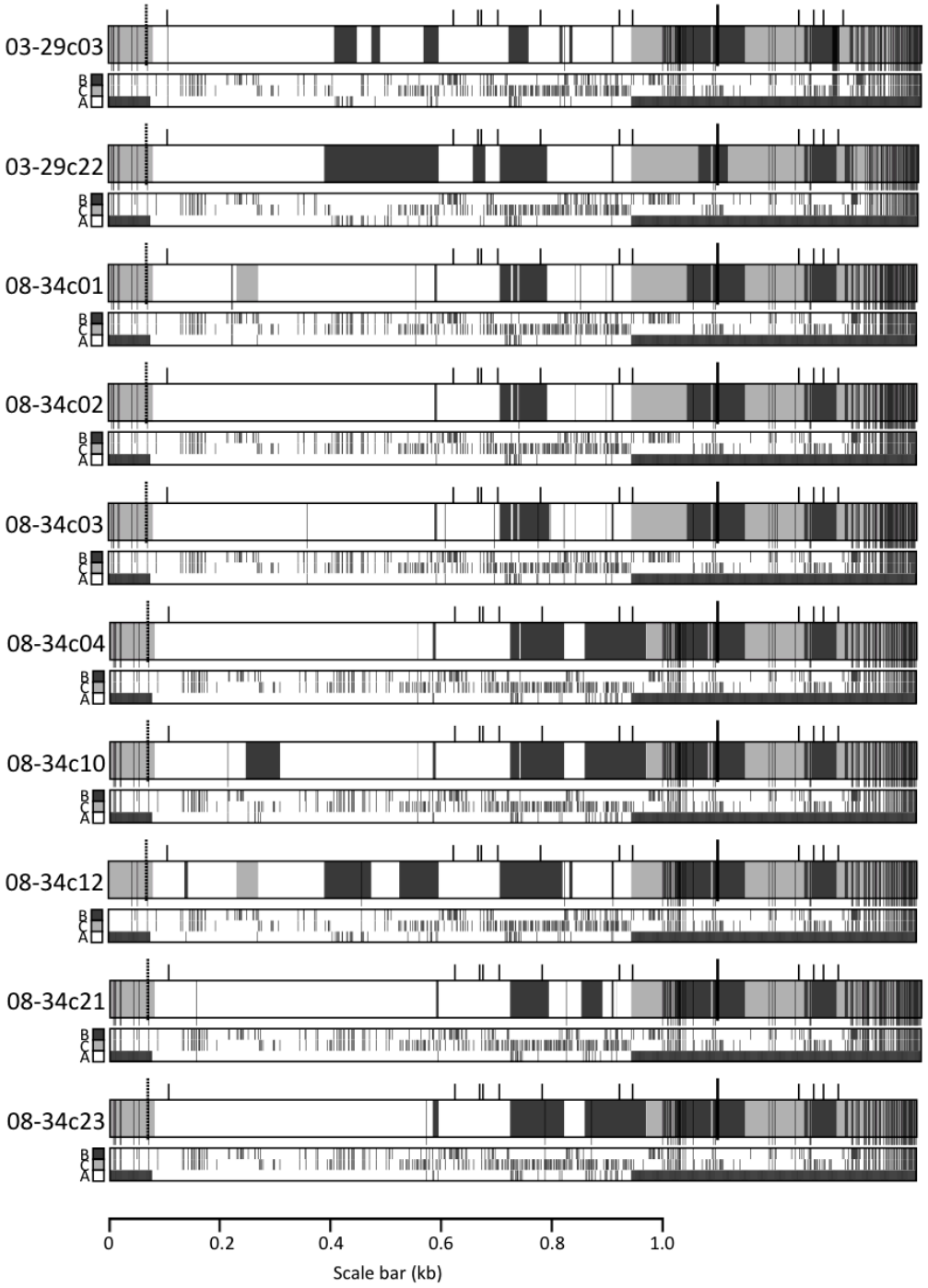


Figure 7.12 Set_17 mosaics. Identified mosaicism from donors and amongst clones. Donor A = tryp_XI-1058d01_assembly, donor B = Tb11.57.0030, donor C = Tb09.244.0090. Clones not drawn: 03-29c19 and 03-29c11 match 03-29c03; 08-34c05 matches 03-34c01; 08-34c09, 04-34c11, 08-34c15 and 08-34c20 match 08-34c02; 08-34c07, 08-34c08, 08-34c14 and 08-34c22 match 08-34c04; 09-34c03, 09-34c08, 09-34c10 and 09-34c14 match 09-34c01; 09-34c09 and 09-34c12 match 09-34c07; 10-30c07 and 12-30c09 match 10-30c05 (thus it is possible that this particular mosaic variant was present in feeder mouse F3). Donor A is an incomplete sequence.

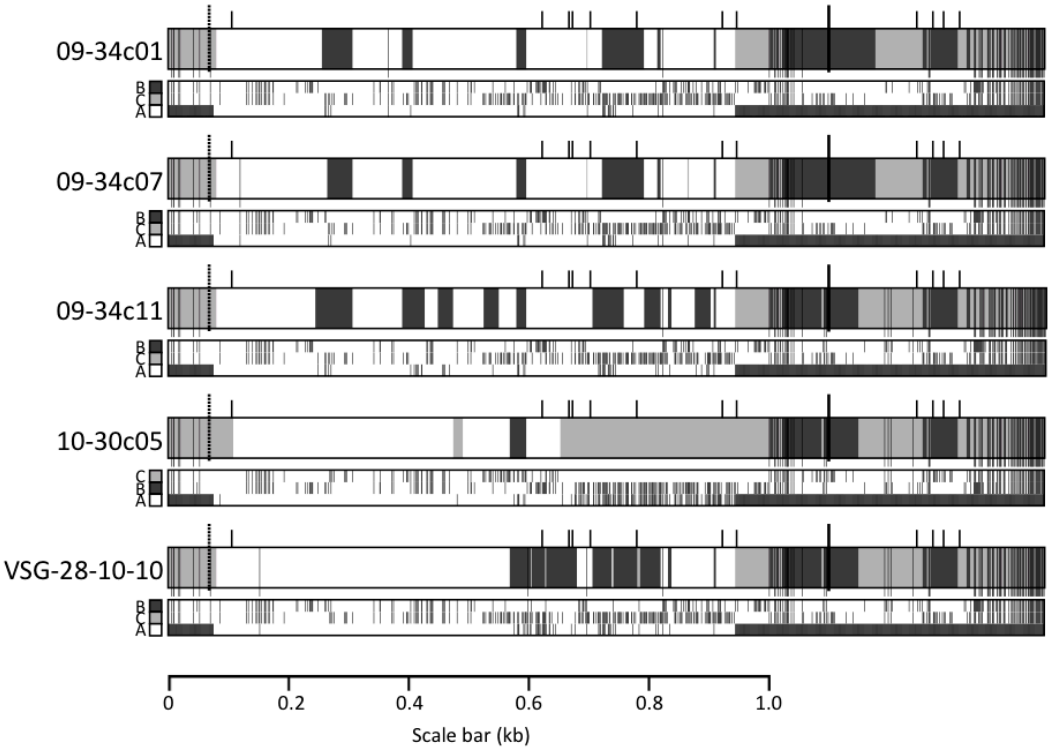


Figure 7.12 (continued)

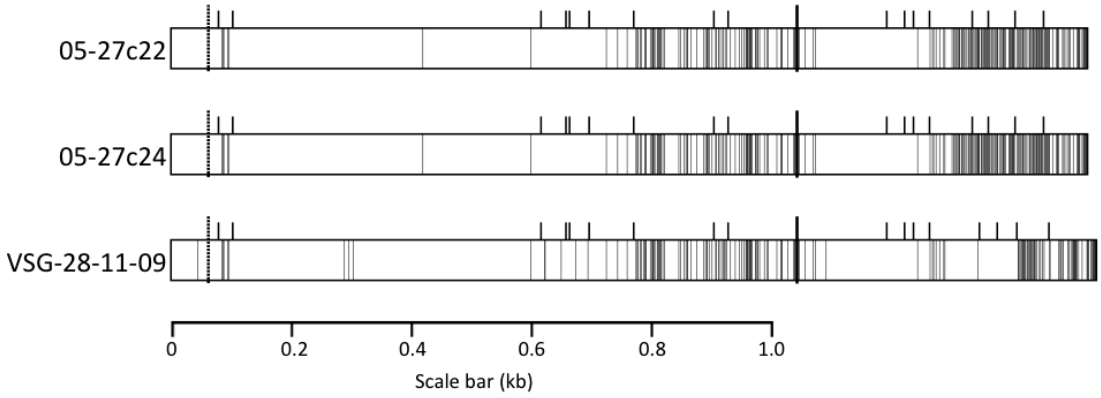


Figure 7.13 Set_20 mosaics. Mosaicism identified from donor only. Identified donor: Tb927.7.6530. It was not possible to find the additional donor read referred to by L. Marcello (Ph.D. Thesis, 2006, University of Glasgow). Most patterns of mismatch with the identified donor are conserved across infections, thus it is likely that mosaicism was not formed over the course of these infections and was pre-existing in feeder mouse F.

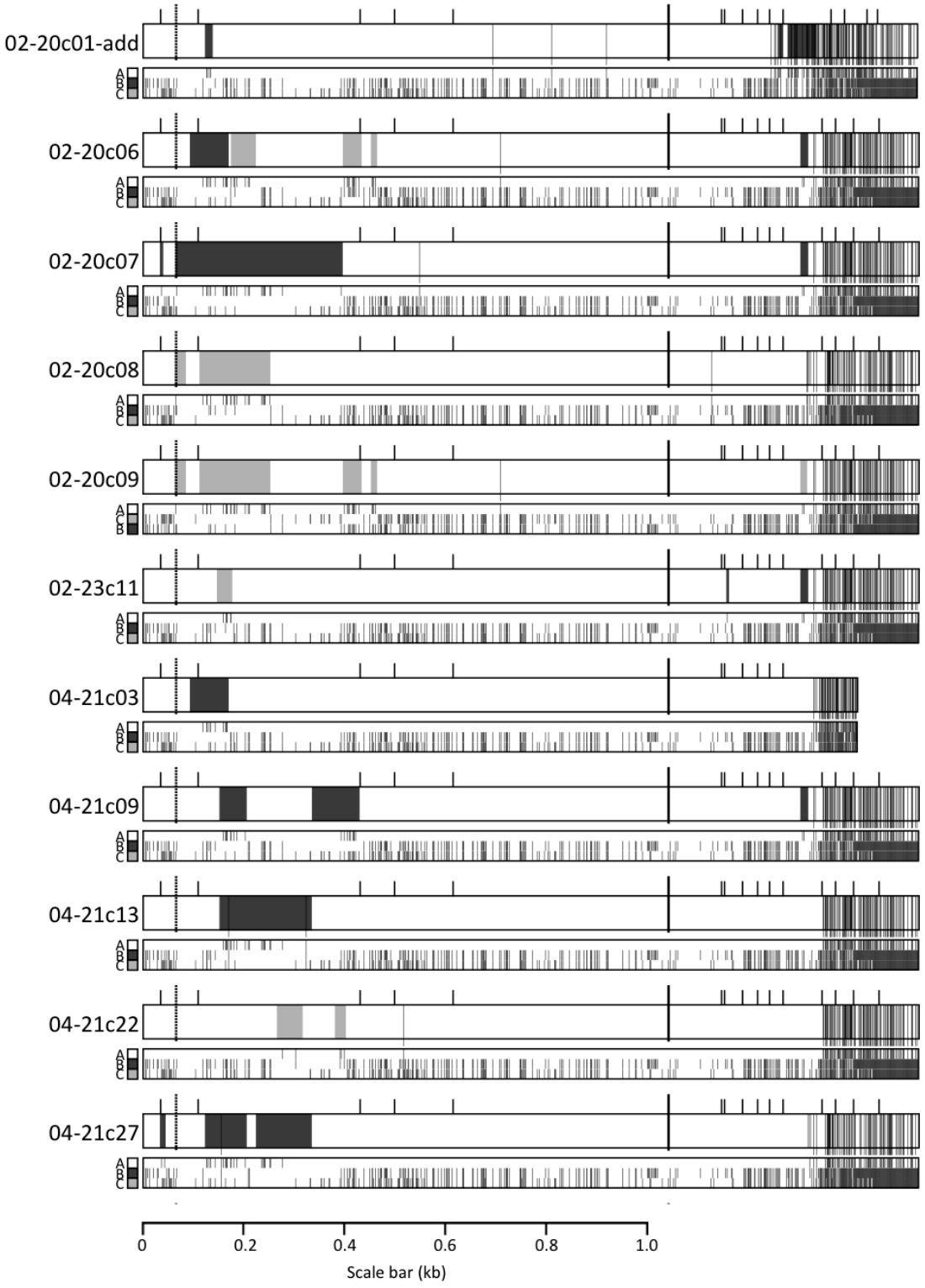


Figure 7.14 Set_22 mosaics. Identified mosaicism from donors and amongst clones. Donor A = Tb05.5K5.330, donor B = Tb11.21.0004, donor C = Tb9275.4840. Clones not drawn: 02-23c01, 02-23c02, 02-23c04 and 02-23c12 match 02-20c06; 02-20c11 and 02-23c10 match 02-20c08; 04-21c05, 04-21c11, 04-21c24, 04-21c30 and 04-21c38 match 04-21c03; 04-21c10, 04-21c14, 04-21c17, 04-21c19 and 04-21c36 match 04-21c09; 04-21c23 and 04-21c28 match 04-21c22; 04-21c32 and 04-21c41 match 04-21c27.

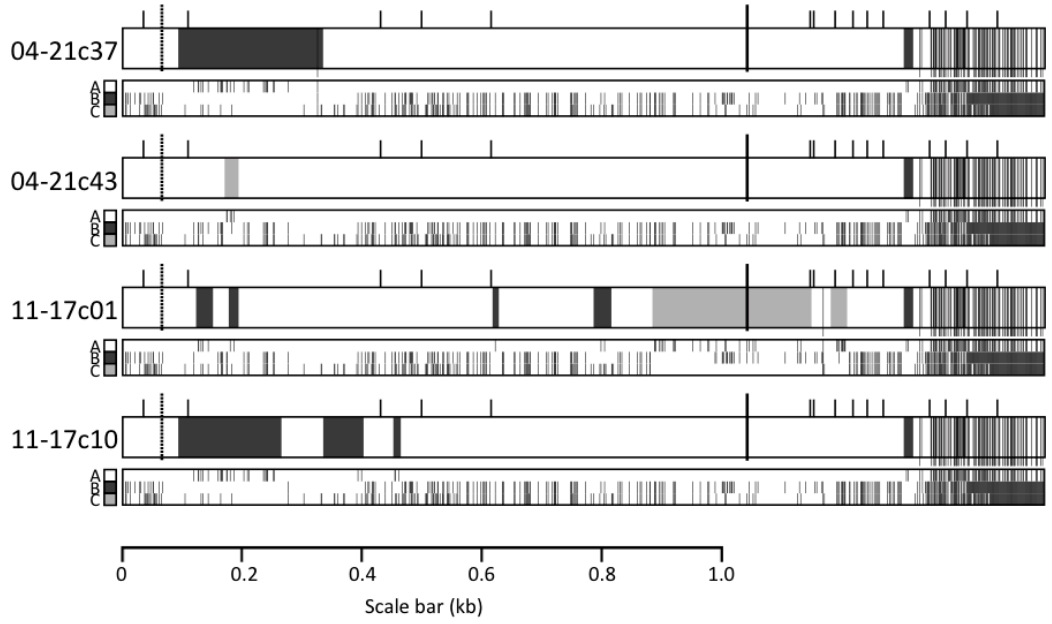


Figure 7.14 (continued)

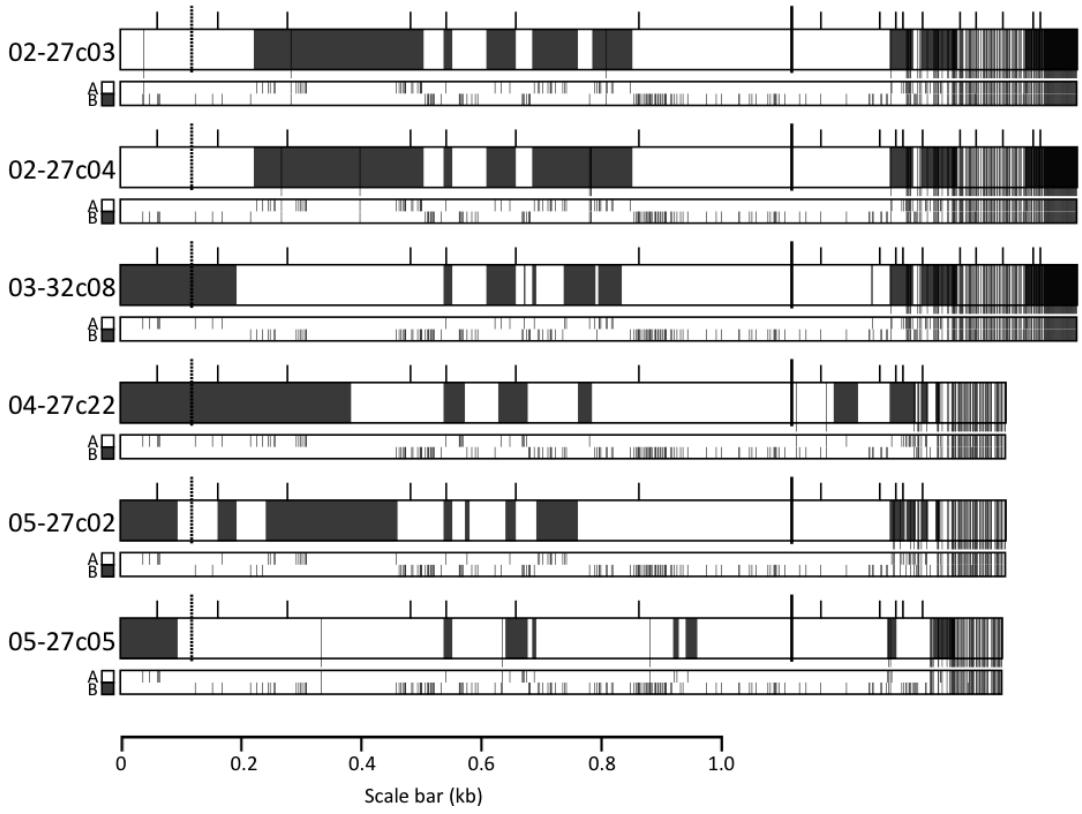


Figure 7.15 Set_64 mosaics. Mosaicism identified from donors and amongst clones. Donor A = Tb927.8.420, donor B = Tb09.244.0360. Clones not drawn: 02-27c05 matches 02-27c03; 04-29c07 matches 04-27c22; 05-27c26 matches 05-27c05; 05-29c26 matches 05-27c02.

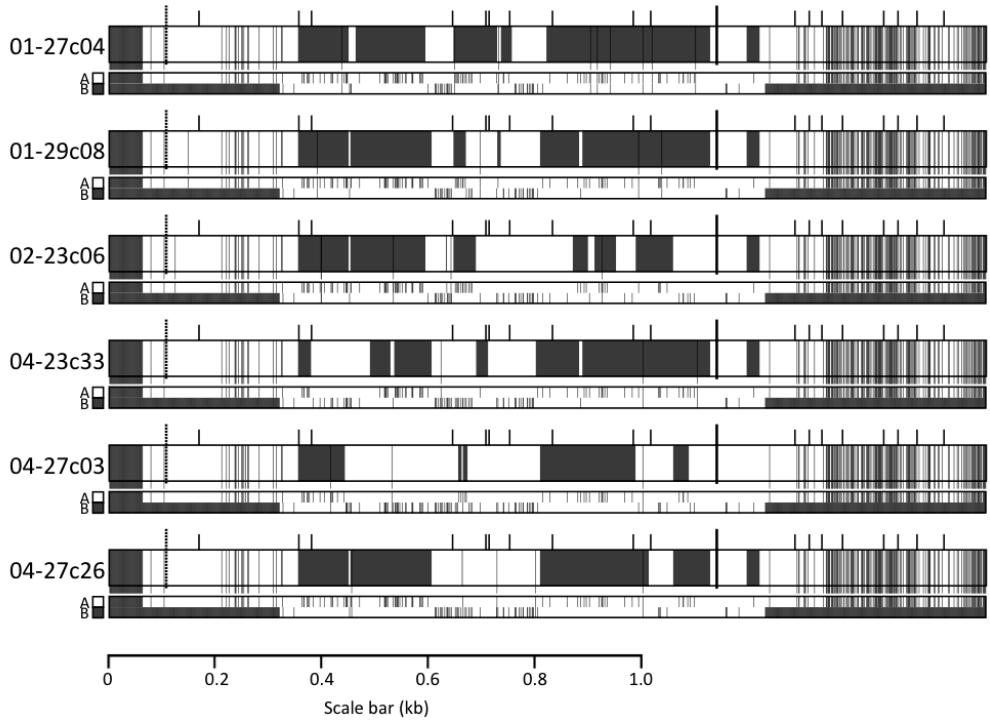


Figure 7.16 Set_29 mosaics. Mosaicism identified from donors and amongst clones. Donor A = Tb09.244.0150, donor B = 927mc_IV-5e11_assembly. Clones not drawn: 02-27c07 matches 02-23c06. Donor B is an incomplete sequence.

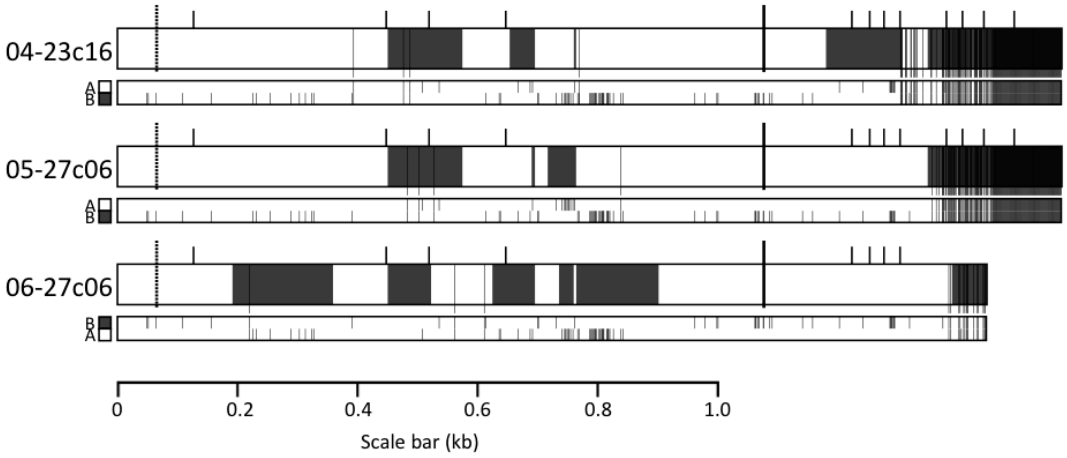


Figure 7.17 Set_31 mosaics. Mosaicism identified from donors and amongst clones. Donor A = Tb09.354.0060, donor B = Tb927.6.5370. Clones not drawn: 04-23c50 and 04-27c06 match 04-23c16; 05-29c32, 05-31c14, 05-31c18 and 05-31c21 match 05-27c06.

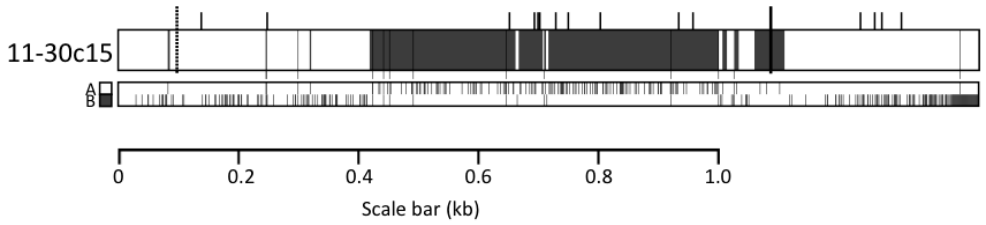


Figure 7.18 Set_33 mosaics. Mosaicism identified from donors only. Donor A = Tb10.v4.0031, donor B = Tb09.v4.0071.

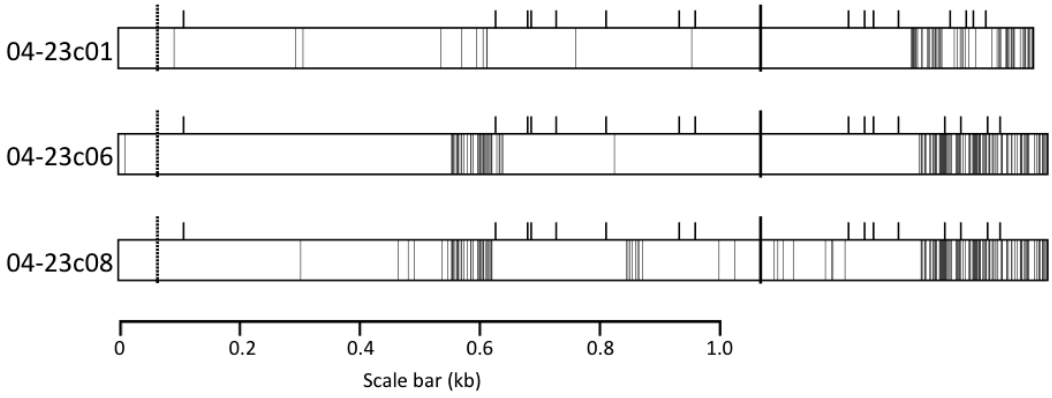


Figure 7.19 Set_37 mosaics. Mosaicism identified amongst clones only. Identified donor: Tb09.v4.0123.

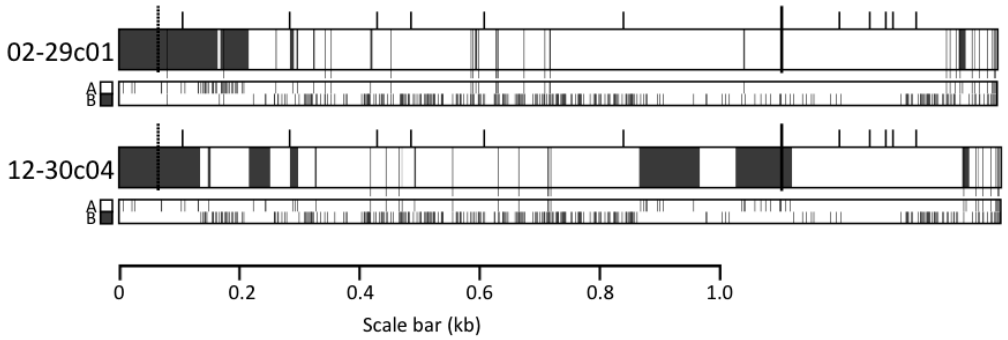


Figure 7.20 Set 43 mosaics. Mosaicism identified from donors and amongst clones. Donor A = Tb10.v4.0102, donor B = Tb10.v4.0081.

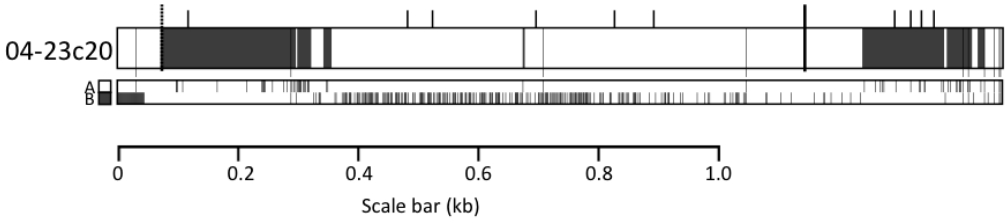


Figure 7.21 Set_46 mosaic. Mosaicism identified from donors only. Donor A = Tb11.16.0003, donor B = Tb10.v4.0095.

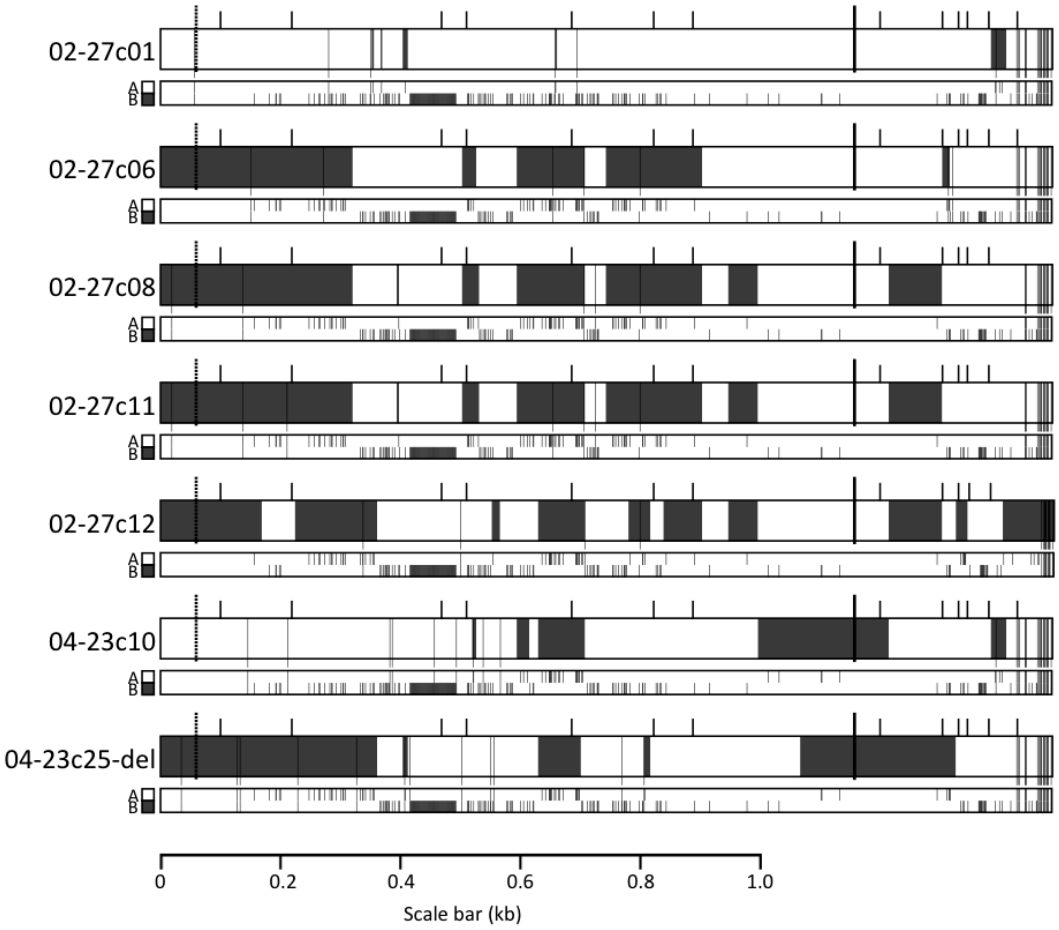


Figure 7.22 Set_40 mosaics. Mosaicism identified from donors and amongst clones. Donor A = Tb11.1480, donor B = Tb10.v4.0063.

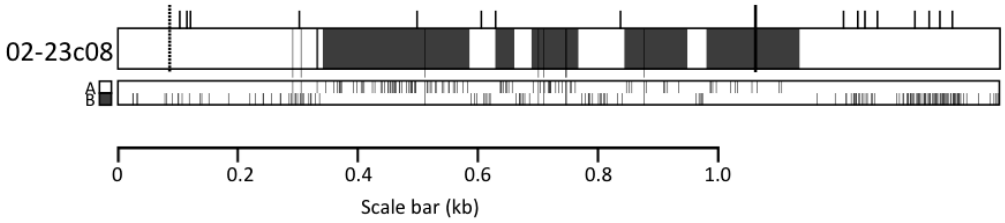


Figure 7.23 Set_42 mosaics. Mosaicism identified from donors only. Donor A = Tb927.8.170, donor B = Tb10.v4.0070.

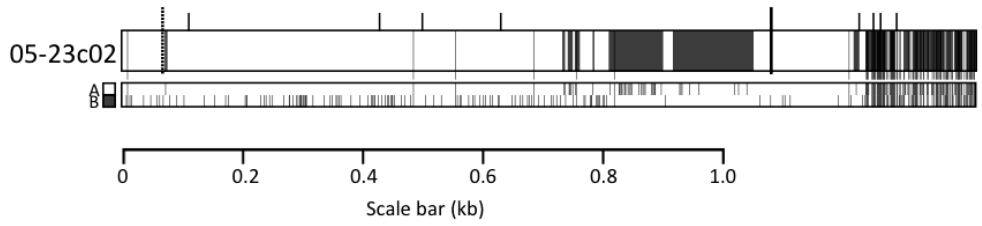


Figure 7.24 Set_48 mosaic. Mosaicism identified from donors only. Donor A = Tb11.35.0001, donor B = Tb09.244.1360.

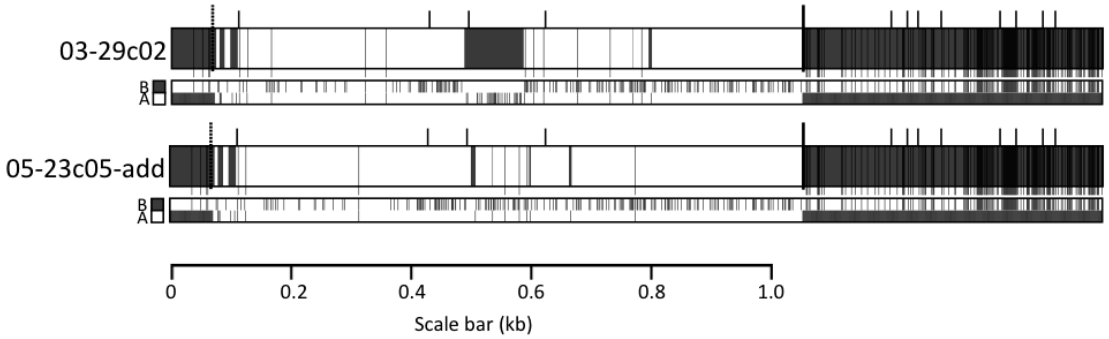


Figure 7.25 Set_51 mosaics. Mosaicism identified from donors and amongst clones. Donor A = tryp_IXa-20a05_assembly, donor B = Tb11.57.0023. Clones not drawn: 05-23c22, 05-23c23 and 05-23c25 match 05-23c05-add; 03-29c12 matches 03-23c02. Donor A is an incomplete sequence.

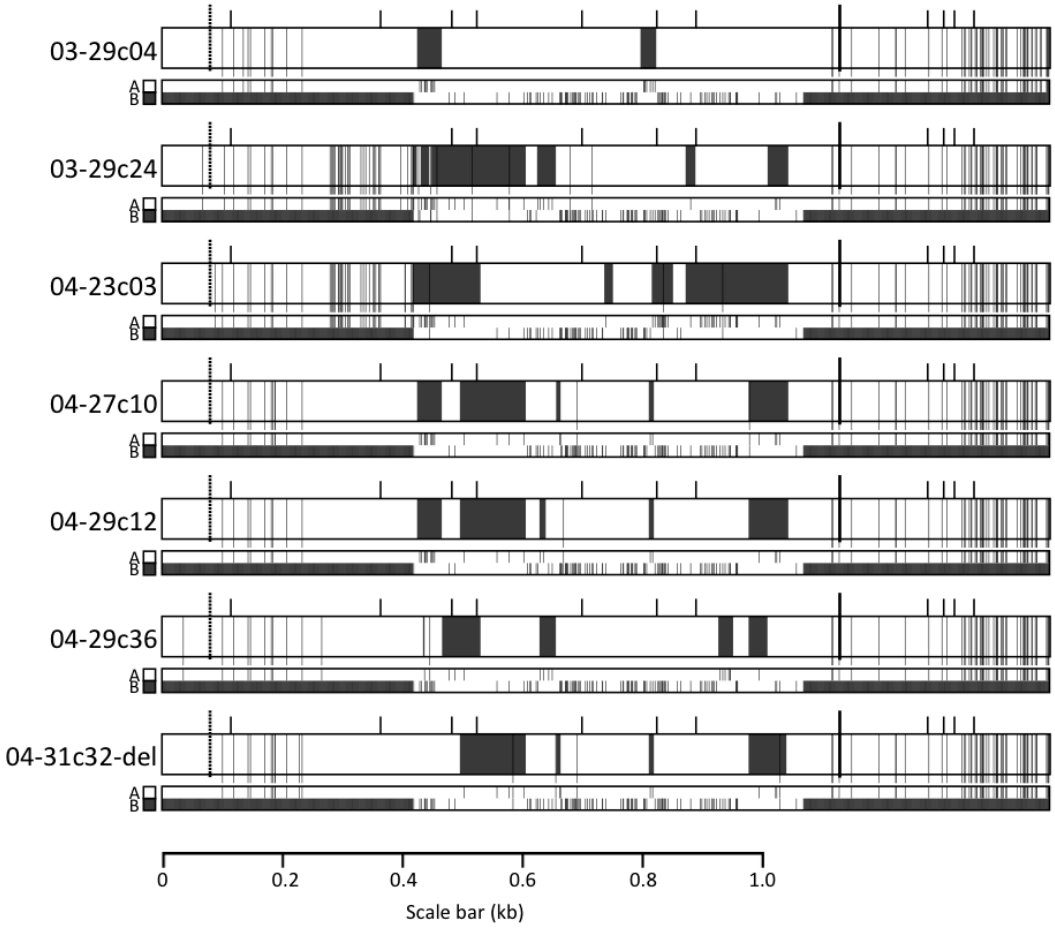


Figure 7.26 Set_49 mosaics. Mosaicism identified from donors and amongst clones. Donor A = Tb11.43.0002, donor B = tryp_XI-909g03.q1k-rev. Clones not drawn: 04-29c01, 04-29c03 and 04-29c05 match 04-27c10; 04-31c41 matches 04-23c03. Donor B is an incomplete sequence.

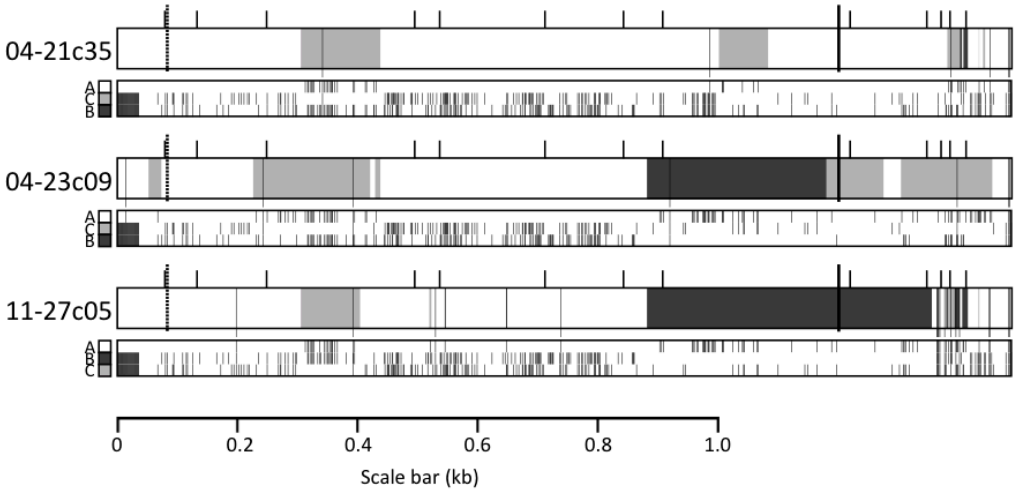


Figure 7.27 Set_62 mosaics. Mosaicism identified from donors and amongst clones. Donor A = Tb927.5.4670, donor B = Tb09.v4.0061, donor C = Tb05.5K5.320.

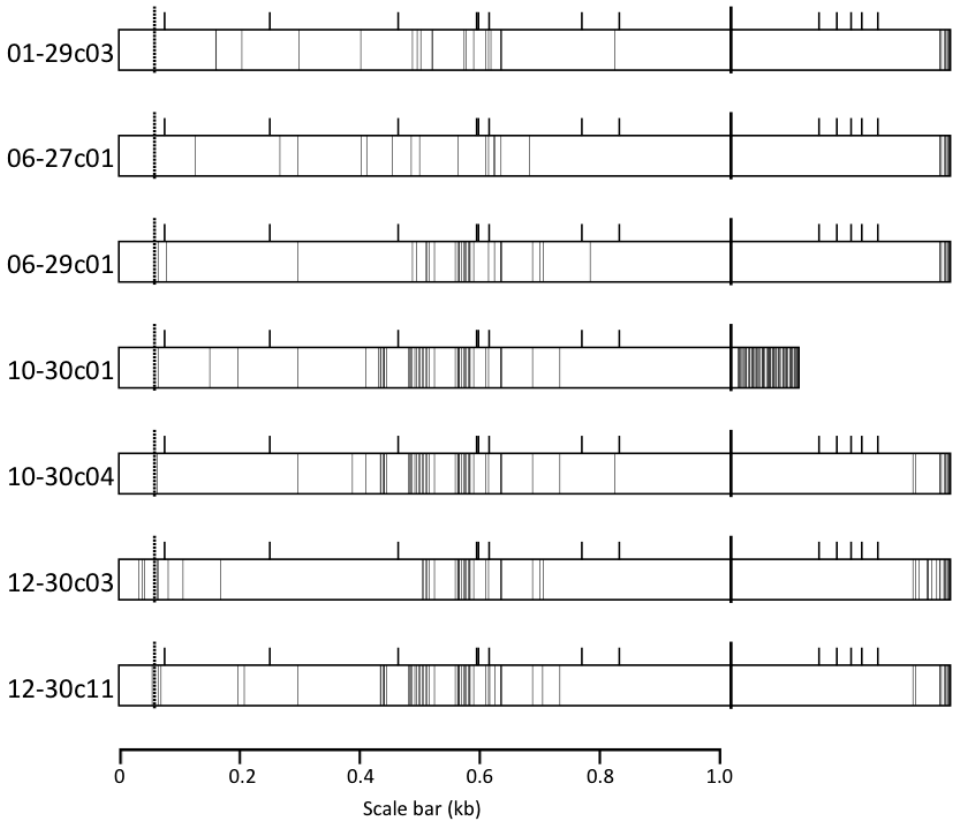


Figure 7.28 Set_84 mosaics. Mosaicism identified amongst clones only. Identified donor = 927mc_VIII-14d09_assembly. Clones not drawn: 06-27c02, 06-27c03, 06-27c04, 06-27c05, 06-27c07, 06-27c08, 06-27c09, 06-27c10, 06-27c11, 06-29c02, 06-29c03, 06-29c04, 06-29c06, 06-29c09 and 06-29c12 match 06-27c01; 10-30c08, 10-30c09, 10-30c12, 12-30c02 and 12-30c10 match 10-30c04. Clone 10-30c01 has lost a fragment of the CTD-encoding region, possibly through the process of plasmid cloning and amplification.

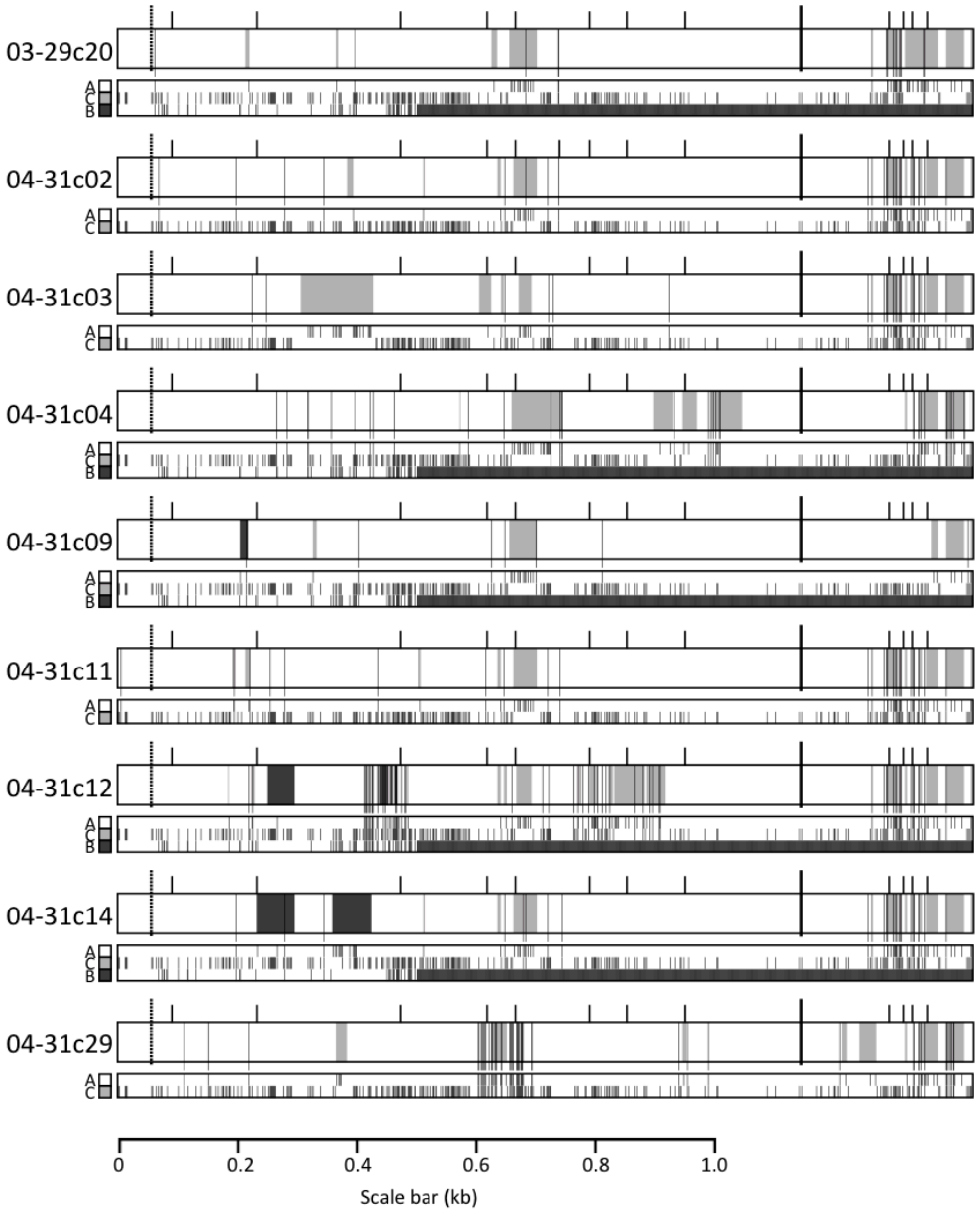


Figure 7.29 Set_54 mosaics. Mosaicism identified amongst clones only (too many patches of mismatch for confidence in the additional donors). Donor A = Tb11.v4.0029, donor B = tryp_IXa-29e12_assembly, donor C = tryp_X-99f10_assembly. Donor B is an incomplete sequence.

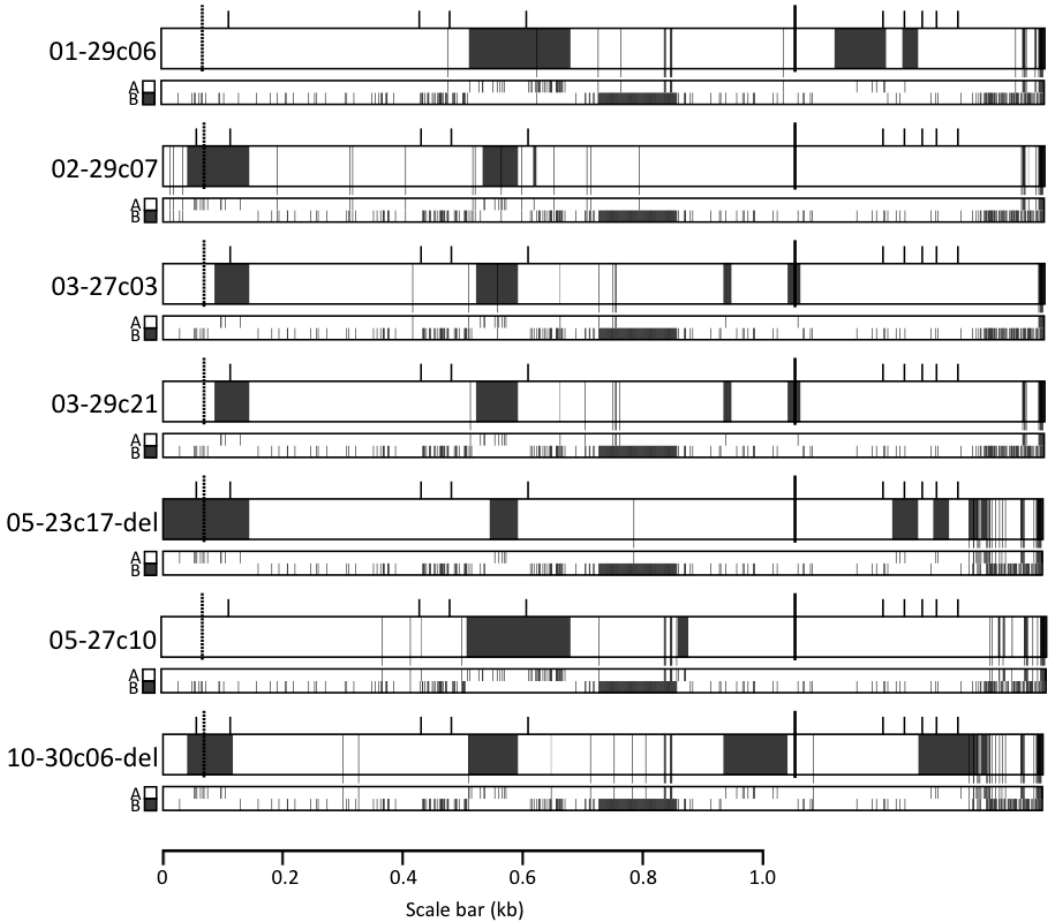


Figure 7.30 Set_59 mosaics. Mosaicism identified from donors and amongst clones. Donor A = Tb927.3.270, donor B = tryp_XI-1157a04_assembly. Clones not drawn: 05-23c21 matches 05-23c17. Donor B was made by from two non-contiguous sequences and thus is incomplete.

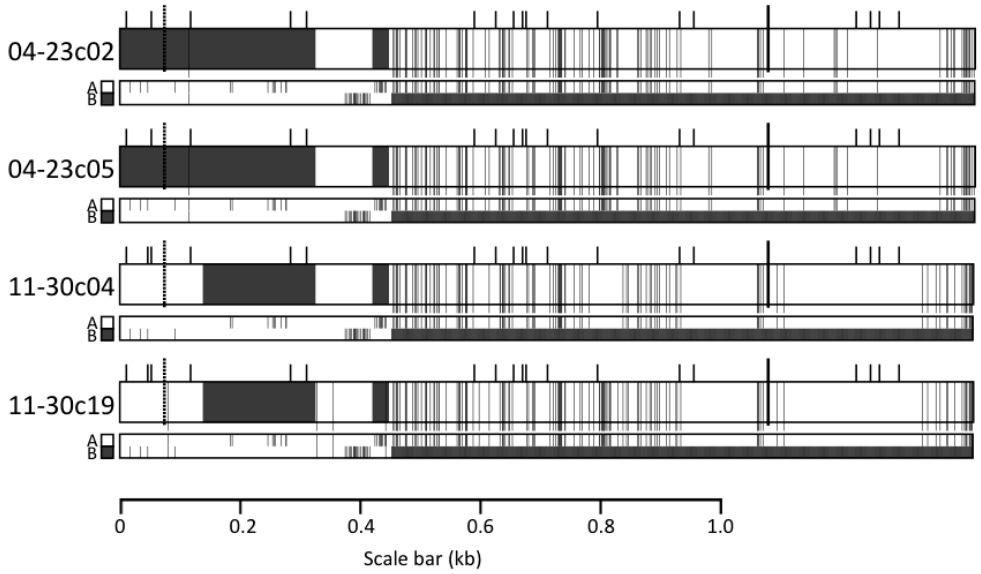


Figure 7.31 Set_63 mosaics. Mosaicism identified amongst clones only (insufficient donor B sequence). Donor A = Tb927.5.4930, donor B = tryp_XI-1007d10. Clones not drawn: 04-23c05 matches 04-23c02.

7.4 Note added in proof

7.4.1 Introduction

Experiments were conducted following submission of this thesis that, whilst not changing the overall argument and conclusions, do resolve some outstanding issues and affect some of the details and figures presented within. The key findings are summarized below.

7.4.2 Additional PCR reactions were consistent with segmental gene conversion occurring during infection

Further to Chapter 4, section 4.5.2, PCR reactions were set up to test whether segmental gene conversion events were genuine and occurred over the course of infection. The design of the primers is described in Figure 7.32 and Table 7.9, and the results of the reactions are shown in Figure 7.33Figure 7.37. Balancing sensitivity and specificity meant that in some cases it was difficult to obtain a single specific product for a primer pair (see for example Set_04 A-B reaction, Figure 7.37). In addition, attempts to identify segmental gene conversion junctions failed in cases where the mosaic *VSG* had been expressed earlier on in infection, since the segmentally-converted *VSG* was absent, or present only at very low abundance, in the terminal samples (see for example Set_10 A-B reactions on gDNA templates from mouse 01, mouse 04 and mouse 05, Figure 7.36). However, the overall pattern is that the segmental gene conversion junctions tested appeared over the course of infection, and represent neither experimental artefact nor pre-existing genomic variation.

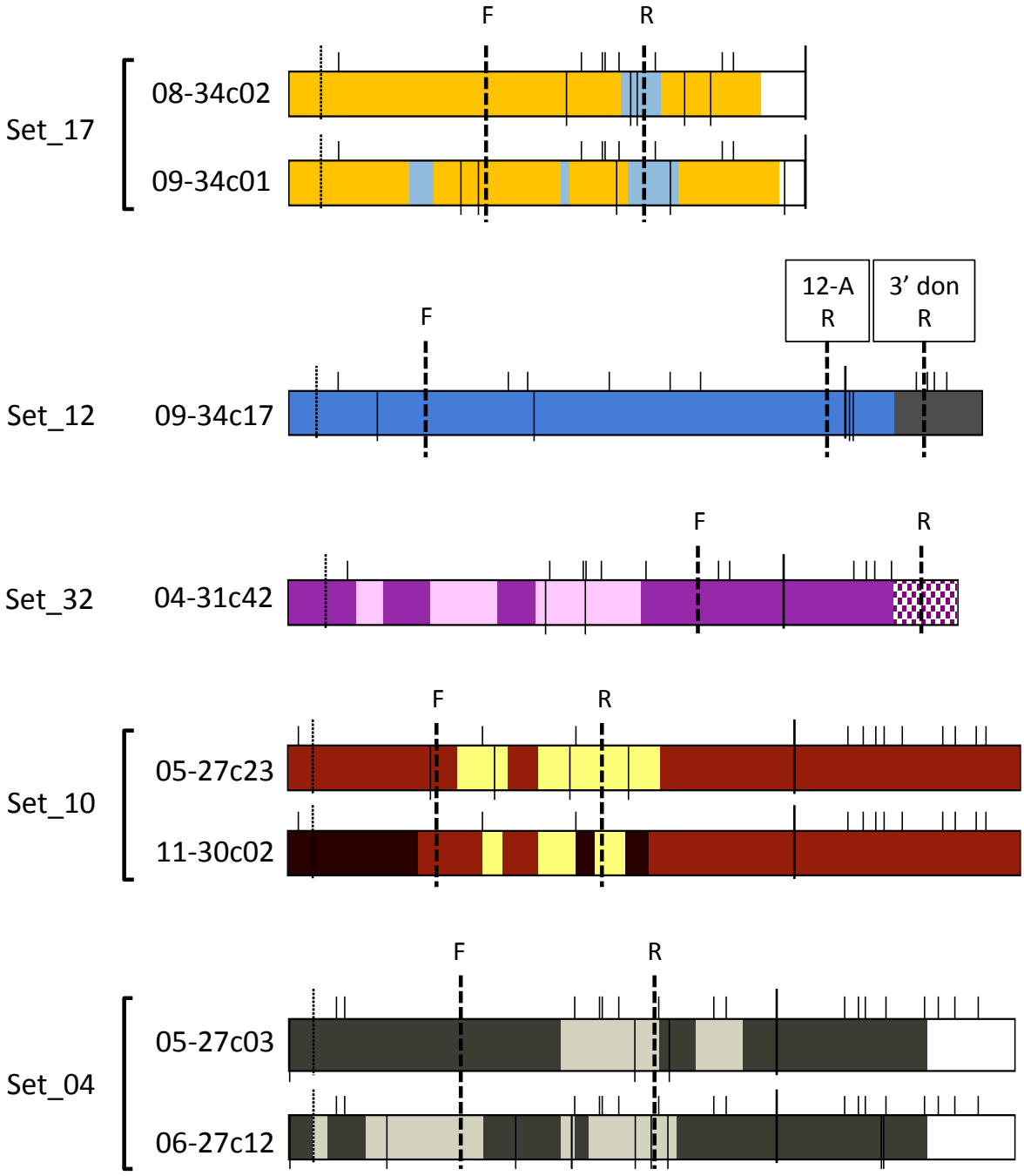


Figure 7.32 Further tests for segmental gene conversion. These diagrams were drawn as those in Appendix 7.3.4, with the different segments coloured according to the mosaic diagrams presented in Chapter 4. Primers were designed for five sets, Set_17, Set_12, Set_32, Set_10 and Set_04. Forward (F) and reverse (R) primers were designed to bind the donor sequences at the positions indicated. In the case of Set_12, the test was for 3' donation and the forward primer was the same in both reactions.

Set to be tested	Primer Name	Primer Sequence
Set_04	04-A_387F	CCTACTAAGCGCCAACGAC
	04-B_387F	CCTACTAAGTACCGACGCC
	04-A_712R	CGGTGACGCCCGAGTC
	04-B_712R	CTGCCACGCCCGAGGT
Set_10	See Table 7.1	
Set_12	12-A_276F	AGTAACAGACAGGGCCAACG
	12-A_1177F	CTGCTGTTTCGATCTCTTCT
	3_end_1382R	TTACCTTCTTCAGCGCCTG
Set_17	17-A_410F	TCGAACGGACAATTGCCC
	17-B_410F	TCGACCGAAAACTTGCAA
	17-B_762R	TGTATTTATGCTGGGTTGGC
Set_32	32-B_587F	CGACAACGGGAAGCAATG
	32-A_587F	CCACATCGGCGACCAACA
	32-B_984R	TTGGTGTGCTTTAGGTGGG
	32-A_984R	TTGGTGTGCTTTAGGTGCA

Table 7.9 Oligonucleotides used for testing of segmental gene conversion.

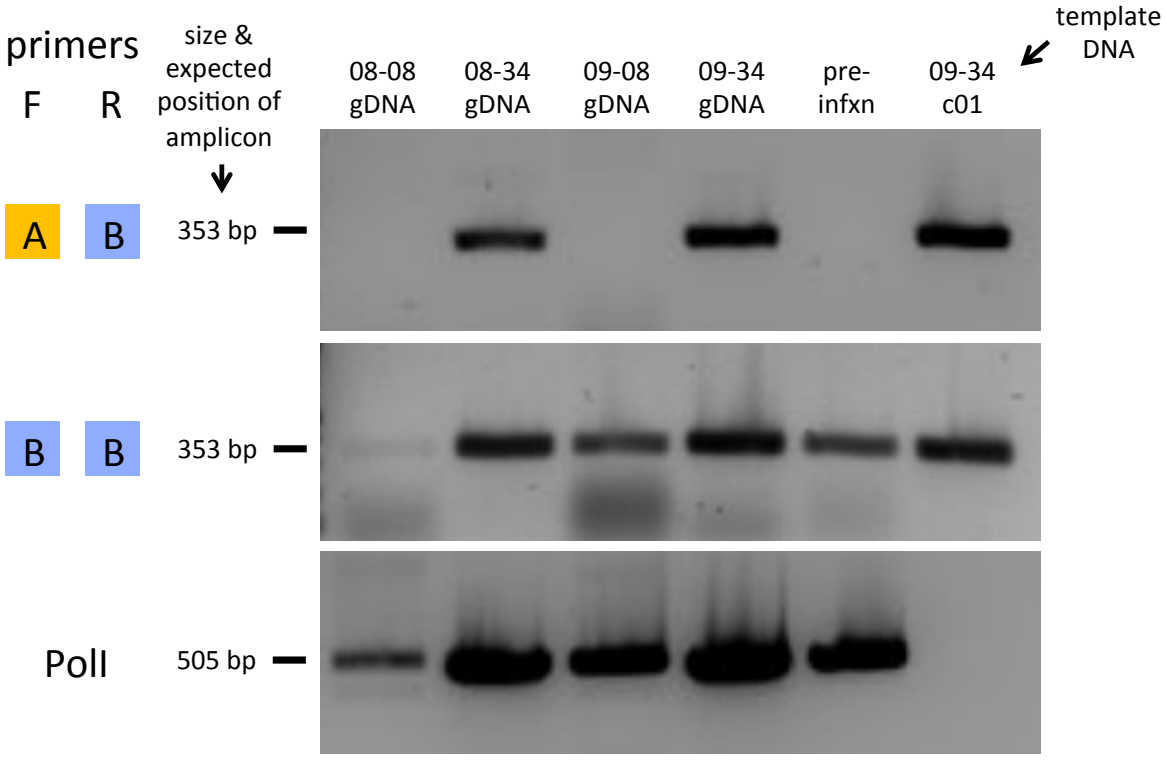


Figure 7.33 PCR tests for Set_17 mosaicism. PCR was applied to genomic DNA obtained from both the first peak of parasitaemia (day 8), and the terminal sample (day 34), of primary clonal infections. gDNA from mouse 08 at day 8 was at low concentration, contributing to the low abundance of product from the B-B and PolI reactions. In both mouse 08 and mouse 09, segmentally-converted product was present only in the later sample, although it is also possible that the low gDNA concentration from mouse 08 at day 8 prevented detection of a product in the A-B reaction.

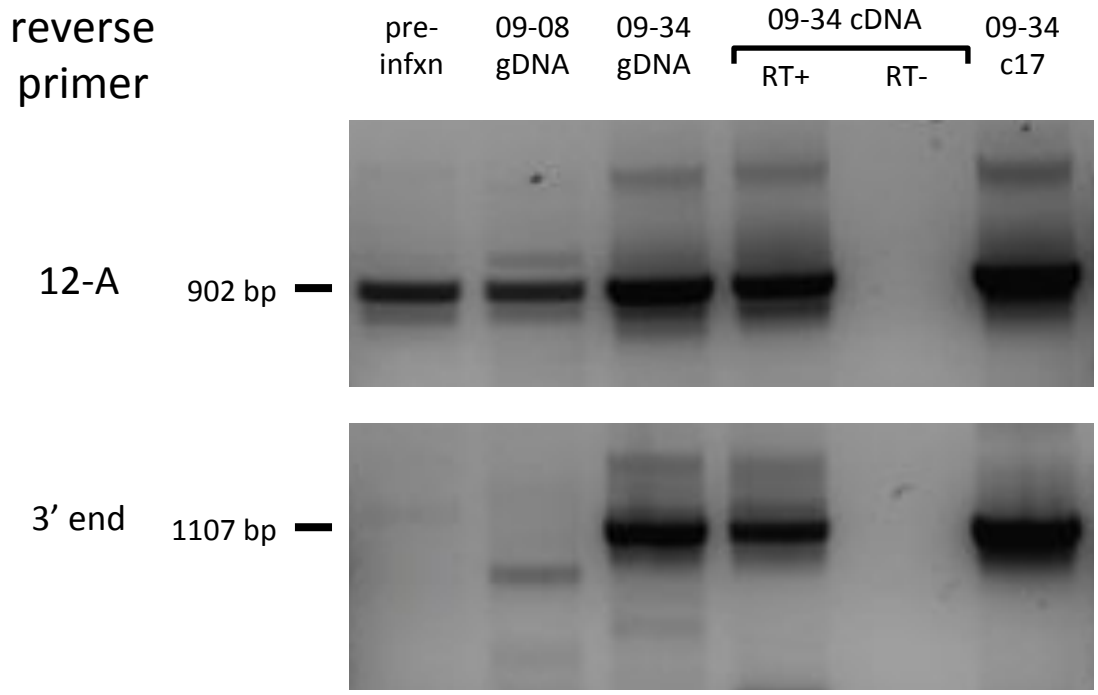


Figure 7.34 PCR tests for Set_12 3' donation. PCR was performed on samples as described in Figure 7.33. No product of the correct size is observed in pre-infection gDNA, or gDNA from the first peak of parasitaemia (day 8).

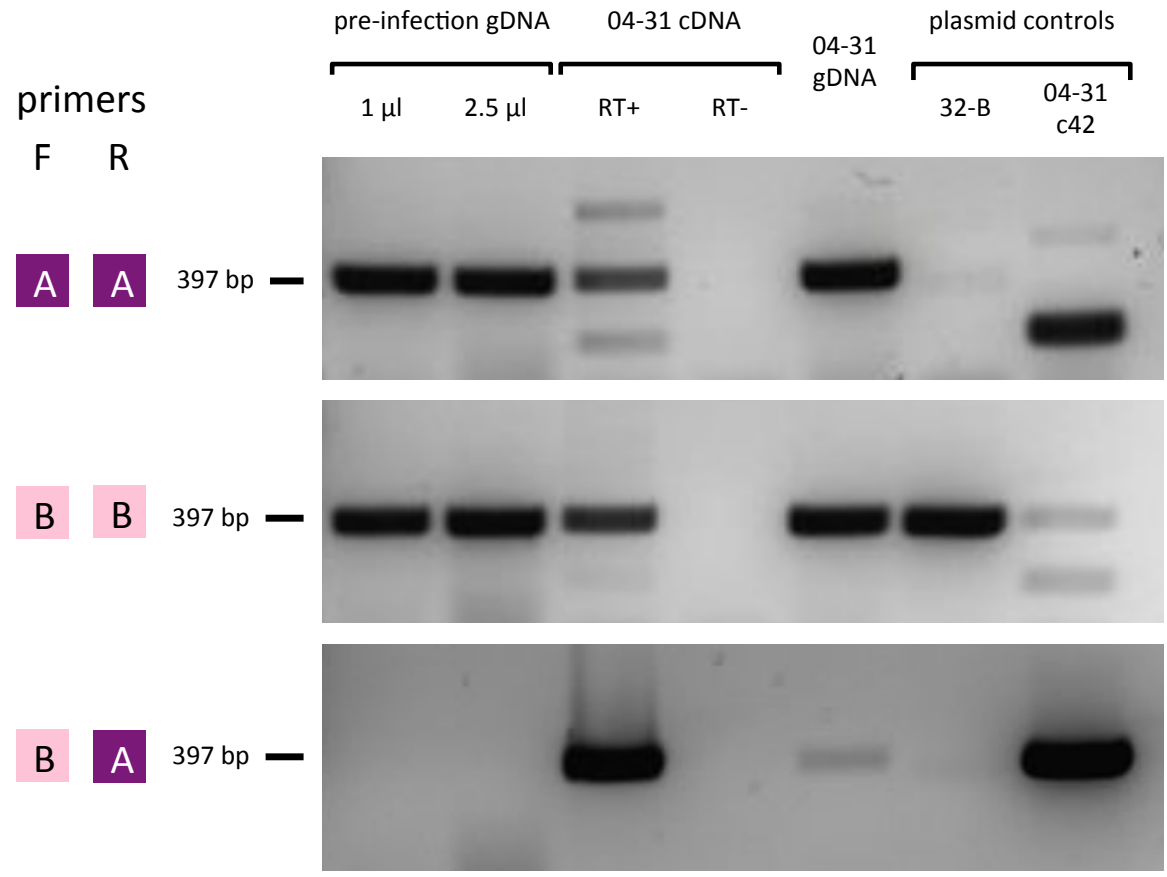


Figure 7.35 PCR tests for Set_32 mosaicism. PCR was performed as described in Figure 7.33. For this reaction, cDNA was also tested, to ascertain whether the mosaic junction was expressed by parasites in the population. Reactions were performed on the products of a reaction to which no reverse transcriptase had been added, as a control for gDNA contamination.

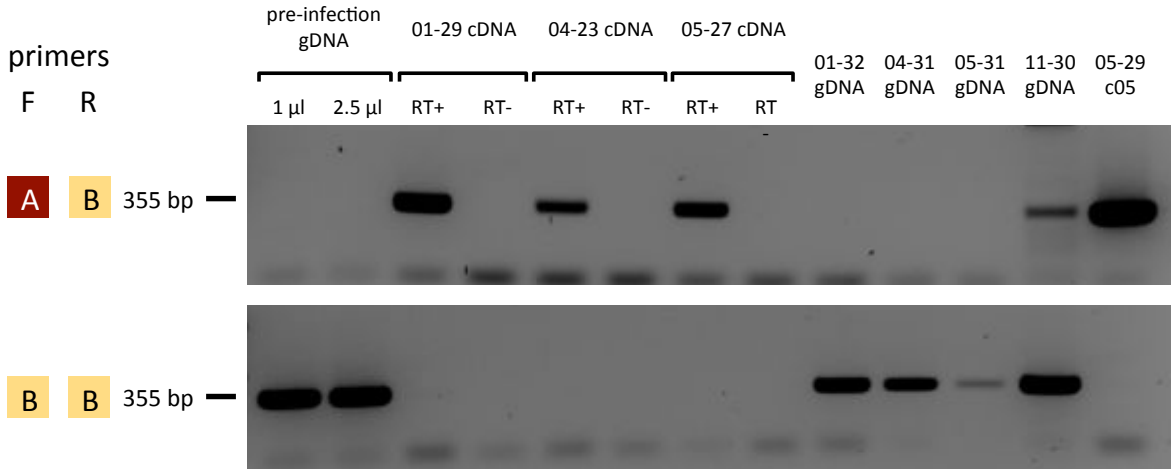


Figure 7.36 PCR tests for Set_10 mosaicism. Reactions were performed as in Figure 4.11 and Figure 7.35. For mouse 01, mouse 04 and mouse 05, gDNA was only available from the terminal sample, from which no Set_10 mosaics were detected. It is therefore likely that if the mosaic junction remained present at day 31/32, it was at such low abundance that it was undetectable. Set_10 mosaics were detectible in mouse 11 at day 30, as evidenced by the clone shown in Figure 7.32 and the product seen here, however, there was insufficient cDNA remaining to test its presence by PCR.

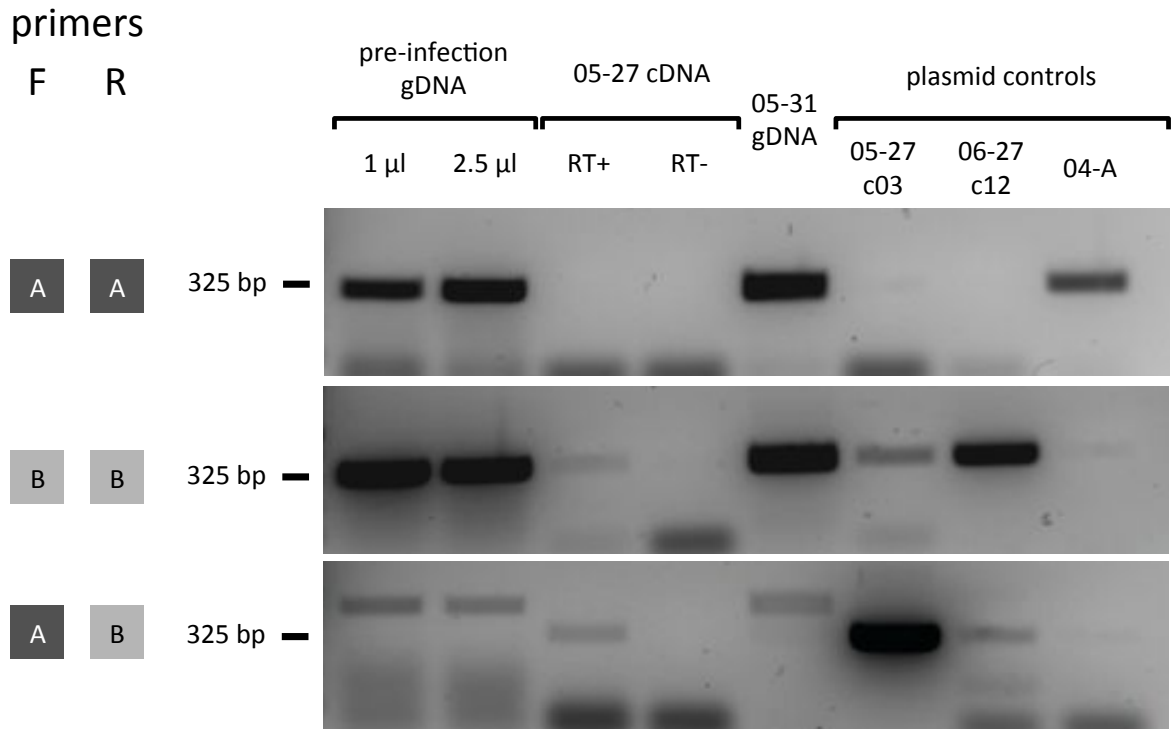


Figure 7.37 PCR tests for Set_04 mosaicism. Reactions were performed as in Figure 7.35. There was insufficient time to optimize reaction conditions to be both sensitive and completely specific, and thus a larger product was amplified from gDNA in the A-B reaction. The low abundance of product from the A-B reaction, when applied to the 05-31 gDNA template, may reflect the fact that Set_04 mosaics were not obtained from the terminal mouse 05 sample.

7.4.3 There were only four Set_14-related VSGs *in the genome*

It was discovered that the cross-reacting bands present in the Southern hybridization (presented in Figure 4.22) were most likely due to the means by which the probe had been made. In section 4.5.10, probe DNA was obtained by amplifying 14-A (Tb11.09.0005) directly from gDNA, and although the product was gel purified it is likely that it was an additional contaminating product that was labelled and hybridized to the blot. Here, a probe was made from DNA amplified from plasmid-isolated copy of 14-A, and was hybridized to another gDNA blot. The results are shown in Figure 7.38. As can be seen in the figure, the unexplained band in the earlier attempt is no longer present. As described before, the single band that does not have an annotated genomic copy is likely to be 14-C, as this band gave the strongest signal when hybridized to a probe made from 14-C DNA (data not shown). This evidence lends support to the premise that the varied Set_14 *VSG*s are constructed by segmental gene conversion, and not by the direct expression of pre-existing genomic copies.

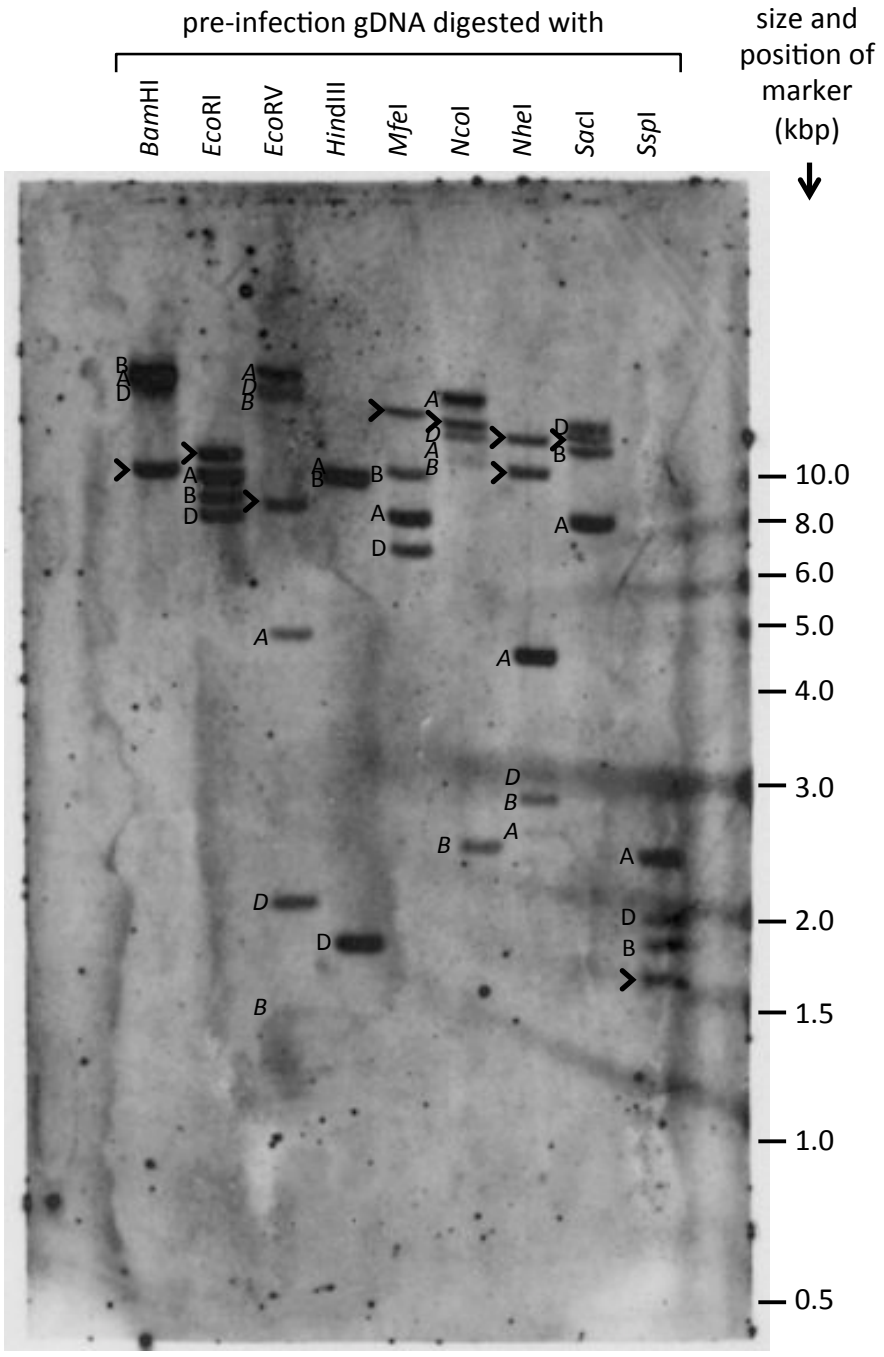


Figure 7.38 Southern hybridization of gDNA for detection of Set_14 donors. The blot was prepared using pre-infection 927 genomic DNA digested with one of nine restriction enzymes as described in Chapter 2, hybridized with a probe complementary to 14-A and washed to medium stringency (55ºC, buffer prepared according to kit instructions). Predicted positions of 14-A (Tb11.09.0005), 14-B (Tb11.13.0003) and 14-D (Tb10.v4.0009) fragments indicated by 'A', 'B' and 'D' respectively. Italics correspond to bands that do not encompass the entire gene. The remaining band indicated with an arrowhead, and corresponds with the band indicated by a ball-and-stick in Figure 4.22. This band is predicted to correspond with 14-C (tryp_XI-1034g11_seq) as it hybridized strongly to a 14-C probe (not shown). Note that the ordering of some of the larger bands could not always be resolved.

7.4.4 Cells referred to as 'variant 21' in Chapter 5 were in fact duplicate variant 23a-expressors

To ensure that the transgenic *VSG* expressors described in Chapter 5 were expressing the exact Set_14 mosaic *VSG*s under consideration, *VSG* cDNAs were cloned and sequenced from these lines. The sequenced clones were as expected for all variants except for variant 21 and variant 29 (VSG 427-4-expressors were not tested). In the case of variant 29, two out of five clones obtained did not match variant 29. As neither of these sequences contained the *Eco*RV site that caused the cutting of all detectible amplicons (Figure 5.4) it is likely that they were at very low abundance in the population, a premise supported by the fact that further variant 29 clones made from post-complement-mediated lysis cDNA samples all matched variant 29 (3 clones sequenced). 'Variant 29' expressors were therefore expressing variant 29. In the case of variant 21, however, all cloned sequences exactly matched variant 23a. To test whether 'variant 21' expressors were in fact expressing variant 23a, *VSG*s amplified from 'variant 21' cDNA were tested by *Sac*I digestion: variant 21 possesses a *Sac*I site, whereas 23a does not. Samples obtained immediately after transfection were completely digested, whereas later samples made from blood-borne parasites, or from recloned cells, were not digested by *Sac*I (data not shown). It is therefore likely that variant 21 was contaminated by 23a-expressing cells at some point after transfection, most probably during attempts to infect animals. Variant 21 was exceptionally difficult to obtain *in vivo* growth: many different infection protocols were attempted (data not shown), and during this period variant 21-expressing and variant 23a-expressing cells were cultured in neighbouring wells in the incubator. The apparently successful infection attempt reported in Chapter 5 required 1 x 10^6 cells, and it was from this infection that cells were recloned for use in experiments. The strong selection for growth and survival that *in vivo* growth placed on the trypanosome population meant that a small contaminating population of 23a expressors could have survived or out-competed the 21 expressors, resulting in the contaminated population on which subsequent experiments were performed.

Stabilates made of variant 21 expressors immediately after transfection were tested by RT-PCR and amplification digestion as described in section 5.2.3, and

the results were consistent with their expressing solely variant 21 (data not shown). Expressed *VSG*s were also cloned and sequenced, and were found to match only variant 21. In addition, SDS-PAGE of crude cell lysate as in section 5.2.4 showed a variant band of the size predicted for variant 21. Therefore, as the post-transfection cells were expressing variant 21, the question of whether variant 21 was antigenically distinct could be addressed using these cells.

7.4.5 Variant 21 was in fact antigenically different from variant 23a

To test the serological properties of genuine variant 21 expressors, attempts were made to raise polyclonal antibodies against this VSG. However, as before, infections with genuine variant 21 expressors failed in the two times they were attempted, despite using the infection protocol that worked for other variants. Variant 21 was therefore tested only with heterologous antiplasma. The results are given in Table 7.10. As can be seen in the table, variant 21 crossreacted with polyclonal antiplasma raised against variants 23a, 23c, 27 and 05-27. In that sense, it acted similarly to variant 23a. This means that the conclusions of the polyclonal antiplasma experiments remain as they did in Chapter 5. However, none of the four monoclonal antibodies bound to variant 21 expressors. This finding resolves the identification of the putative monoclonal antibody epitopes. Figure 5.13 and Figure 5.17 were corrected to give Figure 7.39 and Figure 7.41, and Figure 5.14 and Figure 5.18 were corrected to give Figure 7.40 and Figure 7.42. An updated version of Table 5.9 is given as Table 7.11.

	VSG expressed	antiplasma raised against																monoclonal antibodies, mAb-										
		427-2			23a			23c			27			29			427-4			21-A				21-B		29-A		29-B		
		I	C	A	I	C	A	I	C	A	I	C	A	I	C	A	I	C	A	I(f)	I	C	A	I(f)	I	I(f)	I	I(f)	I	C
	427-2	+++	5	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Set_14 mosaics	21	-	-	-	++	3	3	++	2	3	+	2	2	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-
	23a	-	-	-	++	3	3	++	3	3	++	2	4	-	-	-	-	-	-	+++	+++	4	4	+++	-	-	-	-	-	-
	23c	-	-	-	++	2	2	+++	3	3	++	2	3	-	-	1	-	-	-	+++	+++	4	4	-	X	-	X	-	-	-
	27	-	-	-	++	2	3	++	2	2	++	3	3	-	-	2	-	-	-	+++	+++	4	4	-	X	-	X	-	-	-
	29	-	-	-	-	-	-	-	-	-	-	-	-	++	3	3	-	-	-	-	-	-	-	-	-	+	-	+++	+	-
	427-4	-	-	X	-	-	X	-	-	X	-	1	X	-	-	X	+++	4	X	X	-	-	-	X	X	X	X	X	-	X

Table 7.10 Updated summary of serological cross-reactivity. Table 5.8 Summary of serological analyses of Set_14 mosaic VSGs. A summary of the results shown in Table 5.3, Table 5.4, Table 5.5, Table 5.6 and Table 5.7. 'I' indicates the results of IFA on live cells, 'C' indicates the results of CML, 'A' indicates the results of agglutination, 'I(f)' indicates the results of IFA on acetone fixed cells. For the immunofluorescence assays, '3' indicates strong, eggshell-like fluorescence. '2' indicates strong fluorescence, but perhaps with a speckled appearance or posterior accumulation. '1' indicates patchy or weak fluorescence. '-' indicates no signal. For CML and agglutination assays, the score is given as the number of threefold dilutions still able to give a signal (that is, 1 = 1/6, 2 = 1/18, 3 = 1/54 and so on). 'X' indicates that that particular test was not carried out with that antibody-VSG combination. 23b was omitted due to inconsistencies in VSG expression (section 5.3.7). The non-reciprocal cross-reaction seen in the agglutination test with variant 29 is discussed in section 5.3.3.

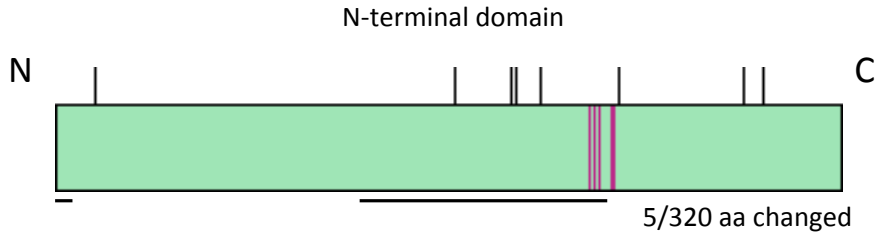


Figure 7.39 Residues contributing to the mAb-21A epitope structure. In a multiple sequence alignment of mature NTD amino acid sequences, positions where the variants that bound mAb-21A (23a, 23c, 27 and the mouse 05 Set_14 variant 05-27) were identical to one another but different from the variant that did not bind mAb-21A (21) are shown in magenta. This is an updated version of Figure 5.13.

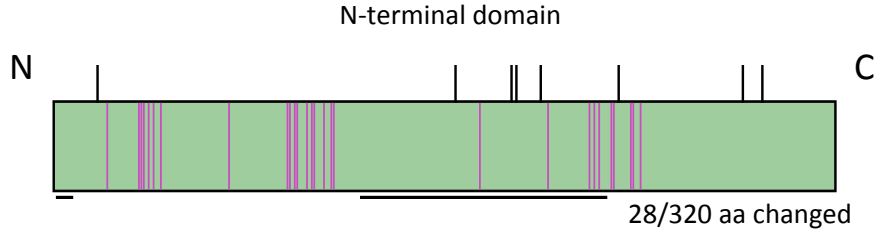


Figure 7.40 Residues contributing to the mAb-21B epitope structure. Positions where the variant that bound mAb-21B (23a) were identical to one another and different from the most similar variant that did not bind (21) are shown in magenta. Figure is annotated as Figure 5.12, with the positions of cysteine residues conserved amongst variants 21, 23a and 23c shown. Although mAb-21B could only bind fixed trypanosomes, it is unlikely that the epitope was contained in the CTD, as the only differences between variants 21 and 23a outside of the predicted NTD were in the GPI-anchor signal sequence.

mAb-21-A		
position	**x-rct**	**21**
218	A	T
220	N	H
222	Q	S
227	I	M
228	E	A

mAb-21-A		
position	**x-rct**	**29**
157	S	N
158	S	T
161	K	T
166	N	S
169	D	S
171	D	Q
199	G	A
205	K	T
206	P	T
209	D	Q
210	A	N
218	A	T
220	N	H
222	Q	S
227	I	M
228	E	A
259	K	S
263	K	P
264	S	T
265	A	Q

mAb-21-B		
pos	**23a**	**21**
20	R	Q
33	V	I
34	A	N
35	D	N
37	V	L
39	Q	E
42	A	T
70	N	K
94	Q	K
95	V	A
97	T	A
98	D	A
102	S	A
104	R	K
105	E	D
109	T	K
112	R	A
113	R	D
173	A	D
201	K	T
218	**A**	**T**
220	**N**	**H**
222	Q	S
227	I	M
228	E	A
235	G	T
236	A	K
239	K	A

Acidic
Basic
Non-polar
Polar

Table 7.11 Amino acid residues associated with mAb escape. Three tables show the amino acid substitutions associated with abrogation of mAb-21-A binding (left and middle, see Figure 7.39) and mAb-21-B binding (right, see Figure 7.40). 'Pos' = amino acid position in mature NTD; 'X-rct' = identities of amino acids at those positions in cross-reacting antigens; '29' or '23-c' = identities of amino acids at those positions in the antigenically distinct antigens. Emboldened rows correspond with positions predicted to be exposed on the top of the molecule. This is an updated version of Table 5.9.

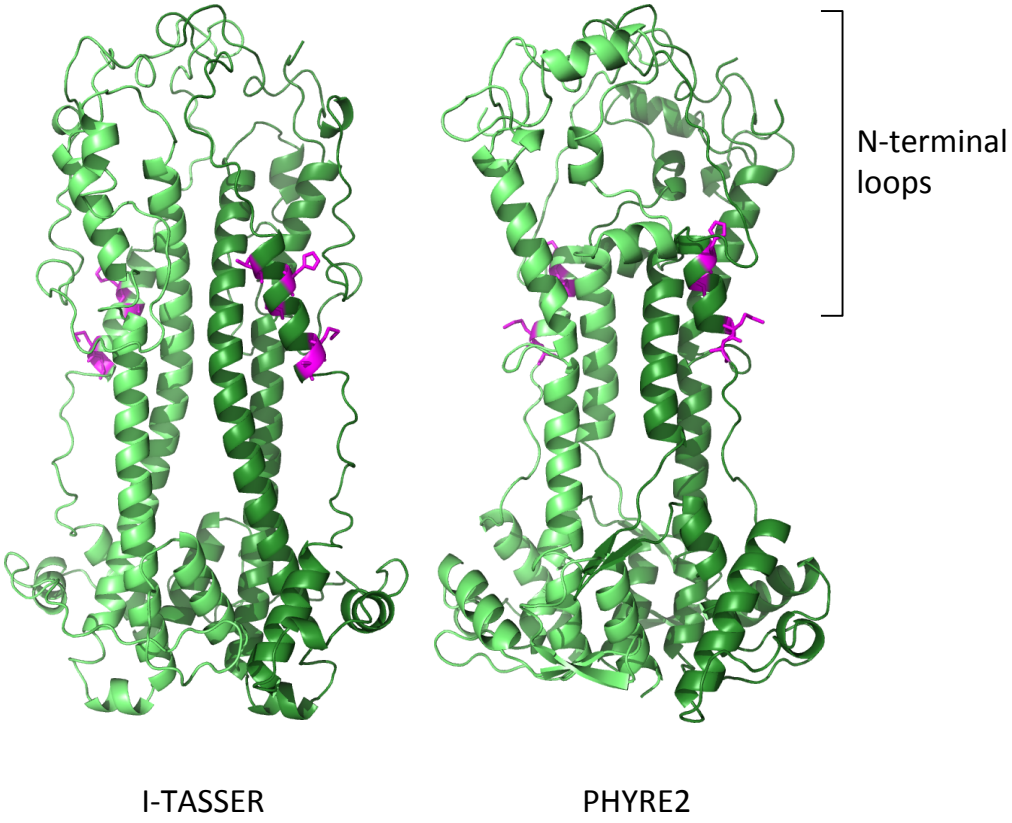


Figure 7.41 Predicted residues contributing to abrogation of mAb-21A binding. Residues identified in Figure 7.39 are indicated on the predicted structure of variant 21 in magenta.

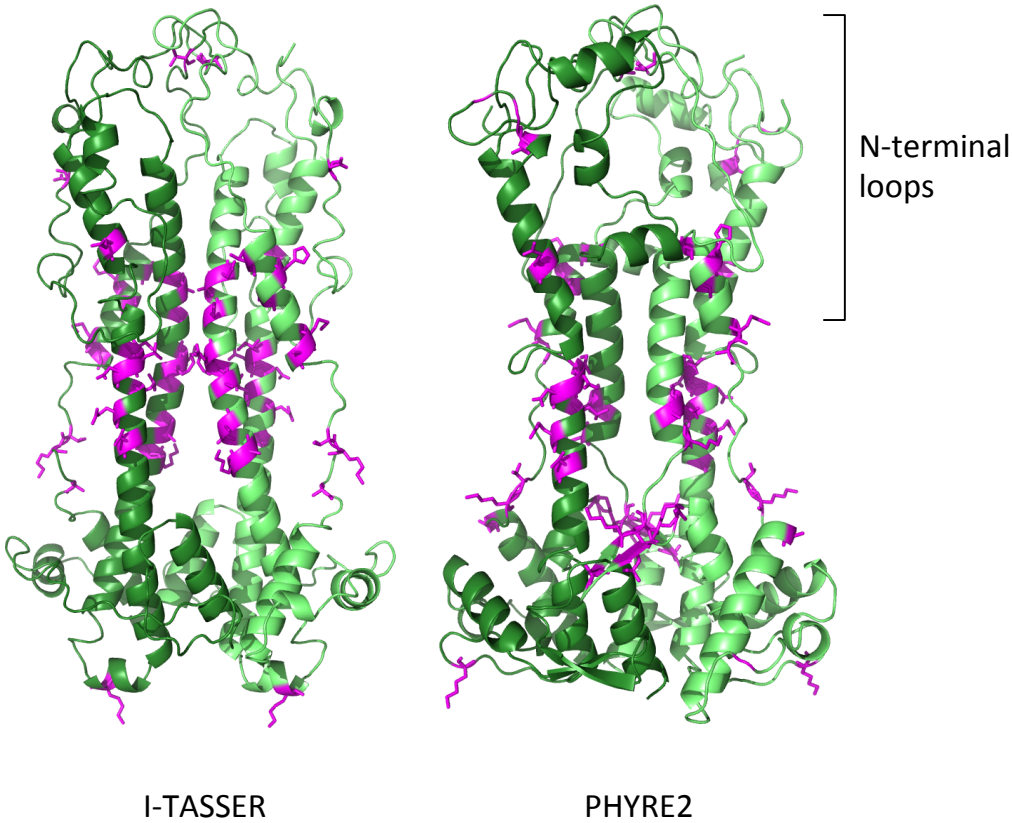


Figure 7.42 Predicted residues contributing to abrogation of mAb-21B binding. Residues identified in Figure 7.40 are indicated on the predicted structure of variant 21 in magenta.

List of references

Adam, R. D., Nigam, A., Sephardi, V., Martens, C. A., Farneth, G. A., Morrison, H. G., Nash, T. E., Porcella, S. F. & Patel, R. (2010). The Giardia lamblia vsp gene repertoire: characteristics, genomic organization, and evolution. *BMC Genomics* **11**, 424.

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

Al-Khedery, B. & Allred, D. R. (2006). Antigenic variation in Babesia bovis occurs through segmental gene conversion of the ves multigene family, within a bidirectional locus of active transcription. *Mol Microbiol* **59**, 402–414.

Aline, R. F., Myler, P. J., Gobright, E. & Stuart, K. D. (1994). Early expression of a Trypanosoma brucei *VSG* gene duplicated from an incomplete basic copy. *J Eukaryot Microbiol* **41**, 71–78.

Aline, R. F., Myler, P. J. & Stuart, K. D. (1989). Trypanosoma brucei: frequent loss of a telomeric variant surface glycoprotein gene. *Exp Parasitol* **68**, 8–16.

Aline, R. F., Scholler, J. K., Nelson, R. G., Agabian, N. & Stuart, K. (1985). Preferential activation of telomeric variant surface glycoprotein genes in Trypanosoma brucei. *Mol Biochem Parasitol* **17**, 311–320.

Aline, R. F. & Stuart, K. (1985). The two mechanisms for antigenic variation in Trypanosoma brucei are independent processes. *Mol Biochem Parasitol* **16**, 11–20.

Aline, R. F. & Stuart, K. (1989). Trypanosoma brucei: conserved sequence organization 3' to telomeric variant surface glycoprotein genes. *Exp Parasitol* **68**, 57–66.

Allen, G. & Gurnett, L. P. (1983). Locations of the six disulphide bonds in a variant surface glycoprotein (VSG 117) from Trypanosoma brucei. *Biochem J* **209**, 481–487.

Alsford, S., Wickstead, B., Ersfeld, K. & Gull, K. (2001). Diversity and dynamics of the minichromosomal karyotype in Trypanosoma brucei. *Mol Biochem Parasitol* **113**, 79–88.

Alsford, S., duBois, K., Horn, D. & Field, M. C. (2012). Epigenetic mechanisms, nuclear architecture and the control of gene expression in trypanosomes. *Expert Rev Mol Med* **14**, –. Cambridge University Press.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Amiguet-Vercher, A., Pérez-Morga, D., Pays, A., Poelvoorde, P., Van Xong, H., Tebabi, P., Vanhamme, L. & Pays, E. (2004). Loss of the mono-allelic control of the *VSG* expression sites during the development of Trypanosoma brucei in the bloodstream. *Mol Microbiol* **51**, 1577–1588.

Arezi, B., Xing, W., Sorge, J. A. & Hogrefe, H. H. (2003). Amplification efficiency of thermostable DNA polymerases. *Anal Biochem* **321**, 226–235.

Askonas, B. A., Corsini, A. C., Clayton, C. E. & Ogilvie, B. M. (1979). Functional depletion of T- and B-memory cells and other lymphoid cell subpopulations-during trypanosomiasis. *Immunology* **36**, 313–321.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., Depledge, D. P., Fischer, S., Gajria, B., & other authors. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res* **38**, D457–62.

Auffret, C. A. & Turner, M. J. (1981). Variant specific antigens of Trypanosoma

brucei exist in solution as glycoprotein dimers. *Biochem J* **193**, 647–650.

Baetselier, P. D., Namangala, B., Noël, W., Brys, L., Pays, E. & Beschin, A. (2001). Alternative versus classical macrophage activation during experimental African trypanosomosis. *Int J Parasitol* **31**, 575–587.

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

Balmer, O., Stearns, S. C., Schötzau, A. & Brun, R. (2009). Intraspecific competition between co-infecting parasite strains enhances host survival in African trypanosomes. *Ecology* **90**, 3367–3378.

Baltz, T., Giroud, C., Bringaud, F., Eisen, H., Jacquemot, C. & Roth, C. W. (1991). Exposed epitopes on a Trypanosoma equiperdum variant surface glycoprotein altered by point mutations. *EMBO J* **10**, 1653–1659.

Baltz, T., Giroud, C., Baltz, D., Roth, C., Raibaud, A. & Eisen, H. (1986). Stable expression of two variable surface glycoproteins by cloned *Trypanosoma equiperdum*. *Nature*, **319**(6054), 602–604.

Barbet, A. F., Davis, W. C. & McGuire, T. C. (1982). Cross-neutralization of two different trypanosome populations derived from a single organism. *Nature* **300**, 453–456.

Barbet, A. F. & Kamper, S. M. (1993). The importance of mosaic genes to trypanosome survival. *Parasitol Today (Regul Ed)* **9**, 63–66.

Barbet, A. F., Lundgren, A., Yi, J., Rurangirwa, F. R. & Palmer, G. H. (2000). Antigenic variation of Anaplasma marginale by expression of MSP2 mosaics. *Infect Immun* **68**, 6133–6138.

Barbet, A. F., Myler, P. J., Williams, R. O. & McGuire, T. C. (1989). Shared surface epitopes among trypanosomes of the same serodeme expressing different variable surface glycoprotein genes. *Mol Biochem Parasitol* **32**, 191–199.

Barbour, A. G. & Restrepo, B. I. (2000). Antigenic variation in vector-borne pathogens. *Emerging Infect Dis* **6**, 449–457.

Barbour, A. G., Dai, Q., Restrepo, B. I., Stoenner, H. G. & Frank, S. A. (2006). Pathogen escape from host immunity by a genome program for antigenic variation. *Proc Natl Acad Sci USA* **103**, 18290–18295.

Barnes, R. L. & 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.

Barrett, M., Burchmore, R., Stich, A. H., Lazzari, J., Frasch, A. C., Cazzulo, J. & Krishna, S. (2003). The trypanosomiases. *The Lancet* **362**, 1469–1480.

Barry, J. D. (1986). Antigenic variation during Trypanosoma vivax infections of different host species. *Parasitology* **92 (Pt 1)**, 51–65.

Barry, J. D., Graham, S. V., Fotheringham, M., Graham, V. S., Kobryn, K. & Wymer, B. (1998). *VSG* gene control and infectivity strategy of metacyclic stage Trypanosoma brucei. *Mol Biochem Parasitol* **91**, 93–105.

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.

Barry, J. D. (1997). The relative significance of mechanisms of antigenic variation in African trypanosomes. *Parasitol Today (Regul Ed)* **13**, 212–218.

Barry, J. D., Ginger, M. L., Burton, P. & McCulloch, R. (2003). Why are parasite contingency genes often associated with telomeres? *Int J Parasitol* **33**, 29–45.

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

Beals, T. P. & Boothroyd, J. C. (1992). Genomic organization and context of a trypanosome variant surface glycoprotein gene family. *J Mol Biol* **225**, 961–971.

Berberof, M., Vanhamme, L., Tebabi, P., Pays, A., Jefferies, D., Welburn, S. & Pays, E. (1995). The 3'-terminal region of the mRNAs for *VSG* and procyclin can confer stage specificity to gene expression in Trypanosoma brucei. *EMBO J* **14**, 2925–2934.

Bernards, A., Van der Ploeg, L. H., Frasch, A. C., Borst, P., Boothroyd, J. C., Coleman, S. & Cross, G. A. (1981). Activation of trypanosome surface glycoprotein genes involves a duplication-transposition leading to an altered 3' end. *Cell* **27**, 497–505.

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. & El Sayed, N.M. (2005). The genome of the African trypanosome Trypanosoma brucei. *Science* **309**, 416–422.

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. & Rudenko, G. (2002). The architecture of variant surface glycoprotein gene expression sites in Trypanosoma brucei. *Mol Biochem Parasitol* **122**, 131–140.

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

Black, S. J., Guirnalda, P., Frenkel, D., Haynes, C. & Bockstal, V. (2010). Induction and regulation of Trypanosoma brucei VSG-specific antibody responses. *Parasitology* 1–9.

Black, S., Hewett, R. S. & Sendashonga, C. N. (1982). Trypanosoma brucei variable surface antigen is released by degenerating parasites but not by actively dividing parasites. *Parasite Immunol* **4**, 233–244.

Blum, M. L., Down, J. A., Gurnett, A. M., Carrington, M., Turner, M. J. & Wiley, D. C. (1993). A structural motif in the variant surface glycoproteins of Trypanosoma brucei. *Nature* **362**, 603–609.

Boni, M. F. (2008). Vaccination and antigenic drift in influenza. *Vaccine* **26**

Suppl 3, C8–14.

Boothroyd, C. E., Dreesen, O., Leonova, T., Ly, K. I., Figueiredo, L. M., Cross, G. A. M. & Papavasiliou, F. N. (2009). A yeast-endonuclease-generated DNA break induces antigenic switching in Trypanosoma brucei. *Nature* **459**, 278–281.

Boothroyd, J. C., Paynter, C. A., Cross, G. A. M., Bernards, A. & Borst, P. (1981). Variant surface glycoproteins of Trypanosoma brucei are synthesised with cleavable hydrophobic sequences at the carboxy and amino termini. *Nucleic Acids Res* **9**, 4735–4743.

Borst, P. & Sabatini, R. (2008). Base J: discovery, biosynthesis, and possible functions. *Annual review of microbiology*, 62, 235–251. doi:10.1146/annurev.micro.62.081307.162750

Bowen, D. G. & Walker, C. M. (2005). Adaptive immune responses in acute and chronic hepatitis C virus infection. *Nature* **436**, 946–952.

Brayton, K. A., Knowles, D. P., McGuire, T. C. & Palmer, G. H. (2001). Efficient use of a small genome to generate antigenic diversity in tick-borne ehrlichial pathogens. *Proc Natl Acad Sci USA* **98**, 4130–4135.

Brayton, K. A., Kappmeyer, L. S., Herndon, D. R., Dark, M. J., Tibbals, D. L., Palmer, G. H., McGuire, T. C. & Knowles, D. P. (2005). Complete genome sequencing of Anaplasma marginale reveals that the surface is skewed to two superfamilies of outer membrane proteins. *Proc Natl Acad Sci USA* **102**, 844–849.

Brayton, K. A., Lau, A. O. T., Herndon, D. R., Hannick, L., Kappmeyer, L. S., Berens, S. J., Bidwell, S. L., Brown, W. C., Crabtree, J., Fadrosh, D., Feldblum, T., Forberger, H. A., Haas, B. J., Howell, J. M., Khouri, H., Koo, H., Mann, D. J., Norimine, J., Paulsen, I. T., Radune, D., Ren, Q., Smith, R. K., Suarez, C. E., White, O., Wortman, J. R., Knowles, D. P., McElwain, T. F. & Nene, V. M. (2007). Genome sequence of Babesia bovis and comparative analysis of apicomplexan hemoprotozoa. *PLoS pathogens*, *3*(10), 1401–1413. doi:10.1371/journal.ppat.0030148

Buelow, R., Overath, P. & Davoust, J. (1988). Rapid lateral diffusion of the variant surface glycoprotein in the coat of Trypanosoma brucei. *Biochemistry* **27**, 2384–2388. American Chemical Society.

Bull, P. C., Buckee, C. O., Kyes, S., Kortok, M. M., Thathy, V., Guyah, B., Stoute, J. A., Newbold, C. I. & Marsh, K. (2008). Plasmodium falciparum antigenic variation. Mapping mosaic var gene sequences onto a network of shared, highly polymorphic sequence blocks. *Mol Microbiol* **68**, 1519–1534.

Bussler, H., Linder, M., Linder, D. & Reinwald, E. (1998). Determination of the disulfide bonds within a B domain variant surface glycoprotein from Trypanosoma congolense. *J Biol Chem* **273**, 32582–32586.

Cahoon, L. A. & Seifert, H. S. (2009). An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. *Science* (New York, NY), 325(5941), 764–767. doi:10.1126/science.1175653

Campbell, D. A., van Bree, M. P. & Boothroyd, J. C. (1984). The 5'-limit of transposition and upstream barren region of a trypanosome *VSG* gene: tandem 76 base-pair repeats flanking (TAA)90. *Nucleic Acids Res* **12**, 2759–2774.

Campbell, G. H. & Phillips, S. M. (1976). Adoptive transfer of variant-specific resistance to Trypanosoma rhodesiense with B lymphocytes and serum. *Infect Immun* **14**, 1144–1150.

Campillo, N., & Carrington, M. (2003). The origin of the serum resistance associated (SRA) gene and a model of the structure of the SRA polypeptide from *Trypanosoma brucei rhodesiense*. *Mol Biochem Parasitol* **127**(1), 79–84.

Capbern, A., Giroud, C., Baltz, T. & Mattern, P. (1977). [Trypanosoma equiperdum: antigenic variations in experimental trypanosomiasis of rabbits]. *Exp Parasitol* **42**, 6–13.

Capewell, P., Veitch, N. J., Turner, C. M. R., Raper, J., Berriman, M., Hajduk, S. L. & Macleod, A. (2011). Differences between Trypanosoma brucei gambiense groups 1 and 2 in their resistance to killing by trypanolytic factor 1. *PLoS Negl Trop Dis* **5**, e1287.

Caporale, L. H. (2003). Natural selection and the emergence of a mutation phenotype: an update of the evolutionary synthesis considering mechanisms that affect genome variation. *Annu Rev Microbiol* **57**, 467–485.

Carrat, F. & Flahault, A. (2007). Influenza vaccine: the challenge of antigenic drift. *Vaccine* **25**, 6852–6862.

Carrington, M. & Boothroyd, J. C. (1996). Implications of conserved structural motifs in disparate trypanosome surface proteins. *Mol Biochem Parasitol* **81**, 119–126.

Carrington, M., Miller, N., Blum, M. L., Roditi, I., Wiley, D. C. & Turner, M. (1991). Variant specific glycoprotein of Trypanosoma brucei consists of two domains each having an independently conserved pattern of cysteine residues. *J Mol Biol* **221**, 823–835.

Caton, A. J., Brownlee, G. G., Yewdell, J. W. & Gerhard, W. U. (1982). The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell* **31**, 417–427.

Cattand, P., Jannin, J. & Lucas, P. (2001). Sleeping sickness surveillance: an essential step towards elimination. *Trop Med Int Health* **6**, 348–361.

Centurion-Lara, A., LaFond, R. E., Hevner, K., Godornes, C., Molini, B. J., Van Voorhis, W. C. & Lukehart, S. A. (2004). Gene conversion: a mechanism for generation of heterogeneity in the tprK gene of Treponema pallidum during infection. *Mol Microbiol* **52**, 1579–1596.

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**, 783–791.

Chattopadhyay, A., Jones, N. G., Nietlispach, D., Nielsen, P. R., Voorheis, H. P., Mott, H. R. & Carrington, M. (2005). Structure of the C-terminal domain from Trypanosoma brucei variant surface glycoprotein MITat1.2. *J Biol Chem* **280**, 7228–7235.

Chaves, I., Rudenko, G., Dirks-Mulder, A., Cross, M. & Borst, P. (1999). Control of variant surface glycoprotein gene-expression sites in Trypanosoma brucei. *EMBO J* **18**, 4846–4855.

Clarke, M. W., Barbet, A. F. & Pearson, T. W. (1987). Structural features of antigenic determinants on variant surface glycoproteins from Trypanosoma brucei. *Molecular Immunology* **24**, 707–713.

Cohen, C., Reinhardt, B., Parry, D. A., Roelants, G. E., Hirsch, W. & Kanwé, B. (1984). Alpha-helical coiled-coil structures of Trypanosoma brucei variable surface glycoproteins. *Nature* **311**, 169–171.

Conway, C., Proudfoot, C., Burton, P., Barry, J. D. & McCulloch, R. (2002). Two pathways of homologous recombination in Trypanosoma brucei. *Mol Microbiol* **45**, 1687–1700.

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

Coutte, L., Botkin, D. J., Gao, L. & Norris, S. J. (2009). Detailed analysis of sequence changes occurring during vlsE antigenic variation in the mouse model of Borrelia burgdorferi infection. *PLoS Pathog* **5**, e1000293.

Criss, A. K., Kline, K. A. & Seifert, H. S. (2005). The frequency and rate of pilin antigenic variation in Neisseria gonorrhoeae. *Mol Microbiol* **58**, 510–519.

Cross, G. A. M. (1975). Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of Trypanosoma brucei. *Parasitology* **71**, 393–417.

Cross, G. A. M. (1990). Cellular and genetic aspects of antigenic variation in trypanosomes. *Annu Rev Immunol* **8**, 83–110.

Cully, D. F., Ip, H. S. & Cross, G. A. (1985). Coordinate transcription of variant surface glycoprotein genes and an expression site associated gene family in Trypanosoma brucei. *Cell* **42**, 173–182. Elsevier.

Dai, Q., Restrepo, B. I., Porcella, S. F., Raffel, S. J., Schwan, T. G. & Barbour, A. G. (2006). Antigenic variation by Borrelia hermsii occurs through recombination between extragenic repetitive elements on linear plasmids. *Mol Microbiol* **60**, 1329–1343.

Das, S., Nikolaidis, N., Goto, H., McCallister, C., Li, J., Hirano, M. & Cooper, M. D. (2010). Comparative Genomics and Evolution of the Alpha-Defensin Multigene Family in Primates. *Mol Biol Evol* **27**, 2333–2343.

De Greef, C. & Hamers, R. (1994). The serum resistance-associated (SRA) gene of T*rypanosoma brucei rhodesiense* encodes a variant surface glycoprotein-like protein. *Mo Biochem Parasitol*, **68**(2), 277–284

De Lange, T. & Borst, P. (1982). Genomic environment of the expression-linked extra copies of genes for surface antigens of Trypanosoma brucei resembles the end of a chromosome. *Nature* **299**, 451–453.

de Raadt, P. (2005). WHO | The history of sleeping sickness. *WHO*. World Health Organization.

Dedet, J. P. & Pratlong, F. (2000). Leishmania, Trypanosoma and monoxenous trypanosomatids as emerging opportunistic agents. *J Eukaryot Microbiol* **47**, 37–39.

Deitsch, K. W., Lukehart, S. & Stringer, J. (2009). Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat Rev Microbiol*.

Delhi, P., Queiroz, R., Inchaustegui, D., Carrington, M. & Clayton, C. (2011). Is there a classical nonsense-mediated decay pathway in trypanosomes? *PLoS ONE* **6**, e25112.

Dero, B., Zampetti-Bosseler, F., Pays, E. & Steinert, M. (1987). The genome and the antigen gene repertoire of Trypanosoma brucei gambiense are smaller than those of T. b. brucei. *Mol Biochem Parasitol* **26**, 247–256.

Doyle, J. J., Hirumi, H., Hirumi, K., Lupton, E. N. & Cross, G. A. (1980). Antigenic variation in clones of animal-infective Trypanosoma brucei derived and maintained in vitro. *Parasitology* **80**, 359–369.

DuBois, K. N., Alsford, S., Holden, J. M., Buisson, J., Swiderski, M., Bart, J.-M., Ratushny, A. V., Wan, Y., Bastin, P., Barry, J. D., Navarro, M., Horn, D., Aitchison, J. D., Rout, M. P. & Field, M. C. (2012). NUP-1 Is a Large Coiled-Coil Nucleoskeletal Protein in Trypanosomes with Lamin-Like Functions. *PLoS Biol* **10**, e1001287.

Dubois, M. E., Demick, K. P. & Mansfield, J. M. (2005). Trypanosomes expressing a mosaic variant surface glycoprotein coat escape early detection by the immune system. *Infection and immunity*, **73**(5), 2690–2697.

Ehlers, B., Czichos, J. & Overath, P. (1987). RNA turnover in Trypanosoma brucei. *Mol Cell Biol* **7**, 1242–1249.

Eisenhaber, B., Bork, P. & Eisenhaber, F. (1999). Prediction of potential GPI-modification sites in proprotein sequences. *J Mol Biol* **292**, 741–758.

Engstler, M., Pfohl, T., Herminghaus, S., Boshart, M., Wiegertjes, G.,

Heddergott, N. & Overath, P. (2007). Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. *Cell* **131**, 505–515.

Engstler, M., Thilo, L., Weise, F., Grünfelder, C. G., Schwarz, H., Boshart, M. & Overath, P. (2004). Kinetics of endocytosis and recycling of the GPI-anchored variant surface glycoprotein in Trypanosoma brucei. *J Cell Sci* **117**, 1105–1115.

Eppig, J. T., Blake, J. A., Bult, C. J., Kadin, J. A., Richardson, J. E., Mouse Genome Database Group. (2012). The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. *Nucleic Acids Res* **40**, D881–6.

Esser, K. M. & Schoenbechler, M. J. (1985). Expression of two variant surface glycoproteins on individual African trypanosomes during antigen switching. *Science* **229**, 190–193.

Ferguson, M. A., Duszenko, M., Lamont, G. S., Overath, P. & Cross, G. A. (1986). Biosynthesis of Trypanosoma brucei variant surface glycoproteins. N-glycosylation and addition of a phosphatidylinositol membrane anchor. *J Biol Chem* **261**, 356–362.

Ferrante, A. & Allison, A. C. (1983). Alternative pathway activation of complement by African trypanosomes lacking a glycoprotein coat. *Parasite Immunol* **5**, 491–498.

Field, M. C., Sergeenko, T., Wang, Y.-N., Böhm, S. & Carrington, M. (2010). Chaperone requirements for biosynthesis of the trypanosome variant surface glycoprotein. *PLoS ONE* **5**, e8468.

Figueiredo, L. M., Janzen, C. J. & Cross, G. A. M. (2008). A histone methyltransferase modulates antigenic variation in African trypanosomes. *PLoS Biol* **6**, e161.

Finlay, B. B. & McFadden, G. (2006). Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. *Cell* **124**, 767–782.

Ford, J. (1969). Control of the African trypanosomiases with special reference to land use. *Bull World Health Organ* **40**, 879. World Health Organization.

Fox, C. H., Johnson, F. B., Whiting, J. & Roller, P. P. (1985). *Formaldehyde fixation. J Histochem Cytochem.*

Frank, S. A. (1999). A model for the sequential dominance of antigenic variants in African trypanosome infections. *Proc Biol Sci* **266**, 1397–1401.

Freymann, D., Down, J., Carrington, M., Roditi, I., Turner, M. & Wiley, D. C. (1990). 2.9 A resolution structure of the N-terminal domain of a variant surface glycoprotein from Trypanosoma brucei. *J Mol Biol* **216**, 141–160.

Frost, S. D. W., Wrin, T., Smith, D. M., Kosakovsky Pond, S. L., Liu, Y., Paxinos, E., Chappey, C., Galovich, J., Beauchaine, J., Petropoulos, C. J., Little, S. J. & Richman, D. D. (2005). Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. *Proc Natl Acad Sci USA* **102**, 18514–18519.

Futse, J. E., Brayton, K. A., Dark, M. J., Knowles, D. P. & Palmer, G. H. (2008). Superinfection as a driver of genomic diversification in antigenically variant pathogens. *Proc Natl Acad Sci USA* **105**, 2123–2127.

Futse, J. E., Brayton, K. A., Knowles, D. P. & Palmer, G. H. (2005). Structural basis for segmental gene conversion in generation of Anaplasma marginale outer membrane protein variants. *Mol Microbiol* **57**, 212–221.

Gibson, W. C. (2007). Resolution of the species problem in African trypanosomes. *Int J Parasitol* **37**, 829–838.

Gibson, W., Peacock, L., Ferris, V., Williams, K. & Bailey, M. (2008). The use of yellow fluorescent hybrids to indicate mating in Trypanosoma brucei. *Parasites & vectors* **1**, 4.

Gjini, E., Haydon, D. T., Barry, J. D. & Cobbold, C. A. (2010). Critical interplay between parasite differentiation, host immunity, and antigenic variation in trypanosome infections. *Am Nat* **176**, 424–439.

Gjini, E., Haydon, D. T., Barry, J. D. & Cobbold, C. A. (2012). The impact of mutation and gene conversion on the local diversification of antigen genes in African trypanosomes. *Mol Biol Evol*.

Glanville, J., Zhai, W., Berka, J., Telman, D., Huerta, G., Mehta, G. R., Ni, I., Mei, L., Sundar, P. D., Day, G. M. R., Cox, D., Rajpal, A., & Pons, J. (2009). Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proc Natl Acad Sci USA* **106**, 20216–20221.

Glover, L. & Horn, D. (2006). Repression of polymerase I-mediated gene expression at Trypanosoma brucei telomeres. *EMBO Rep* **7**, 93–99.

Goto, N., Prins, P., Nakao, M., Bonnal, R., Aerts, J. & Katayama, T. (2010). BioRuby: bioinformatics software for the Ruby programming language. *Bioinformatics* **26**, 2617–2619.

Graham, V. S. & Barry, J. D. (1996). Is point mutagenesis a mechanism for antigenic variation in Trypanosoma brucei? *Mol Biochem Parasitol* **79**, 35–45.

Gray, A. R. (1965). Antigenic variation in a strain of Trypanosoma brucei transmitted by Glossina morsitans and G. palpalis. *J Gen Microbiol* **41**, 195–214.

Guilliams, M., Oldenhove, G., Noel, W., Hérin, M., Brys, L., Loi, P., Flamand, V., Moser, M., De Baetselier, P. & Beschin, A. (2007). African trypanosomiasis: naturally occurring regulatory T cells favor trypanotolerance by limiting pathology associated with sustained type 1 inflammation. *J Immunol* **179**, 2748–2757.

Guirnalda, P., Murphy, N. B., Nolan, D. & Black, S. J. (2007). Anti-Trypanosoma brucei activity in Cape buffalo serum during the cryptic phase of parasitemia is mediated by antibodies. *Int J Parasitol* **37**, 1391–1399.

Günzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L.-C., Chung, H.-M., Lee, P.-T. & Lee, M. G.-S. (2003). RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in Trypanosoma brucei. *Eukaryotic Cell* **2**, 542–551.

Haanstra, J. R., van Tuijl, A., Kessler, P., Reijnders, W., Michels, P. A. M., Westerhoff, H. V., Parsons, M. & Bakker, B. M. (2008). Compartmentation prevents a lethal turbo-explosion of glycolysis in trypanosomes. *Proc Natl Acad Sci USA* **105**(46), 17718–17723. doi:10.1073/pnas.0806664105

Hall, T. & Esser, K. (1984). Topologic mapping of protective and nonprotective epitopes on the variant surface glycoprotein of the WRATat 1 clone of Trypanosoma brucei rhodesiense. *J Immunol* **132**, 2059–2063.

Hamilton, P. B., Gibson, W. C. & Stevens, J. R. (2007). Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. *Mol Phylogenet Evol* **44**, 15–25.

Hamilton, P. B., Stevens, J. R., Gaunt, M. W., Gidley, J. & Gibson, W. C. (2004). Trypanosomes are monophyletic: evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA. *Int J Parasitol* **34**, 1393–1404.

Hamrick, T. S., Dempsey, J. A., Cohen, M. S. & Cannon, J. G. (2001). Antigenic variation of gonococcal pilin expression in vivo: analysis of the strain FA1090 pilin repertoire and identification of the pilS gene copies recombining with pilE during experimental human infection. *Microbiology (Reading, Engl)* **147**, 839–849.

Herbert, W. & Lumsden, W. (1976). Trypanosoma Brucei: a Rapid “Matching”

Method for Estimating the Host's Parasitemia. *Exp Parasitol* **40**, 427–431.

Hertz, C. J., Filutowicz, H. & Mansfield, J. M. (1998). Resistance to the African trypanosomes is IFN-gamma dependent. *J Immunol* **161**, 6775–6783.

Hertz, C. J. & Mansfield, J. M. (1999). IFN-gamma-dependent nitric oxide production is not linked to resistance in experimental African trypanosomiasis. *Cell Immunol* **192**, 24–32.

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. & Berriman, M. (2008). Telomeric expression sites are highly conserved in Trypanosoma brucei. *PLoS ONE* **3**, e3527.

Hill, M. O. (1973). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* **54**, 427–432. Duke University Press.

Hill, S. A. & Davies, J. K. (2009). Pilin gene variation in Neisseria gonorrhoeae: reassessing the old paradigms. *FEMS Microbiol Rev* **33**, 521–530.

Hirumi, H. & 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.

Hoeijmakers, J. H., Frasch, A. C., Bernards, A., Borst, P. & Cross, G. A. M. (1980). Novel expression-linked copies of the genes for variant surface antigens in trypanosomes. *Nature* **284**, 78–80.

Holmes, P. H., MacAskill, J. A., Whitelaw, D. D., Jennings, F. W. & Urquhart, G. M. (1979). Immunological clearance of 75Se-labelled Trypanosoma brucei in mice. I. Aspects of the radiolabelling technique. *Immunology* **36**, 415–420.

Horn, D. & McCulloch, R. (2010). Molecular mechanisms underlying the control of antigenic variation in African trypanosomes. *Curr Opin Microbiol* **13**, 700–705.

Hsia, R., Beals, T. P. & Boothroyd, J. C. (1996). Use of chimeric recombinant polypeptides to analyse conformational, surface epitopes on trypanosome variant surface glycoproteins. *Mol Microbiol* **19**, 53–63.

Hughes, K., Wand, M., Foulston, L., Young, R., Harley, K., Terry, S., Ersfeld, K. & Rudenko, G. (2007). A novel ISWI is involved in *VSG* expression site downregulation in African trypanosomes. *EMBO J* **26**, 2400–2410.

Hutchinson, O. C., Picozzi, K., Jones, N. G., Mott, H. R., Sharma, R., Welburn, S. C. & Carrington, M. (2007). Variant Surface Glycoprotein gene repertoires in Trypanosoma brucei have diverged to become strain-specific. *BMC Genomics* **8**, 234.

Hutchinson, O. C., Smith, W., Jones, N. G., Chattopadhyay, A., Welburn, S. C. & Carrington, M. (2003). VSG structure: similar N-terminal domains can form functional VSGs with different types of C-terminal domain. *Mol Biochem Parasitol* **130**, 127–131.

Inojosa, W. O., Augusto, I., Bisoffi, Z., Josenado, T., Abel, P. M., Stich, A. & Whitty, C. J. M. (2006). Diagnosing human African trypanosomiasis in Angola using a card agglutination test: observational study of active and passive case finding strategies. *BMJ* **332**, 1479.

Iverson-Cabral, S. L., Astete, S. G., Cohen, C. R. & Totten, P. A. (2007). mgpB and mgpC sequence diversity in Mycoplasma genitalium is generated by segmental reciprocal recombination with repetitive chromosomal sequences. *Mol Microbiol* **66**, 55–73.

Hauser, H., Gamble, J., Gilderthorp, R., Marcello, L., McQuillan, J., Otto, T. D., Quail, M. A., Sanders, M. J., van Tonder, A., Ginger, M. L., Field, M.

C., Barry, J. D., Hertz-Fowler, C. & Berriman, M. (2012). Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. *Proc Natl Acad Sci USA* **109**, 3416.

Jackson, A. P., Sanders, M., Berry, A., McQuillan, J., Aslett, M. A., Quail, M. A., Chukualim, B., Capewell, P., Macleod, A., & other authors. (2010). The genome sequence of Trypanosoma brucei gambiense, causative agent of chronic human african trypanosomiasis. *PLoS Negl Trop Dis* **4**, e658.

Jackson, D. G., Owen, M. J. & Voorheis, H. P. (1985). A new method for the rapid purification of both the membrane-bound and released forms of the variant surface glycoprotein from Trypanosoma brucei. *Biochem J* **230**, 195–202.

Jamonneau, V., Ilboudo, H., Kaboré, J., Kaba, D., Koffi, M., Solano, P., Garcia, A., Courtin, D., Laveissière, C., & other authors. (2012). Untreated Human Infections by Trypanosoma brucei gambiense Are Not 100% Fatal. *PLoS Negl Trop Dis* **6**, e1691. Public Library of Science.

Jennings, F. W., Whitelaw, D. D., Holmes, P. H., Chizyuka, H. G. & Urquhart, G. M. (1979). The brain as a source of relapsing Trypanosoma brucei infection in mice after chemotherapy. *Int J Parasitol* **9**, 381–384.

Johnson, J. G. & Cross, G. A. M. (1979). Selective cleavage of variant surface glycoproteins from Trypanosoma brucei. *Biochem J* **178**, 689–697.

Johnson, P. L. F., Kochin, B. F., Ahmed, R. & Antia, R. (2012). How do antigenically varying pathogens avoid cross-reactive responses to invariant antigens? *Proc Biol Sci*.

Jones, N. G., Nietlispach, D., Sharma, R., Burke, D. F., Eyres, I., Mues, M., Mott, H. R. & Carrington, M. (2008). Structure of a glycosylphosphatidylinositol-anchored domain from a trypanosome variant surface glycoprotein. *J Biol Chem* **283**, 3584–3593.

Kahler, C. M. & Stephens, D. S. (1998). Genetic basis for biosynthesis, structure, and function of meningococcal lipooligosaccharide (endotoxin). *Crit Rev Microbiol* **24**, 281–334.

Kamper, S. M. & 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.

Kelley, L. A. & Sternberg, M. J. E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363–371.

Kennedy, P. G. E. (2007). Animal models of human African trypanosomiasis--very useful or too far removed? *Trans R Soc Trop Med Hyg* **101**, 1061–1062.

Kim, J. H., Skountzou, I., Compans, R. & Jacob, J. (2009). Original Antigenic Sin Responses to Influenza Viruses. *J Immunol* **183**, 3294–3301.

Kosakovsky Pond, S. L. & Frost, S. D. W. (2005). Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol Biol Evol* **22**, 1208–1222.

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. W. (2006). Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* **23**, 1891–1901.

Krafsur, E. S. (2008). Tsetse flies: genetics, evolution, and role as vectors. *Infect Genet Evol* **9**, 124–141.

Kyes, S. A., Rowe, J. A., Kriek, N. & Newbold, C. I. (1999). Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with Plasmodium falciparum. *Proc Natl Acad Sci USA* **96**, 9333–9338.

Kyes, S. A., Kraemer, S. M. & Smith, J. D. (2007). Antigenic variation in Plasmodium falciparum: gene organization and regulation of the var multigene family. *Eukaryotic Cell* **6**, 1511–1520.

La Greca, F. & Magez, S. (2011). Vaccination against trypanosomiasis: can it be done or is the trypanosome truly the ultimate immune destroyer and escape artist? *Hum Vaccin* **7**, 1225–1233.

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

Landeira, D., Bart, J.-M., Van Tyne, D. & Navarro, M. (2009). Cohesin regulates *VSG* monoallelic expression in trypanosomes. *J Cell Biol* **186**, 243–254.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.

Laurent, M., Pays, E., Delinte, K., Magnus, E., Van Meirvenne, N. & Steinert, M. (1984a). Evolution of a trypanosome surface antigen gene repertoire linked to non-duplicative gene activation. *Nature* **308**, 370–373.

Laurent, M., Pays, E., Van der Werf, A., Aerts, D., Magnus, E., Van Meirvenne, N. & Steinert, M. (1984b). Translocation alters the activation rate of a trypanosome surface antigen gene. *Nucleic Acids Res* **12**, 8319–8328.

Laveran, A. & Mesnil, F. (1902). Le Traitement et la Prévention du Nagana. *Annales de l'Institute Pasteur* **16**, 785.

Lee, M. G. & Van der Ploeg, L. H. (1987). Frequent independent duplicative transpositions activate a single *VSG* gene. *Mol Cell Biol* **7**, 357–364.

Legros, D., Ollivier, G., Gastellu-Etchegorry, M., Paquet, C., Burri, C., Jannin, J. & Büscher, P. (2002). Treatment of human African trypanosomiasis--present situation and needs for research and development. *Lancet Infect Dis* **2**, 437–440.

Leinster, T. & Cobbold, C. A. (2012). Measuring diversity: the importance of species similarity. *Ecology* **93**, 477–489.

Lin, T., Gao, L., Edmondson, D. G., Jacobs, M. B., Philipp, M. T. & Norris, S. J. (2009a). Central role of the Holliday junction helicase RuvAB in vlsE recombination and infectivity of Borrelia burgdorferi. *PLoS Pathog* **5**, e1000679.

Lin, Y., Hubert, L. & Wilson, J. H. (2009b). Transcription destabilizes triplet repeats. *Mol Carcinog* **48**, 350–361.

Linardopoulou, E. V., Williams, E. M., Fan, Y., Friedman, C., Young, J. M. & Trask, B. J. (2005). Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* **437**, 94–100.

Liu, A. Y., Michels, P. A., Bernards, A. & Borst, P. (1985). Trypanosome variant surface glycoprotein genes expressed early in infection. *J Mol Biol* **182**, 383–396.

Liu, A. Y., Van der Ploeg, L. H., Rijsewijk, F. A. & 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.

Livingstone, D. (1858). Missionary Travels and Researches in South Africa. *Project Gutenberg*, http://www.gutenberg.org/ebooks/1039

Lu, Y., Alarcon, C. M., Hall, T., Reddy, L. V. & Donelson, J. E. (1994). A strand bias occurs in point mutations associated with variant surface glycoprotein gene conversion in Trypanosoma rhodesiense. *Mol Cell Biol* **14**, 3971–3980.

Lukes, J., Hashimi, H, & Zíková, A. (2005). Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Current genetics*, **48**(5), 277–299. doi:10.1007/s00294-005-0027-0

Lyons, M. (2008). From "Death Camps" to Cordon Sanitaire: The Development of Sleeping Sickness Policy in the Uele District of the Belgian Congo, 1903-1914. *The Journal of African History* **26**, 26.

Lythgoe, K. A., Morrison, L. J., Read, A. F. & Barry, J. D. (2007). Parasite-intrinsic factors can explain ordered progression of trypanosome antigenic variation. *Proc Natl Acad Sci USA* **104**, 8095–8100.

Ma, L., Jensen, J. S., Myers, L., Burnett, J., Welch, M., Jia, Q. & Martin, D. H. (2007). Mycoplasma genitalium: an efficient strategy to generate genetic variation from a minimal genome. *Mol Microbiol* **66**, 220–236.

MacGregor, P. & Matthews, K. R. (2010). New discoveries in the transmission biology of sleeping sickness parasites: applying the basics. *J Mol Med*.

MacGregor, P., Savill, N. J., Hall, D. & Matthews, K. R. (2011). Transmission Stages Dominate Trypanosome Within-Host Dynamics during Chronic Infections. *Cell Host Microbe* **9**, 310–318.

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

Magez, S. & Caljon, G. (2011). Mouse models for pathogenic African trypanosomes: unravelling the immunology of host-parasite-vector interactions. *Parasite Immunol* **33**, 423–429.

Magez, S., Radwanska, M., Beschin, A., Sekikawa, K. & De Baetselier, P. (1999). Tumor necrosis factor alpha is a key mediator in the regulation of experimental Trypanosoma brucei infections. *Infect Immun* **67**, 3128–3132.

Magez, S., Schwegmann, A., Atkinson, R., Claes, F., Drennan, M., De Baetselier, P. & Brombacher, F. (2008). The role of B-cells and IgM antibodies in parasitemia, anemia, and *VSG* switching in Trypanosoma brucei-infected mice. *PLoS Pathog* **4**, e1000122.

Majiwa, P. A., Young, J. R., Englund, P. T., Shapiro, S. Z. & Williams, R. O. (1982). Two distinct forms of surface antigen gene rearrangement in Trypanosoma brucei. *Nature* **297**, 514–516.

Majumder, H. K., Boothroyd, J. C. & Weber, H. (1981). Homologous 3'-terminal regions of mRNAs for surface antigens of different antigenic variants of Trypanosoma brucei. *Nucleic Acids Res* **9**, 4745–4753.

Mansfield, J. M. & Paulnock, D. (2005). Regulation of innate and acquired immunity in African trypanosomiasis. *Parasite Immunol* **27**, 361–371.

Marcello, L. & Barry, J. D. (2007a). 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.

Marcello, L. & Barry, J. D. (2007b). From silent genes to noisy populations-dialogue between the genotype and phenotypes of antigenic variation. *J Eukaryot Microbiol* **54**, 14–17.

Marcello, L., Menon, S., Ward, P., Wilkes, J. M., Jones, N. G., Carrington, M. & Barry, J. D. (2007). VSGdb: a database for trypanosome variant surface glycoproteins, a large and diverse family of coiled coil proteins. *BMC Bioinformatics* **8**, 143.

Martinsohn, J. T., Sousa, A. B., Guethlein, L. A. & Howard, J. C. (1999). The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics* **50**, 168–200.

Masterson, W. J., Taylor, D. & Turner, M. J. (1988). Topologic analysis of the

epitopes of a variant surface glycoprotein of Trypanosoma brucei. *J Immunol* **140**, 3194–3199.

Matthyssens, G., Michiels, F., Hamers, R., Pays, E. & Steinert, M. (1981). Two variant surface glycoproteins of Trypanosoma brucei have a conserved C-terminus. *Nature* **293**, 230–233.

McConnell, J., Gurnett, A. M., Cordingley, J. S., Walker, J. E. & Turner, M. J. (1981). Biosynthesis of Trypanosoma brucei variant surface glycoprotein. I. Synthesis, size, and processing of an N-terminal signal peptide. *Mol Biochem Parasitol* **4**, 225–242.

McCulloch, R. & Barry, J. D. (1999). A role for RAD51 and homologous recombination in Trypanosoma brucei antigenic variation. *Genes Dev* **13**, 2875–2888.

McCulloch, R. & Horn, D. (2009). What has DNA sequencing revealed about the *VSG* expression sites of African trypanosomes? *Trends Parasitol* **25**, 359–363.

McCulloch, R., Rudenko, G. & Borst, P. (1997). Gene conversions mediating antigenic variation in Trypanosoma brucei can occur in variant surface glycoprotein expression sites lacking 70-base-pair repeat sequences. *Mol Cell Biol* **17**, 833–843.

McLintock, L. M., Turner, C. M. R. & Vickerman, K. (1993). Comparison of the effects of immune killing mechanisms on Trypanosoma brucei parasites of slender and stumpy morphology. *Parasite Immunol* **15**, 475–480.

Meeus, P. F. M., Brayton, K. A., Palmer, G. H. & Barbet, A. F. (2003). Conservation of a gene conversion mechanism in two distantly related paralogues of Anaplasma marginale. *Mol Microbiol* **47**, 633–643.

Mehlert, A., Bond, C. S. & Ferguson, M. A. J. (2002). The glycoforms of a Trypanosoma brucei variant surface glycoprotein and molecular modeling of a glycosylated surface coat. *Glycobiology* **12**, 607–612.

Mehlert, A., Wormald, M. R. & Ferguson, M. A. J. (2012). Modeling of the N-glycosylated transferrin receptor suggests how transferrin binding can occur within the surface coat of Trypanosoma brucei. *PLoS Pathog* **8**, e1002618.

Melchers, F., Boekel, ten, E., Seidl, T., Kong, X. C., Yamagami, T., Onishi, K., Shimizu, T., Rolink, A. G. & Andersson, J. (2000). Repertoire selection by pre-B-cell receptors and B-cell receptors, and genetic control of B-cell development from immature to mature B cells. *Immunol Rev* **175**, 33–46.

Melville, S. E. (1998). The african trypanosome genome project:focus on the future. *Parasitol Today (Regul Ed)* **14**, 129–131.

Merritt, S. C., Tschudi, C., Konigsberg, W. H. & Richards, F. F. (1983). Reverse transcription of trypanosome variable antigen mRNAs initiated by a specific oligonucleotide primer. *Proc Natl Acad Sci USA* **80**, 1536–1540.

Michels, P.A., Liu, A.Y., Bernards, A., Sloof, P., Van der Bijl, M.M., Schinkel, A.H., Menke, H.H., Borst, P., Veeneman, G.H., Tromp, M.C., and van Boom, J.H. (1983). Activation of the genes for variant surface glycoproteins 117 and 118 in Trypanosoma brucei. *J Mol Biol* **166**, 537–556.

Michiels, F., Matthyssens, G., Kronenberger, P., Pays, E., Dero, B., Van Assel, S., Darville, M., Carvador, A., Steinert, M. & Hamers, R. (1983). Gene activation and re-expression of a Trypanosoma brucei variant surface glycoprotein. *EMBO J* **2**, 1185–1192.

Miller, E. N., Allan, L. M. & Turner, M. J. (1984a). Mapping of antigenic determinants within peptides of a variant surface glycoprotein of Trypanosoma brucei. *Mol Biochem Parasitol* **13**, 309–322.

Miller, E. N., Allan, L. M. & Turner, M. J. (1984b). Topological analysis of antigenic determinants on a variant surface glycoprotein of Trypanosoma brucei. *Mol Biochem Parasitol* **13**, 67–81.

Miller, E. N. & Turner, M. J. (1981). Analysis of antigenic types appearing in first relapse populations of clones of Trypanosoma brucei. *Parasitology* **82**, 63–80.

Molyneux, D., Ndung'u, J. & Maudlin, I. (2010). Controlling sleeping sickness--"when will they ever learn?." *PLoS Negl Trop Dis* **4**, e609.

Morrison, L. J., Majiwa, P., Read, A. F. & Barry, J. D. (2005). Probabilistic order in antigenic variation of Trypanosoma brucei. *Int J Parasitol* **35**, 961–972.

Morrison, L. J., Marcello, L. & McCulloch, R. (2009). Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity. *Cell Microbiol* **11**, 1724–1734.

Morrison, W. I., Black, S. J., Paris, J., Hinson, C. A. & Wells, P. W. (1982). Protective immunity and specificity of antibody responses elicited in cattle by irradiated Trypanosoma brucei. *Parasite Immunol* **4**, 395–407.

Mosser, D. M. & Roberts, J. F. (1982). Trypanosoma brucei: recognition in vitro of two developmental forms by murine macrophages. *Exp Parasitol* **54**, 310–316.

Moxon, R., Bayliss, C. & Hood, D. (2006). Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet* **40**, 307–333.

Mulla, A. F. & Rickman, L. R. (1988). How do African game animals control trypanosome infections? *Parasitol Today (Regul Ed)* **4**, 352–354.

Munang'andu, H. M., Siamudaala, V., Munyeme, M. & Nalubamba, K. S. (2012). A Review of Ecological Factors Associated with the Epidemiology of Wildlife Trypanosomiasis in the Luangwa and Zambezi Valley Ecosystems of Zambia. *Interdisciplinary Perspectives on Infectious Diseases* **2012**, 1–13.

Muñoz-Jordán, J. L., Davies, K. P. & Cross, G. A. M. (1996). Stable expression of mosaic coats of variant surface glycoproteins in Trypanosoma brucei. *Science* **272**, 1795–1797.

Myler, P. J., Aline, R. F., Scholler, J. K. & Stuart, K. D. (1988). Multiple events associated with antigenic switching in Trypanosoma brucei. *Mol Biochem Parasitol* **29**, 227–241.

Namangala, B., Noël, W., de Baetselier, P., Brys, L. & Beschin, A. (2001). Relative contribution of interferon-gamma and interleukin-10 to resistance to murine African trypanosomosis. *J Infect Dis* **183**, 1794–1800.

Nantulya, V. M., Musoke, A. J. & Moloo, S. K. (1986). Apparent exhaustion of the variable antigen repertoires of Trypanosoma vivax in infected cattle. *Infect Immun* **54**, 444–447.

Nantulya, V. M., Musoke, A. J., Rurangirwa, F. R. & Moloo, S. K. (1984). Resistance of cattle to tsetse-transmitted challenge with Trypanosoma brucei or Trypanosoma congolense after spontaneous recovery from syringe-passaged infections. *Infect Immun* **43**, 735–738.

Navarro, M. & Gull, K. (2001). A pol I transcriptional body associated with *VSG* mono-allelic expression in Trypanosoma brucei. *Nature* **414**, 759–763.

Nei, M. & Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* **39**, 121–152.

Niang, M., Yan Yam, X. & Preiser, P. R. (2009). The Plasmodium falciparum STEVOR multigene family mediates antigenic variation of the infected erythrocyte. *PLoS Pathog* **5**, e1000307.

Noormohammadi, A. H., Markham, P. F., Kanci, A., Whithear, K. G. & Browning, G. F. (2000). A novel mechanism for control of antigenic variation in the haemagglutinin gene family of mycoplasma synoviae. *Mol Microbiol* **35**, 911–923.

O'Beirne, C., Lowry, C. M. & Voorheis, H. P. (1998). Both IgM and IgG anti-VSG antibodies initiate a cycle of aggregation-disaggregation of bloodstream forms of Trypanosoma brucei without damage to the parasite. *Mol Biochem Parasitol* **91**, 165–193.

Oberle, M., Balmer, O., Brun, R. & Roditi, I. (2010). Bottlenecks and the Maintenance of Minor Genotypes during the Life Cycle of Trypanosoma brucei. *PLoS Pathog* **6**, e1001023.

Odelberg, S. J., Weiss, R. B., Hata, A. & White, R. (1995). Template-switching during DNA synthesis by Thermus aquaticus DNA polymerase I. *Nucleic Acids Res* **23**, 2049–2057.

Ota, T. & Nei, M. (1994). Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol Biol Evol* **11**, 469–482.

Overath, P., Stierhof, Y. D. & Wiese, M. (1997). Endocytosis and secretion in trypanosomatid parasites - Tumultuous traffic in a pocket. *Trends Cell Biol* **7**, 27–33.

Pal, A., Hall, B. S., Jeffries, T. R. & Field, M. C. (2003). Rab5 and Rab11 mediate transferrin and anti-variant surface glycoprotein antibody recycling in Trypanosoma brucei. *Biochem J* **374**, 443–451.

Palmer, G. H., Bankhead, T. & Lukehart, S. A. (2009). "Nothing is permanent but change-" antigenic variation in persistent bacterial pathogens. *Cell Microbiol* **11**, 1697–1705.

Palmer, G. H., Futse, J. E., Knowles, D. P. & Brayton, K. A. (2006). Insights into mechanisms of bacterial antigenic variation derived from the complete genome sequence of Anaplasma marginale. *Ann N Y Acad Sci* **1078**, 15–25.

Palmer, G. H., Futse, J. E., Leverich, C. K., Knowles, D. P., Rurangirwa, F. R. & Brayton, K. A. (2007). Selection for simple major surface protein 2 variants during Anaplasma marginale transmission to immunologically naïve animals. *Infect Immun* **75**, 1502–1506.

Pan, W., Ogunremi, O., Wei, G., Shi, M. & Tabel, H. (2006). CR3 (CD11b/CD18) is the major macrophage receptor for IgM antibody-mediated phagocytosis of African trypanosomes: diverse effect on subsequent synthesis of tumor necrosis factor alpha and nitric oxide. *Microbes Infect* **8**, 1209–1218.

Parsons, M., Nelson, R. G., Watkins, K. P. & Agabian, N. (1984). Trypanosome mRNAs share a common 5' spliced leader sequence. *Cell* **38**, 309–316.

Pavri, R. & Nussenzweig, M. C. (2011). AID targeting in antibody diversity. *Adv Immunol* **110**, 1–26.

Pays, E., Lips, S., Nolan, D., Vanhamme, L. & Pérez-Morga, D. (2001). The *VSG* expression sites of *Trypanosoma brucei*: multipurpose tools for the adaptation of the parasite to mammalian hosts. *Mol Biochem Parasitol* **114**(1), 1–16.

Pays, E., Houard, S., Pays, A., Van Assel, S., Dupont, F., Aerts, D., Huet-Duvillier, G., Gomés, V., Richet, C. & Degand, P. (1985a). Trypanosoma brucei: the extent of conversion in antigen genes may be related to the DNA coding specificity. *Cell* **42**, 821–829.

Pays, E., Lheureux, M. & Steinert, M. (1981a). Analysis of the DNA and RNA changes associated with the expression of isotypic variant-specific antigens of trypanosomes. *Nucleic Acids Res* **9**, 4225–4238.

Pays, E. (1989). Pseudogenes, chimaeric genes and the timing of antigen variation in African trypanosomes. *Trends Genet* **5**, 389–391.

Pays, E., Guyaux, M., Aerts, D., Van Meirvenne, N. & Steinert, M. (1985b). 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. & Steinert, M. (1983). Gene conversion as a mechanism for antigenic variation in trypanosomes. *Cell* **34**, 371–381.

Pays, E., Van Meirvenne, N., Le Ray, D. & Steinert, M. (1981b). Gene duplication and transposition linked to antigenic variation in Trypanosoma brucei. *Proc Natl Acad Sci USA* **78**, 2673–2677.

Peacock, L., Cook, S., Ferris, V., Bailey, M. & Gibson, W. (2012). The life cycle of Trypanosoma (Nannomonas) congolense in the tsetse fly. *Parasites & vectors* **5**, 109.

Penchenier, L., Alhadji, D., Bahébégué, S., Simo, G., Laveissière, C. & Cuny, G. (2005). Spontaneous cure of domestic pigs experimentally infected by Trypanosoma brucei gambiense. Implications for the control of sleeping sickness. *Vet Parasitol* **133**, 7–11.

Petersen, T. N., Brunak, S., Heijne, von, G. & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Meth* **8**, 785–786.

Pinder, M., van Melick, A. & Vernet, G. (1987). Analysis of protective epitopes on the variant surface glycoprotein of a Trypanosoma brucei brucei (DiTat 1.3.) using monoclonal antibodies. *Parasite Immunol* **9**, 395–400.

Podlipaev, S. A., Sturm, N. R., Fiala, I., Fernandes, O., Westenberger, S. J., Dollet, M., Campbell, D. A. & Lukes, J. (2004). Diversity of insect trypanosomatids assessed from the spliced leader RNA and 5S rRNA genes and intergenic regions. *J Eukaryot Microbiol* **51**, 283–290.

Poisson, G., Chauve, C., Chen, X. & Bergeron, A. (2007). FragAnchor: a large-scale predictor of glycosylphosphatidylinositol anchors in eukaryote protein sequences by qualitative scoring. *Genomics Proteomics Bioinformatics* **5**, 121–130.

Prucca, C. G. & Lujan, H. D. (2009). Antigenic variation in Giardia lamblia. *Cell Microbiol* **11**, 1706–1715.

Radwanska, M., Guirnalda, P., De Trez, C., Ryffel, B., Black, S. & Magez, S. (2008). Trypanosomiasis-induced B cell apoptosis results in loss of protective anti-parasite antibody responses and abolishment of vaccine-induced memory responses. *PLoS Pathog* **4**, e1000078.

Radwanska, M., Magez, S., Michel, A., Stijlemans, B., Geuskens, M. & Pays, E. (2000). Comparative analysis of antibody responses against HSP60, invariant surface glycoprotein 70, and variant surface glycoprotein reveals a complex antigen-specific pattern of immunoglobulin isotype switching during infection by Trypanosoma brucei. *Infect Immun* **68**, 848–860.

Rask, T. S., Hansen, D. A., Theander, T. G., Gorm Pedersen, A. & Lavstsen, T. (2010). Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes--divide and conquer. *PLoS Comput Biol* **6**.

Rausch, S., Shayan, P., Salnikoff, J. & Reinwald, E. (1994). Sequence determination of three variable surface glycoproteins from Trypanosoma congolense. Conserved sequence and structural motifs. *Eur J Biochem* **223**, 813–821.

Recker, M., Buckee, C. O., Serazin, A., Kyes, S., Pinches, R., Christodoulou, Z., Springer, A. L., Gupta, S. & Newbold, C. I. (2011). Antigenic Variation in Plasmodium falciparum Malaria Involves a Highly Structured Switching Pattern. *PLoS Pathog* **7**, e1001306.

Reinitz, D. M. & Mansfield, J. M. (1988). Independent regulation of B cell responses to surface and subsurface epitopes of African trypanosome variable surface glycoproteins. *J Immunol* **141**, 620–626.

Reuner, B., Vassella, E., Yutzy, B. & Boshart, M. (1997). Cell density triggers

slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. *Mol Biochem Parasitol* **90**, 269–280.

Rice-Ficht, A. C., Chen, K. K. & Donelson, J. E. (1981). Sequence homologies near the C-termini of the variable surface glycoproteins of Trypanosoma brucei. *Nature* **294**, 53–57. Nature Publishing Group.

Rice-Ficht, A. C., Chen, K. K. & Donelson, J. E. (1982). Point mutations during generation of expression-linked extra copy of trypanosome surface glycoprotein gene. *Nature* **298**, 676–679.

Rivero, F. D., Saura, A., Prucca, C. G., Carranza, P. G., Torri, A. & Lujan, H. D. (2010). Disruption of antigenic variation is crucial for effective parasite vaccine. *Nature Medicine* **16**, 551–7– 1p following 557.

Robertson, M. (1912). Notes on the Life-History of Trypanosoma gambiense, etc. *Proceedings of the Royal Society B: Biological Sciences* **86**, 66–71.

Robinson, N. P., Burman, N., Melville, S. E. & Barry, J. D. (1999). Predominance of duplicative *VSG* gene conversion in antigenic variation in African trypanosomes. *Mol Cell Biol* **19**, 5839–5846.

Roditi, I., Schwarz, H., Pearson, T. W., Beecroft, R. P., Liu, M. K., Richardson, J. P., Bühring, H. J., Pleiss, J., Bülow, R. & Williams, R. O. (1989). Procyclin gene expression and loss of the variant surface glycoprotein during differentiation of Trypanosoma brucei. *J Cell Biol* **108**, 737–746.

Roger, E., Grunau, C., Pierce, R. J., Hirai, H., Gourbal, B. E. F., Galinier, R., Emans, R., Cesari, I. M., Cosseau, C. & Mitta, G. (2008). Controlled Chaos of Polymorphic Mucins in a Metazoan Parasite (Schistosoma mansoni) Interacting with Its Invertebrate Host (Biomphalaria glabrata). *PLoS Negl Trop Dis* **2**, e330.

Roth, C. W., Bringaud, F., Layden, R. E., Baltz, T. & Eisen, H. (1989). Active late-appearing variable surface antigen genes in Trypanosoma equiperdum are constructed entirely from pseudogenes. *Proc Natl Acad Sci USA* **86**, 9375–9379.

Roth, C. W., Longacre, S., Raibaud, A., Baltz, T. & Eisen, H. (1986). The use of incomplete genes for the construction of a Trypanosoma equiperdum variant surface glycoprotein gene. *EMBO J* **5**, 1065–1070.

Rotureau, B., Subota, I., Buisson, J. & Bastin, P. (2012). A new asymmetric division contributes to the continuous production of infective trypanosomes in the tsetse fly. *Development* **139**, 1842–1850.

Rudenko, G., McCulloch, R., Dirks-Mulder, A. & 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.

Russell, J. & Zomerdijk, J. C. B. M. (2005). RNA-polymerase-I-directed rDNA transcription, life and works. *Trends Biochem Sci* **30**, 87–96.

Salmon, D., Hanocq-Quertier, J., Paturiaux-Hanocq, F., Pays, A., Tebabi, P., Nolan, D. P., Michel, A. & Pays, E. (1997). Characterization of the ligand-binding site of the transferrin receptor in Trypanosoma brucei demonstrates a structural relationship with the N-terminal domain of the variant surface glycoprotein. *EMBO J* **16**, 7272–7278.

Salmon, D., Vanwalleghem, G., Morias, Y., Denoeud, J., Krumbholz, C., Lhomme, F., Bachmaier, S., Kador, M., Gossmann, J., Dias, F. B. S., De Muylder, G., Uzureau, P., Magez, S., Moser, M., de Baetselier, P., Van Den Abbeele, J., Boshart, M. & Pays, E. (2012). Adenylate Cyclases of Trypanosoma brucei Inhibit the Innate Immune Response of the Host. *Science*.

Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle,

B., Gysin, J. & Lanzer, M. (1998). Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in Plasmodium falciparum. *EMBO J* **17**, 5418–5426. Nature Publishing Group.

Schwede, A. & Carrington, M. (2006). Bloodstream form trypanosome plasma membrane proteins: antigenic variation and invariant antigens. *Parasitology* **-1**, 1–11.

Schwede, A., Jones, N., Engstler, M. & Carrington, M. (2011). The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes. *Mol Biochem Parasitol* **175**, 201–204.

Seed, J. R. & Sechelski, J. (1988). Growth of pleomorphic Trypanosoma brucei rhodesiense in irradiated inbred mice. *J Parasitol* **74**, 781–789.

Seed, J. R., Seed, J. R., Wenck, M. A. & Wenck, M. A. (2003). Role of the long slender to short stumpy transition in the life cycle of the african trypanosomes. *Kinetoplastid Biol Dis* **2**, 3.

Sendashonga, C. N. & Black, S. J. (1982). Humoral responses against Trypanosoma brucei variable surface antigen are induced by degenerating parasites. *Parasite Immunol* **4**, 245–257.

Seyfang, A., Mecke, D. & Duszenko, M. (1990). Degradation, recycling, and shedding of Trypanosoma brucei variant surface glycoprotein. *J Protozool* **37**, 546–552.

Shafikhani, S. (2002). Factors affecting PCR-mediated recombination. *Environ Microbiol* **4**, 482–486.

Simpson, A. G. B., Stevens, J. R. & Lukes, J. (2006). The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol* **22**, 168–174.

Smith, D. J., Lapedes, A. S., de Jong, J. C., Bestebroer, T. M., Rimmelzwaan, G. F., Osterhaus, A. D. M. E. & Fouchier, R. A. M. (2004). Mapping the antigenic and genetic evolution of influenza virus. *Science* **305**, 371–376.

Smith, T. K., Vasileva, N., Gluenz, E., Terry, S., Portman, N., Kramer, S., Carrington, M., Michaeli, S., Gull, K. & Rudenko, G. (2009). Blocking variant surface glycoprotein synthesis in Trypanosoma brucei triggers a general arrest in translation initiation. *PLoS ONE* **4**, e7532.

Sodroski, J. G., Wyatt, R., Kwong, P. D., Desjardins, E., Sweet, R. W., Robinson, J. & Hendrickson, W. A. (1998). The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393**, 705–711.

Soltys, M. (1963). Immunity in African trypanosomiasis. *Bull World Health Organ* **28**, 753–761.

Songa, E. B., Hamers, R., Rickman, R., Nantulya, V. M., Mulla, A. F. & Magnus, E. (1991). Evidence for widespread asymptomatic Trypanosoma rhodesiense human infection in the Luangwa Valley (Zambia). *Trop Med Parasitol* **42**, 389–393.

Ssenyonga, G. S. Z., & Adam, K. M. G. (1974). The number and morphology of trypanosomes in the blood and lymph of rats infected with *Trypanosoma brucei* and *T. congolense*. *Parasitology*, **70**(02), 255–261. doi:10.1017/S0031182000049714

Stanne, T. M. & Rudenko, G. (2010). Active *VSG* expression sites in Trypanosoma brucei are depleted of nucleosomes. *Eukaryotic Cell* **9**, 136–147.

Steverding, D. (2008). The history of African trypanosomiasis. *Parasites & vectors* **1**, 3.

Steverding, D. (2010). The development of drugs for treatment of sleeping sickness: a historical review. *Parasites & vectors* **3**, 15. BioMed Central Ltd.

Strickler, J. E., Binder, D. A., L'Italien, J. J., Shimamoto, G. T., Wait, S. W.,

Dalheim, L. J., Novotny, J., Radding, J. A. & Konigsberg, W. H. (1987). Trypanosoma congolense: structure and molecular organization of the surface glycoproteins of two early bloodstream variants. *Biochemistry* **26**, 796–805. American Chemical Society.

Taylor, J. E. & Rudenko, G. (2006). Switching trypanosome coats: what's in the wardrobe? *Trends Genet* **22**, 614–620.

Taylor, K. A. (1998). Immune responses of cattle to African trypanosomes: protective or pathogenic? *Int J Parasitol* **28**, 219–240.

Theodos, C., Reinitz, D. M. & Mansfield, J. M. (1990). Regulation of B cell responses to the variant surface glycoprotein (VSG) molecule in trypanosomiasis. I. Epitope specificity and idiotypic profile of monoclonal antibodies to the VSG of Trypanosoma brucei rhodesiense. *The Journal of Immunology* **144**, 4011.

Thon, G., Baltz, T. & Eisen, H. (1989). Antigenic diversity by the recombination of pseudogenes. *Genes Dev* **3**, 1247–1254.

Thon, G., Baltz, T., Giroud, C. & Eisen, H. (1990). Trypanosome variable surface glycoproteins: composite genes and order of expression. *Genes Dev* **4**, 1374–1383.

Tilley, H. (2008). Ecologies of Complexity: Tropical Environments, African Trypanosomiasis, and the Science of Disease Control in British Colonial Africa, 1900-1940. *Osiris* 21.

Timmers, H. T., De Lange, T., Kooter, J. M. & Borst, P. (1987). Coincident multiple activations of the same surface antigen gene in Trypanosoma brucei. *J Mol Biol* **194**, 81–90.

Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* **302**, 575–581. Nature Publishing Group.

Trowsdale, J. & Parham, P. (2004). Mini-review: defense strategies and immunity-related genes. *Eur J Immunol* **34**, 7–17.

Turner, C. M., Aslam, N. & Dye, C. (1995). Replication, differentiation, growth and the virulence of Trypanosoma brucei infections. *Parasitology* **111 (Pt 3)**, 289–300.

Turner, C. M., Hunter, C. A., Barry, J. D. & Vickerman, K. (1986). Similarity in variable antigen type composition of Trypanosoma brucei rhodesiense populations in different sites within the mouse host. *Trans R Soc Trop Med Hyg* **80**, 824–830.

Turner, C. M., Sternberg, J., Buchanan, N., Smith, E., Hide, G. & Tait, A. (1990). Evidence that the mechanism of gene exchange in Trypanosoma brucei involves meiosis and syngamy. *Parasitology* **101 Pt 3**, 377–386.

Turner, C. M. R. (1990). The use of experimental artefacts in African trypanosome research. *Parasitol Today (Regul Ed)* **6**, 14–17.

Turner, C. M. R. (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. R. (1999). Antigenic variation in Trypanosoma brucei infections: an holistic view. *J Cell Sci* **112 (Pt 19)**, 3187–3192.

Turner, C. & Barry, J. (1989). High frequency of antigenic variation in Trypanosoma brucei rhodesiense infections. *Parasitology* **99**, 67–75.

Uilenberg, G. (1998). *A field guide for THE DIAGNOSIS, TREATMENT AND PREVENTION OF AFRICAN ANIMAL TRYPANOSOMOSIS*. ISBN 92-5-104238-1, published online at http://www.fao.org/DOCREP/006/X0413E/X0413E00.htm

Ulbert, S., Chaves, I. & Borst, P. (2002). Expression site activation in Trypanosoma brucei with three marked variant surface glycoprotein gene expression sites. *Mol Biochem Parasitol* **120**, 225–235.

Van der Ploeg, L. H., Valerio, D., De Lange, T., Bernards, A., Borst, P. & Grosveld, F. G. (1982). An analysis of cosmid clones of nuclear DNA from Trypanosoma brucei shows that the genes for variant surface glycoproteins are clustered in the genome. *Nucleic Acids Res* **10**, 5905–5923.

van der Woude, M. W. & Bäumler, A. J. (2004). Phase and antigenic variation in bacteria. *Clin Microbiol Rev* **17**, 581–611– table of contents.

Van Meirvenne, N., Magnus, E. & Buscher, P. (1995). Evaluation of variant specific trypanolysis tests for serodiagnosis of human infections with Trypanosoma brucei gambiense. *Acta Trop* **60**, 189–199.

Vanhamme, L., Paturiaux-Hanocq, F., Poelvoorde, P., Nolan, D. P., Lins, L., Van Den Abbeele, J., Pays, A., Tebabi, P., Van Xong, H., & other authors. (2003). Apolipoprotein L-I is the trypanosome lytic factor of human serum. *Nature* **422**, 83–87.

Vanhamme, L., Poelvoorde, P., Pays, A., Tebabi, P., Van Xong, H. & Pays, E. (2000). Differential RNA elongation controls the variant surface glycoprotein gene expression sites of Trypanosoma brucei. *Mol Microbiol* **36**, 328–340.

Vanhollebeke, B., Truc, P., Poelvoorde, P., Pays, A., Joshi, P. P., Katti, R., Jannin, J. G. & Pays, E. (2006). Human Trypanosoma evansi infection linked to a lack of apolipoprotein L-I. *N Engl J Med* **355**, 2752–2756.

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.

Vincendeau, P., Daulouède, S., Veyret, B., Darde, M. L., Bouteille, B. & Lemesre, J. L. (1992). Nitric oxide-mediated cytostatic activity on Trypanosoma brucei gambiense and Trypanosoma brucei brucei. *Exp Parasitol* **75**, 353–360.

Vink, C., Rudenko, G. & Seifert, H. S. (2011). Microbial antigenic variation mediated by homologous DNA recombination. *FEMS Microbiol Rev*.

Walder, J., Eder, P., Engman, D., Brentano, S., Walder, R., Knutzon, D., Dorfman, D. & Donelson, J. E. (1986). The 35-nucleotide spliced leader sequence is common to all trypanosome messenger RNA's. *Science* **233**, 569.

Wang, Q.-P., Kawahara, T. & Horn, D. (2010). Histone deacetylases play distinct roles in telomeric *VSG* expression site silencing in African trypanosomes. *Mol Microbiol* **77**, 1237–1245.

Weirather, J. L., Wilson, M. E. & Donelson, J. E. (2012). Mapping of *VSG* similarities in Trypanosoma brucei. *Mol Biochem Parasitol* **181**, 141–152.

Welburn, S. C., Maudlin, I. & Simarro, P. P. (2009). Controlling sleeping sickness - a review. *Parasitology* **136**, 1943–1949.

West, S. C. (2003). Molecular views of recombination proteins and their control. *Nat Rev Mol Cell Biol* **4**, 435–445.

Wheeler, R. J. (2010). The trypanolytic factor-mechanism, impacts and applications. *Trends Parasitol* **26**, 457–464.

WHO. (2012). Fact Sheet N°259: Human African trypanosomiasis (sleeping sickness). *WHO Fact Sheet*. World Health Organization.

Wickstead, B., Ersfeld, K. & Gull, K. (2004). The small chromosomes of Trypanosoma brucei involved in antigenic variation are constructed around repetitive palindromes. *Genome Res* **14**, 1014–1024.

Wirtz, E., Leal, S., Ochatt, C. & Cross, G. A. M. (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. *Mol Biochem Parasitol* **99**, 89–101.

Wolinsky, S. M., Korber, B. T., Neumann, A. U., Daniels, M., Kunstman, K. J., Whetsell, A. J., Furtado, M. R., Cao, Y., Ho, D. D. & Safrit, J. T. (1996).

Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* **272**, 537–542.

Xong, H. V., Vanhamme, L., Chamekh, M., Chimfwembe, C. E., Van Den Abbeele, J., Pays, A., Van Meirvenne, N., Hamers, R., De Baetselier, P. & Pays, E. (1998). A *VSG* expression site-associated gene confers resistance to human serum in Trypanosoma rhodesiense. *Cell* **95**, 839–846.

Young, J. R., Donelson, J. E., Majiwa, P. A., Shapiro, S. Z. & Williams, R. O. (1982). Analysis of genomic rearrangements associated with two variable antigen genes of Trypanosoma brucei. *Nucleic Acids Res* **10**, 803–819.

Young, J. R., Miller, E. N., Williams, R. O. & Turner, M. J. (1983). Are there two classes of *VSG* gene in Trypanosoma brucei? *Nature* **306**, 196–198.

Young, J. M., Endicott, R. M., Parghi, S. S., Walker, M., Kidd, J. M. & Trask, B. J. (2008a). Extensive copy-number variation of the human olfactory receptor gene family. *Am J Hum Genet* **83**, 228–242.

Young, R., Taylor, J. E., Kurioka, A., Becker, M., Louis, E. J. & Rudenko, G. (2008b). Isolation and analysis of the genetic diversity of repertoires of *VSG* expression site containing telomeres from Trypanosoma brucei gambiense, T. b. brucei and T. equiperdum. *BMC Genomics* **9**, 385.

Zaphiropoulos, P. G. (2002). Template switching generated during reverse transcription? *FEBS Lett* **527**, 326.

Zhang, J. R., Hardham, J. M., Barbour, A. G. & Norris, S. J. (1997). Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. *Cell* **89**, 275–285.

Zhang, J. R. & Norris, S. J. (1998). Genetic variation of the Borrelia burgdorferi gene vlsE involves cassette-specific, segmental gene conversion. *Infect Immun* **66**, 3698–3704.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40.

Zhuang, Y., Futse, J. E., Brown, W. C., Brayton, K. A. & Palmer, G. H. (2007). Maintenance of antibody to pathogen epitopes generated by segmental gene conversion is highly dynamic during long-term persistent infection. *Infect Immun* **75**, 5185–5190.

Ziegelbauer, K. & Overath, P. (1993). Organization of two invariant surface glycoproteins in the surface coat of Trypanosoma brucei. *Infect Immun* **61**, 4540–4545.